Language Input for Word Discovery

ISBN: 90-76203-07-5
Cover illustration: Joost van de Weijer
Printed and bound by Ponsen & Looijen bv, Wageningen

Language Input for Word Discovery

Een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 11 januari 1999
des namiddags om 3.30 uur precies

door

Joost Christian van de Weijer
geboren op 31 oktober 1966 te Eindhoven

Promotor:
Prof. Dr. A. Cutler

Manuscriptcommissie:
Prof. Dr. C. Gussenhoven
Dr. P. Brown (MPI, Nijmegen)
Prof. Dr. P. Jusczyk (Johns Hopkins University, Baltimore)

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

Dankwoord

Veel mensen zijn betrokken geweest bij de totstandkoming van dit proefschrift. Hierbij wil ik iedereen bedanken voor de wetenschappelijke, technische of morele ondersteuning die ik gekregen heb. Ik wil echter de volgende mensen apart noemen.

Ten eerste mijn begeleiders, Anne Cutler en Ken Drozd. Ik bedank Anne voor haar enthousiaste begeleiding en de vele aanwijzingen die ik van haar gekregen heb die hebben bijgedragen aan dit proefschrift en mijn wetenschappelijke ontwikkeling. Ik bedank Ken voor vele nuttige aanwijzingen en voor de prettige samenwerking ook tijdens het editten van het Annual Report.

Ten tweede bedank ik de leden van de leescommissie voor hun kritische commentaar dat tot een significante verbetering van de tekst heeft geleid.

Niet minder dank ben ik verschuldigd aan allen die betrokken zijn geweest bij de dataverzameling voor dit project, met name de ouders, de oppas, en de directie en medewerkers van het kinderdagverblijf. De data zijn uniek en alleen door de inzet van alle betrokkenen heeft het corpus zo'n complete verzameling kunnen worden.

Hierbij bedank ik ook Michael Brent. Hij heeft me betrokken bij zijn onderzoek waar ik enthousiast voor was en waardoor ik mijn eerste conferentie heb bijgewoond.

Verder wil ik hier een aantal mensen bedanken voor het kritisch lezen van eerdere versies van delen van het proefschrift en voor het commentaar dat ze gegeven hebben: Jos van Berkum, Shanley Allen, James McQueen, Laura Walsh-Dickey, Herb Clark, Victoria Johansson, Jeroen van de Weijer en Nicole Cooper.

Dankzij de andere Ph.D.-studenten zijn de afgelopen vier jaar een prettige tijd geweest, niet alleen tijdens het werk maar ook daarnaast. Apart bedank ik Arie van der Lugt voor het schrijven van een programma waardoor de orthographische transcripties automatisch konden worden omgezet in fonologische transcripties.

Tenslotte bedank ik familie en vrienden die allen op verschillende manieren een bijdrage hebben geleverd aan dit proefschrift.

Nijmegen, november 1998

Table of Contents

Chapter 1 Introduction 1

1.0 Introduction 1
1.1 The Development of Speech Perception during Infancy 3
1.1.1 Methodology of Research 4
1.1.2 The Development of Sensitivity to Sound Structure 5
1.1.3 Infants' Knowledge of Consonants 7
1.1.4 Infants' Knowledge of Vowels 9
1.1.5 The Relation between Input and Perceptual Changes . . 10
1.1.6 The Importance of Suprasegmental Structure 11
1.1.7 Summary 14
1.2 Lexical Segmentation 14
1.2.1 Segmentation in Adults 15
1.2.2 Segmentation in Infants 18
1.2.3 Summary 23
1.3 Motivation for the Present Research 23
1.4 Outline of the Thesis 24

Chapter 2 Methodology 27

2.0 Introduction 27
2.1 Corpus Collection 28
2.2 Speaking Time 32
2.2.1 Methodology 33
2.2.2 Results 36
2.2.3 Discussion 38
2.3 Transcriptions 39
2.3.1 Format of the Transcriptions 39
2.3.2 Classification of Utterance Types 41
2.3.3 Phonological Transcriptions 42
2.3.4 Description of the Transcribed Corpus 43
2.3.5 Summary 43
2.4 Fragment of the Corpus 44

Chapter 3 Structural Aspects as a Function of Addressee 49

3.0 Introduction 49
3.1 Utterance Length 49
3.1.1 Distribution of Utterance Types 51

3.1.2 Methodology 52
3.1.3 Results 53
3.1.4 Summary 56
3.2 Vocabulary Size 57
3.2.1 Methodology 57
3.2.2 Results 58
3.2.3 Summary 61
3.3 Metrical Structure 61
3.3.1 Methodology 62
3.3.2 Results 63
3.3.3 Summary 66
3.4 Phonotactic Structure 67
3.4.1 Methodology 68
3.4.2 Results 69
3.4.3 Summary 71
3.5 Word Embedding 72
3.5.1 Methodology 72
3.5.2 Results 73
3.5.3 Summary 75
3.6 Conclusions 75

Chapter 4 Suprasegmental Structure 79

4.0 Introduction 79
4.0.1 Pitch 80
4.0.2 Intonation 81
4.0.3 Tempo 82
4.0.4 Focus of this Chapter 85
4.1 Selection of the Material 85
4.2 Pitch 86
4.2.1 Methodology 86
4.2.2 Results 87
4.2.3 Summary 91
4.3 Intonation 92
4.3.1 Methodology 92
4.3.2 Results 93
4.3.3 Summary 95
4.4 Speech Rate 96
4.4.1 Methodology 96
4.4.2 Results 97
4.4.3 Summary 98
4.5 Conclusions 99

Chapter 5 Experiments on Language Learning 103

5.0 Introduction 103
5.0.1 Word Segmentation
in the Absence of Lexical Knowledge 104
5.0.2 Preceding Studies 107
5.0.3 The Present Study 110
5.1 Experiment 1 110
5.1.1 Methodology 110
5.1.2 Results 113
5.1.3 Discussion 116
5.2 Experiment 2 116
5.2.1 Methodology 117
5.2.2 Results 119
5.2.3 Discussion 122
5.3 Conclusions 123

Chapter 6 Conclusions 127

6.0 Introduction 127
6.1 Summary of the Results 128
6.2 Implications of the Results 130
6.3 Implications of the Results for Word Segmentation 132
6.4 Modelling Word Discovery between Six and Nine Months . 133
6.5 Future Research 136

References 139

Appendices 148

Samenvatting 152

Chapter 1
Introduction

In this chapter, two theoretical issues that served as the background for this thesis are discussed: phonological development during the prelingual period and lexical segmentation. Although these are two distinct issues, they are also related since phonological structure may guide children into the initial stages of word segmentation, a notion that is now often referred to as 'prosodic bootstrapping'. After the theoretical part, the goals and the motivation of the present research are described. Finally, the outline of the remainder of the thesis is given.

1.0 Introduction

An important part of language acquisition, especially in the early stages, involves building up a lexicon. Considerable attention has been given to lexical development by researchers in acquisition. Some major issues that have often been (and still are) addressed include: how do children learn to relate words to concepts, or, how do children learn to assign the appropriate grammatical category to words (see Bloom, 1994 for an overview)? An issue that has received relatively little attention is: how do children learn to derive the words of the language spoken around them, of the linguistic input which they hear?

This is not a trivial issue, since deriving the words from the input is not a straightforward matter. The problem lies in the fact that words are usually presented in context, rather than in isolation (Aslin, 1993), and that researchers, so far, have failed to demonstrate reliable cues that indicate word boundaries in spoken language. Comparable markers of word boundaries to those that exist in written language (the spaces between the words) are absent in the acoustic signal. Words in utterances are usually not separated by pauses, for example,

which makes the speech signal a continuous stream. Furthermore, assimilation of phonemes tends to apply across word boundaries, which can make the separation of two words even more difficult. Thus, isolating the words from the input is not a straightforward matter. The adult listener experiences the same problem when listening to a foreign, unknown language.

Still, listeners have to divide utterances into words. It would be impossible to store all the possible combinations of words in memory, given the number of words that occur in a language. This process of locating word boundaries in spoken language is called *lexical segmentation.*

The question of how adult listeners are able to segment spoken language into words is addressed in many psycholinguistic studies (see for example: Cutler, 1995). Relatively few studies focus on infants. Yet, how infants learn to locate the word boundaries at the right places is a basic question for the theory of language acquisition and the theory of speech processing in adult listeners. In the initial stages of language acquisition, infants have no lexicon yet. In the absence of lexical knowledge, prelingual infants have to find a way of dividing the language input into smaller units.

As will be described in this chapter, infants show the ability to segment words out of fluent speech as early as at seven and a half months of age. Furthermore, it is around this age that representations of aspects of the phonological structure of the native language are established. This development involves acquiring knowledge about the sounds of the native language, how these sounds can be ordered, how the metrical structure is defined. It is now often assumed that infants use this knowledge of the phonological structure for the segmentation of words out of context.

These two aspects of language comprehension, being able to segment words out of fluent speech and having representations of the sound structure of the native language, are prerequisites for later lexical development. And, as will also be described in a later section, it is not surprising that early word comprehension follows directly after acquisition of the two 'skills' outlined above.

In order to obtain a better appreciation of the infant's acquisition of these skills it is necessary to have a well-based knowledge of what the input looks like. The infant, as an apparently passive listener to the speech around him, has to analyze what at first sight may seem like an immense amount of information (how many words per day?), that is produced at an incredibly high rate (three, four words per second), with numerous factors causing noise (for instance, the probably high number of speakers the infant hears, all speaking at different speech rates, with different voices, using different words, etc.). Still, the infant appears to succeed well, and, at a very young age. Nonetheless, a complete picture of what the input may look like is thus far lacking.

In this thesis, an attempt is made to provide a complete picture of language input. Therefore, almost all the language that an infant heard between the age of six and nine months was recorded. Before going into detail about this project, however, the early 'prelexical' development that has been shown during this period of life will be described first in this chapter. Section 1.1 focuses on the development of representations of a number of aspects of the phonological structure: the phonetic, phonotactic and metrical structure. The development of representations of these aspects is reflected in a change of speech perception during infancy. Next, section 1.2 focuses on the issue of word segmentation. First, some results of studies with adults listeners are described first, and then some results of studies with infants. In section 1.3, the motivation for the present research is given. Finally, section 1.4 is an outline of the remainder of the dissertation.

1.1 The Development of Speech Perception during Infancy

Children must develop knowledge about the sound structure of their language. They have to learn which sounds occur in the language, and how these sounds can be ordered. Building up representations of the sounds of the native language is not a straightforward matter since the acoustic realization of phonemes varies depending, for example, on speaker, speech rate or phonetic context, a phenomenon that has been called *phonemic variability* (e.g., Miller and Eimas, 1994). So far, it remains unclear how listeners are able to recognize that different realizations of one sound belong to the same category, a concept which is referred to as *phonemic constancy*.

The acquisition of representations of speech sounds is necessary so that the listener can cope with the non-linguistic variability in the speech signal. The studies on development of speech perception seek to answer the questions of how and when phoneme categories are established, and how and when sensitivity to aspects of the structure of the native language (e.g., the ordering of the speech sounds) develops.

Before describing this course of development, three frequently used methods of testing infants' speech perception are explained in section 1.1.1. Subsequently, the development of sensitivity to aspects of the sound structure of the language is described in the sections 1.1.2 to 1.1.4, followed by a brief description of the role of the input on early perceptual development in section 1.1.5. Finally, the importance of the suprasegmental structure of the input is discussed in section 1.1.6, since suprasegmental features appear to play a distinct role in early language development.

1.1.1 Methodology of Research

Generally, three different methods are used to test infants' perception: the head turn procedure, the High Amplitude Sucking procedure, and the preferential listening procedure. These procedures are also described in Jusczyk (1997).

The *head turn procedure* (e.g., Werker and Tees, 1984a; Grieser and Kuhl, 1989) is used for infants between six and twelve months old and is used to test infants' discrimination capacities. During this procedure the infant hears repetitions of one stimulus (for example a CV-syllable) coming from a loudspeaker on one side of the infant. This stimulus is called the background stimulus. At irregular intervals this stimulus is temporarily changed and a few times (usually four) another stimulus (the contrasting stimulus) is presented. In the initial phase of the procedure, the training phase, a visual reinforcer is presented simultaneously with this auditory change so that the infant learns to associate the visual reinforcer with the auditory change. Subsequently, in the test phase, the visual reinforcer is presented only after a correct head turn, that is, after a head turn caused by the presentation of the contrasting stimulus. An instance of a correct discrimination is held to have occurred when the infant reaches a certain number of correct head turns within a limited number of trials.

With younger infants, the *High Amplitude Sucking (HAS)* procedure is used (for example by Bertoncini, Bijeljac-Babic, Blumstein and Mehler, 1987). This procedure is based on the fact that the presentation of a new stimulus leads to increased sucking behaviour of the infant. There are two phases: a pre-shift phase, and a post-shift phase. In the pre-shift phase the infant listens to the repeated presentation of a speech stimulus. Sucking behaviour is registered through a wired nipple. After several presentations the infant gets habituated to the stimulus and sucking rate declines to a baseline. At that moment, a new stimulus is presented. If the infant hears the difference between the new and the old stimulus, the sucking rate will increase again. If not, there will be no change.

The *preferential listening procedure* was originally designed by Fernald (1985). The procedure can be used to test infants with a wide age range, but is usually used with infants between six and twelve months. In this procedure, the infant is seated on the parent's lap. A green light in front of the infant draws the infant's attention towards the middle. Then, two red lights, mounted above two loudspeakers on the right side and the left side of the infant, start to blink simultaneously, and the green light is extinguished. When the infant orients towards one of the lights, one of two versions of speech stimuli associated with that light starts to play (e.g., infant-directed speech or adult-directed speech). This continues until the child looks away from the loudspeaker, after which the procedure is restarted until a certain number of trials is run. An observer

registers the number of times the infant orients towards the right side or the left side. If the infant orients more often to either side, then the conclusion is drawn that the stimulus associated with that side is the preferred one (see: Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss and Kennedy, 1987).

When this procedure was used in other experiments (e.g., Kemler Nelson, Hirsh-Pasek, Jusczyk and Wright Cassidy, 1989; Hirsh-Pasek et al., 1987) the frequency with which the child oriented towards the left side or the right side did not prove to be a measure that was sensitive enough to detect preferences in the child's behaviour. Therefore, it was decided to measure the time that the child oriented towards the left or the right loudspeaker rather than the frequency with which the child oriented.

A number of changes were then made in the experimental procedure. As in the original version, the infant's attention was drawn towards the middle by means of a green light. Then a red light on one side started to blink. When the infant oriented towards this side the stimulus associated with this side started to play. The time that the child looked towards this side was recorded (either on line, or in a later stadium, with a video recorder for a more objective measurement). When the child looked away for longer than two seconds the speech stimuli stopped and the procedure was repeated. If the child looked away for less than two seconds, the trial was not interrupted. The short time that the infant looked away was not included in the total looking time however.

This procedure yielded better results than the original procedure and was used in many experiments afterwards (see below). The procedure is described in detail in Kemler Nelson, Jusczyk, Mandel, Myers, Turk and Gerken (1995). A variation on the preferential listening procedure is described in Swingley, Pinto and Fernald (1998). In this version, which is typically used to test word recognition in young children between one and two years old, not only the time that the child orients towards a visual target is measured, but also the speed (reaction time) with which the child orients towards the target.

1.1.2 The Development of Sensitivity to Sound Structure

During the first year of life, infants acquire knowledge about the sound structure of their native language. Different languages have different sound systems: Not all possible sounds occur in all languages (phonetic constraints), and not all possible orderings of sounds are allowed in all languages (phonotactic constraints). Furthermore, different languages have different metrical (rhythmic) structures: The metrical structure of English, for example, is determined by the alternation of strong and weak syllables, but the metrical structure of French is

determined by the syllable (see section 1.2.1 for further explanation). Infants' sensitivity to these aspects of the sound structure of the language was examined in different experiments using the preferential listening procedure. The main goal of the experiments was to establish the course of development during which this sensitivity develops.

Jusczyk, Friederici, Wessels, Svenkerud and Jusczyk (1993b) tested whether infants are sensitive to the phonetic and the phonotactic structure of their native language. They presented lists of Dutch and English words to American infants of six and nine months old. The lists differed in the sounds that occurred in the words or in the ordering of the sounds. In other words: the Dutch words violated some English constraints, and the English words violated some Dutch constraints. The nine-months old infants listened significantly longer to the English words, but the six-months-olds did not. The preference of the American infants disappeared, however, when the segmental information was removed from the speech signal by means of low-pass filtering. Since the nine-month-olds did not listen longer to the filtered versions of the stimuli, Jusczyk et al. concluded that it was the segmental information that caused the infants' preference (see also: Friederici and Wessels, 1993).

In the same series of experiments Jusczyk et al. also compared Dutch infants with American infants. This time, the stimulus sets were composed of Dutch or English words that differed only on the basis of phonotactic constraints. Thus, both sets contained segments that were permissible in both languages, but differed in the ordering in which these segments occurred. The American infants showed a preference for the English words, while the Dutch infants showed a preference for the Dutch words. Again, this effect disappeared when the stimuli were low-pass filtered. Since the nine-month-olds preferred to listen to words that conformed to the segmental structure of their native language, but six-month-olds did not, Jusczyk et al. concluded that it is in this period that infants develop a sensitivity to the phonotactic structure of the native language.

Jusczyk, Luce and Charles-Luce (1994) extended the results from the previous study by showing that infants are sensitive to the distribution of phonotactic patterns within the language. In this study, infants heard CVC-nonwords that had either high or low phonotactic probabilities. Thus, all the words were constructed of combinations of sounds that are permissible in English, but half of them contained sound sequence that occur only rarely, whereas the other half contained sound sequences that occur frequently. Nine-month-old infants listened significantly longer to the high-probability nonwords, but six-month-olds did not. Thus, at nine months of age, infants are sensitive to the distribution of the phonotactic patterns in their native language.

This result was further confirmed in a study by Kuijpers and Coolen (1996). They considered an alternative explanation of the results that were obtained by Jusczyk et al. (1994), namely that the infants' preference was determined by syllable frequency rather than by phonotactic structure. However, the results of Kuijpers and Coolen showed that nine-month-old infants did not prefer to listen to high- or to low-frequency syllables that were matched for internal transitional or positional phoneme probabilities.

Jusczyk, Cutler and Redanz (1993a) tested whether English infants are sensitive to the metrical structure of their native language. The metrical structure of English is determined by the alternation between strong and weak (reduced) syllables. Words with a strong-weak stress pattern (e.g., *table*) occur much more frequently than words with a weak-strong stress pattern (e.g., *balloon*). Jusczyk et al. tested whether infants are sensitive to this regularity. They presented lists of bisyllabic words to infants of nine months and to infants of six months. Half of the words had a strong-weak stress pattern, and the other half had a weak-strong stress pattern. The nine-month-olds did indeed prefer to listen to the strong-weak words, but the six-month-olds did not. Therefore it seems that infants develop this sensitivity to the metrical structure of the native language between six and nine months. Jusczyk et al. then tested whether the metrical structure, and not the segmental information, was responsible for the infants' preference. They replicated the experiment with low-pass filtered stimuli. The infants' preference did not disappear after filtering. Therefore, it was concluded that the metrical information rather than the segmental information was indeed the cause of the infants' preference.

1.1.3 Infants' Knowledge of Consonants

At a very early age, infants can discriminate a whole range of consonantal contrasts, including contrasts from their native language, as well as contrasts from a foreign language. Bertoncini, Bijeljac-Babic, Blumstein and Mehler (1987) used the High Amplitude Sucking procedure to show that infants, only four days old, discriminated CV-syllables that had either different consonants or different vowels. The infants displayed higher sucking rates when either the vowel or the consonant of the syllable changed in the post-shift phase of the experiment. Similarly, Trehub (1976) showed that the ability to discriminate consonantal contrasts is not limited to contrasts taken from the native language, and thus does not depend on pre-exposure to the sounds. In this experiment, Canadian infants, approximately ten weeks of age, discriminated the Czech consonantal contrast /z/-/ř/, whereas adults readily confused the two sounds.

Furthermore, infants can discriminate phonetically relevant contrasts (i.e., ones that are used as phonemes) better than contrasts that are not relevant. This phenomenon, that differences between phonemes that belong to one phonemic class are perceived less well than differences between phonemes that belong to two phonemic classes, is called *categorical perception.* Categorical perception in infants was first demonstrated by Eimas, Siqueland, Jusczyk and Vigorito (1971). In their study, it was shown that infants, only one month old, perceived differences within a /p/ or a /b/ category better than differences between these two categories. In a similar study by Werker and Lalonde (1988), infants between six and eight months old also had to discriminate CV-syllable pairs that had either two consonants taken from one phonetic category (e.g., two variants of /t/) or from two phonetic categories (e.g., /t/ and /d/). Care was taken that the acoustic differences between the two consonants were the same in both cases. The infants were not able to discriminate the pairs when the consonants were taken from one category, but their discrimination improved significantly when the consonants were taken from two categories.

Hohne and Jusczyk (1994) tested infants' ability to discriminate allophonic realizations of a single phoneme because allophonic variation can be a cue to word boundaries. A voiceless plosive in initial position followed by a stressed vowel is aspirated in English, but a final voiceless plosive is not, as in [phæt]. In this experiment, the High Amplitude Sucking procedure was used. Two-month-old infants heard several tokens of the word *nitrate* in the pre-shift phase. In the post-shift phase, the infants heard a word that was phonemically different from the initial stimulus (e.g., *nikrate*), allophonically different (e.g., *night rate*), or identical. The infants perceived the allophonic differences and the phonemic differences equally well.

However, between the age of eight and ten months, the ability to discriminate non-native consonant contrasts declines. This was demonstrated by Werker and Tees (1984a). They used different tokens of a glottalized velar and a glottalized uvular stop contrast, taken from the American Indian language Nthlakampx. They tested whether English-learning infants, English-speaking adults, and Nthlakampx-speaking adults could discriminate this contrast. Infants between six and eight months old performed well on the test, and so did the Nthlakampx-speaking adults. The English adults however, were unable to discriminate the two sounds. Another group of older infants was subsequently tested to determine the age of change in perception. Infants between ten and twelve months were, like the English-speaking adults, no longer able to perceive the difference between the two consonants. Therefore it seemed that, between eight and ten months, infants lose the ability to discriminate non-native consonantal contrasts.

What these results show is that infants have particular perceptual capacities at the beginning of life. They show the capacity to perceive consonantal contrasts categorically since differences between two phonetic categories are perceived better than differences within a phonetic category. Furthermore, infants can perceive a whole range of consonantal contrasts in the early stages of life, including contrasts that are not phonemic in their native language. However, the ability to discriminate non-native consonantal contrasts disappears around the age of ten months, which suggests that infants of that age have acquired knowledge about the consonantal system of their native language.

1.1.4 Infants' Knowledge of Vowels

There is a development in the perception of vowels comparable to that demonstrated in the perception of consonants. Thus, at an early age, infants can discriminate a whole range of vowel contrasts including contrasts that are not relevant in their native language. But at a certain stage during development, the ability to discriminate non-native vowel contrasts declines.

This was, for instance, shown in a study by Kuhl (1979). In this study, infants, six months of age, had to categorize different tokens of the same vowel. The tokens that were used differed in pitch, speaker, or pitch and speaker. The results showed that the infants discriminated two different vowels, in spite of different realizations of the same vowel. Kuhl concluded that infants have established representations of the vowels at this age.

Further evidence for this conclusion comes from a study by Polka and Werker (1994). They examined the development of vowel perception during infancy. In Polka and Werker's study, English-learning infants of various ages had to discriminate the German vowel contrasts /ʏ/-/ʊ/ and /y/-/u/. Infants, between four and six months, were able to discriminate these contrasts, but infants between six and eight months performed worse, and infants between ten and twelve months performed even worse. It seems that the ability to discriminate non-native vowel contrasts disappears around the age of six months.

This conclusion has to be considered cautiously, however, as was demonstrated by Polka and Bohn (1996). They conducted a similar study as Polka and Werker with English and German-learning infants but found no differences in the performance of infants six to eight months old and infants ten to twelve months old.

Further research on the perception of vowels suggests that the representations of vowels that infants and adults have are very specific. Given the considerable variation in the realization of speech sounds, it has been predicted

that human representations of speech sounds are organized around so-called 'prototypes' (Grieser and Kuhl, 1989; Kuhl, 1991). Prototypes are the 'best' exemplars of a category, for example, vowel exemplars that have average formant frequencies. Adult listeners classify prototypical vowels as 'sounding better' than non-prototypical vowels. Grieser and Kuhl (1989) argued that prototypes resemble the other members of the same category more than non-prototypes. Therefore they predicted that discrimination of a prototype with other tokens of the same category would be worse than discrimination of a non-prototype with other tokens of that category. This is indeed what they found: six months olds' discrimination between a prototypical /i/ and surrounding /i/'s was significantly worse than discrimination between a non-prototypical /i/ and surrounding /i/'s. This effect has been termed the *perceptual magnet effect.*

Sussman and Lauckner-Morano (1995) who had doubts about the perceptual magnet effect replicated Kuhl's (1991) study. They demonstrated that the vowel that was used in the original study as a non-prototypical /i/ was classified as a vowel from another phonetic category (/ɛ/ or /ɪ/) in more than 90% of all cases (see also: Lively, 1993). For that reason, they chose a vowel that was closer to the prototypical /i/ than the one chosen by Kuhl (1991). Using these stimuli, however, they still found a perceptual magnet effect.

The perceptual magnet effect is likely to be influenced by linguistic experience. This was demonstrated in a study by Kuhl, Williams, Lacerda, Stevens and Lindblom (1992). In this study, a group of Swedish infants and a group of American infants were tested on their ability to discriminate vowel contrasts within two vowel categories. One vowel category /i/ is a phoneme in English, the other vowel category /y/ is a phoneme in Swedish. The American infants showed the effect only when tested on contrasts within the /i/ category, whereas the Swedish infants only showed the effect when tested on contrasts within the /y/ category. However, Polka and Bohn (1996) failed to find language-specific differences in a comparable experiment with English and German infants, and therefore conclude that more research on the nature of the perceptual magnet effect is needed.

1.1.5 The Relation between Input and Perceptual Changes

The reason for the decline in the ability to discriminate non-native phoneme contrasts is the exposure to the native language: discrimination deteriorates because the information is not available in the input. However, it is not clear how we should construe this process of decline. Is the ability to discriminate

non-native contrasts actually lost, or can it still be retrieved under special conditions?

Werker and Tees (1984b) tried to answer this question. They showed that, under specific conditions, non-native contrasts can still be discriminated. For example, when the acoustic information that distinguishes two consonants was isolated from the rest of the speech signal, adults were able to hear the differences. Similarly, when the stimuli are presented very quickly after each other (500 ms), adults were also able to hear the differences. Werker and Tees interpreted these results as follows: Listeners develop a 'phonemic' kind of listening strategy and are no longer able to focus on properties of the speech signal that are not relevant for the perception of the native language. The relevant information is processed, but disappears within a very short time.

However, Best et al. (1988) have pointed out that not all non-native contrasts are equally difficult. Two non-native phonemes are only difficult to discriminate when the two assimilate to one consonant in the native language (e.g., Hindi dental and retroflex stops both assimilate to the English alveolar stop). When two non-native sounds do not assimilate to one in the native language, discrimination is much easier. Best et al. showed that English-learning infants between ten and twelve months of age, as well as English-speaking adults, were able to discriminate a Zulu click contrast, which does not resemble any of the phonemes of the English system.

Recent research by Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskuna, Stolyarova, Sundberg and Lacerda (1997) suggests that infant-directed speech plays an important role in the acquisition of vowel categories. They examined acoustical properties of the vowels /i,a,u/ in speech addressed to infants and to adults in three languages (Russian, American and Swedish). Vowels in infant-directed speech were produced with more clearly separated spectra, than vowels in adult-directed speech. This makes the perceptual distance between vowels larger and therefore, presumably, makes it easier for infants to build up representations of the vowels.

1.1.6 The Importance of Suprasegmental Structure

The term suprasegmental structure covers aspects of the speech signal that are not restricted to a single segment but extend over larger units, such as phrases and clauses. A more detailed description of suprasegmental structure is given in Chapter 4. Some aspects of the suprasegmental structure of a speech signal can be isolated through low-pass filtering (i.e., removing all the frequencies that are above a cut-off frequency in the speech signal). After low-pass filtering,

fundamental frequency, speech rate, pauses, etc. are still audible but segmental information is lost. There is evidence that suprasegmental structure plays an important role in the child's early perceptual development.

Firstly, the results of studies on infant speech perception suggest that at a very early age children are attentive to the suprasegmental structure of the language. Mehler, Jusczyk, Lambertz, Halsted, Bertoncini and Amiel-Tison (1988) showed that French newborns were able to distinguish their native language from another language based on suprasegmental information. The infants' discrimination was tested using the High Amplitude Sucking procedure. Speech samples in French and Russian were recorded, produced by a bilingual speaker. The sucking rate of the infants who heard French in the pre-shift phase was higher than that of the infants who heard Russian. Thus, overall, the infants preferred to listen to French. Furthermore, the infants who heard Russian in the pre-shift phase showed a significant increase in sucking rate when the language changed to French in the post-shift phase, whereas the group that listened to French first and then to Russian did not show this increase. Similar results were found when the speech samples were low-pass filtered. This suggests that the infants' preference was determined by differences in the suprasegmental structure of the two languages.

Secondly, infants show a preference to listen to 'motherese', the speech style that is adopted by adults when talking to children. This speech style is characterized by an exaggerated suprasegmental structure, such as high pitch and large pitch variations. Fernald (1985) showed that four-month-old infants prefer to listen to motherese rather than to normal adult-directed speech. In this study, the preferential listening procedure was used to test whether infants preferred to listen to speech samples spoken with an infant-directed speech style or with an adult-directed speech style. The infants made significantly more head turns towards the infant-directed speech samples, which suggests that they preferred this speech style.

Fernald and Kuhl (1987) subsequently showed that the large pitch variations determined the infants' preference. They synthesized new samples based on the utterances of Fernald's (1985) study. In the new samples only the intonation contours were audible. Again, four-month-olds oriented significantly more often towards the infant-directed samples.

Thus, infants are attentive to suprasegmental aspects of the speech signal, but suprasegmental structure also functions to organize the input into linguistically relevant units. For example, utterances are delimited by intonation contours and/or pauses. Using the preferential listening procedure, Hirsh-Pasek et al. (1987) tested whether infants are sensitive to suprasegmental information that signals clause boundaries. In that case, they argued, infants should prefer to

listen to spoken fragments that are interrupted at clause boundaries, rather than to fragments that are interrupted within clauses. They presented recorded fragments of texts, that were spoken to a young child, to infants with an average age of eight months. Two versions of the texts were used: *natural* versions that had one second pauses introduced at clause boundaries, and *unnatural* versions that had the same pauses introduced within the clauses. The infants made the same number of head turns towards the unnatural as to the natural versions, but they spent significantly more time listening to the natural versions of the fragments. Thus it seems that infants of eight months old are sensitive to the suprasegmental correlates of clause boundaries, at least when the cues are conveyed in infant-directed speech style.

In a second study, Kemler Nelson et al. (1989) did the same experiment, but used adult-directed speech instead of infant-directed speech. This time they failed to replicate the preference that was found in the previous study. Now, the infants did not listen significantly longer to the natural versions than to the unnatural versions. The second study strongly suggests that boundaries of clauses are more clearly demarcated in infant-directed speech than in adult-directed speech. This finding lends support to the hypotheses that the suprasegmental exaggerations that are found in infant-directed speech have a facilitative effect on language learning in the early stages of language development.

The results of the previous studies led to a further exploration of infants' sensitivity to suprasegmental structure. Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwoz (1992) tested whether infants were also sensitive to acoustic correlates of phrasal units in spoken language. Infants of nine months of age listened to fragments of text that were interrupted by a one-second break either at the boundary of a phrasal unit (i.e., between the subject phrase and the predicate) or within a phrasal unit. The infants preferred to listen to the versions with the coincident breaks rather than to the versions with the non-coincident breaks. This effect was found when the stimuli consisted of spontaneous speech as well as read speech, and when the stimuli were presented with an infant-directed speech style as well as with an adult-directed speech style. The effect also remained after removing all the segmental information through low-pass filtering. Infants with an average age of six months were also tested, but did not show any signs of preference. Therefore, Jusczyk et al. concluded that the sensitivity to this acoustic information develops between six and nine months. The acoustic cues that were responsible for the results of this study were a significant drop in fundamental frequency over the last three syllables at the end of the unit, and a significant increase in duration of the vowel at the end of the unit.

1.1.7 Summary

In essence, during the first year of life, infants acquire important knowledge about the structure of their native language. They learn which sounds occur in the language, how these sounds can be ordered, and what the metrical structure of the language is. From the experiments, it appears that much of this knowledge develops between the age of six and nine months. Furthermore, it was shown that infants are born with the capacity to discriminate consonant and vowel contrasts. This ability is not restricted to contrasts taken from the native language but has also been demonstrated for contrasts taken from other languages. However, the ability to discriminate nonnative consonant and vowel contrasts declines. Between the age of eight and ten months, the ability to discriminate non-native consonants declines and there is evidence that the ability to discriminate non-native vowel contrasts declines between the age of six and eight months. Furthermore, it has been demonstrated that infants' vowel representations are organized around so-called prototypes. The development in speech perception can be characterized as a change from language-general perception capacities to language-specific perception capacities. This development is caused by exposure to the ambient language. Linguistic input shapes the way the listener perceives speech. Finally, it was shown that, from a very early age on, infants are attentive to suprasegmental characteristics of the language, and therefore it is often assumed that the suprasegmental structure of the input plays an important role in the development of speech perception.

1.2 Lexical Segmentation

Lexical segmentation has been the subject of numerous psycholinguistic studies, some of which deal with adult subjects (e.g., Mehler, Dommergues, Frauenfelder and Segui, 1981; McQueen, Norris and Cutler, 1994; Cutler, 1994; Cutler and Norris, 1988), some of which deal with infants (e.g., Jusczyk and Aslin, 1995; Morgan, 1994).

The question that is addressed in studies with adults is: 'How can listeners segment words from the continuous speech signal?'. The results of this research is described in section 1.2.1.

More recently, attention has also focused on infants because infants also have to deal with segmentation problems, but lack the lexical knowledge that adults can rely on. The studies with infants investigate the early abilities to segment words out of continuous speech. The results of research with infants will be described in section 1.2.2.

1.2.1 Segmentation in Adults

Cole and Jakimik (1980) presented a model according to which adults rely on their lexical knowledge and general knowledge for word segmentation (lexically driven segmentation). The recognition of a word would result in the location of the onset of the next word. However, there are several problems with this assumption (McQueen, Norris and Cutler, 1994). For example, longer words often contain shorter lexical items (e.g., *bed* in *embedded*), and many words become unique only after their offset (McQueen, Cutler, Briscoe and Norris, 1995).

Since segmentation based on lexical information does not seem able to solve these problems, it has been proposed that listeners explicitly segment the speech signal. Explicit segmentation means that the listener locates places in the speech signal where word boundaries are likely to occur. At these places lexical access is triggered. According to this view, speech segmentation is a prelexical process that helps the listener to identify the meaningful units in the speech stream. It should be noted that explicit segmentation is not incompatible with lexically driven segmentation, but is used by the listener to opt for one segmentation rather than an alternative.

But what information can listeners use to locate word boundaries? Various sources of information have been proposed, of which the most important ones will be described here: metrical structure, phonotactic structure, allophonic variation, and statistical information.

Metrical Structure

One proposed source of information that listeners might use is metrical (or rhythmic) structure. It is however, not easy to define metrical structure, and it is not the same in every language (Cutler, 1990). In some languages, like English or Dutch, rhythmic structure is determined by the alternation of strong and weak syllables. A strong syllable is a syllable that contains a full (non-reduced) vowel, and has primary or secondary stress, such as the first syllable of the word *table*. A weak syllable has a reduced vowel, and has no stress, as the first syllable of the word *balloon*. Languages that have a metrical structure determined by vowel quality are called stress-timed languages. There are several indications that metrical structure is useful for word segmentation.

Cutler and Carter (1987) examined the distribution of strong and weak syllables in English words. They showed that English has more content words starting with a strong syllable than with a weak syllable. Moreover, the words that start with a strong syllable are more frequent than the words starting with a weak syllable. A word like *table* is more common than a word like *balloon*,

for example. Cutler and Carter estimated that about 73% of all English content words start with a strong syllable. Furthermore, they estimated that about 90% of all tokens of content words in a corpus of spoken English started with a strong syllable. Therefore, they concluded that there is a correlation between vowel quality and word onsets and that listeners locate word boundaries before every strong syllable.

This hypothesis was tested experimentally by Cutler and Norris (1988) who investigated whether adults listeners used the presence of strong syllables for the location of word boundaries. They tested the ability of subjects to spot monosyllabic words (e.g., *mint*) which were the beginning of bisyllabic nonsense words with either two strong syllables (e.g., *mintayv*), or bisyllabic words with a strong first syllable and a weak second syllable (e.g., *mintev*). The prediction was that the subjects would segment the words before the second strong syllable, which would interfere with recognition of the target word, and consequently that the subjects would be slower in spotting *mint* in *mintayv* than in *mintev*. This is indeed what they found.

Other studies have suggested that the metrical structure of Dutch is similar to that of English. For example, Vroomen and de Gelder (1995) found that 87.5% of Dutch lexical words also start with a strong syllable (see also: Baayen and Schreuder, 1994). Furthermore, Vroomen, van Zon and de Gelder (1996) replicated Cutler and Norris's (1988) word spotting experiment with Dutch materials and Dutch listeners and found similar results as in the original study.

Taken together, these results indicate that vowel quality is a possible cue for word segmentation. However, this can only be so in languages that distinguish between strong and weak syllables, such as English or Dutch. Vowel quality cannot be used as a cue to word boundaries in a language that does not make this distinction, for example French. In French, vowels are essentially all full rather than reduced. The metrical structure of French, instead, seems to be determined by the syllable, and, therefore, French is called a syllable-timed language.

Evidence that French is segmented by syllable rather than by vowel quality comes from an experiment by Mehler, Dommergues, Frauenfelder and Segui (1981). In this study, French subjects had to detect CV syllables (e.g., *pa*) or CVC-syllables (e.g., *pal*) in bisyllabic words of which the first three phonemes were the same, but which had different syllabic structure (e.g., *pa-lace* and *pal-mier*). The results showed that the targets were detected faster when they were aligned with the syllabic boundaries. Thus, *pa* was detected faster in *pa-lace*, whereas *pal* was detected faster in *pal-mier*. Mehler et al. concluded that French listeners segment words syllable-wise.

Phonotactic Structure

A second proposed source of information which can be used for the location of word boundaries is the phonotactic structure of the language (Church, 1987). The phonotactic structure determines the permissible ordering of speech sounds in a language. Dutch words, for example, can start with /kn/, but English words cannot. An English speaker will infer that /kn/ spans a syllable or a word boundary, but a Dutch speaker will not. Therefore, some sound sequences may signal word or syllable boundaries, but others do not.

McQueen (1998) tested experimentally whether adult listeners use their knowledge of the phonotactic structure of their native language for the location of word boundaries. In a word-spotting experiment, similar to that of Cutler and Norris (1988), Dutch subjects had to detect monosyllabic words in bisyllabic nonsense words. The embedded words were either aligned with the syllabic boundary (e.g., *rok*, 'skirt' in *fiem-rok*), or they were misaligned (e.g., *rok* in *fie-drok*). The subjects detected the embedded words significantly faster, and made significantly fewer errors when the word boundaries were aligned to the syllable boundary, than when they were not. McQueen concluded from this experiment that adult listeners are aware of the phonotactic constraints of their language, and use this knowledge for the location of word boundaries.

Allophonic Variation

Phonemes vary in the way they are realized depending on the context, a phenomenon which is called allophonic variation. For instance, the final /d/ of the word *did* in *did you* is often palatalized under the influence of the following phoneme. Allophonic variation can be useful for lexical segmentation, as was argued by Church (1987). For example, in English, stops are sometimes aspirated when they are in syllable-initial position but not when they are in syllable-final position. Similarly, stops in word-final position can be realized as glottal stops, but stops in word-initial position cannot.

Statistical Information

Finally, statistical (distributional) information can be used for segmentation (e.g., Saffran, Newport and Aslin, 1996; Brent and Cartwright, 1996). Recently, connectionist modelling has shown that language input contains many regularities which can be useful for various aspects of language acquisition (Plunkett, 1998). Furthermore, the results of these studies show that combined

cues generally yield better results than single cues (e.g., Shi, Morgan and Allopenna, 1998; Christiansen, Allen and Seidenberg, 1998).

An example of statistical information that is useful for word segmentation is transitional probability. Simply stated, transitional probability refers to the idea that word internal sequences of syllables or phonemes are more frequent than sequences that span a word boundary. A somewhat different approach is taken by Brent and Cartwright (1996) who proposed that the optimal segmentation of an utterance results in a low number of items that have to be stored in a lexicon and reduces the length of these items as much as possible. Since Brent's model and Saffran's et al. experiment will be explained in more detail in chapter 5, they are not described in further detail in this section.

In sum, several kinds of cues to word boundaries have been proposed. There is psycholinguistic evidence that listeners are aware of these cues and use them for segmentation. The various sources of information are not contradictory. It is very well possible that adults use multiple sources of information for lexical segmentation. In fact, Saffran et al. (1996) showed that listeners' learning of an artificial miniature language improved significantly when statistical *and* prosodic information were available. Similarly, Brent's model of word segmentation not only relies on distributional regularities but also on phonotactic information.

1.2.2 Segmentation in Infants

In the previous section it was shown that adults use explicit segmentation strategies in order to divide continuous speech into meaningful units. It was further shown that multiple sources of information may be useful for segmentation and that these sources of information can differ across languages. But what about infants? Do infants segment continuous speech? What kind of information do they use? And when do they display the ability to segment words from continuous speech?

The Input

A number of researchers have argued that the language input to infants is not such that the infants do not have segmentation problems during the course of language development. The arguments have two perspectives.

First, in child-directed speech words very often occur in the context of other words. In a study by Aslin (1993), mothers were asked to teach their children of twelve months old words that were unfamiliar to the children. The results showed that there was considerable variability in how often the mothers

presented the target words in isolation: some of them did it frequently, others never did. Therefore, Aslin concluded, infants must find a way of segmenting these unfamiliar words from the context. In this study, a number of helpful cues were found which assisted the child in isolating the target words. For example, the mothers presented the target words very often in the final position of the utterance or highlighted the target words using emphatic stress. Furthermore, the target words were preceded by a variety of other words. On the other hand, the mothers did not avoid using words that easily assimilated with the target words (e.g., *your wrist*).

A second reason why infants have to deal with segmentation problems, is that much of the input children receive is that of language used among mature users (Cutler, 1994). Thus, although there is reason to believe that word segmentation might be easier in speech addressed to infants, much of their language input consists of spontaneous speech between adults, from adults to older siblings, etc. This is true in our Western cultures, but is even more true in other cultures where adults do not tend to speak to children until the children are about one year old, such as the Kaluli in Papua New Guinea (Schieffelin, 1985) or the Tzeltal in Mexico (Brown, 1996). Thus a great deal of input that children receive is from language among mature users, which is not adjusted to the child's stage of acquisition.

Stage of Development

When, during the course of language development, do infants display the ability to segment words from continuous speech?

The beginning of receptive vocabulary building starts before the age of ten months (Benedict, 1979). In this study the early production and comprehension of children between nine months and 1;8 years old were examined. The results of this study were based on a combination of parental diaries, observations, and tests. The children were visited twice a week during the first half of the period. During the second half, they were visited twice a month. Lists of words that each child produced and understood were compiled. On the average, children understood between zero and 20 words when they were ten months old. By the time they were 1;1 years old, they understood, on the average, 50 words. At this time the children produced about ten words. The first 50 words in comprehension and in production comprised different word classes, including nominals, action words, modifiers and 'personal-social' words. The nominals and the action words dominated. Furthermore, in the first 50 words, action words were relatively more frequent in comprehension compared to production.

Thus, around the age of nine months, infants show limited comprehension of some words. Since words are usually not presented in isolation it seems likely that infants, before the age of nine months, have developed some kind of segmentation strategy. This idea is supported in a number of studies that focus on infants' ability to segment words out of context.

Jusczyk and Aslin (1995) investigated whether infants were able to recognize a word in a text after having heard this word several times in isolation. In this experiment, infants, approximately seven and a half months of age, listened to two monosyllabic words (e.g., *dog* and *cup*) for 30 seconds. After that, the infants listened to four different texts of six sentences. In two of these texts, the familiar words occurred in initial, medial, or final position of each sentence. In the other two texts, the words did not occur. The infants listened significantly longer to the texts containing the familiar words, suggesting that they had detected the words in the texts. Younger infants, six months of age, did not show any signs of recognition.

Jusczyk and Aslin subsequently investigated how detailed the infants' representations of the words were. Therefore, they replaced the target words in the texts with nonwords that differed only minimally from the familiar words (e.g., *tup* instead of *cup*). This time, the infants did not show any preference, which suggests that their representations of the words were quite accurate. Finally, Jusczyk and Aslin showed that the infants were also able to recognize words presented in isolation, when these words were first presented in context. Thus, the order of presentation of the stimuli was now reversed: the infants heard sentences with the word in different positions first and then they listened to words that occurred in these sentences or other words. Again, the infants listened significantly longer to the familiar words.

Information that Infants use for Word Segmentation

Newsome and Jusczyk (1995) extended Jusczyk and Aslin's results. Instead of monosyllabic words, they presented bisyllabic words to the infants. They found that, when infants were familiarized with strong-weak bisyllabic words (e.g., *kingdom*), the infants listened significantly longer to texts in which these words occurred as in the original study. However, when the infants were familiarized with weak-strong bisyllabic words (e.g., *device*) they did not show this preference.

To explore this result further, Newsome and Jusczyk familiarized the infants with only the strong syllables (e.g., *vice*) of the weak-strong words. This time, the infants did listen significantly longer to the texts with the words with the target syllable. Newsome and Jusczyk gave the following interpretation to

the results: Infants consider strong syllables to be the onset of words. In certain situations, however, where this is not true segmentation errors occur.

Kuijpers, Coolen, Houston and Cutler (1998) tried to replicate Newsome and Jusczyk's results with Dutch infants. They predicted the same results since English and Dutch have similar metrical structures. However, Dutch seven-and-half-month-olds did not prefer to listen to texts with target strong-weak words. Nine-months-olds did show a trend to listen longer to texts with the familiar words. The difference between the behaviour of the American and the Dutch infants is explained in terms of a difference in acoustic realization of strong and weak syllables in Dutch and in English. In English, strong vowels are realized with a higher average pitch and with more pitch variation than weak syllables. In Dutch, this difference is less clear. Therefore the Dutch infants were not able to recognize the strong-weak words in the texts. Interestingly, American infants, nine months of age, were also tested with the same Dutch materials (Houston, Jusczyk, Kuijpers, Coolen and Cutler, submitted). It was shown that the American infants were able to pick the Dutch words out of the Dutch sentences. The difference between the Dutch and the American infants is explained by the model that is provided by their native language input. American English might be a better model to derive a metrical segmentation strategy which the American infants also apply when listening to a foreign language.

Goodsitt, Morgan and Kuhl (1993) also investigated infants' segmentation abilities from a different perspective. They tested whether infants would cluster syllables on the basis of transitional probability. They used a discrimination maintenance technique. In this technique, infants are first trained to discriminate two syllables (e.g., /di/ and /te/, the contrast syllables). After that, the syllables are presented in the context of two or more syllables (e.g., /kogadi/ and /kogate/). The experimental question is whether the two target syllables are still distinguished.

In the experiment of Goodsitt et al., the context syllables were presented in such a way that they could either be clustered easily because the order was always the same (high transitional probability), or not easily because the order constantly varied (low transitional probability). Goodsitt et al. predicted that discrimination of the contrast syllables would be better when the context syllables could be grouped together easily than when the context could not be grouped easily. Infants, seven months of age, were tested in three consecutive sessions. In the first session, they were trained to discriminate the contrast syllables. In the following two sessions, they were tested on their ability to discriminate the contrast syllables in the context of two other syllables. The results showed that discrimination improved over the two test sessions only when the two context syllables always occurred in the same order. The

researchers concluded that infants were able to cluster the two context syllables on the basis of transitional probability. In other words: the infants segmented the strings of three syllables, resulting in two rather than three units.

Morgan (1994) subsequently tested whether rhythm also has an important contribution for the clustering of syllables, and, if so, whether transitional probability and rhythm are equally important. The experiments were based on two properties of a unit: *external individuality* and *internal coherence*. External individuality is the property that the behaviour of a unit is similar to the behaviour of individual segments; internal coherence is the property that the elements of a unit resist interruption. Two different experimental techniques were used to test these properties in two separate experiments.

In the first experiment, the discrimination maintenance technique was used. Eight-month-old infants were trained to discriminate the two syllables /di/ and /te/. After the training phase, these two contrast syllables were presented in the context of two other syllables (/ko/ and /ga/, the context syllables), resulting in strings of three syllables. The context syllables were either long or short. Infants were assigned to three conditions depending on the presentation of the context syllables. In the first condition, the context syllables were always presented in the same order, and always with a long first syllable followed by a short second syllable. In the second condition the order of the syllables was varied, but the first one was always long and the second one was always short. In the third condition, the context syllables were not ordered in any systematic way. The results of this study were that infants in the first condition and in the second condition had higher percentages of correct responses than infants in the third condition. Moreover, after repeated testing, infants in the first condition improved their performance significantly, but infants in the second condition did not. Thus, both transitional probability and rhythmic regularities were important for the grouping of the syllables, and the infants integrated the two different sources of information.

In the second experiment, a noise detection technique was used. Infants were first trained to respond to a click sound. This sound was presented within strings of three syllables similar to those of the first experiment. The clicks were presented either within the two context syllables or between the context syllables and the contrast syllable. The rationale behind this design was that responses to clicks within strings of two syllables that can be grouped together would be slower than responses to clicks that occur within syllables that are not grouped together. The results showed that the responses to clicks within syllables that could be grouped together because of their rhythmic and distributional predictability, were significantly slower than responses to clicks in the other contexts.

Finally, Morgan and Saffran (1995) extended the previous results. In a series of experiments, using the noise detection technique and the discrimination maintenance technique, it was observed that nine-month-olds use both sequential and rhythmic information, whereas six-month-olds use rhythmic information only. The performance of the nine-month-olds was significantly better when the context syllables were grouped on the basis of their sequential *and* rhythmic predictability than when the context syllables could only be grouped on the basis of their sequential predictability *or* their rhythmic predictability. The six months old infants, on the contrary, performed equally well in the conditions where the context syllables could be grouped on the basis of both sequential and rhythmic predictability or on the basis of rhythmic predictability only.

1.2.3 Summary

In sum, experimental evidence suggests that adult listeners use explicit segmentation strategies. These strategies depend on structural regularities in the native language, for example, metrical structure, phonotactic structure, or statistical regularities. Infants show signs of the ability to segment words out of context before they are nine months old. Like adults, they rely on various sources of information, including metrical structure, and transitional probability.

The development of the capacity to attend to structural regularities of the native language and using this sensitivity to organize the linguistic input into units (utterances, phrases, words) has been termed *prosodic bootstrapping*. This term was initially used because of infants' attention to prosodic structures that correlate with syntactic structures (Morgan, 1986). However, nowadays the term *bootstrapping from the signal* is preferred because more than just prosodic cues (e.g., phonological regularities or distributional regularities) are implicated in the search for cues to word boundaries (Morgan and Demuth, 1996).

1.3 Motivation for the Present Research

The goal of this thesis is to create a complete picture of the language input to an infant between six and nine months old. This age range was chosen because of the developmental changes that take place during this period. These changes are clearly 'input-driven'. In order to get a complete picture, an attempt was made to record all the language input to an infant of this age.

As was outlined in the beginning of this chapter, the main reason for carrying out this project was to obtain a better appreciation of the infant's

accomplishment of acquiring the necessary skills for later lexical development. After all, the issue whether language development, or, more generally, the human language capacity has an innate component or is merely learned behaviour remains an unsolved issue.

However, there were additional reasons for carrying out this project. The first concerned the nature of the data. The data used in previous studies on language input consisted mostly of parent-child interactions in either laboratory situations or controlled home situations (e.g., Snow and Ferguson, 1977; Gallaway and Richards, 1994). The data of the present study reflect various real-life situations, with more people present, where no constraints were put on interaction, and recorded over a longer period of time. The nature of the present data is thus different from that of the data used in previous studies. This provided the opportunity to see whether the results of previous studies could be extended to the present study.

A second reason for carrying out this project was that specific questions about the amount of language input could be addressed. Since all the input was recorded, it was possible to ask how much language the infant heard, and which proportion of the total input was directly addressed to the infant and which proportion of the input was addressed to others. The reader should realize at this point that the amount of work needed to estimate the amount of language via processing the entire period of three months was surely too much (at least for one person within the limited time of a Ph.D period).

A third reason for carrying out this project was that, although the period between six and nine months is a period during which children exhibit developmental changes in speech perception, systematic studies of language input to infants of this specific age range are lacking. The majority of studies on language input describe language spoken to children older than one year because these children have started producing words and, therefore, changes in the input can be related relatively easily to the language of the child. Relatively few studies focus on preverbal infants and most of those concern the suprasegmental structure of the input (e.g., Stern, Spieker and MacKain, 1982; Fernald and Simon, 1984), or structural aspects (e.g., Phillips, 1973; Snow 1977b).

1.4 Outline of the Thesis

In Chapter 2, the speakers, recording set-up, location, and the recording times are described. Further, the processing of the total collection of material is described. This included extracting spoken language from the recordings, and transcribing language that the infant heard. This chapter also gives an estimate

of the amount of language that the infant heard, and concludes with a description of the transcribed material.

In Chapter 3, five structural aspects of the language input are described. For this purpose, the transcribed material was used. Utterance length, vocabulary size, metrical structure, phonotactic structure, and word embedding were analyzed. These aspects were chosen because they are related to the word segmentation problem. Each aspect is described in language spoken to the infant, language spoken to older children, and language spoken to adults. The results in these three conditions are compared, and implications for word segmentation are discussed.

In the following chapter, three aspects of the suprasegmental structure of the input are examined: pitch, speech rate, and intonation. The question addressed in this chapter is whether the language that was addressed to the infant was produced in a so-called 'motherese' speech style, i.e., the speech register that mothers (but also other caretakers) adopt when speaking to children (Newport, 1977). The reason for analysing aspects of the suprasegmental structure was that these aspects play an important role in the early speech development. The aspects were analyzed in the same three addressee conditions as used in Chapter 3. The results are compared across the conditions, and the possible functions of modifications in the suprasegmental structure of infant-directed speech are discussed.

In Chapter 5, two experiments with adult listeners are described. These experiments model the infant's task of segmenting longer strings into smaller units and are therefore a supplementary contribution to the analyses of the corpus reported in Chapters 2 to 4. The experiments were run in collaboration with Michael Brent from Johns Hopkins University, Baltimore. They are based on his model of speech segmentation in the absence of lexical knowledge (Brent and Cartwright, 1996).

The final chapter summarizes and discusses the results of this project. Furthermore, a model of the development during the period between six and nine months is given, as well the factors that are likely to influence this development. The chapter concludes with some directions for future research.

Chapter 2
Methodology

In this chapter, a global picture of what the language input to the infant looked like is created. In the initial part of the project, we attempted to record all the language that the infant was exposed to during the period of three months. I describe the way in which this corpus was collected: The method of recording, the recording times, the main speakers, and the main situations. Next, the speaking time during a sample of 18 days taken from the total collection was calculated to estimate the overall amount of language input and to investigate how the input was distributed across different speakers and different addressees. The results showed that the total speaking time was 2.56 hours per day on average, of which 14% was language addressed to the infant. Finally, the transcription of part of the material is described.

2.0 Introduction

Many previous studies have examined language input to children (e.g., Snow and Ferguson, 1977; Gallaway and Richards, 1994). Collection of data in these studies normally involved making recordings in laboratory situations or controlled home situations. In these situations the child and one of the caretakers is present and is asked to interact with the child as normally as possible.

The data that served as the basis of the present research differed from that used in previous studies of language input. In the present study, recording continued non-stop for a period of three months, whenever the infant was awake. Thus, the language input was recorded in a variety of situations, and with a variety of people present, and not only language to the infant was recorded but also all the other language that the infant presumably heard.

The resulting corpus primarily offered the opportunity to create a complete picture of the total language input. After all, language input is not only the language that is addressed to the child but also contains language between the parents, to siblings, to other people, etc. In this chapter estimates are given of: *How much language input did the infant receive overall?* and *How much of the total input was directed to the infant, and how much other language did the infant hear?*

Furthermore, the corpus of the present study could be used to test whether claims about child-directed speech in previous studies could be extended to naturalistic settings as represented in the present data. These issues are not addressed in this chapter but in Chapters 3 and 4 in which a number of aspects are analyzed in speech to the infant and compared with speech to other addressees.

This chapter is divided into three main parts. In the first part (section 2.1), the collection of the data for the present study is described, including the situations in which the recordings were made, the participants in this study, and the times during which the recordings were done. In the second part (section 2.2), an estimate of total amount of input is given as measured in speaking time. In the third part (section 2.3), the transcriptions of a part of the input are described.

The chapter ends with a sample of the transcribed corpus (section 2.4). The sample is used to illustrate some choices that were made during the process of selection of the material, and the codes that were used in the transcriptions. References to the sample are in the text with the line numbers.

2.1 Corpus Collection

Equipment

A portable DAT-recorder (SONY walkman) was used for the recordings. The recorder was carried along in a small basket whenever the infant was moved from one place to another. A wired microphone was clipped to the infant's chair or hung over the edge of the basket (see Figure 2.1). The recorder had a built-in clock so that time and date of the recordings were also registered on the tapes. DAT-tapes of two hours each were used. This made it possible to record non-stop for four hours when the recorder was in a *long play* mode, so that the tapes had to be changed only once or twice a day.

The resulting quality of the recordings with this set-up was adequate for the purposes of this study. Most language spoken close to the infant was under-

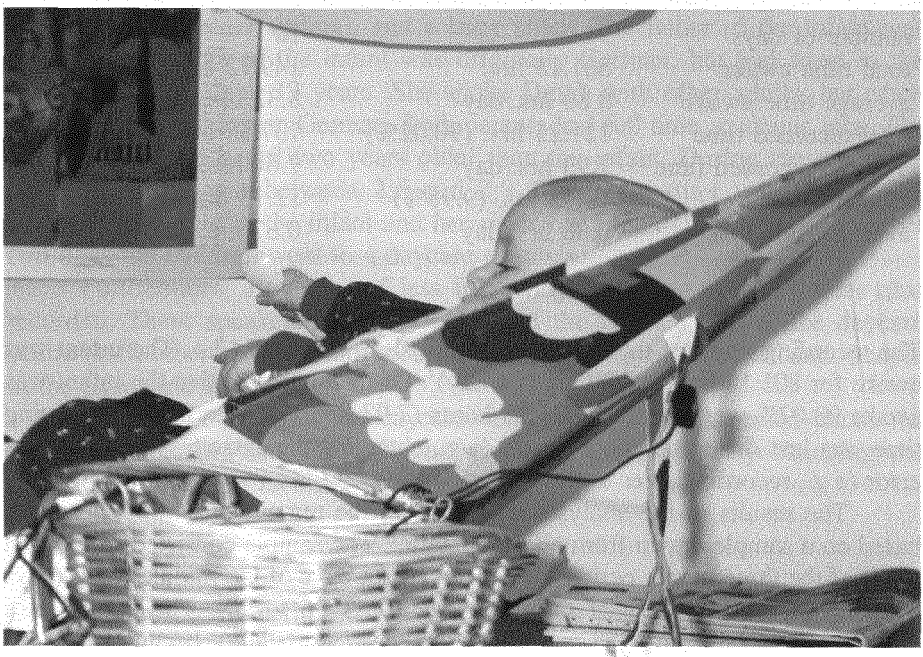


Figure 2.1: Recording set up. The infant is shown in her chair. The microphone is clipped to the chair and connected to the recorder in the basket.

standable and transcribable. However, the setting in which the recordings were made cannot, of course, be compared to an ideal situation such as a sound-proof studio. There are several other noise sources, such as the acoustics of the room, movements of the microphone, blowing wind outside, etc. Furthermore, there was great variability in how close the speakers were to the microphone. Usually, language directed to the infant was spoken at a relatively short distance, but language between adults was often produced at a greater distance. This made the infant-directed speech generally clearer than most of the other speech.

Recording Time

A total of 91 days was recorded. The infant was exactly six months old on the first day of the recording. The recorder was turned on when the infant woke up in the morning, and was turned off when she went to sleep in the evening or during the day. The exact times were written down in a diary by the mother.

Table 2.1: Recording times of the entire period and of the sample.

number of days	91	18
total time awake	801.18 hrs.	149.72 hrs.
average time awake	8.80 hrs./day	8.32 hrs./day
total recorded time	721.25 hrs. (90.02%)	137.58 hrs. (91.89%)
average recorded time	7.93 hrs./day	7.64 hrs./day
total time lost	79.93 hrs. (9.98%)	12.13 hrs. (8.10%)
average time lost	0.88 hrs./day	0.67 hrs./day

The recording times of the entire period are shown in Table 2.1. The infant was awake for 801.18 hours (8.80 hours a day)[1]. Of the total time that the infant was awake 90.02% was recorded: 721.25 hours (7.93 hours per day). The remaining time was lost due to practical causes (e.g., changing the cassettes) or accidental error (e.g., recording over a previous recording).

The results of the analyses described in the remainder of this thesis are based on a sample drawn from all the material. The sample consisted of the first week of the three-month period (days 1 to 7); six days of the middle week (days 39 to 44); and five days of the last week (days 86 to 90). The choice of a sample was made during the process of analysis (see below), because it turned out to be too time-consuming to process all the material. The time that the infant was awake, the recorded time, and the lost time in this sample are also given in Table 2.1.

Thus, about 20% of the total material was further processed and analyzed. The description of the participants and the situations below is based on the sample of 18 days. With few exceptions, the speakers and the situations in the intervening days were the same as those described for the selected sample.

Participants

The infant was the youngest daughter of a Dutch family. She showed no signs of health problems or hearing problems, and her language (though not measured with standardized tests) developed normally after the recording period. During the 18 selected days, she was not seriously ill but had to stay at home two days, instead of going to the daycare centre, because of a fever (days 3 and 42).

[1]Hours are given as decimal numbers. For example, 2.8 hours is two hours and 48 minutes.

The main speakers were: the other members of the family (the father, the mother, and an older sibling), and a baby sitter. The father of the infant was Dutch. The mother of the infant was originally German, but had lived in the Netherlands for about 12 years. She spoke Dutch with other adults, but often addressed her children in German. Both parents had full time academic jobs. The older sibling was a girl two years older than the infant. During the recording period she thus was between 2;6 and 2;9. The baby sitter was Dutch. She normally looked after the infant and her sister for three days a week. The other two days the children went to a daycare centre.

Furthermore, a number of less frequently occurring speakers were also recorded. These speakers included: people at the daycare centre (the daycare centre staff, other children, and their parents); the baby sitter's husband; people in the street and in shops.

The baby sitter and the daycare centre staff were informed about the purpose of the project. They all cooperated in the project by changing the cassettes, turning the recorder on and off, etc.

Situations

During the recording period, the mother and other adults kept a diary in which they wrote down which persons were around, at what times the infant went to sleep, and which activities took place, and so on. Based on this diary, three main situations could be distinguished during the 18 selected days:

1. The infant was at home with one or both parents and usually also the older sister. This was the normal situation in the weekends, and during the week in the early mornings and the evenings. About 97 hours (70%) were recorded in this situation, spread out over all of the 18 days in the sample.
2. The infant was with the baby sitter and the older sister. About 29 hours (21%) were recorded in this situation, spread out over seven days (days 4 to 6, 39 to 41, and 88).
3. The infant was at the daycare centre with the daycare centre staff and the other children. About 7 hours (5%) were recorded in this situation, spread out over 2 days (days 7 and 87). In selecting the material I avoided choosing this situation because the number of speakers and the nature of the situation made it very hard to analyze the spoken language.

Thus, most of the recordings were made in the infant's home situation. Except for these three main situations a few other situations also occurred during the 18 days. A number of times, the parents or the baby sitter took the children out to

go shopping or to go for a walk. This happened 13 times, usually not more than 30 minutes each time. In addition, the baby sitter and her husband paid two visits to the family during the 18 days, and the family paid one visit to them.

Summary

In sum, 90% of all the time that the infant was awake during a period of three months was successfully recorded. During this period, the infant was most often at home with one or two adults (her parents or a baby sitter), and usually also with her older sister. Most often, the language she was exposed to was Dutch, but a substantial part of the input she received was in German (see section 2.3.4 for an estimate of the proportion of the input that was in German).

The following section will describe how the language spoken during the 18 selected days was processed, in order to allow calculation of the total amount of language that the infant heard every day, and calculation of the proportion of the total language that was addressed to the infant or to others.

2.2 Speaking Time

Previous studies have focused on the amount of input children receive and how the amount of input affects language acquisition. For instance, Huttenlocher, Haight, Bryk, Seltzer and Lyons (1991) showed that the amount of words that mothers produced to their children explained a significant proportion of the variance that was observed in the rate with which the children acquired new words between the age of 16 and 26 months. Similarly, Barnes, Gutfreund, Satterly and Wells (1983) found that children's gain scores were positively correlated with the amount of input that they received in a number of selected language samples during a period of nine months. However, as in many other studies of language input, the focus of Huttenlocher et al.'s and Barnes et al.'s studies was only the language spoken to the children.

In this section, I aim to give a more global estimate of the overall amount of language input that the child received than has been done in previous studies. That is, not only the amount of language spoken to the infant, but also the amount of all the other language that the infant heard will be estimated.

The overall amount of language input has not been estimated in any study that I am aware of. Normally, only the language spoken to the infant is examined. We know, however, that there are cultural differences in how much adults speak to children. In modern western cultures adults talk to infants and to young children, but this is not necessarily the case in other cultures, as for

example the Kaluli in Papua New Guinea (e.g., Schieffelin, 1985) or the Tzeltal in Mexico (Brown, 1996). In these cultures, adults do not typically talk to prelingual infants because they believe that there is no point in speaking to someone who cannot understand. Thus, in these cultures, almost all the language input that children receive is between mature language users (adults to adults, adults to older children).

In modern western cultures, it is likely that a large part of the input to infants or children is not directed to the child but is between mature language users (as was, for instance, suggested by Cutler, 1994). It is unclear, however, which proportion of the total input is language addressed to the infant, and how much 'other' language the input consists of.

The purpose of this section is to give estimates of the total amount of input and of the proportion of the total input that was language addressed to the infant and language addressed to others. For this purpose, the language that was spoken during the 18 selected days was extracted from the tapes and classified as to speaker and addressee, as will be described below. The measure that was chosen to estimate the amount of input was the speaking time (the time in which the subject speaks). This is a crude measure that is not often used (an exception is Wagner, 1985, who used speaking time to estimate how much a child says in a day). The amount of total speaking time, as well as the proportion of total speaking time in different addressee conditions were compared across the three weeks (first, middle, last).

2.2.1 Methodology

All the language that was spoken during the 18 days was extracted from the tapes. This was done in the following way. Consecutive chunks of 30 minutes of the original recording were converted to a readable signal format with a sample frequency of 16 kHz. This signal was visualized on the computer screen with a speech editor (the ESPS-waves speech editing program). On average, chunks of ten seconds were displayed at a time, although in some instances a shorter section was displayed to enable greater precision in placing the marks.

Markers were set at the beginning and the end of stretches of spoken language with the speech editor on the basis of the visual information on the screen and the auditory information through the headphones[2].

[2]The markers were set as close to the signal as possible. However, given the length of the signal that was displayed at the time, it was not possible to be very precise in marking the beginning and end of a stretch of speech. A small margin at the beginning and the end

Table 2.2: Classification of spoken language as to speaker and addressee.

Label	Meaning
AA	Adult to Adult
AI	Adult to Infant
AC	Adult to Child
CA	Child to Adult
CI	Child to Infant
CC	Child to Child
VN	'Vocal Noise'

The marked stretches of speech were labelled as to speaker and addressee, using the classification displayed in Table 2.2. In the fragment at the end of this chapter, the labels are displayed, followed by the speaker and the utterances that were produced in this segment.

Adult referred to any of the adults described above: the parents, the baby sitter, etc. *Child* referred usually to the older sister, but it was also used for the children at the day care centre, children in the street, and so on. The label *vocal noise* was used when I could not make a satisfactory choice out of one of the other labels, e.g., when it was not clear whether the speaker was a child or an adult, or when it was not clear to whom the utterance was addressed, or when more persons were talking at the same time to different addressees. It was also occasionally used for television conversation. Note that this label was not used when I could not hear what exactly was said, but when it was clear who the speaker and the addressee were.

Choosing one label or another was usually not difficult, because it was clear which participants were present, which participant spoke, and to which participant the utterance was addressed. However, specific choices were made in the following situations:

1. In normal everyday speech there is considerable overlap between speakers. This posed a problem in choosing a label when the overlapping utterances belonged to two different categories. However, sometimes, one label was

---

of each segment that contained no speech was usually left over. The duration of these margins was estimated based on a sample of 5250 labelled segments. These segments were those that were used for the calculation of speech rate, described in Chapter 4, section 4.4. The average duration of the two margins together in this sample was 0.243 seconds. This value was subtracted from every segment when total speaking time was calculated in order to give a more precise estimate.

more appropriate than another label: for instance, when one of the utterances was much longer than the other one, or when one of the utterances was in the foreground and the other one was in the background. In these cases, the label appropriate for the long utterance, or the one in the foreground was selected. If the two utterances were more or less equal, the label VN was selected. Examples of instances where another speaker produced an utterance simultaneously can be found on lines 25, 59, 92, 149 and 163 in the fragment in section 2.4. The utterances in the background are in italics.

2. In some situations, e.g., outside in the street, it was impossible to know who was speaking or who the addressee was, but sometimes one option was highly likely because of the content of the utterance, or the voice of the speaker. In these instances, this likely option was chosen. In other instances, the label VN was chosen.
3. Sometimes people moved away from the room where the infant was, e.g., after dinner when the parents went to the kitchen for cleaning up and the infant stayed in the living room. Spoken language was audible but not understandable. This meant that it was not possible to select a label on the basis of the content of the utterance. However, the reaction of the addressee could make it clear which label to select.
4. Sometimes adults addressed other adults in the presence of a child, in a way as if they were speaking to the child. For example: when the mother left in the morning, the baby sitter would say *dag mama*, ('bye mama'), to stimulate the child to say the same. These utterances could have been labelled either AA or AC. In these instances, the label AC was chosen. An example can be found on lines 203 and 209-213. Based on the content, these utterances would have been labelled AA, but the manner with which the utterances were produced made it clear that they were intended to be child-directed utterances.
5. In a few cases, people talked to themselves. Depending on the person who was in the room I labelled this other person as the addressee. So for example, when the child was playing in the room and an adult was reading the newspaper, and the child would say something to herself, that was labelled CA.

Most of these difficult cases, however, were incidentally occurring utterances. In the majority of cases, it was clear who the speaker and the addressee were.

Finally, the labelled segments were edited out of the original files with a special labelling program and were stored on datatapes. The speaking time was calculated as the sum of the durations of the labelled segments in each category.

2.2.2 Results

A total of 85068 stretches of speech were marked. Of the total, 9096 were labelled as AA, 29780 as AC, 13024 as AI, 31110 as CA, 396 as CC, 669 as CI, and 993 as VN. Table 2.3 is an overview per day of the recorded time, the total speaking time (the sum of the durations of all the labelled segments) with the percentage of the recorded time that was speaking time, and the percentage of speaking time in each category. The day number is in the first column, the recorded time in the second. The third column lists the amount of speaking time in hours. The fourth column lists the percentages of recorded time that was speaking time. The columns five to eleven list the percentages of the total speaking time in each of the seven categories. The average values are given at the bottom of the table.

On average, the total speaking time was 2.56 hours per day, which was somewhat less than 34% of the total recorded time. The average percentage was 34% during the first week, 31% during the middle week, and 38% during the last week.

Table 2.3: Overview of speaking times.

Day	Rec. Time (hrs)	Speaking Time (hrs)	(%)	AA	AC	AI	CA	CC	CI	VN
1	9.27	2.83	(30.5)	25.5	18.9	23.5	19.9	0.4	2.6	9.3
2	7.37	2.22	(30.1)	19.7	27.1	15.0	37.1	--	0.2	1.0
3	6.37	1.55	(24.3)	26.2	24.7	21.3	27.4	--	0.5	--
4	8.30	3.20	(38.6)	14.7	36.6	12.0	35.7	0.2	0.9	--
5	8.10	3.08	(38.0)	14.4	41.6	18.8	24.4	0.3	0.4	--
6	8.02	3.11	(38.8)	19.5	37.7	11.1	29.8	0.1	0.7	1.0
7	5.57	2.21	(39.7)	21.3	39.2	10.3	24.6	1.2	0.1	3.3
39	6.50	2.46	(37.8)	23.4	34.2	13.6	27.6	--	1.1	--
40	4.45	1.61	(36.2)	3.6	34.4	16.6	44.8	--	0.5	--
41	6.72	2.32	(34.5)	7.6	43.7	12.9	34.0	--	1.9	--
42	8.43	1.72	(20.4)	30.0	15.7	35.4	18.2	--	0.0	0.5
43	9.23	2.02	(21.9)	17.4	28.0	18.1	30.4	--	0.4	5.8
44	9.22	3.46	(37.5)	29.7	24.0	8.0	31.9	0.1	0.3	6.1
86	7.15	2.26	(31.6)	27.9	25.3	3.7	35.6	--	3.1	4.3
87	8.15	4.38	(53.7)	14.3	17.4	3.2	19.2	8.1	0.2	37.5
88	8.17	2.87	(35.1)	17.7	33.1	13.3	32.2	--	1.3	2.3
89	7.77	2.22	(28.6)	17.3	18.9	4.1	34.9	--	0.7	24.1
90	8.82	2.63	(29.8)	15.3	31.2	7.4	44.8	0.0	0.7	0.4
	7.65	2.56	(33.7)	19.2	29.6	13.8	30.7	1.3	0.9	8.0

A one-way analysis of variance showed that these percentages were not significantly different: $F[2,15] = .8948$, $p > .05$. Thus, the total amount of spoken language was roughly constant per week. Furthermore, about 30% of the total speaking time was in the AC category (0.87 hours); 31% was in the CA category (0.89 hours); 18% (0.52 hours) was in the AA category, and 14% (0.38 hours) was in the AI category. The remaining 7% (0.20 hours) was distributed over the other three categories. The average proportions are illustrated in Figure 2.2 below.

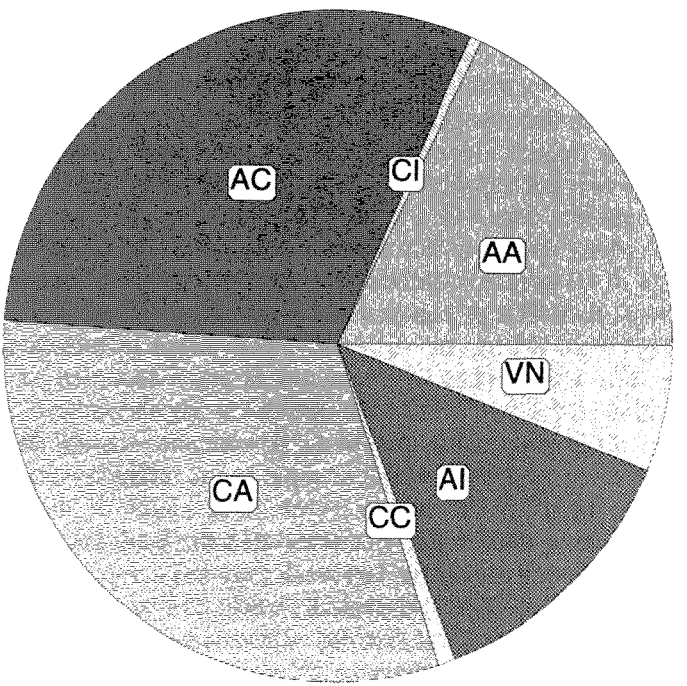


Figure 2.2: Distribution of speaking time. Each portion of the graph represents the average percentage of the total speaking time in a category.

The percentages of speaking time in each category were compared across the three weeks with separate analyses of variance. There were significant differences in the percentages of AI speech. In the first week the average percentage was 16%, in the middle week it was 18%, but in the last week it was only 6%. The differences were significant: $F[2,15] = 4.26$, $p < .05$. Thus, in the last week, the relative amount of input to the infant was less than in the first or the middle week. None of the percentages in the other categories showed a similar pattern: AA: $F[2,15] = 0.09$, $p > .05$; AC: $F[2,15] = 0.99$, $p > .05$; CA: $F[2,15] = 0.58$, $p > .05$; CC: $F[2,15] = 1.57$, $p > .05$; CI: $F[2,15] = 0.54$, $p > .05$; VN: $F[2,15] = 1.13$, $p > .05$.

2.2.3 Discussion

The results showed that there was about 2.56 hours of speaking time per day. This amount was fairly constant over the 18 days. About 30% of the total speaking time was from adult to child, and another 30% was from child to adult. About 18% was from adult to adult, and 14% was from adult to infant.

Thus, the largest part of the speaking time was between adult and child. Most of the speech in this category was conversation between the parents or the baby sitter with the infant's elder sister. The low percentage of speaking time in the CC category is a consequence of the fact that most of the time there was only one child present (except for the daycare centre days, of course, but in that situation, much of the speech in this category was lost in the category VN). The low percentage of CI speech indicates that the elder child did not address her baby sister very often. The percentage in the category VN was generally low, except in a few days. During those days the percentage in this category rose because there was a relatively large amount of input recorded in situations outside the house: in the daycare centre, outside during shopping.

The percentage of speaking time in the AI category was not consistent. In the last week of the period, the amount of input to the infant was much smaller than in the first week or the middle week. One cause of this decline in percentage of speaking time to the infant was that the baby sitter took care of the children only one day during the last week (day 88) compared to three days in the first week and three days in the middle week. The percentage of speaking time in the AI category during that one day was relatively high compared to the other days of the last week. However, the percentages of speaking time addressed to the infant during the other days of the last week were still relatively low compared to similar days in the first and the middle week. Thus, the parents spoke less to the infant in the last week than in the first week or the middle

week. Possibly, practical reasons played a role in this change. For instance, the parents may have been more busy than usual that week. There were no significant changes in any of the other categories.

In conclusion, we can say that, based on the present data, the infant was exposed to a considerable amount of speech that was not directed to her. Expressed as a percentage of the total speaking time, only 14% was language addressed to the infant. This is consistent with Cutler's (1994) hypothesis that much of the language that infants hear is among more mature language users.

Of course, many factors influence the composition of the input: whether there are siblings, whether the parents work, whether there is usually more than one adult present, etc. Therefore, the results can only be interpreted in the light of the situation in which the recordings were made.

2.3 Transcriptions

The language spoken in the AI category, the AC category, and the AA category was transcribed orthographically. A considerable proportion of the spoken language in the AI and the AC categories was in German. The transcriptions of the German utterances were checked by the mother of the infant and corrected where necessary.

The transcriptions were formatted according to CHILDES conventions (MacWhinney, 1995) and several additional codes were added to the transcripts (see Appendix A for an overview of these codes). The advantage of formatting transcriptions in this way was that a number of tools could be used for the analysis of the transcriptions (CLAN tools), for instance, to calculate mean length of utterance, or to count the frequency with which words occurred. Furthermore, the transcriptions can be made public eventually as part of the CHILDES database. In the following subsection, the format of the transcriptions is described. The line numbers refer to the example in section 2.4.

2.3.1 Format of the Transcriptions

According to CHILDES conventions, each utterance has to be on a separate line, the main tier. The main tier starts with a three-letter speaker identification, and ends with an utterance delimiter. Utterances were distinguished on the basis of sentence structure and suprasegmental structure. Examples of utterances are: complete sentences (line 10); incomplete sentences (line 47); phrases that were delimited by prosodic marking such as pauses, intonational movements (line 31)

or two or more sentences connected by one or more conjunctions (line 7). Run on sentences (with no break in between) were treated as separate utterances.

There were three main speakers: the mother, the father, and the baby sitter. These three speakers each had their own speaker identification in the transcriptions. Father: **FAT*, baby sitter: **BBS*, mother: **MOD* (when the utterance was in Dutch), **MOG* (when the utterance was in German), **MOT* (when the utterance could not be classified as either German or Dutch, as was the case in short utterances like *ja*, *oh*, *zo*, *hm*, etc., or when she only used the infant's or the child's name, or for *baby talk* (see below) utterances). If an utterance could be either Dutch or German because of the overlap between the languages, e.g., *hier*, (here), it was assumed that the utterance was in German. The mother spoke only Dutch to other adults, and in the adult-directed speech she was always coded as **MOT*. Furthermore, the baby sitter's husband who appeared regularly on the recordings had a separate speaker identification: **MEN*. All other speakers were coded as **OTH*.

Words that were not understandable were coded as *xxx* (line 18). The number of words in every utterance was added between squared brackets at the end of every utterance. Compounds, or combinations of words that in normal written language are written with a space between them, were joined by a plus sign (e.g., *'s+avonds* 'in the evening', *T+shirt*) and were counted as one word in the calculation of utterance length. When the number of words could not be determined, the utterance length was set to a missing value (lines 18, 92).

Especially in the adult-directed speech, utterances were often interrupted by an utterance from a second speaker. In these cases, the first speaker's utterance was divided over two lines, separated by the utterance of the second speaker. The end of the first part of the interrupted utterance was coded as +..., and the beginning of the last part of the interrupted utterance was coded as +,. An interrupted utterance was counted as one utterance. For example, lines 4 and 6 were interrupted by line 5 from the second speaker.

Also, especially in the adult-directed speech, the speaker sometimes restarted his or her own utterance or continued it in a different way. A restart was coded as *[\\]*. Utterances with a restart were also treated as one utterance. Examples of utterances with restarts are on lines 6, 34, 102, 170.

There were also a few instances where the speaker used words from other foreign languages, or where both German and Dutch words were used in the same sentence. In these cases the foreign words were tagged with a @ followed by the first letter of the language that the word was taken from (e.g., *dat kwam op BBC@e*, 'that was on BBC'). Some non-words (e.g., when the speaker stopped speaking in the middle of a word) were coded as *@x* (e.g., *hon@x [\\] honderden keren*, 'hun@x [\\] hundreds of times').

Additional information was coded in so-called 'constant headers' throughout the files: the date of the recording (*@Date:*); the sound file that the transcripts were taken from (*@tape:*). See, for example, the beginning of the corpus fragment.

Finally, a so-called 'local depfile' was created. This is a file containing all the relevant information about the specific codes and symbols that were used in the transcripts. The depfile is mainly necessary in order to be able to use the CLAN tools for analysing the transcripts. In the depfile a number of codes that were specific for the present transcription were defined. The local depfile is given in Appendix D.

2.3.2 Classification of Utterance Types

Every utterance was classified as to its structure and content. The main reason for classifying the utterances was that previous studies have reported that the distribution of utterance types in child-directed speech is different from that in adult-directed speech. For instance, Snow et al. (1976) found that a relatively high proportion of utterances in a sample of language directed to two-year-old children were interrogatives. The distribution of utterance types in the present study will be described in Chapter 3. Furthermore, the classification was useful for some of the analyses reported in Chapter 3. *Baby talk* utterances, for instance, could easily be excluded from the calculation of vocabulary size.

In the transcriptions, the code for utterance type was added to the main tier: the utterance was placed between angled brackets, followed by the type code between squared brackets. In some cases one utterance consisted of a combination of types. The following classification was used: declaratives ([DEC]), interrogatives ([QUE]), imperatives ([IMP]), fillers ([FIL]), tags ([TAG]), social expressions ([SOC]), vocatives ([VOC]), routines ([ROU]), baby talk ([BBT]).

A *declarative* was an utterance with a declarative word order (line 21) or a declarative intonation (for instance in the case of a single word utterance, line 109). An *interrogative* was an utterance with an interrogative word order (line 84) or an interrogative intonation (lines 61, 88). An *imperative* was a subjectless utterance with an imperative verb (line 104) or an infinitive used as imperative (line 151). *Fillers* were discourse markers, interjections, small words that 'keep the conversation going' (lines 5, 80, 85, 152). *Tags* were small words at the ends of longer utterances, usually with a rising intonation (lines 132, 128, 153). Note, that declarative utterances followed by tags were still coded as declaratives (line 118) although these combinations usually sound like questions. Utterances with

a declarative word order and an interrogative intonation, on the other hand, were classified as questions. *Social expressions* included greetings, excuses, warnings (lines 67, 76, 112). *Vocatives* were ways to address the listener by the proper name or otherwise (line 83). *Baby talk* utterances were contentless talk as imitations of the baby's own productions (e.g., *ta ta ta*, etc.) or utterances that are typically produced to babies (line 223). *Routines* included nursery rhymes, songs, story telling.

Several combinations of these types occurred, for instance declaratives plus tags (line 231), interrogatives or imperatives followed by a vocative (lines 142, 225), social expressions followed by vocatives, etc. Combinations of declaratives, interrogatives, or imperatives that were produced without a break were usually separated in the transcriptions, except when the two utterances were connected by a conjunction. In some cases, however, the combination was not separated in the transcription, for example in the case were a short imperative was followed by a declarative (e.g., *kom, dan gaan we eten*). However, some instances of these combinations did occur. A number of utterances could not be classified because the content of the utterance could not be transcribed reliably. These utterances were coded as [X] (line 11).

2.3.3 Phonological Transcriptions

The orthographic transcriptions were converted to phonological transcriptions. The conversion was done partly automatically and partly by hand. The words that occurred in the original transcripts were looked up in the online CELEX database (Centre for Lexical Information). This database contains orthographic word forms, phonological transcriptions, and information about word frequency and syllable structure. This information is available for English, Dutch and German (CELEX, 1993, 1990, 1995). The phonological transcriptions are in computer phonetic codes (DISC characters which assign a single unique code to each phonetic segment in the standard sound systems of Dutch, English and German, see: Burnage, 1990). These CELEX transcriptions were used for the conversion. A number of words that did not occur in this database were transcribed manually. Many of these words included proper names (e.g., *Schiphol*, 'name of the Dutch national airport', *NRC*, 'name of a Dutch newspaper'), and several words that are typical for spoken language (e.g., *oh, ah, au,* etc.). Both Dutch and German utterances were converted in this way. The phonological transcriptions were stored in separate files, which were in the same format and contained the same codes as the files with the orthographic transcriptions.

2.3.4 Description of the Transcribed Corpus

Table 2.4 shows the total number of utterances per speaker in each condition. A total of 16242 utterances (41611 words) in the AI condition were transcribed, 43772 in the AC condition (130324 words), and 21379 in the AA condition (89354 words). Thus, the sample of AC speech was the largest, the sample of AI speech was the smallest. The number of utterances produced by the main speakers and the other speakers are also displayed. 57.1% of the AI utterances were produced by the mother, 9.7% by the father, 30.7% by the baby sitter, 2.5% by other speakers. 41.7% of the utterances in the AC condition were produced by the mother, 17.6% by the father, 35.6% by the baby sitter, 5.0 by other speakers. Finally, 45,5% of the utterances in the AA condition were produced by the mother, 25.1% by the father, 16.4% by the baby sitter, and 13.1% by other speakers. Across the three conditions, most of the utterances were produced by the mother. In the AI and the AC conditions, the baby sitter produced most of the utterances, after the mother. In the AA condition the father produced most of the utterances after the mother. Most of the speech in the AA condition was between the parents. The baby sitter was alone with the children most of the time. 3913 utterances in the AI condition were true German utterances (24%). 10727 utterances in the AC condition were true German utterances (25%).

Table 2.4: Number of utterances per speaker. Percentages are given between parentheses. The total corpus size is displayed in the bottom row.

	AI		AC		AA	
Mother	9282	(57.1)	18264	(41.7)	9695	(45.3)
Father	1573	(9.7)	7694	(17.6)	5358	(25.1)
Baby Sitter	4989	(30.7)	15611	(35.7)	3516	(16.4)
Other	398	(2.5)	2203	(5.0)	2810	(13.1)
Nr of utterances	16242		43772		21379	
Nr of words	41611		130324		89354	

2.3.5 Summary

The total transcribed corpus consisted of 81393 utterances. There were 16242 utterances from adults to the infant (19.8%), 43772 from adults to other children (53.4%), and 21379 from adults to adults (26.7%). Given the 138 hours of

recorded time, this amounts to approximately 590 utterances of input per hour, of which about 118 were to the infant. Furthermore, the corpus consisted of 261289 words. This amounts to about 1890 words per hour, of which about 300 to the infant. Of course, the entire input was much bigger, because only a part of everything was transcribed.

Most of the transcribed utterances were produced by three main speakers: the mother, the father, and the baby sitter. Of these three, most of the utterances were produced by the mother. About 24% of the speech in the AI category was in German, and about 25% of the speech in the AC category was in German. There was no German in the AA category.

2.4 Fragment of the Corpus

```
        @Begin
        @Participants: MOG Mother (German), MOD Mother (Dutch), MOT mother, FAT father,
                       BBS babysitter, CHI child
        @tape  t940z.txt
        @Date: March 23, 1994 (Wednesday). Second day, middle week, day nr. 40.
   AA1  *BBS: < toen had ik nog dat nummer op willen zoeken maar ik denk nou dan doe ik
                dat vanavond wel > [DEC] . [19]
   AA2  *MOT: < xxx > [X] . [0]
        *BBS: < dat sofinummer want > [DEC] +...
 5      *MOT: < ja > [FIL] . [1]
        *BBS. +, < ik wist [\\] ja Menno die weet precies waar 't ligt > [DEC]   [13]
        *BBS: < ik denk dan moet ik gaan zoeken en eh kom ik helemaal te laat > [DEC]
                [14]
   CA1  *CHI: < ikke ook > [DEC] . [2]
10 AC1  *BBS: < moet je ook 'n zakdoek > [QUE] ? [5]
        *MOD: < xxx morgen > [X] . [0]
        *BBS: < heb je ook 'n vieze neus > [QUE] ? [6]
   AC2  *BBS: < he > [FIL] ? [1]
   AC3  *MOT  < ja > [FIL] . [1]
15 AA3  *MOT: < misschien is die wel 't zelfde voor jullie allemaal denk ik > [DEC] .
                [12]
        *MOT. < of niet > [QUE] ? [2]
        *MOT: < xxx > [X] . [0]
   AA4  *BBS: < ik denk dat we ieder 'n eigen nummer hebben > [DEC]   [9]
20      *MOT: < ja > [FIL] ? [1]
        *BBS: < ja > [FIL] < ik denk 't wel > [DEC] . [5]
   AI1  *BBS: < och > [FIL] < wou je ook 'n zakdoek > [QUE] ? [6]
        *BBS. < hoe is 't nou met jou > [QUE] ? [6]
        *BBS: < is 't beter > [QUE] ? [3]
25      *MOD: < ja > [FIL] < is goed > [DEC]   [3]
   AA5  *MOT: < xxx > [X] . [0]
   AA6  *BBS  < nee > [FIL]   [1]
        *BBS. < ze heeft zo geslapen gister > [DEC] < he > [TAG]   [6]
   AA7  *MOT: < ik zag 't > [DEC] . [3]
30      *MOT  < ik dacht nou hoe kan dat er maar een bandje is > [DEC]   [11]
        *BBS: < een bandje > [DEC] . [2]
        *BBS  < ja nou > [FIL] . [2]
        *MOT: < xxx > [X]   [0]
        *BBS  < ze werd om eh drie uur [\\] bij drie uur wakker en kwart over vier lag
35              ze er al weer in > [DEC]   [20]
        *MOT  < oh > [FIL] < xxx eentje > [QUE] < he > [TAG] ? [0]
   AA8  *BBS: < zo > [FIL] . [1]
        *MOT. < eentje > [DEC] . [1]
        *BBS: < een > [DEC] . [1]
40      *BBS: < goed zo > [DEC] . [2]
        *BBS. < doen we die d'r weer in > [DEC] < he > [TAG] . [7]
        *BBS  < voor straks > [DEC] . [2]
```

```
        *MOT: < ja > [FIL] . [1]
    AA9 *BBS: < en om kwart over vier lag ze er al weer in > [DEC] . [11]
 45     *BBS: < en gelijk duim en ze was vertrokken > [DEC] . [7]
   AA10 *BBS: < zo moe was ze gister > [DEC] . [5]
        *MOT: < ze heeft echt eh > [DEC] . [4]
        *BBS: < en hangerig en > [DEC] . [3]
        *MOT: < ja > [FIL] . [1]
 50     *BBS: < niks vond ze leuk en > [DEC] . [5]
        *MOT: < nee > [FIL] . [1]
   AI2  *BBS: < he > [FIL] ? [1]
   AI3  *BBS: < alleen banaantjes eten > [DEC] . [3]
        *BBS: < dat vond jij leuk > [DEC] < he > [TAG] ? [5]
 55     *CHI: < xxx > [X] . [0]
   AA11 *MOT: < gegeten heeft ze toch wel > [DEC] . [5]
        *BBS: < ja > [FIL] . [1]
        *BBS: < 'n grote fles melk nog gedronken > [DEC] . [6]
        *CHI: < xxx > [X] . [0]
 60 AI4  *BBS: < he > [FIL] . [1]
   AI5  *BBS: < vieze neus > [QUE] ? [2]
        *BBS: < nou is hij niet meer vies > [DEC] < he > [TAG] . [7]
   AI6  *BBS: < nou heb je 'm gepoetst > [DEC] . [5]
        *MOT: < ja > [FIL] . [1]
 65 AA12 *MOT: < ja > [FIL] < ik weet niet > [DEC] . [4]
        *MOT: < ik hoop dat 't weer beter gaat vandaag > [DEC] . [8]
   AA13 *FAT: < goedemorgen > [SOC] . [1]
   AA14 *BBS: < ah > [FIL] < goedemorgen > [SOC] . [2]
        *FAT: < wat zeg je > [QUE] ? [3]
 70 AA15 *MOT: < ik zei tegen An dat ik hoopte dat 't beter was met Josje vandaag > [DEC]
                . [14]
   AA16 *MOT: < die was zo slaperig en hangerig gisteren > [DEC] . [7]
   AA17 *BBS: < gister > [DEC] . [1]
        *BBS: < die heeft zoveel geslapen > [DEC] < he > [TAG] . [5]
 75     *BBS: < zo echt hangerig > [DEC] . [3]
   AI7  *MOT: < foei foei foei > [SOC] . [3]
   AI8  *MOT: < foei foei foei > [SOC] . [3]
   AI9  *BBS: < da's lekker > [DEC] . [2]
   AI10 *BBS: < da's lekker > [DEC] . [2]
 80 AI11 *BBS: < zo > [FIL] . [1]
   AI12 *MOG: < da > [DEC] . [1]
   AI13 *BBS: < zo > [FIL] . [1]
   AI14 *MOD: < vreetzakje > [VOC] . [1]
   AC4  *BBS: < en heb jij ook goed geslapen > [QUE] ? [6]
 85 AA18 *MOT: < jawel > [FIL] . [1]
   AC5  *BBS: < heb jij ook goed geslapen > [QUE] ? [5]
   AA19 *MOT: < twaalf uur moest ze 'n boterham met leverworst eten > [DEC] . [9]
   AC6  *BBS: < twaalf uur 'n boterham eten > [QUE] ? [5]
   AA20 *MOT: < ja > [FIL] . [1]
 90 AC7  *BBS: < hoe kan dat dan > [QUE] ? [4]
   AA21 *FAT: < oh > [FIL] < maar d'r is geen > [DEC] . [5]
        *BBS: < xxx honger > [DEC] . [0]
        *FAT: < d'r is niet voldoende brood meer om te lunchen > [DEC] . [9]
   AA22 *BBS: < oh > [FIL] < dan haal ik wel eh > [DEC] . [6]
 95 AA23 *FAT: < wil je 'n pakje zuurkool meenemen > [QUE] ? [6]
   AA24 *BBS: < ja > [FIL] . [1]
        *BBS: < hoeveel > [QUE] ? [1]
        *FAT: < 'n pond > [DEC] . [2]
        *BBS: < oh > [FIL] < zo'n [\\] zo'n pak > [QUE] ? [4]
100 AA25 *FAT: < op is op > [DEC] . [3]
        *BBS: < of van die verse > [QUE] ? [4]
   AA26 *MOT: < hebben ze die [\\] hebben ze die vers dan > [QUE] ? [8]
   AA27 *BBS: < ja > [FIL] < bij de groenteman > [DEC] < he > [TAG] . [5]
   AA28 *MOT: < doe maar verse dan > [IMP] . [4]
105 AA29 *BBS: < uit 'n pak > [DEC] . [3]
   AA30 *FAT: < ja > [FIL] . [1]
        *MOT: < ja > [FIL] . [1]
   AA31 *BBS: < goed > [DEC] . [1]
   AA32 *MOT: < lekker > [DEC] . [1]
110 AA33 *BBS: < zuurkool > [DEC] . [1]
        *BBS: < brood > [DEC] . [1]
   AA34 *FAT: < hai > [SOC] . [1]
        *MOT: < brood > [DEC] . [1]
        *BBS: < melk denk ik ook > [DEC] < he > [TAG] . [5]
```

```
115       *FAT: < tot ziens > [SOC] < he > [TAG] . [3]
          *MOT: < melk > [DEC] . [1]
     AA35 *MOT: < staat nog 'n half > [DEC]   [4]
     AA36 *BBS: < zullen we even kijken > [DEC] < he > [TAG] . [5]
     AA37 *MOT: < xxx > [X] . [0]
120  AA38 *BBS: < oh > [FIL] < die heb je nog wel > [DEC] . [6]
          *MOT: < die staat er nog 'n half > [DEC] . [6]
     AA39 *BBS: < zo > [FIL] . [1]
     AA40 *MOT: < ja > [FIL] . [1]
     AC8  *BBS. < xxx jongedames > [QUE] ? [0]
125  AC9  *BBS. < doe je even naar papa zwaaien > [QUE] ? [6]
     AC10 *BBS: < moet je niet even naar papa zwaaien > [QUE] ? [7]
     CA2  *CHI: < xxx > [X] . [0]
     AC11 *BBS: < hm > [FIL] ? [1]
     AC12 *BBS: < moet je niet even naar papa zwaaien > [QUE] ? [7]
130  CA3  *CHI: < zak > [DEC] . [1]
     AC13 *BBS: < zakken > [QUE] ? [1]
          *BBS: < je hebt geen zak nou > [DEC] < he > [TAG] ? [6]
     AC14 *BBS  < moet die tussen 't elastiek > [QUE] ? [5]
     AC15 *BBS: < zo > [FIL]   [1]
135  AC16 *BBS: < tussen 't elastiek stoppen we 'm > [DEC] . [6]
     AC17 *BBS: < zo > [FIL]   [1]
          *BBS: < 'n eindje tussen > [DEC] . [3]
          *BBS: < dan blijft hij zitten > [DEC] . [4]
          *BBS: < zie je dat > [QUE] ? [3]
140  AC18 *BBS: < tussen 't elastiek he doen we de zakdoek > [DEC] . [8]
     AC19 *BBS: < ja > [FIL]   [1]
     AC20 *BBS: < en heb je nu ook goed gegeten > [QUE] < Bente > [VOC] ? [8]
     AA41 *MOT: < nou > [FIL] < niet zo bijzonder > [DEC] . [4]
     AC21 *BBS: < nee > [FIL] ? [1]
145  AC22 *BBS: < doe je vanmiddag weer veel boterhammen eten > [QUE] ? [7]
          *BBS: < he > [FIL] ? [1]
     CA4  *CHI. < xxx jam > [DEC]   [0]
     AC23 *BBS: < je thee opdrinken > [IMP] . [3]
          *CHI: < xxx > [X] . [0]
150  AC24 *BBS: < zeg > [IMP]   [1]
          *BBS: < en je thee opdrinken > [IMP] . [4]
     AC25 *BBS: < hee > [FIL] . [1]
     AA42 *BBS: < An had de broek gister geknipt helemaal > [DEC] < he > [TAG]   [8]
          *MOT: < mooi > [DEC]   [1]
155       *BBS: < ze had 'm d'r afgehaald maar ze had geen naaimachine > [DEC] < he > [TAG]
                . [11]
     AA43 *BBS: < nee > [FIL] < jij hebt die oude naaimachine niet meer > [DEC] < he >
                [TAG] ? [9]
     AA44 *MOT: < nee > [FIL] . [1]
160  AA45 *BBS: < nee > [FIL] . [1]
     AA46 *MOT: < nee > [FIL] . [1]
     AC26 *BBS: < anders had ik 'm gisteren al kunnen stikken zeg maar > [IMP] . [10]
          *MOD: < xxx > [X] . [0]
     AC27 *BBS  < ja > [FIL] . [1]
165  AC28 *BBS. < met zakken d'rin > [DEC] < he > [TAG] . [4]
     AA47 *MOT  < ehm > [FIL] . [1]
     AC29 *BBS: < zakken d'rin > [DEC] < he > [TAG] . [3]
     AA48 *MOT: < xxx > [X]   [0]
          *BBS. < ja > [FIL]   [1]
170  AA49 *BBS: < de voetje zat los [\\] zat los > [DEC] < he > [TAG] . [7]
          *MOT: < xxx > [X]   [0]
     AA50 *MOT: < en ik dacht > [DEC] . [3]
     AA51 *BBS. < ging niet echt goed > [DEC] . [4]
          *BBS: < nee > [FIL] . [1]
175  CA5  *CHI: < eten > [DEC]   [1]
          *MOD  < xxx > [X]   [0]
     AA52 *MOT: < je kunt wel iets als je eventjes een xxx > [DEC] . [0]
     AA53 *BBS: < oh ja > [FIL] . [2]
          *BBS. < maar als je echt iets eh > [DEC]   [6]
180  AA54 *BBS  < nee > [FIL]   [1]
          *MOT. < xxx > [X] . [0]
     CA6  *CHI  < ho > [FIL]   [1]
     AC30 *BBS  < oh oh > [FIL]   [2]
     AC31 *MOT  < oh oh > [FIL]   [2]
185  CA7  *CHI. < xxx > [X] . [0]
     AC32 *BBS. < thee > [DEC] . [1]
```

```
        *BBS: < is die heet > [QUE] ? [3]
    AC33 *BBS: < is heet > [DEC] < hoor > [TAG] . [3]
        *BBS: < doe maar niet > [IMP]  [3]
190 CA8  *CHI: < op > [DEC]  [1]
    AC34 *BBS: < ja > [FIL] < is die op > [QUE] ? [4]
    CA9  *CHI: < ja > [FIL] . [1]
    CA10 *CHI: < xxx > [X] . [0]
        *MOD: < xxx werken > [X] . [0]
195 AC35 *BBS: < ja > [FIL] < dan gaan we zwaaien en dan gaan wij naar boven > [DEC]
                [11]
    AC36 *BBS: < gaan we lekker badderen > [DEC]  [4]
    AC37 *MOT: < ja > [FIL] . [1]
    AC38 *BBS: < gaan we alletwee badderen > [DEC] . [4]
200     *BBS: < en dan gaan we heleboel was van zolder afhalen > [DEC] . [9]
        *BBS: < en dan gaan we 'n heleboel vouwen en strijken > [DEC] . [9]
    AC40 *BBS: < heel veel werk hebben we > [DEC]  [5]
    AC41 *BBS  < gaat Bente heel lief spelen altijd mama > [DEC]  [7]
    AC42 *BBS: < he > [FIL] ? [1]
205     *MOT: < ja > [FIL] . [1]
        *BBS: < bij An > [DEC] < he > [TAG] ? [3]
    CA11 *CHI: < ikke spelen > [DEC] . [2]
        *BBS: < ja > [FIL] . [1]
    AC43 *BBS: < kan ze zo lief spelen > [DEC]  [5]
210 AC44 *BBS  < gaat ze bij An zitten > [DEC]  [5]
        *BBS: < op de grond > [DEC] < he > [TAG] ? [4]
        *BBS: < bouwen > [DEC] < he > [TAG] ? [2]
    AC45 *BBS: < ja > [FIL]  [1]
        *BBS: < met de beestjes > [DEC] . [3]
215 AA55 *MOT: < xxx boven 'n beetje los > [QUE] ? [0]
    AA56 *BBS: < nou nee > [FIL] < 't ligt er al aan > [DEC] . [7]
        *BBS: < als zij slaapt dan doen we 't gewoon beneden > [DEC] . [9]
    CA12 *CHI: < nee > [FIL] . [1]
    AA57 *BBS: < en anders dan eh > [DEC] . [4]
220 AA58 *BBS: < vorige week heb ik boven gedaan > [DEC]  [6]
    AA59 *BBS  < 't ligt er al aan > [DEC] < he > [TAG] . [6]
    AC46 *BBS  < nee > [FIL]  [1]
        *BBS: < hoepekee > [BBT]  [1]
    AC47 *BBS: < zo > [FIL]  [1]
225 AI15 *BBS: < kom 's hier > [IMP] < Josje > [VOC] . [4]
        *BBS: < geven we jou ook 'n zakdoek > [DEC]  [6]
        *MOD: < xxx > [X] . [0]
        *BBS: < doen we jouw zakdoek in die zak > [X] . [0]
    AI16 *BBS: < allemaal 'n eigen zakdoek > [DEC] < he > [TAG] . [5]
230 AI17 *BBS: < voor dat kleine snotpinneke > [DEC]  [4]
        *BBS: < vind jij niet leuk > [DEC] < he > [TAG] ? [5]
        @end
```

Chapter 3
Structural Aspects as a Function of Addressee

The focus of this chapter is on five aspects of the structure of the language input: Utterance length, vocabulary size, metrical structure, phonotactic structure, and word embedding. These five aspects were analyzed in language addressed to the infant (AI), language addressed to other children (AC), and language addressed to adults (AA). The results showed differences with respect to all these aspects in the three conditions. Generally, the structure of the AI and the AC speech was simpler than that of the AA speech. The implications for word discovery are discussed.

3.0 Introduction

In this chapter, five aspects of the language input are described: utterance length, vocabulary size, metrical structure, phonotactic structure, and word embedding. These aspects were analyzed because they are all more or less related to the issue of word segmentation. The five variables are each compared across the three addressee conditions: infant-directed speech (AI condition), child-directed speech (AC condition), and adult-directed speech (AA condition).

3.1 Utterance Length

The first aspect that was analyzed was Mean Length of Utterance (MLU). It has been hypothesized that children have to learn explicit segmentation strategies because adults do not tend to present words in isolation to children. This has been argued for in particular by Aslin (1993). To test this hypothesis, utterance length in general was analyzed, and the distribution of one-word utterances in

particular. A relatively high proportion of one-word utterances and relatively short utterances would make the process of word discovery considerably easier than when words were presented only in the context of other words.

Utterance length has often been a topic of research on child-directed speech. Generally, studies focus on the factors that influence MLU. For instance, Snow, Arlman-Rupp, Hassing, Jobse, Joosten and Vorster (1976) examined the speech of Dutch mothers of three social classes to their children with an age range of 18 to 38 months. A number of variables were measured in a free play and a book reading situation. The results showed that the mothers produced short, grammatically simple utterances (MLU 4.16). The differences between the social classes were very small, but language used in the book reading situation was more complex than that used in the free play situation.

Snow (1977b) investigated the development of MLU in the speech of two mothers to their children at various time points when the children were between three and 20 months of age. She found that MLU (approximately four words per utterance) did not change over this period. Similarly, Phillips (1973) found no significant difference in the length of utterances in the speech of mothers to their children at the age of eight months (MLU 3.56) or 18 months (MLU 3.47). However, there was more variability in length of utterance in speech to the 8-month-olds compared to speech to the older children. Thus, the mothers of the 8-month-olds produced very long sentences as well as very short sentences, whereas the mothers of the older children produced only short sentences. In the same study, MLU of utterances addressed to children with an average age of 28 months was 4.01, and that of utterances addressed to adults was approximately 8.40 words.

More recently, Ratner and Rooney (1993) measured utterance length in a corpus of language spoken to children in the age range of 13 to 20 months. MLU was 3.5 words, 24% of all the utterances were one-word utterances, 16% were two-word utterances, and 19% were three-word utterances.

In sum, utterances to children are short, and as the child grows older the utterances become longer. This development, which is related to the child's developing language, has been called 'fine-tuning' (e.g., Snow, 1995). However, from the prelinguistic period until children are about 18 months old, there do not appear to be significant differences in MLU dependent of the age of the addressee. Thus, it seems that from birth on, up to a certain age, MLU does not change much. The language development in this period does not seem to be sufficient to have any effect on the caretaker's MLU.

In this section, the following two questions will be addressed: (1) Were there differences in MLU between the three addressee conditions? (2) Was there a difference between the three addressee conditions in the proportion of one-

word utterances? Based on the results of previous studies, utterance length was expected to be shorter in the AC speech and the AI speech than in the AA speech. It was difficult to predict whether there would be a difference between MLU in the speech to the older sibling (30-33 months old) and MLU in speech to the infant, and, if there was a difference, whether MLU in the AI speech would be higher or lower than that in the AC speech. None of the studies mentioned above compared MLU in speech to children in this age range.

Before describing the methodology and the results of the analysis of MLU, the distribution of utterance types is considered first in the next subsection (the types that were distinguished are described in section 2.3.2). There were mainly two reasons for doing this. First, MLU seemed to be closely related to utterance type. Some types were very short but occurred very frequently (especially, utterances that consisted only of a social expression, a vocative or a filler). Second, it seemed reasonable to exclude certain utterance types from the computations (for instance, utterances that were produced during story telling and were actually utterances read from a book, and utterances that consisted only of baby talk).

3.1.1 Distribution of Utterance Types

The classification of utterance types described above in Chapter 2 (section 2.3.2) was used. Based upon this classification, utterances were divided into six main categories: declaratives, interrogatives, imperatives, baby talk, routines, other. The 'other' category included all utterances that consisted only of fillers (e.g., *ja, nee, hm, oh*), vocatives (e.g., *Josje, meisje*), or social expressions (e.g., *goede-morgen, hai, hallo*).

Note that in every condition there were a number of utterances that consisted of a combination of a declarative and an interrogative, or an imperative. In the transcriptions, these combinations were often separated, but not always. In particular, short imperatives in combination with a declarative were often not separated (e.g., *kom, dan gaan we eten* 'come, we are going to eat').

In total, there were 16242 utterances in the AI speech, 43772 in the AC speech, and 21379 in the AA speech (see also Table 2.4). The distribution of the utterance types in every addressee condition is shown in Table 3.1. Because of the overlap between declaratives, interrogatives, and imperatives, the numbers do not add up exactly to the total number of utterances. Furthermore, in each condition, a number of utterances could not be classified because the content was unclear. These utterances are listed as 'missing'.

Table 3.1: Distribution of utterance types in each addressee condition. Percentages are given between parentheses.

	AI		AC		AA	
Declaratives	3814	(23.5)	15552	(35.6)	11000	(53.9)
Interrogatives	1582	(9.7)	5170	(11.8)	2098	(9.8)
Imperatives	1360	(8.4)	6092	(13.9)	199	(0.9)
Baby Talk	1031	(6.3)	480	(1.1)	2	(0.0)
Routines	58	(0.4)	1113	(2.6)	0	(0.0)
Other	7317	(45.0)	11700	(26.7)	5903	(27.6)
Missing	1174	(7.2)	3720	(8.5)	2193	(10.3)
Total Nr	16336		43827		21395	

The majority of utterances in the AI speech consisted of 'other' utterances. The percentages of these other utterances were much lower both in the AC speech and in the AA speech. The percentages of baby talk and routines were relatively low in the AI and the AC condition. Two utterances were classified as baby talk in the AA speech, where an adult imitated child's speech. About 42% of all the AI utterances were imperatives, declaratives, interrogatives. In the AC speech, this was about 61%, and in the AA speech, this was about 64%. Within these three utterance types, the declarative was the most frequent type in every condition: 56% in the AI speech, 58% in the AC speech, and 83% in the AA speech. Furthermore, interrogatives occurred relatively equally often in every condition (around 10% of all utterances). Finally, imperatives occurred relatively often in the AI speech and the AC speech (around 10%), but in the AA speech, only 2% of all utterances were classified as imperatives.

Thus, the main differences between the three conditions were caused by the frequency of occurrence of 'other' utterances, imperatives and declaratives. The AI speech was different from the AC speech and the AA speech because of the high proportion of other utterances. The AA speech was different from the AI and the AC speech because of the low proportion of imperatives. The proportion of declaratives was lowest in the AI speech, higher in the AC speech, and highest in the AA speech.

3.1.2 Methodology

Utterance length was computed twice: once over all utterances except baby talk and routines, and once over declaratives, interrogatives and imperatives only.

The reason for excluding baby talk and routines was that these utterances do not occur in all conditions, and furthermore that they often are not 'spontaneous' utterances (as in the case of nursery rhymes, or songs for example). Both German and Dutch utterances were included. Compounds or combinations that occurred often were counted as one word (e.g., *Rolls+Royce*, *T+shirt*). Contractions (e.g., *da's* or *di's*, 'that's') were also counted as one word. Utterance length was computed for the three main speakers separately (the mother, the father, and the baby sitter), and for all other speakers.

3.1.3 Results

After exclusion of only baby talk and routines and of utterances of which the exact number of words was not audible, 13978 utterances remained in the AI condition, 37380 in the AC condition, and 16963 in the AA condition. Table 3.2 shows the MLU values of the speakers in the three conditions.

Table 3.2: MLU values over all utterance types except baby talk and routines. Standard deviations are given between parentheses.

	AI	AC	AA
Mother	2.33 (1.69)	2.58 (1.84)	4.54 (4.48)
Father	2.69 (1.99)	3.54 (2.54)	3.98 (4.13)
Baby sitter	3.17 (2.25)	3.42 (2.34)	5.16 (4.57)
Other	2.79 (2.14)	3.89 (2.88)	4.47 (4.62)
Mean	2.66 (1.97)	3.13 (2.27)	4.51 (4.45)

The average values of the speakers were separately subjected to analyses of variance. Post hoc comparisons were done following Tukey's HSD procedure. The MLU values of the mother were significantly different across the three conditions: $F[2,68318] = 1830.36$, $p < .05$. MLU of the AI speech was significantly lower than that of the AC speech, and the MLU of the AC speech was significantly lower than that of the AA speech. The MLU values of the father were also significantly different across the conditions: $F[2,11955] = 90.12$, $p < .05$. MLU of the AI speech was significantly lower than that of the AC speech, and MLU of the AC speech was significantly lower than that of the AA speech. The MLU of the baby sitter was significantly different across the three conditions: $F[2,22115] = 596.26$, $p < .05$. Again, MLU of the AI speech was

significantly lower than that of the AC speech, and the MLU of the AC speech was significantly lower than that of the AA speech. Finally, the MLU of the other speakers was also significantly different across the three conditions: $F[2,4425] = 34.18$, $p < .05$. Again, MLU in the AI condition was significantly lower than that in the AC condition, and MLU in the AC condition was significantly lower than that in the AA condition. Thus, MLU was consistently lower in the AI condition than in the AC condition, and MLU in the AC condition was consistently lower than in the AA condition.

The number of one-word utterances in every condition was determined. The utterances were collapsed over all speakers. 5527 utterances in the AI speech were one-word utterances (39.5%), 11801 utterances in the AC speech utterances were one-word utterances (31.6%), and 5323 utterances in the AA speech were one-word utterances (31.4%). The frequencies were compared using a chi-square test, and the differences between the conditions were significant: $\chi^2(2) = 323.66$, $p < .05$. Two additional comparisons were carried out by partitioning the chi-square value into two components (Everitt, 1977). The first component (χ^2_1) was associated with the difference between the AC and the AA condition. The second component (χ^2_2) was associated with the difference between the AC and the AA conditions on the one hand, and the AI condition on the other hand. The first comparison was not significant: $\chi^2_1(1) = 0.19$, $p > .05$, but the second comparison was: $\chi^2_2(1) = 323.46$, $p < .05$. Thus, the AC speech contained about as many one-word utterances as the AA speech, but the AI speech contained significantly more one-word utterances.

MLU was once more determined with the exclusion of all the 'other' utterances from the previous set. After exclusion, 6681 utterances remained in the AI condition, 25720 utterances in the AC condition, and 11061 utterances in the AA condition. Table 3.3 shows MLU per speaker in each addressee condition.

Table 3.3: MLU values of all utterance types except baby talk, routines, and other utterances. Standard deviations are given between parentheses.

	AI	AC	AA
Mother	3.73 (1.80)	3.40 (1.91)	6.29 (4.66)
Father	4.02 (2.08)	4.40 (2.50)	5.82 (4.53)
Baby sitter	4.30 (2.13)	4.14 (2.24)	6.57 (4.54)
Other	4.07 (2.12)	4.77 (2.80)	6.24 (4.90)
Mean	4.01 (2.00)	3.97 (2.26)	6.24 (4.65)

The average values are higher than in the previous table, and the differences between the conditions are not as consistent as they were in the previous analysis. The values of the speakers were subjected to separate analyses of variance. Post hoc comparisons were carried out, using Tukey's HSD procedure. The differences in MLU of the mother were significant across the three conditions: $F[2,16753] = 1558.08$, $p < .05$. MLU of the AI speech was significantly higher than that of the AC speech, and MLU of the AA speech was significantly higher than that of the AI speech. The differences in MLU of the father were also significant across the three conditions: $F[2,7675] = 172.70$, $p < .05$, but the pattern was not the same as that of the mother. MLU of the AI speech was significantly *lower* than that of the AC speech, and MLU of the AC speech was significantly lower than that of the AA speech. The MLU of the baby sitter was also significantly different over the three conditions: $F[2,16005] = 797.37$, $p < .05$. The pattern across conditions was the same as that of the mother. The MLU of the AI speech was significantly *higher* than that of the AC speech, and MLU of the AA speech was significantly higher than that of the AC speech. The MLU of the other speakers was also significantly different between the conditions: $F[2,3017] = 61.74$, $p < .05$. The post hoc comparisons showed that the MLU in the AI condition was not significantly different from that in the AC condition, but the MLU in the AA condition was significantly higher than both the AI and the AC conditions.

Thus, MLU was generally highest in the AA speech, but the differences between the AI speech and the AC speech were not the same across the speakers. The mother and the baby sitter both showed a higher MLU in the AI speech, but the father showed a higher MLU in the AC speech.

Finally, the utterances were collapsed over the speakers, and the percentages of one-word utterances were determined. Now the AI speech consisted of 462 one-word utterances (6.9%), the AC speech consisted of 2923 one-word utterances (11.4%), and the AA speech consisted of 831 one-word utterances (7.5%). Thus, the percentages of one-word utterances were now much lower than when the 'other' utterances were included in the counts. The frequencies were compared using a chi-square test, and the result showed that the differences were significant: $\chi^2(2) = 200.93$, $p < .05$. The overall chi-square was once more partitioned into two components. The first component (χ^2_1) was associated with the difference between the AI and the AA speech, the second component (χ^2_2) was associated with the AC speech on the one hand, and the AI and the AA speech on the other hand. The results showed that the difference between the AI speech and the AA speech was not significant: $\chi^2_1(1) = 1.70$, $p > .05$, but the second comparison was: $\chi^2_2(1) = 199.23$, $p < .05$. Thus, the proportion of one-word utterances was significantly higher in the AC condition

than in the AI and the AA conditions, whereas the difference between the AI and the AA conditions was not significant.

3.1.4 Summary

The general question that was addressed here was whether there were differences in utterance length between the three conditions, and more specifically, whether the percentages of one-word utterances were different between the conditions. In the first analysis, almost all utterances were included except baby talk and routines. The results showed that MLU was consistently lowest in the AI speech (MLU = 2.66), higher in the AC speech (MLU = 3.13), and highest in the AA speech (MLU = 4.51). Furthermore, the percentages of one-word utterances were high in every condition: almost 40% in the AI condition, and somewhat more than 30% in both the AC and the AA conditions.

However, when utterances classified as 'other' were excluded, the results changed. The overall MLU values were higher. MLU of the AI speech (MLU = 4.01) was now even slightly higher than that of the AC speech (MLU = 3.97), but MLU of the AA speech remained highest (MLU = 6.24). Furthermore, the percentages of one-word utterances decreased dramatically to about 7% in the AI and the AA conditions, and about 11% in the AC condition.

Thus, most of the one-word utterances were utterances which were classified as other. These are words that children do not usually produce early, nor are they words that are tested in comprehension. Still, they also occur often as parts of larger utterances. Assuming that the child is able to recognize these parts, they could be segmented from longer utterances relatively easily, leaving only a one or a two word utterance.

However, the fact that MLU of the AI speech was not lower than that of the AC speech (after exclusion of the 'other' utterances) indicates that the adults in this study did not adapt their speech style when addressing the infant, in a way to produce only very short utterances. In that case, it would have been expected that the MLU of the AI speech would have been lower than that of the AC speech, and the proportion of one-word utterances would have been higher. However, in both cases the reverse was true. In fact, the proportion of one-word utterances in the AC condition was significantly higher than that in the AI condition. A possible explanation for this difference is that a lot of one-word utterances that the adults produced to the older child were repetitions of the child's own utterances. However, since the Child-to-Adult speech was not transcribed this hypothesis cannot be tested at present.

3.2 Vocabulary Size

The second aspect that was analyzed concerned vocabulary size. The question that was addressed was whether there were differences in vocabulary size across the three addressee conditions. The importance of vocabulary size for word segmentation is obvious. A small repetitive vocabulary makes word segmentation easier than a diverse large vocabulary.

The vocabulary diversity of language addressed to children has been analyzed in many previous studies (e.g., Phillips, 1973; Broen, 1972; Snow, 1972). For instance, Broen (1972) compared vocabulary diversity in mothers' speech addressed to young children (mean age of 21 months), older children (mean age of 60 months), and to adults. The type-token ratio (number of word types divided by number of word tokens) was used as a measure for vocabulary diversity. She found significantly lower type-token ratios in language to the younger children compared to language to the older children or to the adults. Phillips (1973) calculated type-token ratios over samples of 300 words. The average type-token ratios were 0.31, 0.34, 0.41, and 0.51 in speech to eight-month-olds, 18-month-olds, 28-month-olds, and adults respectively. The difference between the eight-month-olds condition and the 18-month-olds condition was not significant, but the averages in the 28-month-olds and in the adult condition were significantly higher.

There are probably two causes for the low vocabulary diversity that is found in child-directed speech. Firstly, the content of the child-directed speech is very much limited to a restricted number of subjects (Snow, 1977a). Secondly, as demonstrated in particular by Snow (1972), mothers use utterance repetitions frequently when talking to children. In this section, an estimate of the vocabulary size of the input is given.

3.2.1 Methodology

Vocabulary size was investigated in two ways. First, the vocabulary size as a function of the sample size was determined in every addressee condition. For this purpose, the vocabulary size was determined (using the CHILDES software FREQ command) for the first 500 words, then for the first 1000 words, the first 1500 words, etc.

Second, the type-token ratio was used as a measure for repetitiveness of the vocabulary: the number of different word types in a sample divided by the number of word tokens in that sample. A relatively high number of word types results in a high ratio (i.e., a non-repetitive sample), a relatively low number of

types results in a low ratio (i.e., indicating a repetitive sample). There is, however, an important drawback in the use of type-token ratios. This drawback lies in the fact that variation in the number of tokens causes variation in the ratio. Ratios that are based on large samples are relatively low compared to ratios that are based on small samples (Richards, 1987; Richards and Malvern, 1997). In other words, two ratios can only be compared when they are based on the same number of word tokens. Therefore, in this study, ratios were determined over consecutive chunks of 300 words and then averaged.

A problem in estimating the vocabulary size in this study was caused by the fact that input to the infant and to the older child drew on the vocabulary of two languages. Since the aim of the first analysis was to give a global description of the total vocabulary sizes in each condition, both German and Dutch were included in that analysis. The German words were excluded from the second analysis, so that this analysis could serve as a more direct comparison of the three conditions.

3.2.2 Results

The total vocabulary size of the AI speech consisted of 2081 types distributed over 41622 word tokens. Of the total number of word tokens, 2757 occurred in routines and baby talk utterances, 11541 occurred in German utterances from the mother. The total vocabulary size of the AC speech consisted of 4808 types distributed over 130329 word tokens, of which 1187 occurred in baby talk utterances and routines, and 30713 word tokens occurred in German utterances. The total vocabulary of the AA speech consisted of 5523 types distributed over 89355 word tokens. In the AA speech there were no German utterances, no routines and only two baby talk utterances.

To illustrate differences in vocabulary size, the vocabulary size was plotted as a function of the sample size. In order to prevent Dutch and German words that are spelled the same way (c.f., Dutch *schoen* /sxun/, 'shoe', and German *schoen* /ʃøn/, 'nice') from being counted as one word, the counts were done on the phonological transcriptions instead of the orthographic transcriptions (this slightly changes the total number of word types, since some word tokens that are spelled differently are pronounced the same). The result is shown in Figure 3.1. The figure shows the sample size (N) on the X-axis and the vocabulary size (V) on the Y-axis. The sample size of speech to the infant (lower line) is smallest, and the vocabulary size grows slowly. The vocabulary size in the AA speech grows fastest, and the AC curve falls in between the other

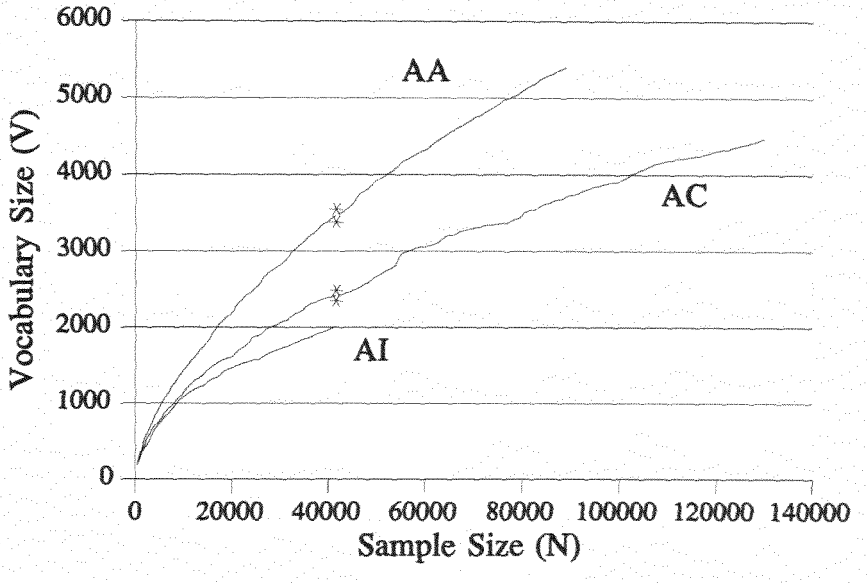


Figure 3.1: Vocabulary size (*V*) is displayed as a function of the sample size (*N*). The top line represents the AA speech, the middle line the AC speech, the lower line the AI speech. The stars mark the vocabulary size in the AC speech and the AA speech plus and minus two standard deviations at sample size 41622.

two. The AI and the AC curves would, of course, have been much lower if the input had been monolingual.

The three curves were compared statistically, using a simple procedure (taken from Chitashvili and Baayen, 1993). The variance of *V* can be estimated by subtracting the vocabulary size at sample size *N* from the vocabulary size at sample size *2N*. Taking the square root of the difference results in the standard deviation (*s*). This enabled a comparison to be made of the three curves in the following way. The total sample size of the AI curve was 41622 word tokens. This was the basis used to calculate the standard deviations of the AC curve and the AA curve. The vocabulary size at sample size *N* (41622) and at sample size *2N* (83244) was estimated in the AC and the AA curves, and the differences between the two values were computed. Taking the square root of both differences resulted in the standard deviations of the AC and the AA curves. The results are displayed in Table 3.3.

Table 3.3: Estimation of standard deviations of the AC and the AA curves.

	AC	AA
$V(N)$	2417	3466
$V(2N)$	3571	5223
$s^2 = V(2N)-V(N)$	1154	1757
s	33.97	41.92

The difference between V of the AI speech and the AC speech at sample size N = 41622 is equal to 2417 – 2014 = 403. At this point, the standard deviation of the AC curve is equal to 33.97. Thus, the vocabulary size of the AI speech is considerably more than two standard deviations below the AC speech. Similarly, the difference between the vocabulary size of the AC speech and the AA speech at sample size 41622 is equal to 3466 – 2417 = 1049. The standard deviation of the AA curve at this sample size is equal to 41.92. Thus, the vocabulary size of the AC speech is again more than two standard deviations below that of the AA speech. In Figure 3.1, the vocabulary size in the AC and the AA curves plus and minus two standard deviations at sample size 41622 is marked in each case by stars.

A more conventional way of expressing vocabulary size is the calculation of type-token ratios. German utterances, baby talk and routines were now excluded so that every corpus consisted of the same utterance types. After exclusion, the total vocabulary size of the AI speech comprised 1138 types distributed over 27371 word tokens, the total vocabulary size of the AC speech comprised 2874 types distributed over 94333 word tokens, and the total vocabulary of the AA speech consisted of 5507 types distributed over 89347 word tokens. The samples were divided into consecutive chunks of 300 tokens, and for every chunk the number of types was determined, and the ratio was calculated. Table 3.4 shows the results of the calculation. The last row (N) shows the number of ratios that were calculated in each condition. The highest ratios are found in the AA speech, and the lowest ratios are found in the AI speech. The differences between the conditions were tested in an analysis of variance. The main effect of addressee was significant: $F[2,701] = 545.6$, $p < .01$. Tukey post hoc comparisons indicated that the average type-token ratio in the AI speech was significantly lower than that in the AC speech, and that the average type-token ratio in the AC speech was significantly lower than that in the AA speech. Thus, the AI speech was characterized by a more restricted and more repetitive vocabulary than the AC speech, and the AC speech was characterized by a more restricted vocabulary than the AA speech.

Table 3.4: Average number of types per 300 tokens, and average type-token ratios. Standard deviations are given between parentheses.

	AI	AC	AA
Av. nr. of types	96.59 (16.81)	116.88 (13.25)	141.43 (10.31)
Av. TT ratios	0.322 (0.056)	0.390 (0.043)	0.471 (0.034)
N	91	315	298

3.2.3 Summary

The second aspect that was analyzed concerned the vocabulary size. It was expected that the size of the vocabulary in the AI condition would be smallest, that it would be bigger in the AC condition, and biggest in the AA condition. This is indeed what the results showed. A direct comparison of the total vocabulary sizes was difficult, because the AI and the AC conditions contained both German and Dutch words. However, significant differences were found even when the German words were included in the comparisons. The differences are expected to be even bigger if the input had been monolingual, and if all the words with the same meaning would be counted only once instead of twice.

Furthermore, significant differences in type-token ratios were also observed. The type-token ratios offered a somewhat better comparison of the three conditions because all the German words could be excluded without affecting the average ratios very much. The values in the AI speech and the AC speech were comparable to those reported by Phillips (1973). The values in the AA speech are somewhat lower in the present study, but not significantly so.

Why the differences? As has been stated in previous studies, adults talk about 'here and now' to children, and therefore the vocabulary is limited (e.g., Snow, 1977). Furthermore, the vocabulary size in the AI speech was certainly affected by the high proportion of 'other' utterance types (as was described in the previous section). These utterances made up a relatively high proportion of all the word tokens, but only a very small proportion of all the word types. Thus in effect, the infant heard very many repetitions of a few utterances which had little content.

3.3 Metrical Structure

The importance of metrical structure for word segmentation was described in Chapter 1 (section 1.2.1). To review: the metrical structure of English and Dutch

is determined by the alternation of strong and weak syllables. There are three reasons to assume that metrical structure is important for word segmentation.

First, distributional analysis has shown that most English and Dutch content words start with a strong syllable. Cutler and Carter (1987) estimated that about 90% of all lexical items in a corpus of spoken English started with a strong syllable, whereas only 9.5% of the grammatical words in that corpus started with a strong syllable. Furthermore, of all strong syllables 74% were the first syllable of a content word. Thus, in English, the quality of the syllable seems to be a fairly reliable predictor of a word boundary. In Dutch, about 87.5% of all content words start with a strong syllable according to estimates by Vroomen and de Gelder (1995). Baayen and Schreuder (1994) report that about 60% of the lexical words that start with a weak syllable are words with an initial weak prefix (e.g., *gevonden*, /xəvɔndə/ 'found').

Second, experimental research has shown that Dutch and English listeners use stress for the segmentation of spoken language (Cutler and Norris, 1988; Vroomen, van Zon and de Gelder, 1996).

Finally, young infants are sensitive to the predominance of words with initial strong syllables (Jusczyk et al., 1993a). Furthermore, they seem to use this knowledge for word segmentation. Newsome and Jusczyk (1995) showed that infants that were familiarized with bisyllabic strong-weak words listened significantly longer to passages that contained these words, whereas infants that were familiarized with weak-strong bisyllabic words, did not show a preference.

Little research has been done on the metrical structure of child-directed speech. Kelly and Martin (1994) report that in a corpus of language input to children between 15 and 25 months, 95% of the strong syllables corresponded with word boundaries. This percentage is higher than the one reported by Cutler and Carter (1987) in their corpus of adult-directed speech.

The present section aims to give a detailed description of the metrical structure of the input to the infant in this study. The methodology that will be described follows the study done by Cutler and Carter (1987). Thus, a distinction between grammatical words and lexical words will be made. The question that is addressed is whether there were differences in the metrical structure in the three conditions. The effect of possible differences on word segmentation will be discussed.

3.3.1 Methodology

For the estimation of the distribution of strong and weak syllables, a list of words that occurred in the transcriptions was compiled in each condition. These

lists contained the words and their frequencies. The word lists were created with the CHILDES software FREQ-command. This command lists the word types in a text with the token frequency of each type. German utterances, *baby talk* and *routines* were excluded for the compilation of the lists. The following information was then added to the lists: the lexical status of each item (lexical or open-class word versus grammatical or closed-class word) and the quality of the syllables of each word (strong or weak).

The lexical status of each word was determined on the basis of a list of Dutch closed-class items (van Wijk and Kempen, 1980). Some words that did not occur in the list were classified as a closed-class item based on analogy to words that did occur in the list, for example *zometeen* (shortly) was not in the list, but *zostraks* (shortly) was. A number of words were classified as *other*. These were mainly words that are typical for spoken language, such as *goedemiddag* (good afternoon), *au* (excl.), *oei* (excl.), etc., and words from other languages. Proper names were coded as lexical items, numbers were coded as closed class items.

The stress pattern of the words was determined on the basis of lexical stress and vowel quality. Syllables with schwa were coded as weak (e.g., the first syllable of the word *gedaan*, 'done'). Stressed syllables were coded as strong (e.g., the second syllable of the word *gedaan*) except when the syllable contained a schwa. This was the case in some monosyllabic function words, such as *de*, 'the' *'t*, 'the', etc. Note that the transcriptions distinguished between a number of function words that were realized with a full vowel (e.g., *zij*, /zɛi/, 'she'), or with a reduced vowel (e.g., *ze*, /zə/, 'she'). All monosyllabic lexical items were coded as strong. Unstressed syllables were either coded as strong when the syllable contained a full vowel (e.g., the first syllable of the word *moment*, /momɛnt/ 'moment'), or as weak when the syllable contained a vowel that was reduced or could easily be reduced (e.g., the last syllable of the word *weinig*, /wɛinəx/ 'little'). In each addressee condition, the number of lexical items that started with a strong syllable was determined, as well as the number of grammatical items that started with a strong syllable. Finally, the frequency of occurrence of strong syllables at other places than at the onsets of lexical items was also determined.

3.3.2 Results

The resulting corpus consisted of 22871 word tokens in the AI speech (1077 types), 86341 word tokens in the AC speech (2728 types), and 83224 word tokens in the AA speech (5270 types). The word tokens in the AI speech

consisted of 14808 closed-class items (64.7%) and 8063 open-class items (35.3%); the word tokens in the AC speech consisted of 57466 closed-class items (66.6%) and 28875 open-class items (33.4%); the word tokens in the AA speech consisted of 59242 closed-class items (71.2%) and 23982 open-class items (28.8%).

Table 3.5 shows the percentages of monosyllabic lexical items, polysyllabic lexical items with strong initial syllables, and polysyllabic lexical items with weak initial syllables in each condition. The percentage of word tokens with initial strong syllables (monosyllables plus strong-initial polysyllables) was 97.2 in the AI condition, 96.4 in the AC condition, and 88.3 in the AA condition. The frequencies were compared across the three conditions with a chi-square test. The overall chi-square was significant: $\chi^2(2) = 1581.41$, $p < .05$. Two additional comparisons were carried out by partitioning the overall chi-square into two components. The first component (χ^2_1) was associated with the difference between the AI and the AC conditions, the second component (χ^2_2) with the AI and the AC conditions on the one hand, and the AA condition on the other hand. The results showed that both differences were significant: $\chi^2_1(1) = 6.34$, $p < .05$, $\chi^2_2(1) = 1575.08$, $p < .05$. Thus, the differences in percentages between the AI and the AC conditions were relatively small but statistically significant. Furthermore, the proportion of lexical items in that started with a strong syllable was significantly lower in the AA condition.

Table 3.5: Distribution of lexical items starting with a strong (S) syllable or with a weak (W) syllable. Percentages are provided between parentheses.

	AI		AC		AA	
Monosyllables	2974	(36.9)	11186	(38.7)	8676	(36.2)
S-initial polysyl.	4860	(60.3)	16640	(57.6)	12503	(52.1)
W-initial polysyl.	229	(2.8)	1049	(3.6)	2803	(11.7)
Total nr.	8063		28875		23982	

Next, the distribution of strong and weak syllables in the grammatical words was determined. Table 3.6 shows the results. In the AI speech, 82.6% of all grammatical words started with a strong syllable, 79.5% of all the grammatical words in the AC speech started with a strong syllable, and 85.0% of all the grammatical items in the AA speech started with a strong syllable. In other words: The stress pattern of the grammatical words was similar to that of the lexical words because most grammatical words also started with a strong syllable.

Table 3.6: Distribution of grammatical items with initial strong (S) or weak (W) syllables. Percentages are provided between parentheses.

	AI		AC		AA	
W *monosyl.*	2542	(17.2)	11259	(19.6)	8286	(14.0)
S monosyl.	11764	(79.3)	42583	(74.1)	45513	(76.8)
S-initial polysyl.	473	(3.2)	3098	(5.4)	4856	(8.2)
W-initial polysyl.	29	(0.2)	526	(0.9)	587	(1.0)
Total nr.	14808		57466		59242	

The frequencies were again compared using a chi-square test. The overall chi-square was significant: $\chi^2(2) = 614.48$, $p < .05$. Two additional comparisons were carried out by dividing the overall chi-square value into two components. The first component (χ^2_1) was associated with the difference between the AI and the AC conditions, the second component (χ^2_2) with the AI and the AC conditions on the one hand, and the AA condition on the other hand. Both differences were significant: $\chi^2_1(1) = 80.11$, $p < .05$, $\chi^2_2(1) = 534.37$, $p < .05$. Thus, the percentage of grammatical items starting with a strong syllable was lowest in the AC speech, it was significantly higher in the AI speech, and it was significantly higher again in the AA speech than in the AI speech. However, although significant, the differences between the conditions were relatively small.

Thus it seemed that a high proportion of strong syllables also occurred at places other than the beginning of lexical items - the majority of grammatical words also began with strong syllables. This is confirmed when we look at Table 3.7 which shows the number of strong syllables that occurred as the onsets of lexical items and the number of strong syllables that occurred at other places.

Table 3.7: Distribution of location of strong syllables. Percentages are given between parentheses.

	AI		AC		AA	
Onset lex. item	7834	(36.4)	27826	(35.1)	21179	(25.3)
Other places	13716	(63.6)	51493	(64.9)	62447	(74.7)
Total Strong	21550		79319		83626	
Total Weak	7879		32527		28902	
Total overall	29429		111846		112528	

In the AI speech, 63.7% of all strong syllables did not occur as the beginnings of lexical items, but as the beginnings of grammatical items, or as non-initial syllables in lexical items or grammatical items. In the AC speech this was 64.9%, in the AA speech it was 74.7%. The conditions were again compared using a chi-square test. The overall differences were significant: $\chi^2(2) = 2169.27$, $p < .05$. The overall chi-square was partitioned into a component associated with the difference between the AI and the AC condition (χ^2_1), and a component associated with the AI and the AC condition on the one hand and the AA condition on the other hand (χ^2_2). Both differences were significant: $\chi^2_1(1) = 12.90$, $p < .05$; $\chi^2_2(1) = 2156.41$, $p < .05$. Thus, the percentage of strong syllables that occurred as the onsets of lexical items was significantly lower in the AA speech than in the AC speech, and it was significantly lower in the AC speech than in the AI speech. However, although statistically significant, the difference between the AI and the AC condition was relatively small.

3.3.3 Summary

The results showed that, overall, a very high percentage of lexical items started with a strong syllable. This percentage was highest in the AI speech and lowest in the AA speech. An informal inspection of the word lists suggested that one of the main causes of the differences was that the AA speech contained many more words occurred starting with a weak initial prefix (e.g., *ge-*, *be-*, *ver-*). In the AI and the AC conditions, these prefixed words (usually inflected verb forms) seemed to occur less frequently.

However, a very high proportion of all strong syllables also occurred as initial syllables of grammatical items or as non-initial syllables of lexical items. In the AI and the AC speech, this was a little less than 65%, and in the AA speech it was a little less than 75%.

Suppose a metrical segmentation strategy to be applied to the present input data. Every strong syllable would thus be considered as a potential onset of a lexical item. This would go wrong more often than right: in the AI speech and in the AC speech, more than three out of five syllables would be false alarm, and in the AA speech, a little less than three out of four syllables would be false alarm.

The results presented here differ from those reported for English by Cutler and Carter (1987) mainly because the number of closed-class items starting with strong syllables was much higher in the present corpus. In the present study, 80% or more of the grammatical word tokens started with a strong syllable. In the study by Cutler and Carter, this was only about 10%. Therefore the predicted

rate of false alarms by the metrical segmentation strategy is much higher in this study. Furthermore, these results extend the ones reported for Dutch by Vroomen and de Gelder (1995) who did not examine the distribution of strong syllables in places other than the onset of lexical items.

Of course, the results of the analyses of metrical structure depend on the distinction between lexical and grammatical words that was made a priori. If this distinction had not been made, would the results look very different? Adding the results from Tables 3.5, 3.6 and 3.7 it is possible to determine the number of strong syllables that occurred at (lexical and grammatical) word onsets, and the number of strong syllables that occurred at other places. In the AI condition, 20071 strong syllables were word onsets (93.14% of all strong syllables), in the AC condition, 73507 (92.67%) of all strong syllables were word onsets; in the AA condition, 71548 (85.56%) of all strong syllables were word onsets[1].

Thus, in this perspective, the results indeed look much different: It would be more successful to apply a metrical segmentation strategy to the AI and the AC speech than to the AA speech because the presence of a strong syllable is more like to be the onset of a word in these two conditions.

3.4 *Phonotactic Structure*

The importance of phonotactic structure for word segmentation was also discussed in Chapter 1. To review, there are two kinds of evidence that phonotactic structure plays a role in the segmentation of continuous speech. First, the phonotactic structure of the language provides potential evidence for syllable boundaries (Church, 1987). Some sound sequences can only occur at the beginning of words or syllables, whereas others can only occur at the ends of words or syllables. The string /kn/, for example, can occur at the beginning of a Dutch word, but not at the end. In English, however, the sequence /kn/ cannot be the beginning or the end of a word. Therefore there must be a word or a

[1]These frequencies were compared using a chi-square test. Overall, the differences were significant: $\chi^2(2) = 2536.66$, $p < .05$. Subsequently, the overall chi-square was partitioned into a component associated with the difference between the AI and the AC condition (χ^2_1), and a component associated with the AI and the AC condition on the one hand and the AA condition on the other hand (χ^2_2). The first component was significant: $\chi^2_1(1) = 3.89$, $p < .05$; $\chi^2_2(1) = 2156.41$, $p < .05$. Thus, the number of strong syllables that occurred word-initially was significantly lower in the AA condition than in the AC and the AI condition, and it was significantly lower in the AC condition than in the AI condition. However, the difference between the AI and the AC conditions was very small.

syllable boundary between the /k/ and the /n/. Thus, knowledge of these regularities can be useful for segmentation.

McQueen (1998) showed that adult listeners indeed seem to use this knowledge for word segmentation. In McQueen's study, adult listeners spotted real words embedded in nonsense words significantly faster when the embedded words were aligned with syllable boundaries cued by phonotactics than when the embedded words were misaligned with such boundaries.

Second, Jusczyk et al. (1993b) showed that infants are sensitive to phonotactic patterns at a very early age. Furthermore, Friederici and Wessels (1993) demonstrated that infants also use this knowledge, since, under specific conditions, infants preferred to listen to strings of syllables consisting of phonotactically legal sequences rather than to strings of syllables containing phonotactically illegal sequences.

The question that is addressed in this section is whether the phonotactic structure of the AI speech was less complex than that of the AC and the AA speech. For this purpose, the number of possible consonant combinations that occurred in word onset and word offset was determined. Clearly, a lower number of possible word onsets and word offsets would result in a less complex word structure. Then the question was asked whether differences in phonotactic structure would be useful for segmentation. For this purpose, it was determined how often word boundaries could reliably have been located on the basis of the phonotactic structure of the language in each addressee condition.

3.4.1 Methodology

The same word lists as described in section 3.3.1 were used. The phonological transcriptions of these words (taken from CELEX, or supplied by hand) were used for the present analysis. The word onsets and the word offsets were isolated from the rest of the words. Possible onsets and offsets were one or more consonants or a vowel. (No distinction was made between long and short vowels. Although Dutch words normally do not have short vowels in word-final position, many exclamations (*he*, *bah*, *goh*, excl.), typical of spoken language do (Booij, 1995). Since many of these exclamations occurred in this corpus, it was decided that this would not be an extra constraint for a word boundary.)

Next, word boundaries within utterances, and syllable boundaries within words in the transcriptions were isolated so that only the consonants that surrounded the boundary were left over. For instance, in the utterance *'dat is fijn'*, /dɑtɪsfɛin/ ('that's nice'), the word boundaries /t/ and /sf/ were isolated, and from the word *omdraaien* /ɔmdrajə/ the syllable boundaries /mdr/ and /j/ were

isolated. German utterances, baby talk utterances, and utterances with words from foreign languages were excluded for the compilation of the lists of word boundaries and syllable boundaries.

Next, every word boundary was classified as 'clear' or 'unclear'. A word boundary was classified as clear when, on the basis of the word onsets and word offsets and the possible syllable boundaries that occurred in that condition, it was clear where the boundary was. For instance, the word boundary /sf/ from the example above, taken from the AI speech, was a clear boundary: no words in the AI speech started or ended with /sf/ and no words had an internal syllable boundary /sf/. Therefore, it was clear that there was a word boundary between the /s/ and the /f/. However, if this boundary occurred in the AA condition it was classified as an unclear boundary. This was because the AA condition did include words which started with /sf/ (as in *sfeervol*), or /sf/ occurred word-internally at a syllable boundary (as in *pasfoto*). The classification was done in two steps: first, on the basis of possible word onsets and word offsets only, and second, on the basis of word onsets and offsets and on word-internal syllable boundaries.

3.4.2 Results

In the AI speech 49 possible word onsets and 44 possible word offsets were identified. In the AC speech 54 possible word onsets and 62 possible word offsets were identified. In the AA speech, 60 possible word onsets and 75 possible word offsets were identified. Thus, the total number of word onsets and word offsets was lowest in the AI condition, higher in the AC condition, and highest in the AA condition. In the AI speech, the number of onsets was larger than the number of offsets, whereas in the AC and the AA speech the reverse was true. Table 3.8 lists the 20 most frequent (according to a token count) word onsets and offsets in each condition (a complete list of onsets and offsets is given in Appendix C). As can be seen from the table, these 20 onsets and offsets accounted for more than 95% of all the onsets and offsets in every condition.

A total of 147773 word boundaries were isolated from the entire corpus: 16026 in the AI speech (548 types), 63019 in the AC speech (928 types), and 68728 in the AA speech (1097 types). Table 3.9 lists the numbers of word boundaries that were classified as 'clear' on the basis of the possible word offsets and word onsets only. As the table shows, about half of the boundaries were clear in each condition. The differences between the conditions were tested using a chi-square test. The resulting chi-square was significant: $\chi^2(2) = 49.29$, $p < .05$. However, although significant, the differences were very small.

Table 3.8: The 20 most frequent word onsets and offsets in each addressee condition. Cumulative percentages are given between parentheses. V = vowel.

AI				AC				AA			
Onsets		Offsets		Onsets		Offsets		Onsets		Offsets	
j	(20.5)	V	(53.4)	V	(19.2)	V	(43.5)	V	(25.4)	V	(38.0)
V	(35.8)	t	(61.4)	j	(29.8)	n	(53.8)	d	(37.5)	n	(50.6)
h	(49.6)	r	(68.9)	m	(40.4)	t	(63.7)	j	(45.6)	t	(62.8)
m	(59.2)	s	(74.5)	d	(50.0)	r	(71.6)	m	(53.6)	r	(70.7)
z	(65.6)	n	(80.1)	h	(57.2)	m	(77.7)	h	(60.3)	k	(75.6)
d	(70.9)	j	(83.9)	n	(64.1)	s	(83.0)	w	(66.2)	s	(80.3)
n	(75.1)	x	(87.2)	z	(69.7)	k	(85.9)	n	(71.8)	x	(83.5)
w	(79.2)	k	(89.4)	b	(75.1)	x	(88.8)	z	(77.0)	l	(86.5)
k	(83.0)	m	(91.5)	w	(79.5)	p	(91.0)	v	(80.8)	p	(88.1)
l	(86.3)	p	(92.9)	k	(83.5)	l	(93.2)	x	(84.6)	nt	(89.6)
b	(89.5)	l	(94.2)	x	(87.1)	nt	(94.2)	b	(87.5)	m	(90.7)
x	(91.6)	nt	(95.4)	v	(89.2)	j	(94.7)	k	(90.1)	f	(91.7)
v	(93.5)	ʃ	(96.3)	l	(91.2)	f	(95.3)	t	(92.2)	xt	(92.6)
t	(94.7)	st	(97.2)	p	(92.6)	xt	(95.7)	l	(93.5)	rt	(93.3)
sl	(95.3)	xt	(97.5)	t	(93.9)	st	(96.2)	p	(94.2)	ft	(94.0)
kl	(95.9)	ns	(97.8)	st	(94.7)	ft	(96.6)	st	(94.7)	ls	(94.6)
sx	(96.4)	rt	(98.0)	sx	(95.2)	rt	(96.9)	r	(95.3)	st	(95.1)
p	(96.9)	f	(98.3)	kl	(95.8)	rst	(97.2)	tw	(95.7)	ŋ	(95.6)
xr	(97.3)	pt	(98.4)	br	(96.1)	ls	(97.5)	s	(96.0)	ts	(96.0)
f	(97.6)	ks	(98.6)	xr	(96.4)	w	(97.7)	xr	(96.3)	lf	(96.4)

Next the boundaries that were classified as 'clear' in the previous step were classified as 'unclear' if this boundary also occurred word-internally. Table 3.10 shows the results. The number of clear boundaries obviously decreased. This was true in every condition, but the decline in the AI condition was much lower

Table 3.9: Word boundaries constrained by possible word onsets and offsets only. Percentages are given between parentheses.

	AI		AC		AA	
Clear	8207	(51.2)	33007	(52.4)	34669	(50.4)
Unclear	7819	(48.8)	30012	(47.6)	34059	(49.6)
Total nr.	16026		63019		68728	

than the decline in the AC and the AA speech. 25% of all the word boundaries were still clear in the AI condition, about 15% were still clear in the AC condition, and about 10% were still clear in the AA condition.

The differences were tested with a chi-square test. The overall chi-square was significant: $\chi^2(2) = 2439.54$, $p < .05$. Two additional comparisons were carried out, by partitioning the overall chi-square value into two components. The first component (χ^2_1) was associated with the difference between the AI and the AC condition. The second component (χ^2_2) was associated with the AI and the AC conditions on the one hand, and the AA condition on the other hand. The results showed that both components were significant: $\chi^2_1 = 942.25$, $p < .05$ and $\chi^2_2 = 1487.28$, $p < .05$.

Table 3.10: Word boundaries constrained by possible word onsets, offsets, and by word-internal syllable boundaries. Percentages are given between parentheses.

	AI		**AC**		**AA**	
Clear	3995	(24.9)	9750	(15.5)	7121	(10.4)
Unclear	12031	(75.1)	53269	(84.5)	61607	(89.6)
Total nr.	16026		63019		68728	

3.4.3 Summary

This section reported an analysis of the phonotactic structure of the words in the three addressee conditions. The results of this analysis showed that the number of possible word onsets and offsets was relatively low in the AI speech. It was somewhat higher in the AC speech, and highest in the AA speech. However, in each condition, the 20 most frequent word offset and word onset types accounted for more than 95% of all word onset and offset tokens. This suggests that the differences between the conditions are really relatively small.

Furthermore, it was estimated how well word boundaries could be located on the basis of the possible onsets and offsets in every condition. Relatively small differences were found. About half of all the word boundaries could be located in every condition. However, when the word boundaries that could also occur as word-internal syllable boundaries were excluded, there was a clear advantage in the AI condition: About 25% of all the boundaries were still clear, whereas in the AC speech this was about 15% and in the AA speech, it was 10%. Thus, the results suggest that the phonotactic structure of the words in the AI speech was simpler than that of the words in the AC or the AA speech.

3.5 Word Embedding

The final aspect of the linguistic structure of the corpus that was analyzed was word embedding: the occurrence of short words in longer words (e.g., *bed* in *embedded*). The frequency of word embedding is a possible confounding factor for listeners and in particular for a child learning to recognize words in fluent speech. When a short word occurs as a substring of a longer word, the listener might infer that the total input string was two or three words instead of one.

Previous studies have shown that word embedding is common. For instance, McQueen, Cutler, Briscoe and Norris (1995) showed that more than 80% of English polysyllabic words in a dictionary contain one or more embedded words. Furthermore, McQueen et al. showed that the embeddings occurred more often at word onsets than at later positions.

Similarly, Cutler, McQueen, Baayen and Drexler (1994) showed that word embedding is also common in spoken language. They demonstrated that more than 70% of the polysyllabic words in a corpus of spoken English contained embedded words with aligned syllable boundaries. Moreover, when syllable boundaries were not taken into consideration in the analysis, more than 90% of all the words contained embedded words.

In this section, the following question is addressed: how often did words that occurred in each condition also occurred as substrings of longer words in that condition?

3.5.1 Methodology

As in the Cutler et al. (1994) study, word embedding was examined twice: a *phoneme analysis* and a *syllable analysis*. In the phoneme analysis, words that occurred as substrings of longer words (the matrix words) were counted as embedded words irrespective of syllable boundaries. For instance, *bang* /bɑŋ/ 'scared' and bank /bɑŋk/ 'bench' were both counted as embedded words in *bankje* /bɑŋk-jə/ 'little bench' or *oor* /or/ 'ear' in *hoort* /hort/ 'hears'). In the syllable analysis, only words that were exactly aligned with syllable boundaries were counted. Thus, bank /bɑŋk/ 'bench' in /bɑŋk-jə/ 'little bench' was counted as an embedded word, but *bang* /bɑŋ/ 'scared' was not. The phoneme analysis naturally results in a higher proportion of embedded words. Many of these embedded words are potentially not very confounding for the listener because either the remaining string of the matrix word starts with an illegal phonotactic sequence (as /kjə/ in /bɑŋk-jə/ when /bɑŋ/ is taken out), or the remaining string is not a possible word (as /h/ or /t/ in /hort/ when /or/ is taken out).

For the phoneme analysis, all word types that occurred in each condition were used. For the syllable analysis, only polysyllabic words were used. Single phoneme words (e.g., *oh*, *ah*, etc.) were excluded from both analyses.

3.5.2 Results

Phoneme Analysis

In the phoneme analysis, a total of 1118 word types in the AI condition were used, 2835 in the AC condition, and 5418 in the AA condition. Each word was paired with the embedded words that it contained. The results, displayed in Table 3.11, showed that 355 word types in the AI condition (31.8%) contained no embedded words, and 763 (68.2%) contained one or more embedded words. In the AC condition, 541 word types (19.1%) contained no embedded words, and 2294 (80.9%) contained one or more embedded words. In the AA condition, 769 word types (14.2%) contained no embedded words, and 4649 (85.5%) contained one or more embedded words.

Table 3.11: Frequencies of word embedding in the phoneme analysis. Percentages are given between parentheses.

	AI		AC		AA	
0 embedded words	355	(31.8)	541	(19.1)	769	(14.2)
at least 1 embed. word	763	(68.2)	2294	(80.9)	4649	(85.8)
	1118		2835		5418	

The frequencies were compared, using a chi-square test. The overall chi-square was highly significant: $\chi^2(2) = 200.40$, $p < .05$. Two additional comparisons were carried out by partitioning the chi-square value into two components. The first component (χ^2_1) was associated with the difference between the AI and the AC condition. The second component (χ^2_2) was associated with the difference between the AI and the AC conditions on the one hand, and the AA condition on the other hand. Both comparisons were significant: $\chi^2_1(1) = 88.10$, $p < .05$; $\chi^2_2(1) = 112.30$, $p < .05$. Thus the percentage of matrix words that contained no embedded words was significantly higher in the AI speech than in the AC speech, and it was significantly higher in the AC speech than in the AA speech.

Syllable Analysis

After exclusion of the monosyllabic words, 696 word types in the AI condition remained, 2007 word types in the AC condition, and 4364 word types in the AA condition. Each word was paired with the same possible embedded words that were used in the phoneme analysis.

The results, displayed in Table 3.12, showed that 299 word types (43.0%) in the AI condition contained no embedded words, and 397 word types (57.0%) contained one or more embedded word. In the AC condition, 635 word types contained no embedded words (31.6%), and 1372 word types contained one or more embedded words (68.4%). In the AA condition, 1068 word types (24.5%) contained no embedded words, and 3296 (75.5%) contained one or more embedded words. Note that the differences in percentages of embedded words between the phoneme analysis and the syllable analysis are comparable across the conditions. In the AI condition, the difference was roughly 11%; in the AC condition, it was roughly 12%; in the AA condition, it was roughly 10%.

The frequencies were again compared using a chi-square test. The overall chi-square was highly significant: $\chi^2(2) = 116.17$, $p < .05$. Two additional comparisons were carried out by partitioning the overall chi-square value into two components. The first component (χ^2_1) was associated with the difference between the AI and the AC condition. The second component (χ^2_2) was associated with the difference between the AI and the AC conditions on the one hand, and the AA condition on the other hand. Both comparisons yielded significant differences: $\chi^2_1(1) = 32.62$, $p < .05$; $\chi^2_2(1) = 83.55$, $p < .05$.

Thus the percentage of words that contained no embedded words was significantly higher in the AI condition than it was in the AC condition, and the percentage of words that contained no embedded words was significantly higher in the AC condition than it was in the AA condition.

Table 3.12: Frequencies of word embedding in the syllable analysis. Percentages are given between parentheses.

	AI		AC		AA	
0 embedded words	299	(43.0)	635	(31.6)	1068	(24.5)
1 or more embed. words	397	(57.0)	1372	(68.4)	3296	(75.5)
	696		2007		4364	

3.5.3 Summary

The results showed that word embedding was common in all the conditions including the AI condition. In the phoneme analysis, almost 70% of all the polysyllabic words in the AI speech contained at least one other word, and in the syllable analysis, almost 60% contained at least one word. However, these percentages were still significantly lower than those in the AC and the AA conditions which indicates that the speech in the AI condition consists of a less confusing sample for the infant than speech in the AC or the AA conditions.

The most frequent words that were embedded in the matrix words were short function words such as *te* /tə/ 'to', *je* /jə/ 'you', *'s* /əs/ 'once', etc. which in Dutch can also serve as affixes in longer words, e.g., *te*vreden 'satisfied', boek*je* 'little book'.

The percentages of words containing no embedded words was higher in the syllable analysis than in the phoneme analysis. The main cause of this difference is that in the syllable analysis, the residue of the matrix word must consist of at least a syllable, and, therefore, must contain at least one vowel. The fact that the differences between the results of the phoneme analysis and of the syllable analysis were comparable across the conditions indicates that there was no *extra* advantage in the AI speech in that a smaller number of words was embedded in longer words aligned with syllable boundaries. One could argue that leaving a residue that contains no vowel would be less confusing than a residue that does. The reason is that the presence of a vowel is a strong constraint as to whether a certain sound string is a word or not (Norris, McQueen, Cutler, and Butterfield, 1997). This constraint is, for example, applied to Brent and Cartwright segmentation model (Brent and Cartwright, 1996) described in Chapter 5. If the difference between the results of the phoneme analysis and the syllable analysis had been much bigger in the AI condition than in the other conditions (because fewer words had residues containing a vowel) then word embedding would have been less confusing for lexical segmentation. However, the results showed that this was not the case.

3.6 Conclusions

Five aspects related to word segmentation were analyzed. These aspects were compared across the three addressee conditions. The first aspect was MLU. The results showed that the AI speech consisted for almost 40% of one-word utterances. Although this is a very high proportion, almost all of these one-word utterances were vocatives, fillers, social expressions. When these were excluded,

less than 7% were one-word utterances, a percentage that was about equal to the one found in the AA condition, whereas the percentage in the AC condition was higher. Thus, it seems that the adults in this study did not tend to present words in isolation to the infant, as a kind of word-teaching strategy. It might very well be that this kind of strategy is adopted when child become a little older. As was described earlier, the period between six and nine months is a period that precedes early word comprehension. Caretakers are aware of their children's abilities and do not adapt their speech style. This was further confirmed by the MLU values in the AI and the AC conditions. Calculated over declaratives, interrogatives, and imperatives only ('meaningful' utterances), MLU was not significantly different in the AI and the AC conditions.

The second aspect that was analyzed was vocabulary size. The results showed that in this period of 18 days, the total vocabulary size of speech to the infant consisted of about 2081 word types, distributed over more than 41000 word tokens. Stated otherwise, every word recurs once every 20 words on average. The vocabulary size in the AC condition was also larger, but the sample size in this condition was also larger. However, the growth curves showed that the vocabulary size of the AI speech grew less rapidly than that of the AC speech. The same result appeared in the calculation of the type-token ratios.

Thus, the adults in this study used a very restricted vocabulary when addressing the infant. Obviously, this must facilitate word segmentation. A small vocabulary means that the same strings occur in the input very frequently. If the child is able to recognize these strings, that might be a possible bootstrap to isolate new strings. This proposal has recently been developed in a model of word segmentation (see: Brent and Cartwright, 1996; Chapter 5 of this thesis). The combination of the results of MLU and vocabulary size can have an additive effect: there was a high proportion of one-word utterances which drew on a very restricted vocabulary. These items also occur as part of longer utterances which may therefore be recognized relatively easily. However, it is important to realize that the child has to learn the principle that utterances are built of words which can occur at various places and in different contexts.

The third aspect that was analyzed concerned metrical structure. It was assumed that strong syllables can function as cues to word boundaries, since lexical items more often start with a strong than with a weak syllable. Indeed, a very high proportion of lexical items starting with a strong syllable was found. However, the percentage of all strong syllables that did not occur as the onset of a lexical item (but as non-initial syllables in lexical words, or in grammatical words) was even higher. Therefore, the number of misses would be greater than the number of hits. Thus, the quality of the syllable does not seem to be a cue that is very useful for infants as an initial bootstrap into the lexicon.

The fourth aspect that was analyzed was phonotactic structure. The number of possible word onsets and word offsets was relatively low in the AI condition. There were 49 possible onsets, and 44 possible offsets. These numbers are still quite high, but it was shown that the 20 most frequent onsets and offsets accounted for more than 95% of all the tokens. In order to determine whether these offsets and onsets were potential cues to word boundaries the consonants that surrounded the word boundaries were isolated from the words. Then it was determined whether the boundary could be precisely located on the basis of the onsets and the offsets and on the basis of the word-internal consonant clusters. In the AI speech, 25% of all the utterance-internal word boundaries could thus be located. Although this was much higher than the percentages in the AC and the AA speech, this is still not a very high proportion. A potentially better cue (to be examined in future research) would be the transitional probabilities of consonant clusters. The sequence /sf/, for example, would have a relatively low transitional probability because /s/ can be followed by a variety of other sounds and /sf/ occurs only rarely as the onset of a word. Therefore it would be more plausible, in an uncertain case, to locate a word boundary between the /s/ and the /f/. Adopting this strategy could result in higher percentages of correct predictions of word boundaries than the strategy adopted in the present study.

Finally, the results of the analysis of word embedding showed that words in the AI condition had significantly less often other words within them, but that there was no extra advantage in the AI condition when all the embedded words that were not aligned with syllable boundaries were excluded. Furthermore, the percentage of words that did have embedded words was still considerable. How confusing word embedding is for infants is an experimental question that might be addressed in future research.

Chapter 4
Suprasegmental Structure

In this chapter, three aspects of the suprasegmental structure of the input are described: pitch, intonation, and speech rate. These three aspects were measured in speech of the three main speakers in this study (the mother, the father, the baby sitter) while they addressed the infant, the older child, and other adults. The results revealed that, when compared to the adult-directed speech, the speech to the infant and to the older child showed modifications in each of these three aspects of the suprasegmental structure. The differences between speech to the infant and speech to the older child were not always consistent across the three speakers: some of the modifications were more exaggerated in speech to the infant, some in speech to the older child. Thus, in general, the results are consistent with previous findings. The possible functions of the suprasegmental modifications are discussed.

4.0 Introduction

This chapter deals with the *suprasegmental structure* of the input, that is: the acoustic aspects of the speech signal that are, so to speak, above the segments, because they are not necessarily determined by the characteristics of the utterance itself. Aspects of the suprasegmental structure can be divided into three groups: pitch, duration, and loudness. *Pitch* is the subjective correlate of changes in the fundamental frequency (F0) of the speaker's voice. Women's voices have higher pitch than men's voices, because their vocal cords are smaller and hence vibrate faster. The melodic line that reflects the changes in F0 over an utterance is called the *intonation contour*. Variation in *duration* is the temporal manifesta-

tion of suprasegmental structure. Differences in duration (of segments, words, etc.) can be caused by a number of factors: individual differences (fast speakers versus slow speakers), speech rate, speech style (reading versus spontaneous speaking), and other factors, such as speaking under pressure, or being tired. *Loudness* is the perceptual correlate of amplitude, the acoustic energy of a signal. Like pitch, loudness is determined by individual differences, or differences in situation (the distance between speaker and listener for example), but numerous other factors can be thought of that make someone talk with a loud voice or not.

Thus, here we do not talk about differences in pitch, duration, or loudness that cause changes in lexical meaning, as is the case in shift of lexical accent (voor*ko*men, 'to prevent' versus *voor*komen, 'to occur'), or differences between vowels (short vowels versus long vowels). Nor will we be concerned with the linguistic functions of intonation contours, although contour differences will be analysed. Suprasegmental structure is the aspect of speech that can distinguish different speech styles, different speakers, different emotions, etc. These functions are what will mainly concern us here.

A number of studies have reported differences in the suprasegmental structure of infant-directed speech and adult-directed speech in a variety of languages (see for example: Grieser and Kuhl, 1988; Garnica, 1977; Fernald and Simon, 1984; Fernald, Taeschner, Dunn, Papoušek, Boysson-Bardies and Furui, 1989). These studies focused mainly on pitch, intonation, and tempo. The results are described in the following three subsections.

4.0.1 Pitch

The first and probably the most salient characteristic of the suprasegmental structure of infant-directed speech concerns its pitch characteristics. Garnica (1977) measured pitch in the speech of mothers addressing an adult or a child. The children were divided into a group with an average age of 2;3 years, and a group with an average age of 5;4 years, in order to test whether the mothers adjust their way of speaking depending on the age of the child. The average F0 value of speech to the adult was around 200 Hz, to the five-year-olds it was 206 Hz, to the two-year-olds it was 267 Hz. The average F0 of speech to the five-year-olds was not significantly different from that of speech to the adults, but the average F0 of speech to the two-year-olds was significantly higher. Furthermore, the frequency range was expanded in speech to the children of both age groups. The lower limit of the range was more or less the same, irrespective of the addressee, probably because that value is at the floor of the speakers' pitch

range. The upper limit was shifted upwards in speech to the two-year-olds, and, to a lesser extent, in speech to the five-year-olds. Thus, pitch variation in speech to the children was bigger than pitch variation in speech to the adults.

Fernald and Simon (1984) obtained similar results in a study with German mothers addressing their newborn infants or an adult. The mothers spoke to the infants with a higher average pitch (257 Hz) compared to 203 Hz in speech addressed to the adult. Furthermore, the pitch variability (expressed in semitones per second) was significantly higher in the infant-directed speech than in the adult-directed speech.

Grieser and Kuhl (1988) also measured average frequency and frequency variability in the speech of Chinese-speaking mothers addressing adults and infants. Chinese was studied because it is prosodically very different from English. In spite of this difference, the results showed a very similar pattern to the one that was found in the previous study. The average frequency of the infant-directed speech was 245 Hz, and that of the adult-directed speech was 199 Hz. Furthermore, the frequency range of the infant-directed speech was expanded compared to that of the adult-directed speech.

Finally, Fernald et al. (1989) investigated whether this pattern of pitch modification was generally true across languages. They measured average F0 and F0 variability in the speech of mothers and fathers to preverbal infants and to adults. The languages that were examined were: French, German, British English, American English, Italian and Japanese. The results of this study showed that both mothers and fathers addressed the infants with a higher average F0 and with greater F0 variability when speaking to the infants than when speaking to the adults. This pattern was observed for all speakers and across all the languages, although the size of the modifications found in infant-directed speech was not equally large in every language.

In sum, infant-directed speech is produced with a higher average pitch and with greater pitch variability. This is true of speech directed to children in the age range of zero to five years, and in a variety of languages. That this might not be the case in absolutely every language is suggested by Ratner and Pye (1984) who studied Quiche Mayan, a language spoken in Guatemala. They found *lower* average pitch in language addressed to children, which suggests that the prosodic modification is, at least partly, culturally defined.

4.0.2 Intonation

A second characteristic of the suprasegmental structure of infant-directed speech concerns intonation. A few studies have suggested that intonation contours in

infant-directed speech differ in two ways from those in adult-directed speech (e.g., Stern, Spieker and McKain 1982; Fernald and Simon, 1984; Grieser and Kuhl, 1988). Firstly, the shape of the contours in infant-directed speech is simpler than that of the contours in adult-directed speech. Secondly, the intonational movements of the contours in infant-directed speech are more exaggerated (larger, slower) than those in adult-directed speech.

For instance, Fernald and Simon (1984) compared intonation contours in the speech of the German mothers to the infants and adults. The intonation contours of the utterances were classified as 'expanded' contours or 'non-expanded' contours. An expanded contour was defined as a contour that had a frequency change larger than six semitones per second (a semitone is a logarithmic measure that reflects the perceptual difference between two frequencies). Thus, the measure of semitones per second reflects how quickly F0 changes during the course of an utterance. Using this criterion, Fernald and Simon found that 59% of the intonation contours in speech to the newborns were expanded contours, whereas only 6% of the intonation contours in speech to the adults were expanded. Furthermore, they found that the expanded contours of the infant-directed utterances could be classified according to their shape using only a few different types: rising contours, falling contours, flat (level) contours, bell-shaped contours, complex (multi-directional) contours.

Grieser and Kuhl (1988) tried to replicate these findings in their analysis of the Chinese-speaking mothers. They found that, using the same criterion of six semitones per second, 78% of the contours of the infant-directed utterances were expanded. The shape of these expanded contours could be classified with the following five types: rising, falling, flat, bell-shaped and complex. However, in this study it was found that 56% of the adult-directed utterances were also expanded, a percentage much higher than the one reported by Fernald and Simon. Grieser and Kuhl do not give an explanation for the relatively high percentage of expanded contours in their sample of adult-directed speech. However, the average duration of the adult-directed utterances in their study is lower than that in Fernald and Simon's study, which naturally inflates the ratio of semitones per second.

4.0.3 Tempo

A third characteristic of infant-directed speech concerns tempo. Fernald and Simon (1984) calculated that the articulation rate of speech to the newborns was 4.2 syllables per second, as compared to 5.8 syllables per second in speech to the adults. Furthermore, pauses between utterances in speech to the newborns

were longer (1.5 sec.) than in speech to the adults (0.8 sec.). Papoušek, Papoušek and Haekel (1987) report that the articulation rate of German mothers and fathers to their three-months-old infants was 3.8 syllables per second. Finally, Grieser and Kuhl (1988) also found that pauses between utterances were longer in the infant-directed speech (1.1 sec.) than in the adult-directed speech (0.8 sec.), but they did not measure speech rate.

Thus, these studies have found that tempo of infant-directed speech is relatively slow because the articulation rate is low and because pauses between utterances are much longer in infant-directed speech than in adult-directed speech, thus reducing the number of utterances per minute.

In sum, findings about the suprasegmental structure are very consistent in the reported studies. Infant-directed speech has higher average pitch, more pitch variation, simpler and expanded intonation contours, and slow tempo. But what purpose do these modifications serve? Generally, two explanations for the distinct suprasegmental structure of infant-directed speech are given: an attentional-social explanation and a linguistic explanation. Note that these two explanations are not alternatives. Both might be true at the same time.

Attentional-social Explanation

According to the attentional-social explanation, the prosodic modifications merely serve to elicit infants' attention or are a way to convey positive affect towards the infant. In other words, they are used to interact with the infant, especially with preverbal infants who do not yet respond to segmental structure.

Evidence for this explanation comes from preferential listening studies, such as those described in Chapter 1. Fernald (1985) showed that infants respond differentially to infant-directed speech than to adult-directed speech because of its suprasegmental structure. In a preferential looking paradigm infants preferred to look (listen) to a loudspeaker that played infant-directed speech than to a loudspeaker that played adult-directed speech. Subsequently, Fernald and Kuhl (1987) showed that it was the large pitch variation that caused this preference. If it was the high average pitch, they argued, infants would have preferred to listen to a woman's voice rather than to a man's voice, but no preference was found. Furthermore, infants did not prefer to listen to a high monotone voice over a normal voice.

Kitamura (1994) investigated whether the attentional or the affective component of infant-directed speech is more important. For this purpose, adults rated speech samples addressed to five-month-old and to twelve-month-old infants as to how affectionate these samples sounded to them. The samples varied in average pitch and pitch variation. The results showed that the adults' rating of

perceived affection did not depend on the pitch of the samples, but on the age of the infant that the speech was directed to: The speech samples to the five-month-olds were rated as more affectionate than the samples to the twelve-month-olds. Subsequently, the preference of a group of preverbal infants was tested to listen to selected samples that the adults had rated in the first part of the experiment. The pitch of these samples as well as the affectionate ratings of the adults were systematically varied in order to establish the influence of each on the infants' preference. The results showed that this preference was determined both by ratings of how affectionate the speech sounded and by pitch. Thus, the infants preferred to listen to samples that were rated as more affectionate than to samples that were rated as less affectionate, but this was only true when these samples also had a high average pitch.

Finally, Fernald (1993) found that mothers used different contours to express different attitudes (prohibition and approval) and that infants responded differently to different contours. The infants more often showed a positive reaction (smiling) to the approval utterances and more often a negative reaction (frowning) to the prohibition utterances. Similarly, Stern, Spieker and MacKain (1982) found that mothers used different intonation contours depending on context and utterance type. So, for example, if the infant gazed away mothers would use a rising contour to attract the infant's attention.

Linguistic Explanation

According to the linguistic explanation, the exaggerations in the suprasegmental structure of infant-directed speech might also have a facilitative effect on early language acquisition because they facilitate organization of the input. Since the intonation contours are relatively simple and the utterances are more clearly separated by pauses, the identification of utterance boundaries is relatively simple.

Evidence for the linguistic explanation is given by Kemler-Nelson et al. (1989) who found that infants at the age of nine months perceived clauses as perceptual units. The infants preferred to listen to speech that was interrupted at clause boundaries over speech that was interrupted within clauses. However, this result was only obtained when the stimuli were spoken in an infant-directed speech style and not when the stimuli were spoken in an adult-directed speech style (see also Chapter 1).

Similarly, Fernald and Mazzie (1991) showed that prosodic emphasis was much more consistent in infant-directed speech than in adult-directed speech. In the infant-directed speech, focused words were positioned on pitch peaks more often than in adult-directed speech.

4.0.4 Focus of this Chapter

Are the reported characteristics of the suprasegmental structure of infant-directed speech also true in everyday life, in a great variety of situations? The results of the reported studies are all based on short situations in which the caretakers are asked to interact with their children. The obvious advantage of such a setup is that disturbing factors (noise) can be controlled for as much as possible. Nevertheless, it is not necessarily the case that deliberate interaction for a recording session produces the suprasegmental characteristics of natural speech in uncontrolled situations.

In this chapter the same three aspects of suprasegmental structure are analyzed in the real-life corpus: pitch, intonation, and speech rate. Measurements were taken in these three dimensions for each of the three main speakers in the corpus: the mother, the father, and the baby sitter, while they addressed another adult (AA condition), the infant (AI condition), or the older child (AC condition). For the acoustic measurements, it was necessary to select a number of utterances that were relatively free of background noise. The selection of the utterances is described in section 4.1. This section is followed by an analysis of pitch (section 4.2), intonation (section 4.3), and speech rate (section 4.4).

4.1 Selection of the Material

A quasi-random sample was drawn from the entire database. This was not a straightforward procedure because the choice of the material was restricted by the following four factors: First, there was a lot of overlap between speakers. This was especially true in the AA speech, where, for example, an utterance of a speaker was often interrupted by an utterance of the addressee. Second, there were differences in the acoustic quality of the utterances in each addressee condition. The utterances directed to the infant were mostly spoken close to the microphone which made the acoustic quality of speech in this condition relatively good. The distance between the microphone and an adult speaking to another adult was often much bigger (sometimes even in an adjacent room), which reduced the quality of the speech in this condition. Third, some utterances occurred much more frequently than others. For instance, discourse markers occurred much more frequently than declaratives. This restricted the choice of utterance types. Fourth, long utterances are disrupted more often by noise than short utterances, which restricted the selection of longer utterances.

The material that was used for the analyses in this chapter was selected using the following guidelines: The utterance had to be free of background

noise; approximately 50 utterances per day per speaker in each addressee condition were selected; utterances were selected spread over the day; a variety of different utterances were selected as much as possible.

A total of 5789 utterances were finally selected: 1883 AI utterances, of which 734 were from the mother, 574 from the father, and 575 from the baby sitter; 1412 AC utterances of which 493 were from the mother, 502 from the father, and 417 from the baby sitter; 2493 AA utterances, of which 874 were from the mother, 1113 from the father, and 506 from the baby sitter. These utterances were extracted from the original material and stored as separate digitized sound files for further processing.

4.2 Pitch

Following previous studies, the following measures that describe pitch were determined: average, minimum, maximum, range, and standard deviation of F0. These measures were computed by the speech editor (the ESPS waves speech editing program). Since it is not desirable to average frequency measures of male and female voices, the three speakers will be separated in the presentation of the results. Differences as a function of addressee will be discussed.

4.2.1 Methodology

For the pitch analysis, intonation contours were created with the pitch-tracking algorithm of the speech editor. The intonation contours were created first because the pitch values are calculated based on the estimated F0 trace. Since pitch-tracking algorithms sometimes produce incorrect estimates of the F0, the contour of every utterance was checked, so that only those utterances with correct and intact intonation contours were included in the pitch analysis. The checking of the contours was done as follows.

Intonation contours were visualized on the computer screen aligned with the original wave form of the signal. The estimated F0 was plotted every 10 ms against a linear scale from 50 to 650 Hz. Using the default settings, possible F0-values lie between 50 and 650 Hz. Contours could be inaccurate for the following reasons: the contour was discontinuous (for example when the women's voices exceeded the higher limit of the frequency range); the estimated pitch of the entire contour was wrong (this could be checked by estimating the frequency on the basis of the length of a period from the oscillogram); parts of

the contour that did not occur as a consequence of voiceless segments in the signal or the entire contour were missing.

Where possible, discontinuous contours were corrected by changing the parameters that are used for the estimation of F0. Although it is usually not recommended to manipulate these parameters, in this case it seemed justified for two reasons. The first reason is that infant-directed speech differs from normal speech in that it is produced at a relatively high average pitch with large high pitch excursions. The women's voices were often in the high region of the default region and sometimes even exceeded the maximum default value of the pitch estimator. In these cases, the resulting F0 trace often contained a discontinuity - an 'octave error', in which the estimated pitch dropped by one or two octaves. Therefore, adjusting the parameters provided a correction of the pitch contour. The second reason for changing the F0 parameters is that the acoustic quality of the material was not as good as a studio recording. This resulted a number of times in inaccuracies which could be restored after manipulation. In the cases in which parameter settings were changed, however, care was taken that this did not affect the accurately analysed portions of the original F0 trace.

Errors could usually be corrected by making the frequency range of the pitch extractor wider or narrower. If the contour remained incorrect or incomplete the utterance was not used for the pitch analysis. Contours with incidental outliers were also removed, unless the outlier fell within the frequency range of the rest of the utterance. The average pitch, minimum and maximum, and standard deviation (Hz) were calculated by the speech editor. These values were computed across the voiced portions of the signal only.

4.2.2 Results

A total of 1624 utterances (28%) were excluded because of inaccurate contours: 272 from the mother (13%), 1079 from the father (49%), and 273 from the baby sitter (18%). Clearly, the pitch tracker had more trouble in estimating the father's voice than the women's voices. The father's voice seemed to be relatively soft. This could be either a characteristic of his voice, or, less likely, could reflect the fact that he was usually further away from the microphone. However, the resulting low signal to noise ratio might have been the cause of the fact that in many instances, the estimated pitch contour showed unacceptable inaccuracies.

The results of the calculation of the pitch measures, calculated over the remaining 4165 utterances with good contours are given in Table 4.1. In this table, M stands for mother, F stands for father, and B stands for baby sitter. The

first three rows list the average values of mean F0, minimum F0, and maximum F0, as determined by the speech editor. In the next two rows, two average values for the pitch range are listed, one in Hz (the difference between minimum and maximum), and one in semitones (calculated as: $12 * ^2\log$ F0-max/F0-min). The next row gives the mean values of standard deviation F0, determined by the speech editor.

Table 4.1: Overview of pitch measures, averaged per speaker (M stands for Mother, F for father, B for baby sitter).

	AI			AC			AA		
	M	F	B	M	F	B	M	F	B
Average (Hz)	279	157	303	285	143	282	232	130	225
Minimum (Hz)	228	126	234	227	120	202	196	111	180
Maximum (Hz)	349	190	386	363	173	373	291	157	276
Range (Hz)	122	64	151	136	53	171	95	46	96
Range (semitones)	7.22	7.14	8.89	7.86	6.36	10.8	6.64	5.80	7.27
Std (Hz)	36	19	47	39	15	54	25	13	28
Nr. of utterances	622	282	518	435	293	364	772	536	343

These values were subjected to statistical testing. The analyses were done for each speaker separately, using oneway analyses of variance. Post hoc comparisons were done, using the Tukey HSD procedure ($p < .05$). The results are described for each speaker separately.

Mother's speech

The average pitch of the mother's utterances differed significantly across the three conditions: $F[2,1826] = 244.19$, $p < .01$. The difference between AI speech and AC speech was not significant however, according to the post hoc comparisons. The lower limit of the pitch range (F0 min) differed significantly across the addressee conditions: $F[2,1826] = 145.79$, $p < .01$. Again, F0 min of the AI speech and the AC speech were not significantly different from each other, but F0 min of the AA speech was significantly lower. The upper limit of the range (F0 max) was significantly different across the conditions: $F[2,1826] = 141.78$, $p < .01$. F0 max of the AC speech was significantly higher than that of the AI speech, and F0 max of the AI speech was significantly higher than that of the AA speech. The pitch range (Hz) also differed significantly across the

three conditions: $F[2,1826] = 51.37$, $p < .01$. The AC speech had a significantly bigger range than the AI speech, and the AI speech had a significantly bigger range than the AA speech. The range in semitones revealed the same pattern. The differences were significant across the conditions: $F[2,1826] = 16.15$, $p < .01$. The difference between the AI speech and the AA speech was significant, as well as the difference between the AI speech and the AA speech. Finally, the differences in standard deviation (F0 std) were significant across the conditions: $F[2,1826] = 65.36$, $p < .01$. F0 std of the AC speech was significantly higher than that of the AI speech, and F0 std of the AI speech was significantly higher than that of the AA speech.

In sum, in the mother's speech, the highest average pitch, the highest pitch range, and the highest F0 standard deviation were found in the AC speech. In other words, from these results it seems that the mother addressed her older daughter in a somewhat livelier voice than when she addressed her youngest daughter. The values that were found in her speech to the younger daughter were nevertheless all higher than the values that were found in her speech when she addressed an adult.

Father's speech

The average pitch of the father's speech was significantly different across the three conditions: $F[2,1108] = 81.86$, $p < .01$. The average pitch of the AI speech was significantly higher than that of the AC speech, and the average pitch of the AC speech was significantly higher than that of the AA speech. F0 min was significantly different across the three conditions: $F[2,1108] = 33.51$, $p < .01$. F0 min of the AI speech was significantly higher than that of the AC speech, and F0 min of the AC speech was significantly higher than that of the AA speech. F0 max was also significantly different across the conditions: $F[2,1108] = 61.24$, $p < .01$. F0 max of the AI speech was significantly higher than that of the AC speech, and F0 max of the AC speech was significantly higher than that of the AA speech. The difference in range in speech of the father was also significant: $F[2,1108] = 24.19$, $p < .01$. The range (Hz) of the AI speech was significantly bigger than that of the AC speech, and the range of the AC speech was significantly bigger than that of the AA speech. The range in semitones revealed a similar pattern. The differences were significant across the conditions: $F[2,1108] = 11.33$, $p < .01$. The range of the AA speech was significantly lower than that of the AI speech and the AC speech; the difference between the AC speech and AI speech, however, failed to reach significance. Finally, F0 std was significantly different across the conditions: $F[2,1108] = 31.46$, $p < .01$. F0 std

of the AA speech was significantly *lower* than that of the AC speech, and F0 std of the AC speech was in this case significantly *lower* than that of the AI speech.

Thus, contrary to what was found in the mother's speech, in the father's speech the highest values were found in the AI speech, and not in the AC speech. Thus, the father seemed to address the younger daughter in a somewhat livelier voice than the older daughter.

Baby sitter's speech

The average pitch of the baby sitter was significantly different across the conditions: $F[2,1222] = 86.06$, $p < .01$. The AI speech had a significantly higher average pitch than the AC speech, and the AC speech has a significantly higher pitch than the AA speech. F0 min differed across the three conditions: $F[2,1222] = 54.49$, $p < .01$. F0 min of the AA speech was significantly lower than that of the AC speech, and F0 min of the AC speech was significantly lower than that of the AI speech. F0 max was significantly different across the three conditions: $F[2,1222] = 109.05$, $p < .01$. The AA speech had a significantly lower F0 max than the AI speech and the AC speech, but the difference between the AI speech and the AC speech was not significant. The range (Hz) of the baby sitter's speech was also significantly different across the conditions: $F[2,1222] = 59.40$, $p < .01$. The pitch range of the AC speech was significantly bigger than that of the AI speech, and the range of the AI speech was significantly bigger than that of the AA speech. The range in semitones showed the same pattern: the differences were significant across the conditions: $F[2,1222] = 35.14$, $p < .01$. The range of the AA speech was significantly lower than that of the AI speech, and the range of the AI speech was significantly lower than that of the AC speech. Finally, F0 std was significantly different across the conditions: $F[2,1222] = 60.45$. F0 std of the AA speech was significantly lower than that of the AI speech, and F0 std of the AI speech was significantly lower than that of the AC speech.

In sum, the speech of the baby sitter showed a combination of the patterns that were found in the mother's speech and the father's speech. On the one hand, as with the father, the average pitch, F0 min and F0 max were all shifted upwards in the AI speech compared to the AC speech. On the other hand, as with the mother, the range and F0 std were both higher in the AC speech than in the AI speech. Thus, overall, the baby sitter spoke with a higher voice when addressing the infant, but used more variation when addressing the older sibling. When she addressed other adults, average pitch and pitch variation were all lower than when she addressed the children.

Thus, there were clear differences in average pitch and pitch variation in the three addressee conditions. However, differences in the pitch variation (F0 std) might have been caused by differences in utterance length, since long utterances are likely to have more variation than short utterances. In order to control for this potential confound, F0 std of utterances that were only one syllable long was computed. The results are shown in Table 4.2.

Table 4.2: F0 standard deviations (Hz) of utterances of one syllable length.

	AI	AC	AA
Mother	33.33	35.36	20.18
Father	22.12	11.31	10.77
Baby Sitter	35.83	57.64	16.19

Most of these values were somewhat lower than the values in the previous table. but the overall pattern of results was the same as for the entire sample. F0 std in the mother's speech was significantly different across the conditions: $F[2,482] = 17.41$, $p < .01$. The difference between AI speech and AC speech was not significant, according to post hoc comparisons, but F0 std of the AA speech was significantly lower that of the AI speech and AC speech. The differences between the conditions in the father's speech were also significant: $F[2,357] = 31.03$, $p < .01$. F0 std deviation of the AI speech was significantly higher than that of the AC speech and the AA speech, but the difference between the AC speech and the AA speech was not significant. Finally, the differences between the conditions in the baby sitter's speech were also significant: $F[2,339] = 36.34$, $p < .01$. F0 std of the AC speech was significantly higher than that of the AI speech, and F0 std of the AI speech was significantly higher than that of the AA speech.

4.2.3 Summary

The results showed that average pitch of the AI speech and the AC speech were higher than that of the AA speech. Pitch range of the AI speech and of the AC speech were also larger than that of the AA speech. These findings confirm previously reported results on pitch of child-directed speech. Thus, the main speakers in this study adjusted their speech style when addressing the infant or the older child. The differences between AC speech and AI speech were not

consistent across speakers. The mother spoke with a higher average pitch to the older child than to the infant (although this difference was not statistically significant) but the father and the baby sitter spoke with a lower average pitch to the older child than to the infant. Furthermore, the speech of the mother and the baby sitter to the older child had more pitch variation than speech to the younger child, but the speech of the father to the younger child had more variation than speech to the older child.

4.3 Intonation

In this section, intonation is discussed. Following previous studies, a contour classification was used, based on frequency and durational measures. This classification serves to distinguish contours that are typical for child-directed speech versus contours that are typical for adult-directed speech. The frequency of occurrence of the various types of contours will be considered.

4.3.1 Methodology

Every accurately traced contour (as described in the previous section) was categorized according to its shape and the expansion of the frequency range which it represented. The procedure described by Fernald and Simon (1984) served as an example for the present study. Following this study, contours were classified as 'expanded' contours or 'non-expanded' contours. Expanded contours are typical for infant-directed speech, non-expanded contours are typical for adult-directed speech. A contour is called expanded when it is characterized by one of the following three characteristics: (1) it has a frequency change larger than six semitones per second; (2) it is a falling or rising contour produced at an exceptionally high pitch level; (3) it is a flat (level) contour produced on an exceptionally long vowel[1].

In the original study, no criteria are given for the cases (2) and (3). In this study, the following criteria were adopted: Rising or falling contours that were produced at an average pitch level more than two standard deviations above the speaker's average pitch level in the AA speech were classified as expanded. Furthermore, level contours which were produced with a speech rate less than two syllables per second were classified as expanded. The procedure used to

[1] In the original study, whispered utterances were also included in the category of expanded contours, but this category is not included in this study.

determine speech rate and utterance duration is described in section 4.4. The expanded contours were then classified as to their shape, using the following types based on Fernald and Simon's study (also illustrated in Figure 4.1 on the following page):

1. Uni-directional rising contours
2. Uni-directional falling contours
3. Level contours
4. Bell-shaped contours: contours that have the shape of a bell or a U.
5. Complex contours: multi-directional contours

The frequency of occurrence of expanded and non-expanded contours will be considered in the next section, as well as the frequency of occurrence of each type of expanded contours. The utterances of the three speakers are not divided. The distribution of expanded versus non-expanded contours will first be described, followed by the distribution of the five expanded contour types.

4.3.2 Results

In total, there were 4165 accurately traced contours: 1422 in the AI speech, 1092 in the AC speech, and 1651 in the AA speech. In the AI speech, 1031 contours (72.5%) were classified as expanded, 391 (27.5%) were classified as non-expanded. In the AC speech, 772 contours (70.7%) were classified as expanded, 320 as non-expanded (29.3%). In the AA speech, 948 were classified as expanded (57.4%), and 703 as non-expanded (42.6%). Thus, the percentages of expanded contours in the AI speech and the AC speech were approximately equivalent, and the percentage of expanded contours in the AA speech was slightly (10-15%) lower. The independence of the frequencies with respect to addressee conditions was tested using a chi-square test. The resulting chi-square value was highly significant: $\chi^2(2) = 91.76$, $p < .01$, suggesting that the frequencies were not independent from the addressee conditions. An additional comparison was carried out by partitioning the chi-square value into two components (Everitt, 1977). The first component (χ^2_1) was associated with the difference between the AI and the AC condition. The second component (χ^2_2) was associated with the difference between the AI and the AC condition on the one hand, and the AA condition on the other hand. The first comparison was not significant: $\chi^2_1(1) = 0.900$, $p > .05$, but the second comparison was: $\chi^2_2(1) = 90.861$, $p < .05$. Therefore, contours in the AI and the AC conditions were expanded significantly more often than in the AA condition.

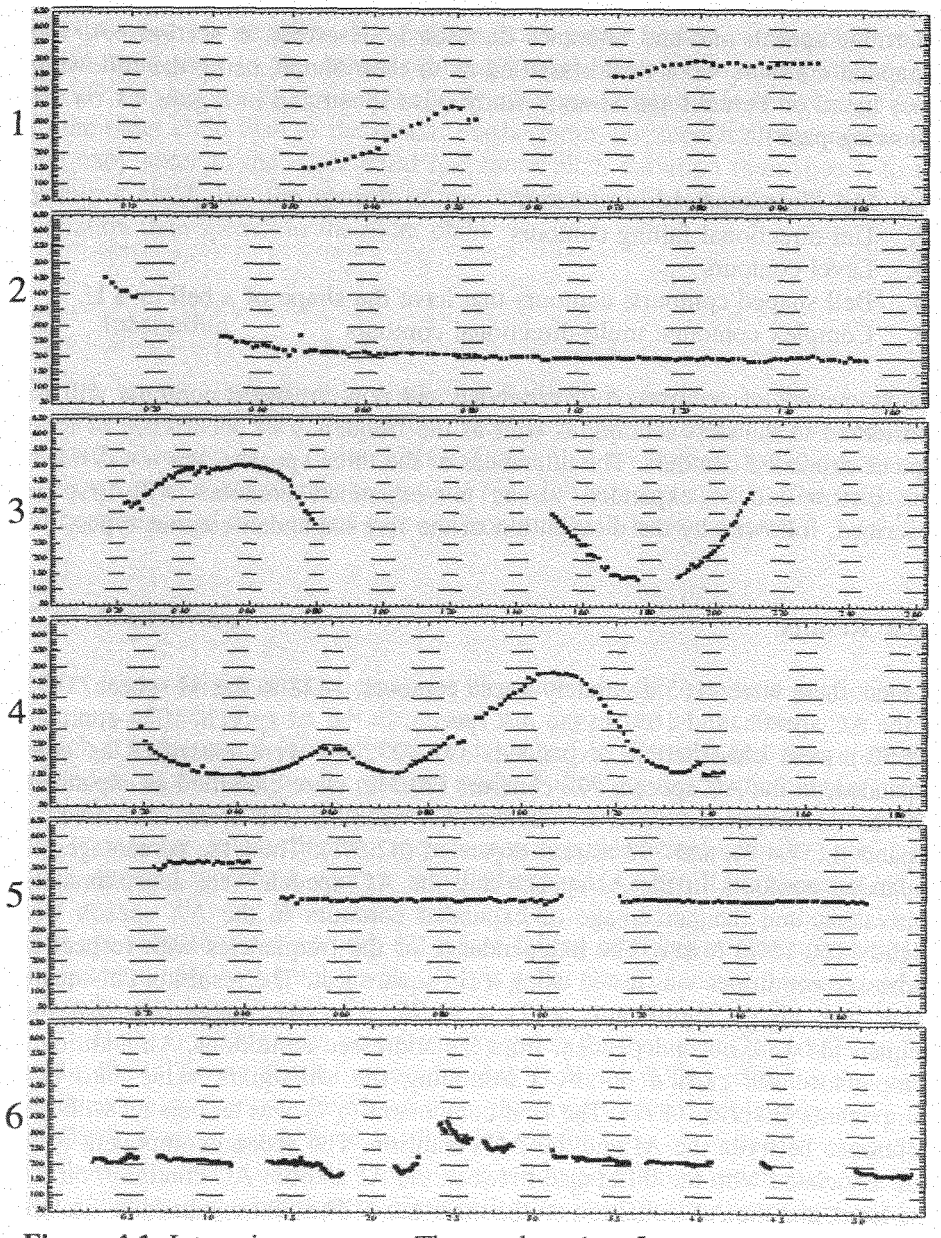


Figure 4.1: Intonation contours. The numbers 1 to 5 are expanded contours, number 6 is a non-expanded contour. The following types are diplayed: 1 rising, 2 falling, 3 bell-shaped (2), 4 complex, 5 level, 6 non-expanded.

Table 4.3: Distribution of expanded contour types. Percentages are given between parentheses.

	AI		AC		AA	
Falling	157	(15.2)	111	(14.4)	292	(30.8)
Rising	309	(30.0)	230	(29.8)	221	(23.3)
Bell-Shaped	178	(17.3)	80	(10.4)	71	(7.5)
Level	63	(6.1)	16	(2.1)	0	(0.0)
Complex	324	(31.4)	335	(43.4)	364	(38.4)
Total nr.	1031		772		948	

Next, the distribution of the five contour types in the expanded contours is examined. Table 4.3 shows the numbers and the percentages of these types in the three speech conditions. The independence of the frequencies and the addressee conditions was tested with a chi-square test. The resulting chi-square value was highly significant: $\chi^2(8) = 213.36$, $p < .01$, suggesting that again the frequencies of the contours were not independent of the addressee.

Overall, the most frequently occurring contour type was a complex (multi-directional) contour. The least frequently occurring contour type was a level contour, which was entirely absent in the AA speech. Bell-shaped contours occurred also relatively infrequently across the three conditions. The two remaining contour types varied across the conditions: In the AI speech and AC speech, the percentage of rising contours was higher than the percentage of falling contours, whereas in AA speech the reverse was true.

4.3.3 Summary

Taken together, a higher percentage of expanded intonation contours was found in the AI speech and the AC speech. The percentage of expanded contours in these two conditions was comparable to that found in previous studies by Fernald and Simon (1984) and Grieser and Kuhl (1988). The percentage of expanded contours in the AA speech, however, was much higher than that reported by Fernald and Simon, but a little lower than that reported by Grieser and Kuhl. Thus, the results of this analysis do not show the clear distinction between intonation contours that are typical for infant-directed speech versus contours that are typical for adult-directed speech that Fernald and Simon found. In their study, expanded contours were almost absent in adult-directed speech, whereas, in the present sample, still 57% of all the intonation contours were

classified as expanded. It is likely that the nature of the material plays an important role in this finding. The material used in Fernald and Simon's studies consisted of adults' responses in an interview situation while in the present study it consists of everyday speech. The utterances in the original study were therefore much longer than those in everyday conversation (the reported utterance duration was more than two seconds, in the present study it is only little more than one second, see next section). In spite of the effort taken to vary the kinds of utterances, the relatively short average duration of the AA utterances is probably partly due to the restrictions that were put on the selection of the material. On the other hand, many utterances in the entire corpus analyzed in this dissertation were very short, and, as will become clear in the next section, the average length of the selected AA utterances was in fact longer than the average length of the AA utterances in the corpus overall (see also Chapter 2).

The distribution of the different expanded contours showed similarities and differences across the conditions. In every condition, complex contours were the most frequently occurring contours. Furthermore, in every condition, bell-shaped contours and level contours occurred relatively infrequently. In the AI speech and the AC speech, rising contours occurred more often than falling contours. In the AA speech, falling contours occurred more often than rising contours. This finding is consistent with previous findings that rising contours occur frequently in child-directed speech.

4.4 Speech Rate

The last issue that will be examined is speech rate. Different measures of speech rate can be used. A global measure is words per minute. A more precise measure is syllables per second. It would be even more precise to exclude the final vowel and subsequent consonants, because final vowel lengthening is a confounding factor in measuring speech rate (Haselager, Slis and Rietveld, 1991). For the purpose of this study, however, a measure of syllables per second seems a sufficient standard on which to base calculation of speech rate.

4.4.1 Methodology

The beginning and the end of each utterance were marked with the speech editor, as were pauses within the utterance. This was done both acoustically and on the basis of the waveform of the speech signal. The utterance duration was defined as the total length of the utterance, pauses included. Vocalization time

was defined as the total length of the utterance minus the length of one or more pauses within the utterance. The number of syllables of each utterance was counted. Common contractions that occur in Dutch (e.g., *da's*, 'that's', instead of *dat is*, 'that is') were counted as one syllable. In other, less obvious, cases (e.g., *trug* instead of *terug*, 'back') the number of 'canonical' syllables was used (i.e., the number of syllables if the word had been pronounced carefully) for the computation of speech rate. When the content of the utterance could not be transcribed reliably, the number of syllables was set to a missing value. In a few cases, although it was not possible to determine what was said exactly, the number of syllables that were produced could still be heard. Speech rate was calculated as the number of syllables divided by the vocalization time in seconds.

4.4.2 Results

Many of the utterances that could not be used for the estimation of pitch and intonation could be used for the estimation of speech rate. Of the total of 5789 utterances in the selection, only 216 had to be excluded because the number of syllables could not be determined. 5573 utterances remained: 1861 in the AI speech, 1401 in the AC speech, and 2311 in the AA speech. A limited number of utterances contained pauses: only 4 utterances in the AI speech (0.2%), only 15 in the AC speech (1.1%), and 94 in the AA speech (4.1%). Thus, in most instances, utterance duration and vocalization time were the same. The average duration of the AI utterances was 0.89 seconds, the average duration of the AC utterances was 0.84 seconds, and the average duration of the AA utterances was 1.03 seconds. A oneway analysis of variance indicated that these differences were significant: $F[2,5786] = 43.44$, $p < .01$. Post hoc comparisons using the Tukey HSD procedure indicated that the difference in duration between AI speech and AC speech was significant, and that the difference between AC speech and AA speech was also significant. The average length of the AI utterances was 3.32 syllables; the average length of the AC utterances was 3.91 syllables; and the average length of the AA utterances was 5.64 syllables. The differences between the conditions were significant: $F[2,5570] = 188.24$, $p < .01$. Again, the difference between AI speech and AC speech was significant, as well as the difference between AC speech and AA speech.

Based on these measures, speech rate was calculated. The results are listed in Table 4.4. The values are broken down by utterance length (first column), since longer utterances tend to have faster speech rate. Speech rate of utterances longer than 10 syllables was averaged.

Table 4.4: Speech rate in syllables/second, as a function of addressee condition.

	AI			AC			AA		
Length (syl)	M	F	B	M	F	B	M	F	B
1	2.44	2.88	2.07	2.50	3.16	2.24	3.33	4.25	3.00
2	3.40	3.84	2.66	3.42	4.33	3.04	4.58	5.83	3.95
3	3.88	4.19	2.88	3.93	5.04	3.94	5.47	5.46	3.89
4	4.45	5.00	3.56	4.71	5.87	4.47	5.66	5.82	4.22
5	4.84	5.44	3.78	5.29	5.80	4.71	6.15	6.30	4.76
6	5.06	5.07	4.33	5.94	6.20	4.98	6.56	6.69	5.15
7	5.04	6.20	4.26	5.79	6.80	5.47	6.66	6.78	5.42
8	5.62	6.33	4.64	6.45	6.72	5.62	6.15	6.87	5.91
9	6.39	6.64	4.86	6.46	7.47	5.49	6.73	7.26	5.95
10 or more	6.20	7.28	5.30	6.19	7.60	5.71	6.43	7.03	6.21
Average	3.81	4.10	3.16	4.00	5.29	3.92	5.27	5.72	4.64

The average values were subjected to statistical testing. Post hoc comparisons were carried out, following Tukey's HSD procedure ($p < .05$). The differences were analysed for each speaker separately.

The differences in speech rate of the mother's speech were significant across the three conditions: $F[2,2009] = 166.32$, $p < .01$. The post hoc comparisons indicated that speech rate of the AI speech was not significantly different from that of the AC speech, but the speech rate of the AA speech was significantly faster. Speech rate of the father's speech was also significantly different across the conditions: $F[2,2075] = 149.58$, $p < .01$. Post hoc comparisons indicated that the speech rate of the AI speech was significantly slower than that of the AC speech, and the speech rate of the AC speech was significantly slower than that of the AA speech. Finally, the speech rate of the baby sitter's speech was also significantly different across conditions: $F[2,1480] = 291.44$, $p < .01$. According to the post hoc comparisons, the speech rate of her AI speech was significantly slower than that of her AC speech, and the speech rate of her AC speech was significantly slower than that of her AA speech.

4.4.3 Summary

The average utterance length and duration was significantly different across the three addressee conditions. The AI utterances had the shortest duration and length. The duration and length of the AC utterances were greater, and those of

the AA utterances were greatest. Across the three speakers, differences in speech rate were also found. The fastest speech rate was measured in the AA speech, and the slowest speech rate was measured in the AI speech. The difference between the AI speech rate and the AC speech rate was significant for two of the three speakers.

4.5 Conclusions

In this chapter, three aspects of the suprasegmental structure of the input were analyzed: pitch, intonation and speech rate. Previous studies have reported that these aspects of the suprasegmental structure are different in language spoken to infants or to young children compared to normal adult-directed speech. More specifically, in child-directed speech, the overall pitch is higher, the pitch range is expanded, and there is more variation in pitch. Furthermore, intonation contours have simple shapes and the F0 movements are bigger. Finally, speech rate of infant-directed speech is slower.

Based on the results of previous work, it was expected to find the same results in this study, since the suprasegmental modifications have been demonstrated in a wide variety of languages, including languages that are prosodically very different from American English and European languages.

In the present study, clear differences were found between the AA speech on the one hand, and the AI and the AC speech on the other hand, but the differences between the AI speech and the AC speech were not always consistent across the speakers. The differences between the AA speech and the AI and AC speech were always in the predicted direction, and therefore confirm existing results of acoustic analyses of child-directed speech. Thus, average pitch of the AA speech was relatively low; pitch range and pitch standard deviation were also relatively low. Furthermore, the percentage of so-called expanded intonation contours was significantly lower in the AA speech and the distribution of the five expanded contour types was different, mainly because of the absence of level contours in the AA speech, and because the percentage of falling contours was higher than that of rising contours. Finally, the AA speech was produced with a significantly faster speech rate than the AI or the AC speech.

The differences between the AI speech and the AC speech were not consistent across the three speakers. Although I did not make any specific predictions as to the direction of the suprasegmental modifications, it would seem more likely that these modifications were more exaggerated in the AI speech, since suprasegmental aspects of the utterance are attractive to infants and, therefore, play an important role in speech perception especially for

preverbal infants. Further support for this prediction comes from a study by Stern, Spieker, Barnett and MacKain (1983) that examined age-related changes. In this study, it was found that modifications were more exaggerated in speech to children at the age of four months compared to speech to children at the age of zero, twelve, or 24 months. The explanation that the authors give is that at four months of age, children are relatively much engaged in social interaction because, contrary to neonates, they can sit up and direct their gaze voluntarily towards the caretaker, and contrary to older infants, they cannot move around freely yet and therefore stay in closer contact to the caretaker.

The results of the present study showed that most aspects of the suprasegmental structure were indeed more exaggerated in the AI speech than in the AC speech. For instance, speech rate of the AI speech was slower than that of the AC speech, and, for two of the three speakers, average pitch was higher, pitch range was bigger, and pitch standard deviation was higher. However, the AC speech of the mother had higher average pitch than the AI speech, the standard deviation of the baby sitter's AC speech was higher than that of the AI speech, and the percentage of expanded intonation contours was more or less equal in both conditions.

Thus, the results point to the conclusion that there were multiple factors that influenced the suprasegmental structure. Some of these factors appear to be individually determined (as was also pointed out in the introduction), but other contextual variables, similar to those mentioned by Stern et al. (1983) also a play a role. It seems likely, for instance, that the distance between the adults and the child in this study was much more variable than the distance between adults and the infant. When the child was further away from the adult, the adult would 'speak up' for example, and raise his or her voice. Furthermore, it seems that pitch was more influenced by these multiple factors than intonation or speech rate since these last two aspects did not show the same variability in the results.

Do the present results support either of the two hypotheses about the function of suprasegmental exaggerations in child-directed speech: the linguistic explanation or the social/attentive explanation? The answer to this question is not an easy one, since predictions can be made in two directions and since the results of this study are not unambiguous.

On the one hand, one might argue that, if the suprasegmental modifications serve a linguistic function, they would be more exaggerated in the AI speech, since the relative contribution to the content or message of the utterance should be much higher in speech to the infant compared to speech to the older child to whom, between the age of 2;6 and 2;9 years, the lexical content of the utterance should have a relatively larger contribution. However, the same prediction can be made if the prosodic modifications serve an attentional or an

affective function. One might argue that these modifications are more exaggerated in the AI speech because, again, the older child could rely on the meaning of the utterance.

However, the results showed that average pitch of two of the three speakers was higher in the AC speech than in the AI speech, and pitch standard deviation was also higher in the AC speech of two of the three speakers. This is a remarkable finding, since pitch variation, rather than average pitch, is the feature that makes infant-directed speech sound attractive to infants (Fernald and Kuhl, 1987). The distribution of intonation contours did not vary remarkably between AI speech and AC speech. The percentage of expanded contours was more or less the same in both conditions, and the distribution of the five expanded contours was also very similar in both conditions. The speech rate of the AI speech was consistently slower in the AI speech.

The final question that needs to be addressed is: can these suprasegmental modifications have a facilitative effect on early word segmentation or on early phonological development? The only answer to this question that we can give at this stage is that the results of the present study showed that suprasegmental modifications were most clearly present in the AI speech. Given the fact that infants are attracted to this speech style, it might draw their attention to this speech and so, indirectly, have a linguistic function during this period of language development: infants listen more carefully to this speech style and thus learn to recognize the sound structure of the native language more easily than if the input had not had these modifications. Furthermore, since the differences between the AI speech and the AC speech were relatively small, the infant in this study could have profited from a large part of the complete input that was not directed to her but to the older child.

A better answer to the question of whether suprasegmental modifications facilitate early perceptual development could be given through experimental work with infants that do not get any input in an infant-directed speech style, either because it is a cultural habit not to adopt that speech style, or it is a cultural habit not to speak to preverbal infants at all. If the suprasegmental modifications facilitate early language development, the infants of these cultures should not show the perceptual development at the same age as the infants who do get this input.

Chapter 5
Experiments on Language Learning

In this chapter, two experiments are described that were designed to test specific predictions derived from a model of segmentation in the absence of lexical knowledge (Brent and Cartwright, 1996). In these experiments, adult listeners heard nonsense words and sentences that were composed of CV-syllables. After this, they were tested on their acquisition of the vocabulary that formed this artificial miniature language. The results of Experiment 1 showed that the listeners were able to use the words they had heard in isolation to isolate the unfamiliar parts of the sentences. The results of Experiment 2, however, suggested that the subjects' representations of the unfamiliar words were not very accurate.

5.0 Introduction

In Chapter 1, a number of segmentation strategies were described on which adults can rely for the location of word boundaries (see section 1.2.1). In addition, adults can of course use their lexical knowledge for the process of segmentation. Most of the models described in the first chapter assume that lexical knowledge is available. Prelingual infants, however, do not yet have this lexical knowledge.

Recently, a possible account of how infants learn to segment words from utterances was presented by Michael Brent (e.g., Brent and Cartwright, 1994, 1996). This model assumes no prior lexical knowledge, but places word boundaries on the basis of distributional regularities in the input, as will be described below.

In this chapter, two experiments are described that were carried out to test the plausibility of the model. Given the limitations of doing research with infants, the experiments were carried out with adult listeners rather than with infants. In order to simulate the infant's task of learning words from the input, the adults had to learn a new language on the basis of a limited amount of input. For this purpose, an artificial miniature language was created with a limited set of 'words' that were combined into 'sentences' from which possible cues to word boundaries were eliminated. In this way, we were able to control the input that the subjects heard, and to test some predictions that were formulated on the basis of the model.

The principles of Brent and Cartwright's segmentation model are described in section 5.0.1. In section 5.0.2, two previous studies comparable to the present experiments are discussed. Next, Experiment 1 is described in section 5.1, and Experiment 2 in section 5.2.

5.0.1 Word Segmentation in the Absence of Lexical Knowledge

A computational model of word segmentation in the absence of lexical knowledge was developed by Brent (e.g., Brent and Cartwright, 1994, 1996). This model draws on two sources of information that are useful for segmentation: *distributional regularity* and *phonotactic constraints*.

The notion of distributional regularity refers to the intuition that "strings that occur often and in a variety of contexts are more likely to be words than those that occur rarely and in few contexts" (Brent and Cartwright, 1994). This notion can be illustrated by means of the following example. Consider the two segmentations given in (1) of the two utterances *The kitty. See the kitty.*

(1)	*Thekitty.*	*Thek itty.*
	See thekitty.	*Seeth ekitty.*

The first segmentation is intuitively very plausible whereas the second is not. The reason why the first one is more plausible, is because the string *thekitty* that occurs in both utterances is not split up in the first segmentation, with the result that only two different word types are postulated. In the second segmentation, however, this string is split up in two different ways, resulting in four different word types.

The principle of distributional regularity has been formalized using three criteria: (a) the number of word types that result from the segmentation should be as small as possible; (b) the number of word tokens that result from the

segmentation should be as small as possible; (c) the sum of the lengths (in letters in case of orthographic input, or in phonemes in the case of phonologically transcribed input) of the word types should be as small as possible. Various segmentations are evaluated by the computer according to these criteria, and the optimal segmentation is the one which results in the smallest value with respect to the criteria. Example 1 is now used to illustrate how the value for a segmentation is determined[1].

In the first segmentation in the example, two word types occur: *thekitty*, which occurs twice, and *see* which occurs only once. The value of the segmentation is equal to the sum of the three criteria described above: the number of word types, the length of the word types, and the number of word tokens. In this case, the number of word types is two, the length of the word types is 11 letters, and the number of word tokens is three. The sum is 2 + 11 + 3 = 16.

In the second segmentation, there are four word types, *thek*, *itty*, *seeth*, and *ekitty*. Each word type occurs only once, so there are four tokens. The length of the word types is 19 letters. The sum is 4 + 19 + 4 = 27.

Various possible segmentations of the utterances can be evaluated in this way, and the one resulting in the lowest value is selected. In the example the first segmentation is selected because the resulting value is lower than the value produced by the second segmentation.

The phonotactic constraints that are used in the model consist of the following criteria: every word must have at least one vowel; word boundaries can only be located before possible word onsets in the language or after possible word offsets in the language. In example (1) the segmentation *see th ekitty* would thus not be considered because there is no vowel in the second word. The possible word onsets and word offsets can either be specified in advance or can be derived from utterance onsets and offsets that occur in the input sample.

In principle, this model can operate in two modes: a batch mode or an incremental mode (Brent, 1995). When the model is operating in the batch mode, an entire sample is read at once and all possible segmentations are evaluated according to the criteria. In the incremental mode, the model starts with the first utterance, evaluates possible segmentations and after that continues with the next utterance. In the incremental mode, entire utterances are stored as single words in a lexicon initially. When familiar strings are found in new utterances, these familiar strings will be separated from the rest of the utterance, and the remaining part will also be stored as a word in the lexicon.

---

[1]The illustration is a simplified version of the real procedure. For more details, see Brent and Cartwright (1996).

An example of the output of the model in incremental mode is given in Table 5.1 (taken from Brent and van de Weijer, 1996). The data are based on the corpus of the present project. The phonological transcriptions served as input to the model. The left side of the table shows the transcribed segmentations (in DISC notation) generated by the model. Every word starts with a dot. The right side shows the original orthographic transcripts. The first seven lines are output generations from utterances in the beginning of a language sample, after only 24 lines have been evaluated; the last five lines are output generations from utterances taken from the end of a language sample, after more than 2700 utterances have been evaluated.

As is shown in the table, the first utterance is stored as one word in the lexicon. This is the optimal segmentation because the resulting lexicon contains only one novel word. If the utterances had been segmented in two or more words, the lexicon would have contained more novel words, and the segmentation would have got a higher value. The same is true for the second utterance. However, in the third utterance, a familiar word is recognized: the string *'dat is mooi'* (that is nice). This string is isolated from the rest of the utterance and the remainder *'Josje'* (name of the infant) is also stored as a word in the lexicon. In the fourth utterance, no familiar strings are found and the entire utterance is stored as one word in the lexicon. In the fifth utterance, a familiar string is found which is isolated from the rest of the utterance, and the remainder of the utterance is again stored. No familiar parts are found in the last two utterances.

Table 5.1: Example of output segmentations generated by the model in the incremental mode. The (phonologically transcribed) output segmentations are on the left side; the original orthographic transcriptions are on the right side. See the text for further explanation.

.dAtIsmoj	dat is mooi
.zo	zo
.dAtIsmoj .jOsj@	dat is mooi Josje
.darkAnj@mojmespel@hE	daar kan je mooi mee spelen he
.zo .IsdAtlEk@rsxAt	zo is dat lekker schat
.IsdAtlEk@r	is dat lekker
.ja	ja
.....	.....
.Al .@s .xut	alles goed
.Ik .zAl .jM .@s .v@r .sxon@	ik zal jou 's verschonen
.j@ .hEpt .slap .hE	je hebt slaap he
.j@ .bEnt .lL .mKsj@	je bent lui meisje
.Ik .dENk .dAt .j@ .tOx .mar .ev@ .mut .wAxt@	ik denk dat je toch maar even moet wachten

The last five lines show that after a considerable amount of input, many more word boundaries are located. The last three utterances are all correctly segmented. In the two preceding utterances, all the word boundaries are, in principle, correctly placed. However, two additional boundaries are introduced within the words *verschonen* (change diapers) en *alles* (everything). These are likely segmentations. Splitting the word *alles* results in two existing words (*al*, 'all', and *'s* 'sometimes'). Splitting the word *verschonen* results in the prefix *ver-* which can occur in combination with various other stems, and the stem *schonen* which is pronounced the same way as the inflected adjective *schone* (clean) which occurred frequently in the sample.

5.0.2 Preceding Studies

Before we go on with the description of the experiments that are the scope of this chapter, two comparable experiments with adult listeners learning an artificial miniature language are described. The first experiment was conducted by Hayes and Clark (1970), the second one by Saffran, Newport and Aslin (1996).

Hayes and Clark

The background idea of the experiment by Hayes and Clark was that segmentation proceeds via a clustering mechanism. This mechanism, which Hayes and Clark called *correlation*, implies that events within units tend to correlate because they occur frequently together in the same pattern, whereas events that span boundaries do not correlate because they do not occur together frequently. Hayes and Clark predicted that listeners are sensitive to this mechanism, and are able to recognize correlating events (phonemes) as units (words). To test these predictions, the authors designed an artificial speech analogue that consisted of 'phonemes' and 'words'. The phonemes were synthetically generated 'consonants' (sounds that changed rapidly) and 'vowels' (sounds that remained relatively stable). Two experiments were run using these stimuli.

In the first experiment, four words were created that each consisted of six or eight phonemes selected from a list of twelve consonants and seven vowels. These words were presented to subjects for a period of 45 minutes, in random order, and without pauses or any other indication of boundaries between the words. After this listening phase, the subjects had to choose between sets of items in which pauses were inserted either within the words or between the words. The subjects chose more often the set that contained the correct words. None of them, however, obtained a perfect score.

Hayes and Clark interpreted the results as indicating that the subjects were able to recognize certain events in the sound stream, and to correlate these events with neighbouring events until a boundary was found.

To investigate this interpretation further, a second experiment was run in which the correlation of events within units was systematically varied. A high correlation was obtained by composing the words out of many different phonemes (16 in total) whereas a low correlation was obtained by composing the words out of few different phonemes (four in total). Composing the words out of many different sounds implies that the sounds that occur at the boundaries tend to be different from those that occur word internally. Composing the words out of few different sounds implies that the sounds that occur at the boundaries tend to be the same as those that occur word internally, thus obscuring the boundaries.

Subjects were tested in three conditions. In the first condition, a group of subjects heard words that were six phonemes long, selected from a list of 16 phonemes. In the second condition, another group of subjects heard words that were also six phonemes long, selected from a list of only four phonemes. In the third condition, another group of subjects heard words that were twelve phonemes long, selected from a list of four phonemes. In the last two conditions word length is different, but the correlations are equal. The experimental procedure was the same as in the first experiment. Only the subjects in the first condition had scores significantly better than chance. The subjects in the other two conditions had scores that were not significantly better than chance.

Thus, the results of this experiment suggest that human listeners are sensitive to a clustering mechanism based on correlating events.

Saffran, Newport and Aslin

A similar approach was adopted by Saffran, Newport and Aslin (1996). Their study was based on the idea that transitional probabilities can be useful for word segmentation. Transitional probability is the frequency with which two individual elements occur together; if a syllable (or another unit) is highly predictive of the following syllable (because they occur often together) then these two syllables are likely to be a unit. For instance, if the syllable /be/ is often followed by the syllable /bi/ in a sample of speech, then the two probably form a word. If the syllable /bi/ is followed by a variety of other syllables, then it is likely that it is followed by a word boundary. Therefore, transitional probability depends both on the frequency of occurrence of the first syllable and the frequency of occurrence of the combination of two syllables.

An experiment was run to investigate whether listeners are sensitive to transitional probabilities in a sample of language input from which all other possible cues to word boundaries had been eliminated. For this purpose, six words of three syllables each were created. The syllables were selected from a list of 12 CV-syllables. Because some syllables occurred more frequently than others, the transitional probabilities within the words were not equal.

The words were presented in random order for a period of 21 minutes with no breaks between them. The input sounded thus like a continuous flow of CV-syllables. After this listening phase, the subjects heard pairs of two strings of three CV-syllables. One of these strings was a 'word', i.e., had occurred in the training phase; the other one was not. The non-words were either sequences of syllables that had not occurred at all in that order in the input, or they were strings of which two syllables were part of a real word and one syllable had been changed.

The subjects chose the correct words significantly more often than would have been predicted by chance. The scores were highest when the alternative non-word was a string which had not occurred in that order in the input (i.e., a string with a transitional probability of zero). Furthermore, words with a high internal transitional probability yielded better performances than words with a relatively low internal transitional probability. An interesting finding was that the subjects tended to make more mistakes when they heard non-words that consisted of two syllables that had been the *last* two syllables of a real word plus an extra syllable than when these two syllables had been the *first* two syllables of a real word. Thus, the listeners seemed to pay more attention to the ends of words.

The results of this experiment were extended through the introduction of a prosodic marker of word boundaries: vowel lengthening. The same experiment was run in three different conditions. In the first condition, the *initial* syllables of half of the words in the input were lengthened. In the second condition, the *last* syllables of half of the words in the input were lengthened. In the third condition none of the syllables were lengthened (as in the first experiment). The results showed that subjects in the second condition had significantly higher accuracy scores than subjects in the other two conditions, whose scores did not differ from each other. This results can be explained because final vowel lengthening is a relatively common phenomenon in languages whereas initial vowel lengthening is not.

Thus, overall the results of these two experiments show that transitional probability is useful for segmentation, and that the co-occurrence of multiple sources of information can enhance performance.

5.0.3 The Present Study

Two experiments were conducted (similar to those of Hayes and Clark, and of Saffran, Newport and Aslin) to test whether adult listeners make use of distributional regularities in the input for word segmentation. The experimental paradigm that was chosen in both experiments was an auditory lexical decision task. As in the experiments described above, subjects underwent a training phase, after which they had to decide whether strings were words of the language they had heard or not. The reaction times of the subjects were measured, as well as their error rates.

5.1 Experiment 1

The experimental question that was addressed in Experiment 1 was whether listeners would use knowledge of familiar words to isolate novel words that had not been presented in isolation to them, as would be predicted by Brent and Cartwright's model. For this purpose, an artificial miniature language was created with a limited set of words. Some of these words were presented a few times in isolation, whereas other words occurred only in context.

5.1.1 Methodology

Materials

The strings that were used in the experiments as 'words' and 'sentences' were combinations of CV syllables. These syllables were controlled for pitch and duration using MBR-PSOLA (Dutoit and Leich, 1993). We used synthesized material in order to eliminate confounding factors that correlated with word boundaries, such as differences in stress, coarticulation or duration.

A total of 20 syllables were necessary to create the stimuli for the first experiment. The syllables were composed of the vowels /o,i,a,u,ɛ/ and the consonants /s,p,k,ʀ,m,f,b,l/. The average syllable duration was 0.202 sec. (sd. 0.009 sec.), the average pitch was 129.90 Hz (sd. 1.63 Hz). The syllables, with their durations and fundamental frequencies, are listed in Appendix B.

Subsequently, the syllables were combined with the speech editor (the ESPS waves speech editing program) to create words and sentences. The combinations were chosen so that they did not resemble existing Dutch words.

Subjects

A total of 24 subjects participated in Experiment 1. All subjects were native speakers of Dutch and had normal hearing. They were recruited from the Max Planck Institute subject pool and were paid for participation.

Procedure

Subjects were tested in a sound proof booth. The material was presented through headphones. The subjects were told that they were going to listen to words and sentences from a non-existing language, and that the words and sentences would sound rather odd.

In the first part of the experiment the subjects heard two words in isolation, a bisyllabic word and a trisyllabic word. They heard each word four times and repeated the words out loud. These two words will now be referred to as the *Familiar* words.

In the second part of the experiment, the subjects heard four 'sentences' which consisted of one of the Familiar words and a remaining string of two or three syllables. Both Familiar words occurred once at the beginning of a sentence and once at the end. Since we hypothesized that the subjects would recognize the Familiar words in the sentences and would infer that the unfamiliar parts were words, we will call these unfamiliar parts the *Inferred* words. The sentences were presented in random order until the subject had heard each sentence six times. Like the words in the first part, the sentences had to be repeated out loud.

The third part of the experiment was the actual test phase. The subjects now had to decide whether a given string was a word in the language or not. There task was to press a Yes button if they thought that the word that they heard had occurred in the language they had been exposed to, and a No button if they thought it had not. Possible candidates for words were (1) the two Familiar words; (2) the four Inferred words; (3) four strings of two syllables that had occurred in the sentences but which spanned a word boundary (these will be referred to as the *Cross-boundary* words); (4) two words consisting of two and three syllables that had not occurred in the sentences at all (these will be referred to as the *New* words).

Table 5.2 summarizes the experimental design and the stimuli. In this table each letter represents an individual syllable. The whole experiment took no longer than 15 minutes.

Table 5.2: Design of Experiment 1. Each letter represents a CV syllable from the set given in Appendix B.

phase	stimuli
vocabulary learning:	*ab cde*
utterance exposure:	*ab·fg hij·ab cde·kl mno·cde*[1]
test:	*ab cde* (Familiar) *fg hij kl mno* (Inferred) *bf ja ek oc* (Cross-boundary) *pq rst* (New)

[1]The dots indicate hypothesized but inaudible word boundaries.

Data Analysis

The following values were calculated: the percentages of Yes responses across the four conditions; three contrasts involving the average reaction times of the Yes responses in the Familiar and the Inferred conditions; the average reaction times of the No responses in the New and the Cross-boundary conditions; the average reaction times of the Yes responses in the Inferred and the Cross-boundary conditions. The reaction times were measured from stimulus offset.

As explained above, the predicted outcomes of the results were: a high percentage of Yes responses in the Familiar condition, and decreasing percentages in the Inferred, Cross-boundary, and New conditions, in that order. With regard to the reaction times, we expected relatively fast reactions in the Familiar conditions compared to the Inferred condition, and relatively fast reactions in the New condition compared to the Cross-boundary condition. There was no specific hypothesis with regard to the Cross-boundary/Inferred contrast. We could not say that a Yes response in the Cross-boundary condition was actually incorrect. However, relatively slow reactions in this condition compared to the Inferred condition would have suggested that the subjects found it difficult to decide whether this item was a word or not. On the other hand, if no such difference was found, it would have seemed as if some subjects (those who said Yes) dealt with the items in the Cross-boundary condition in the same way as with the items in the Inferred condition, and thus did not make the inferences that we predicted.

The design of the experiment was a *repeated measurements* design with items and subjects as random factors. The standard procedure of analysing this

design would have been to calculate two F-ratios: F1 which is obtained after collapsing the reaction times over the items, and F2 which is obtained after collapsing the reaction times over the subjects. However, the number of items within each condition was much smaller than is usually the case in psycholinguistic experiments. Therefore, only F1 values are reported here.

5.1.2 Results

Figure 5.1 shows the percentages of Yes responses across the four conditions. The overall differences between the percentages are highly significant: F1[3,69] = 120.78, MSE = 340.15, $p < .05$. In the Familiar condition 100% Yes responses were found. In the Inferred condition 80.2% Yes responses were found. This difference was significant: F1[1,23] = 17.33, MSE = 271.17, $p < .05$. The relatively high proportion of Yes responses to the Inferred words supports the hypothesis that the subjects were able to filter out the unfamiliar parts of the sentences and recognized these at a later stage when they were presented in isolation. The subjects responded Yes far less often to the Cross-boundary items (37.5%) than to the Inferred items. This difference was also significant: F1[1,23]

Figure 5.1: Percentages of Yes responses. FAM: Familiar words, INF: Inferred words, CRB: Cross-boundary words, New words.

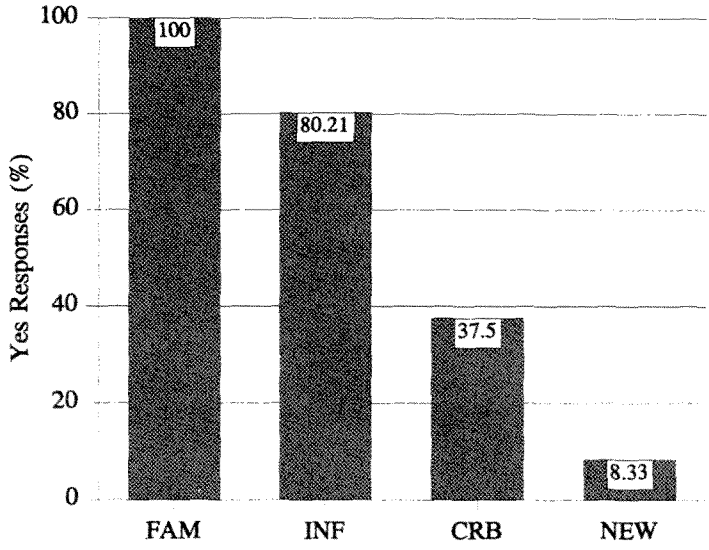
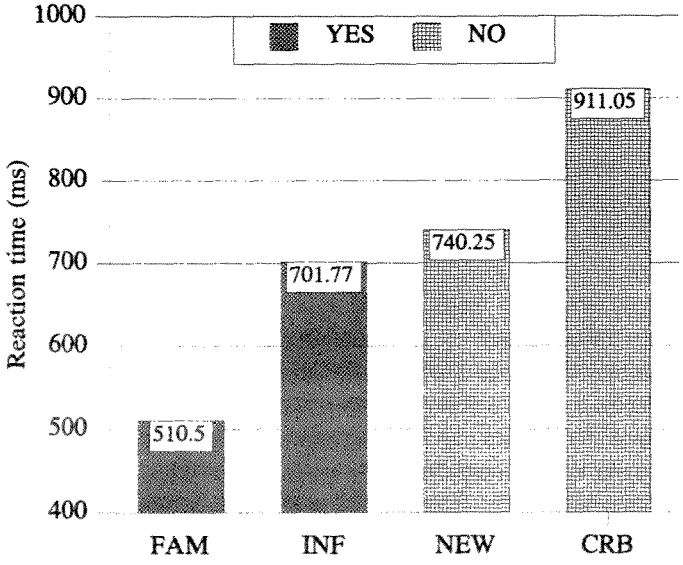


Table 5.3: Average reaction times to Yes responses and to No responses.

	YES	NO
Familiar	510.50 (226.67)	
Inferred	701.77 (279.07)	
Cross-boundary	844.16 (275.57)	911.05 (283.66)
New		740.25 (336.37)

= 41.35, MSE = 529.33, $p < .05$. This is in line with the expectations. Both strings were present in the input, but the Inferred items were aligned with a hypothesized word boundary, whereas the Cross-boundary items were not. Finally, there were only 8.3% Yes responses in the New condition. This was significantly less than the percentage of Yes responses to the Cross-boundary items: F1[1,23] = 27.49, MSE = 371.38, $p < .05$. Again, this is in line with the expectations. The Cross-boundary items might have yielded incorrect responses from the subjects because these items sounded more familiar to them than the completely new items.

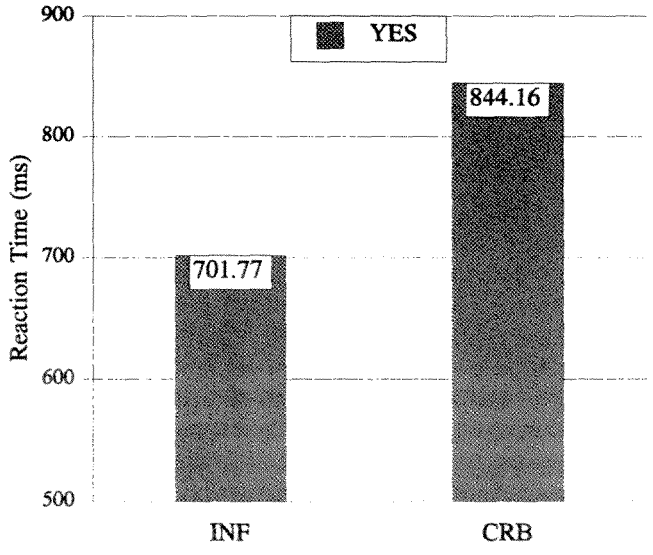
Figure 5.2: Average reaction times (ms) of Yes responses to Familiar and Inferred words, and of No responses to Cross-boundary and New words.



The average reaction times with their standard deviations are given in Table 5.3. The table shows only the reaction times in the conditions that were actually tested. Two contrasts were of first interest: that between Yes responses to the Familiar versus the Inferred items, and that between the No responses to the Cross-boundary versus the New items (the 'correct' or expected responses). These two contrasts are displayed in Figure 5.2. As can be seen from the graph, the No responses take longer than the Yes responses. Furthermore, responses in the Familiar and the New conditions are faster than the responses to the Inferred and the Cross-boundary conditions respectively, suggesting that the former two are relatively easy compared to the latter two.

The first contrast - the difference between the Familiar and the Inferred conditions - is in the predicted direction, and is significant: $F1[1,23] = 12.85$, $MSE = 68321.92$, $p = .002$. The second contrast - the difference between the New and the Cross-boundary conditions - is in the predicted direction, and is also significant: $F1[1,23] = 4.72$, $MSE = 148411.57$, $p = .040$. A third contrast of interest was the difference in reaction time of the Yes responses to the Cross-boundary items versus the Inferred items. It can be argued that a Yes response to a Cross-boundary item is correct, and indeed, a considerable number of Yes responses to these items was found. The difference is presented in Figure 5.3.

Figure 5.3: Average reaction times of Yes responses to the Inferred and Cross-boundary words.



The average reaction times to the Cross-boundary items were slower than to the Inferred items and this difference was significant: F1[1,18] = 4.40, MSE = 80438.64, p = .050. This suggests that subjects who said Yes to the Cross-boundary words hesitated longer to accept the Cross-boundary words than the Inferred words.

5.1.3 Discussion

The results of the first experiment supported the hypothesis that subjects used familiar strings to isolate the unfamiliar strings. Furthermore they supported the hypothesis that the subjects treated the unfamiliar parts of the sentences as words. All the Familiar words were accepted in the test phase by all subjects. About 80 percent of the Inferred words were also accepted, though a little less rapidly than the Familiar words. Cross-boundary words were more often rejected than accepted but the percentage of Yes responses (37.5) to Cross-boundary words was still considerable. We cannot say that Yes is either a correct or an incorrect response, since these strings occurred as such in the sentences. However, the reaction times of the Yes responses to the Cross-boundary items were significantly slower than the Yes responses to the Inferred words. Thus, the subjects hesitated longer when they heard a Cross-boundary item than when they heard an Inferred item.

Finally, the subjects were significantly faster in saying No to the New items, than to the Cross-boundary items, suggesting that they showed some recognition of the Cross-boundary items but then correctly decided that this string was not a word, and that this decision took time.

5.2 Experiment 2

The results of Experiment 1 showed that the subjects could infer words from the sentences. Experiment 2 was designed to examine this inference process in more detail. The second experiment was very similar to the first experiment. However, the Cross-boundary items in the test phase were now replaced by an item that either consisted of an Inferred word plus an extra syllable, or of an Inferred word minus a syllable. This was done to explore how accurate the subjects' representations of the Inferred words were. For the rest, the design and the other test items were essentially the same.

5.2.1 Methodology

Stimuli

The stimuli in the second experiment were composed in the same manner as in Experiment 1. A total of 25 syllables were necessary to create the stimuli for the second experiment. We used the same 20 syllables as for the first experiment, plus an additional five newly composed syllables. The syllables, with their durations and fundamental frequencies, are listed in Appendix B. The average syllable duration was 0.202 sec. (sd. 0.009 sec.), and the average frequency was 129.90 Hz (sd. 1.47 Hz).

Subjects

A total of 27 subjects participated in this experiment, divided into a group of 13 and a group of 14. They were all native speakers of Dutch and had normal hearing. The subjects were recruited from the Max Planck Institute subject pool and were paid for participation. None of them that participated in the second experiment had taken part in the first experiment.

Procedure

In the first part, the subjects heard four words in isolation, two bisyllabic words and two trisyllabic words. Each word was repeated four times. The subjects repeated the words out loud. Again, these four words will be referred to as the *Familiar* words. In the final phase of the experiment, the subjects were tested on only two of these four Familiar items.

In the second part of the experiment, the subjects heard four sentences which they had to repeat out loud. The sentences were composed of one of the Familiar strings and an unfamiliar string of two or three syllables (the *Inferred* words). This part of the experiment differed slightly from that of the first experiment. In the first experiment, each Familiar word occurred in two different sentences in the second phase. In the second experiment each Familiar word occurred only once, either in initial or in final position. Thus the subjects in Experiment 2 heard the Familiar words less often than in those in Experiment 1. The sentences were presented in random order until the subjects had heard every sentence six times.

In the third part of the experiment five different kinds of test stimuli were presented to the subjects: (1) Two of the Familiar words; (2) two of the Inferred words; (3) a trisyllabic word that consisted of an unfamiliar part of a sentence

plus an extra syllable from the adjacent word (we will call these *Long* words); (4) a bisyllabic word that consisted of an unfamiliar part of the sentence minus one syllable (we will call these *Short* words); (5) two *New* words.

The presentation of the stimuli was counterbalanced over the two groups: two of the Familiar words served as test stimuli for one group, the other two for the other group. Similarly, two of the Inferred words served as stimuli for one group, the other two for the other group, etc. The two Inferred words that were not presented in the test phase were used to create a Long word and a Short word. The same pair of New words was presented to both groups. The design is presented in Table 5.4.

Table 5.4: Design of Experiment 2. Each letter represents a CV syllable from the set given in Appendix B.

phase	Group 1	Group 2
voc. learning:	*ab cde mn opq*	*ab cde mn opq*
utter. exposure:	*ab·fg hij·mn cde·kl rst·opq*	*ab·fg hij·mn cde·kl rst·opq*
test:	*ab opq* (Familiar) *hij kl* (Inferred) *bfg* (Long) *st* (Short) *uvw xy* (New)	*mn cde* (Familiar) *fg rst* (Inferred) *ekl* (Long) *ij* (Short) *uvw xy* (New)

Data Analysis

As in the first experiments, the percentages of Yes responses were compared across the conditions. We predicted relatively high percentages of Yes responses to Familiar and Inferred words, and relatively low percentages of Yes responses to Short, Long, and New words. If the representations of the Long and the Short words were quite accurate, the percentages of Yes responses to the Long and the Short words, would be more or less the same.

Furthermore, the reaction times of the Yes responses were analyzed separately from the reaction times of the No responses. The reaction times of the Yes responses to the Inferred items and the Familiar items were compared. We predicted faster reactions for Yes responses to Familiar words than to Inferred words. The reaction times of the No responses to the Long, Short, and New

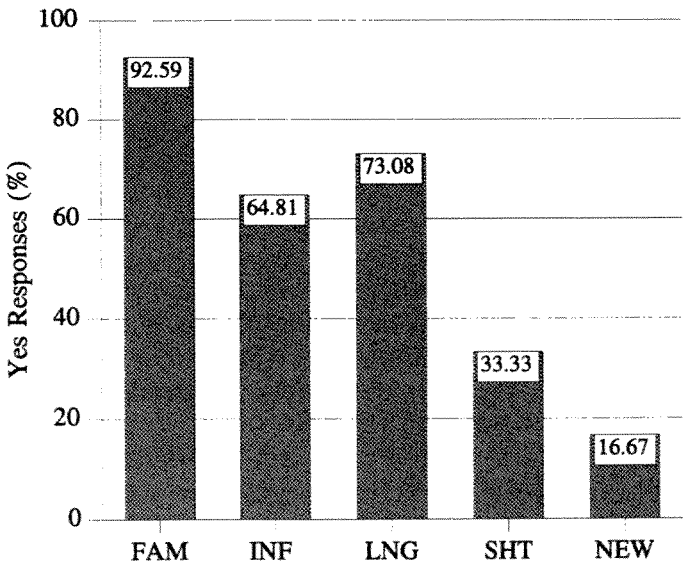
items were also compared. We predicted faster reactions for No responses to New words than to Long or Short words. Finally, the average reaction times of the Yes responses to the Inferred, Long and Short items were also compared. Based on the results of Experiment 1, it was expected that Yes responses to the Short and the Long words would be relatively slow compared to those to the Inferred items.

5.2.2 Results

The percentages of Yes responses of the two groups were compared using a *t*-test. The average percentage of Yes responses produced by Group 1 was 61.5, that produced by Group 2 was 50.7. The results of the *t*-test showed that this difference was not significant: t[132] = 1.37, $p = 0.174$. The outcomes were therefore collapsed over the two groups. Figure 5.4 shows the percentages of Yes responses in each condition.

The pattern of results was similar to that found in the first experiment, although the differences between the conditions were somewhat smaller. For

Figure 5.4: Percentages of Yes responses: FAM: Familiar words, INF: Inferred words; LNG: Long words; SHT: Short words; NEW: New words.



instance, in this experiment the score of 100% correct in the familiar condition was not obtained, and there were more Yes responses to the New items than in the previous experiments. The main effect of condition was significant: $F1[4,100] = 16.70$, $MSE = 1420.77$, $p < 0.05$. The percentage of Yes responses to the Familiar words was significantly higher than that to the Inferred words: $F1[1,26] = 11.61$, $MSE = 897.44$, $p < .05$. An unexpected result was that the subjects responded Yes relatively often to the Long words: More often than to the Short words, and even more often than to the Inferred words. Additional analyses showed that the difference between the percentage of Yes responses to the Long words was significantly higher than the percentage of Yes responses to the Short words: $F1[1,25] = 7.91$, $MSE = 2430.77$, $p < 0.05$, but the percentage of Yes responses to the Long words was not significantly higher than that to the Inferred words: $F1[1,25] = 0.52$, $MSE = 1469.23$, $p = 0.476$. The percentage of Yes responses to the Short words, on the other hand, was significantly lower than that to the Inferred items: $F1[1,25] = 7.67$, $MSE = 1745.01$, $p < .05$. These results, which are not in line with the predictions, will be discussed later. The percentage of No responses to the Short items was lower than that to the New items. This is in line with the predictions, but the difference was not significant: $F1[1,26] = 2.60$, $MSE = 1442.31$, $p > .05$.

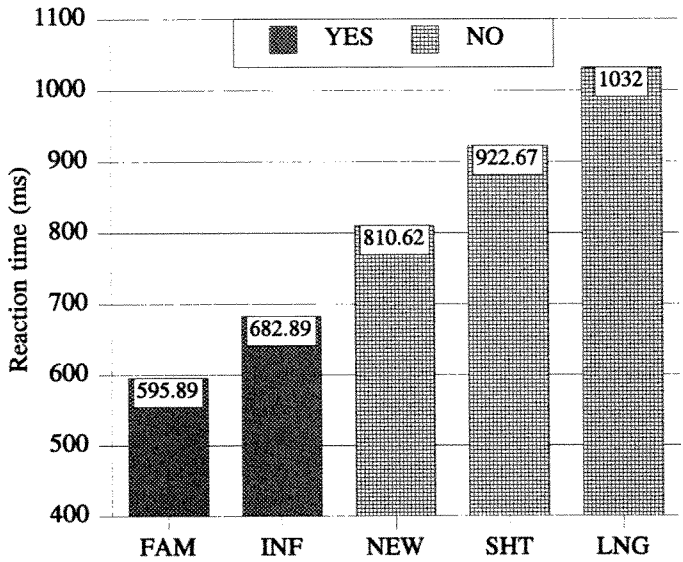
The reaction times of the two groups were also compared with a *t*-test. Although the average reaction time of Group 1 (728 ms.) was somewhat faster than that of Group 2 (799 ms.), this difference was not significant: $t[132] = 1.02$, $p = 0.312$. The reaction times were therefore collapsed over the two groups. Table 5.5 shows the average reaction times of the responses that were subjected to a statistical analysis (see also Figure 5.5).

Table 5.5: Average reaction times to Yes responses and to No responses.

	YES		NO	
Familiar	595.89	(222.72)		
Inferred	682.89	(408.47)		
Long	718.21	(466.70)	1032.00	(534.98)
Short	785.22	(502.84)	922.67	(478.01)
New			810.62	(348.19)

At first sight, the results support the predicted outcomes. Yes responses to the Familiar items were faster than those to the Inferred items. This difference was, however, not significant: $F1[1,22] = 0.71$, $MSE = 92732.37$, $p > 0.05$.

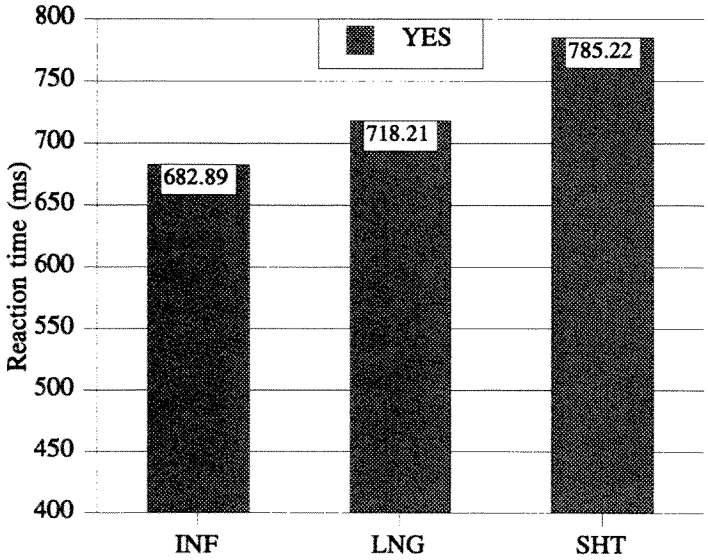
Figure 5.5: Average reaction times of the correct responses.



The second predicted outcome of the experiment was relatively slow reactions of No responses to Long and Short words compared to New words. The results showed that reactions to Long words were slower than reactions to Short words, and that reactions to Short words were slower than reactions to New words. The difference between reaction times to the Short and New items was tested but was not significant: $F1[1,17] = 2.04$, $MSE = 104895.33$, $p > 0.05$. Because of the high proportion of unpredicted (Yes) outcomes in the Long condition, the difference between Long words and Short words, or Long words and New words could not be tested.

As in Experiment 1, we also looked at the reaction times of the Yes responses to the Short and the Long items, and compared these with the reaction times of the Yes responses to the Inferred words. As is shown in Table 5.5 and in Figure 5.6, reactions to Inferred words were fastest, while reactions to Short words were slowest. The difference between Inferred words and Long words is, however, not significant: $F1[1,15] = 1.66$, $MSE = 149531.96$, $p > 0.05$. The differences between Short words and Long words and between Short words and Inferred words were again not tested because of the large number of missing Yes responses in the Short condition.

Figure 5.6: Average reaction times of the Yes responses to the Inferred, Long, and Short items.



5.2.3 Discussion

There were some differences and some similarities in the results of the second experiment compared to those of the Experiment 1. The subjects showed recognition of the Inferred words, but responded Yes significantly less often and were slower in their reactions than to the Familiar words. The difference in reaction times was, however, not significant. Furthermore, the subjects responded more often No to the New words than to either the Long or the Short words, and the reactions to the New words were faster than those to the Long and the Short words. The difference in reaction times was, however, again not significant.

An unpredicted outcome of the second experiment was that the subjects more often responded Yes to Long words than to Inferred words, although this difference was not significant. The reactions of the Yes responses to Long words were a little slower, but this difference was also not significant. It thus seems that the Long words sounded to the subjects more or less equal in acceptability to the Inferred words. Thus, the subjects' representations of hypothesized words in the language were not very accurate. On the other hand, the subjects res-

ponded No more often to the Short words than to the Long words or the Inferred words. The percentage of No responses to the Short words was comparable to that found in Experiment 1 for the Cross-boundary words. A likely explanation for these findings is the following: In Experiment 1 and Experiment 2, respectively, the Cross-boundary items and the Short items occurred within the sentences. They were not delimited by an utterance boundary. The Long items in Experiment 2, however, were delimited (see also Table 5.4), and in that sense they were more comparable to the Inferred words. Thus, it seems that the subjects showed recognition of the Inferred items, but, since the Long items were also accepted as good words their representations did not appear to be very accurate.

5.3 Conclusions

Two experiments were carried out to test the plausibility of a model of word segmentation in the absence of lexical knowledge. This model predicts that listeners will recognize familiar strings presented in an utterance, when they have heard these strings in isolation before. Then they will infer that there is a word boundary between the familiar part and the unfamiliar part of the utterance, and thus they can isolate new words they have never heard before in isolation.

In the experiments, adult listeners heard words from an artificial miniature language in isolation. After that, they heard a few sentences in which these words occurred in combination with novel strings. The subjects' task was to listen to these sentences and to repeat them out loud. After this training phase, the subjects had to respond to auditorily presented items that were words from the language they just heard or not. Possible words were the familiar items or the parts of the utterances that were novel to the subjects (Inferred words). Items that were not words were strings that either had not occurred in the sentences at all (New words), or strings that had occurred in the sentences but were misaligned with a word boundary (Cross-boundary, Long or Short words).

The overall results of Experiment 1 were as follows: the subjects recognized the Familiar words in the test phase as revealed by the high percentage of Yes responses and the fast reactions. They also showed recognition of the Inferred words, but made more mistakes, and were somewhat slower in their reactions. Furthermore, they rejected the Cross-boundary items significantly more often than the Inferred items, suggesting that they segmented the sentences as predicted. The New words and the Cross-boundary words were both more often rejected than accepted, which was also as predicted. However, the subjects made significantly more errors in rejecting the Cross-boundary items

than the New items, and the rejections of the Cross-boundary items were slower than those of the New items.

The results of Experiment 2 showed a similar pattern to that of Experiment 1, but there was also one unpredicted result. The subjects accepted the Familiar items in the test phase, and they were relatively fast in their decisions. The overall percentage was somewhat lower than that found in Experiment 1, probably because there were four instead of two Familiar words in the second experiment, and each Familiar word occurred less often in the utterances of the training phase. As in Experiment 1, the subjects responded Yes relatively often to the Inferred words, but there were significantly more errors in their responses and the responses were slower, although the difference in reaction time was not significant. The unpredicted result of Experiment 2 was that the responses to the Long words were more or less the same as to the Inferred words. There was no significant difference in percentage of Yes responses to Long words and Inferred words, and there was no significant difference in reaction times. The percentage of Yes responses to the Short items, on the contrary, was significantly lower than that of the Inferred items. The subjects more often rejected the Short words and the New words, which is a predicted result. However, they made more errors in rejecting the Short words, and their decisions took more time, although both differences were not statistically significant.

In sum, the results showed that the subjects were able to learn new words that they had not heard in isolation, but only in the context of a Familiar word. These results support the predictions that were made on the basis of the model. The listeners used the Familiar strings to isolate the remainder of the utterance and inferred that the remaining part was also a word.

However, one result suggests that this interpretation has to be considered cautiously. The Long words in Experiment 2 were accepted more or less equally often as the Inferred words. This suggests that the subjects' representations of the Inferred words were not very accurate, but that they responded to more global characteristics of the items. Since the Long words always occurred at the end of an utterance (cf. Table 5.4), it seems as if the subjects recognized the ends of the Inferred words (which were delimited by an utterance boundary) but not the word onsets. We cannot say from the present results whether it is generally easier to isolate an unfamiliar string from a following familiar string than to isolate an unfamiliar string from a preceding familiar string. According to the segmentation model of Brent and Cartwright, both instances should be equally difficult, but this issue remains to be investigated. In any case, the finding that the subjects responded to the ends of the inferred words is interesting for two reasons. First, Saffran et al. (1996) also observed that the listeners seemed to pay more attention to the ends of the items. Second, Slobin (1973) noted that

children also appear to pay attention to the ends of words rather than to the beginnings. Initial syllables are more frequently omitted in child speech than final syllables. Furthermore, relations that are expressed through post-nominal and post-verbal constructions are generally produced earlier than relations that are expressed through prepositions. These observations led Slobin to formulate a so-called 'operating principle' that children pay attention to the ends of words.

The results of this study suggest that distributional regularities in the input are useful for segmentation. The question that arises is whether infants or children use the same regularities in the input to isolate novel words that are presented in context of familiar words, and, not less importantly, at what age they start to use them. The results reported in Chapter 3 suggest that infant-directed speech is particularly well-suited for this kind of segmentation strategy: The vocabulary is repetitive, and utterances are relatively short.

However, infants are not likely to perceive language the same way as adults do. Infants tend to focus on phonological or suprasegmental characteristics of language. The adults in the present study might have used different strategies to arrive at their segmentations, for instance, they might have built orthographic representations of the words they heard. In other words, they might have used multiple sources of information, including some that the infant has not been able to use.

Therefore, the question whether distributional regularities in the input are used by infants for word segmentations remains at this point an unresolved issue. The results of the two experiments reported here, however, indicate that this question is worthwhile investigating.

Chapter 6
Conclusions

In this chapter, I give an overview of the results of the present project, how these results contribute to existing knowledge of language input, and how the results relate to the issue of word segmentation. Next, a general model of how language perception develops between six and nine months is given together with some factors that might influence this development. The chapter concludes with some directions for further research.

6.0 Introduction

The general aim of the present project was to provide a complete description of language input to an infant between six and nine months old. The main reasons for choosing this age range was that between six and nine months representations of aspects of the phonological structure are established, and that children display the ability of segmenting words out of continuous speech. Furthermore, it is assumed that it is not a coincidence that these two changes in development are noticeable at the same time since knowledge of the phonological structure may guide the infant into word discovery, a notion that is now often referred to as 'prosodic bootstrapping' or 'bootstrapping from the signal'.

In this concluding chapter, a short summary of the results is given first. After that, the contribution of the results to the existing knowledge about language input is outlined followed by the way the results have an effect on word segmentation. Then, some implications for a model of development during this period of life are described. Finally, some directions for future research are given.

6.1 Summary of the Results

The methodology of the present research was to record all the language that an infant heard during this three months period, so that a complete picture of what the input looked like could be obtained. Chapter 2 described how the language spoken during a number of 18 selected days was extracted from the original recordings. Based on this sample, the estimated amount of input was 2.56 hours of speaking time per day on average, of which 14% was language addressed to the infant. Furthermore, in Chapter 2, the transcriptions of the language to the infant (AI condition) to other children (AC condition), and to adults (AA condition) were described. The total corpus collected for the present research consisted of more than 81000 utterances, and more than 261000 words. The transcriptions were formatted according to CHILDES conventions so that the text-analysing tools could be used for the analysis of the transcriptions, and that eventually the corpus can be made publicly available.

Five aspects of the input that are related to word segmentation were analyzed. The results are described in Chapter 3. The aspects that were analyzed were Mean Length of Utterance (MLU), vocabulary size, metrical structure, phonotactic structure, and word embedding. Each aspect was compared across the three addressee conditions AI, AC, and AA. It was found that MLU in the AI condition was shortest, but that this result was mainly due to a high proportion of 'other' utterances that were not declaratives, interrogatives, or imperatives. When these 'other' utterances were excluded, MLU in the AI condition was not significantly different from that in the AC condition. Furthermore, the size of the vocabularies in the three conditions were compared. The total vocabulary size of the AI speech was 2081 word types (Dutch and German words). This was significantly smaller than the vocabulary size of the AC and the AA speech. The idea that speech to the infant consisted of a restricted vocabulary was further confirmed when the German words in the corpus were excluded, and the type-token ratios were compared. The average type-token ratio in the AI condition was significantly lower than that in the AC condition. Third, the metrical structure of the speech in the three conditions was compared. The results showed that, although a high number of content words started with a strong syllable, a very high proportion of strong syllables was found at other places than the onsets of content words. In general, the metrical structure of the AI and the AC speech was similar (although there were significant differences, these differences were rather small). Similar proportions of lexical items with initial strong syllables were found, and similar proportions of strong syllables that occurred at other places. The fourth aspect that was analyzed in Chapter 3 was the phonotactic structure. It was shown that the

number of possible word onsets and word offsets in the AI condition was lower than that in the AC and the AA conditions. Furthermore, when the possible onsets and offsets and the possible internal consonant clusters in each condition were taken into consideration, a significantly higher percentage of word boundaries could be located in the AI condition compared to the AC and the AA conditions. Finally, word embedding was analyzed. The results showed that significantly less words in the AI condition contained embedded words than in the AC or the AA conditions. However, the proportion of words that contained embedded words was still considerable and there was no extra advantage in the AI condition when all the embedded words that were not aligned with a syllable boundary were excluded.

In sum, in Chapter 3 it was shown that segmentation of the AI speech would be more successful than segmentation of the AC or the AA speech because the aspects that were analyzed indicated that the structure of the AI speech was simpler than that of the AC or the AA speech.

In Chapter 4, aspects of the suprasegmental structure of the input were analyzed. It was shown that for some of the speakers, the pitch and the pitch variation was significantly higher when they addressed the infant, whereas for the other speakers the reverse was true. Furthermore, the proportion of expanded intonation contours was not significantly different in the AI and the AC conditions but it was significantly lower in the AA speech. Finally, the speech rate was consistently lowest in the AI condition, although this difference was not statistically significant for all the speakers. The results of the analyses described in Chapter 4 were generally consistent with results from previous studies of suprasegmental modifications in infant-directed speech.

In Chapter 5, a different approach was taken with respect to the problem of word segmentation in the absence of lexical knowledge. Two experiments were carried out with adult listeners learning an artificial miniature language. The question that was addressed was whether the listeners were able to segment sentences in this language on the basis of distributional regularities. The results of Experiment 1 were promising. The subjects relatively often accepted items that were, according to the predictions, good candidates for words. Furthermore, the subjects relatively often rejected items that, according to the predictions, were bad candidates for words. In Experiment 2, the subjects' knowledge of the hypothesized vocabulary was examined in more detail. The results of the second experiment were generally similar to those of the first experiment, although the differences were not significant in all cases. However, the results suggested further that the subjects' representations of the hypothesized words were not very accurate because they also relatively often accepted items that resembled the predicted words.

6.2 Implications of the Results

This thesis contributes to the existing studies on language input in three important ways. First, a global picture of the entire input to the infant was given. There are no previous studies that provided such an estimate. The results showed that most of the language that the infant heard was language spoken to other addressees. Of course, this result has to be interpreted in light of the situation in which the recordings were made: a family with academic, full-time working parents and one older sibling. Presumably, the total amount of speaking time would be less if the infant had been the only child in the family because the AC and the CA proportions would be much smaller; it would have been more if there had been more siblings or if the infant had been at the daycare centre every day. In that case, however, the proportion of input to the infant, as well as the absolute amount of input to the infant self, would probably decrease dramatically. Furthermore, it was found in the present study that there was considerable variation in the absolute amount of input to the infant, because it was significantly less in the last week than in the first or the second week.

What can we conclude from the data about the amount of language input for word discovery or language acquisition in general? First, it is important to keep in mind that the language spoken to infants or children is not the only source of information that serves as their database from which they develop their language. In fact, this might only be a small proportion of the entire input that children receive, as was demonstrated in the present study. Further evidence that the indirect input that children receive is also important for language development is provided by the fact that children in some non western cultures do not receive much direct language input up to a certain age (Schieffelin 1985; Brown, 1996). At present, there is no evidence that these children are slower in learning their language or deviate in another way from children in our own cultures.

Second, although the proportion of the total input to the infant was small, the absolute amount of input still looks rather large. Suppose language comprehension starts around nine months (Benedict, 1979). At this age, say 275 days, the infant would have been exposed to approximately 700 hours of speech of which 14% - a little less than 100 hours - was addressed directly to her. 700 hours easily contain millions of words and the estimated amount of words to the infant would be over 600000.

It is an interesting question whether more input facilitates acquisition or not. On the one hand, more input provides the infant with more opportunities to establish knowledge about the language. On the other hand, it seems an impossible task to structure this enormous amount of information and the infant might conceivably be more helped had the input been reduced.

In any case, the enormous amount of information supports the assumption that learning a language is a constrained process, i.e., a process that is guided by innate dispositions that are either or not language specific (e.g., Gleitman and Wanner, 1982). It seems not feasible to pay the same amount of attention to every detail in the input or to store everything in memory. Rather, it appears to be the case that infants pay attention to certain aspects of the input (perhaps due to limited processing or memory capacities at an early age, as has been proposed by Newport, 1990) that are important for the acquisition of the language. Thus, it might be possible, that during a certain period of time, infants pay more attention to the language spoken with an infant-directed speech style than to that produced with an adult-directed speech style. And it likewise seems plausible that they therefore acquire some of the basic knowledge about the structure of the language on the basis of only a small proportion of the overall language input.

The second contribution of the results of this dissertation concerns the age range of the infant to whom the language was addressed. There are no studies of language input to infants of this specific age range. A general trend that can be observed in previous studies is that speech input to preverbal infants does not change significantly during the first year of life and that there is no abrupt change in the input around the age of one year when language production starts to develop (Phillips, 1973; Snow, 1977b).

Although age-related changes were not directly tested in the present study, this trend was indirectly supported by the results. Most of the analyzed aspects were not significantly different in the AI and the AC conditions, or the differences were only small: The percentage of expanded intonation contours, MLU, pitch, metrical structure, and speech rate. The only aspect that differed significantly between the AI and the AC conditions was vocabulary size. Thus, the results of the present study are generally in agreement with the results of studies of language input to either younger or older children.

The third contribution of the results of the present study concerns the nature of the data that the results were based on. Contrary to many previous studies, the data of the present study were collected over a long period of time in a naturalistic setting. Obviously, both methodologies have advantages and disadvantages. The advantage of collecting the data as it was done in the present study is that the results are based on real everyday speech. A disadvantage of the present methodology was that there are relatively many factors causing noise: multiple people talking simultaneously, background noises, poor acoustics, and so on. For the data collected in laboratory situations, the reverse can be said. In these situations, the caretakers might change their speech styles because they were being recorded or the topic of the conversation is much more restricted because the situation does not change. This might affect vocabulary size, for

instance. On the other hand, the advantage of laboratory recordings is that noise can be reduced relatively easily. The caretaker can be left alone with the child, and possible noise factors can be controlled for.

However, the results from the present study did not differ noticeably from those from previous studies. The first two aspects, analyzed in Chapter 3 and the suprasegmental characteristics, analyzed in Chapter 4, were generally consistent with results from previous studies. Since metrical structure and phonotactic structure have not been studied specifically in infant-directed speech, these two aspects cannot be directly compared.

In conclusion, it seems that the data that were used in other studies are not so much different from the data of the present study, and therefore give a good picture of what real input may look like.

6.3 Implications of the Results for Word Segmentation

Although it was observed that the language addressed to the infant was only a small proportion of the total input, a number of differences were found between the speech in the AI condition and speech in the AC and the AA conditions. With respect to other aspects of language acquisition (e.g., syntax or semantics), it has often been argued that modifications in child-directed language provide the child with a better database for acquisition than adult-directed language (see for instance Hoff-Ginsberg and Shatz, 1982, for a discussion of this issue, or Newport, 1977, for an opposite view on this assumption). Are the results of the present study consistent with this hypothesis or not?

In general, the results indicated that segmentation of the AI speech would be easier than segmentation of the AA speech. Mainly, this conclusion is based on the following observations. In the AI speech, utterances were shorter, vocabulary was more restricted, and phonotactic and metrical structure were simpler. Furthermore, the suprasegmental modifications that were demonstrated in Chapter 4 might have the additional effect of attracting the infant's attention towards this simpler speech.

However, two further remarks have to be made. First, the general picture that emerged from the results of the present research showed that each of the analyzed aspects differed in the AI and the AC condition from the AA condition but that the differences between the AI and the AC conditions were not significant or only rather small. The MLU of meaningful utterances was more or less the same in both conditions, the metrical and the phonotactic structure in both conditions were similar, and the suprasegmental modifications in both conditions were also quite similar. The only aspect that showed a clear

difference between the AI and the AC speech was vocabulary size, which was significantly smaller in the AI condition than in the AC condition.

Given these generally small differences between the AI and the AC conditions, the infant in the present study might on the one hand have benefited from the AC speech. This would considerably increase the amount of input that the infant is likely to attend to. On the other hand, however, the small differences between the AI and the AC speech also indicate that the adults did not adjust their language in a number of ways while speaking to the infant compared to when they spoke to the older child. Thus, the speech that the infant heard was similar to that directed to the older child who was much more advanced in her language development. The process of word segmentation would have been easier if the input had been adjusted to the infant's acquisition level.

Second, the AA speech was usually produced in the presence of the two children. The differences between the AI and the AC versus the AA condition might have been bigger if this had not been the case. That is, the presence of the two children might have had the consequence that the adult-directed speech was relatively simple compared to what it had been if the children had not been present. In other words, the measures of the AA speech might be an underestimate of the complexity of AA speech in general, thought they are likely to be representative for AA speech in the presence of children.

6.4 Modelling Word Discovery between Six and Nine Months

The final question that is addressed is: *how should the development of word discovery between six and nine months be modelled?* This period precedes the development of a lexicon. The infant's accomplishment during the period of six to nine months that is related to early lexical acquisition consists, roughly speaking, of two things.

First, the infant learns to identify recurring sound patterns in the input that correspond to adult-like word forms. Evidence that infants are able to do this has been described in Chapter 1 (e.g., Jusczyk and Aslin, 1995).

Second, these sound patterns need to be stored in memory for recurring recognition when they are produced by different speakers or in different contexts. Although there is experimental evidence that for a limited amount of time infants are able to store words that are presented to them during an experiment (e.g., Jusczyk and Aslin, 1995; Newsome and Jusczyk, 1995), only recently, researchers have tackled the question how long infants can retain certain sound patterns in memory. For instance, Jusczyk and Hohne (1997)

demonstrated that eight-months-old infants recognized words that had been presented to them ten days before the day they were tested.

Furthermore, I assume that it is not likely that the infant attaches meaning to the sound patterns at this age. The most often cited evidence that infants start to attach meaning to sound patterns comes from Benedict's (1979) study. According to the results of this study the process of attaching meaning to sound does not start before the age of nine months.

Finally, I assume that, at this stage in development, it is not absolutely necessary that the infant uses the sound patterns in memory to isolate novel patterns that occur in combination with familiar patterns (as has been proposed by Brent and Cartwright, 1996).

In this section, I will present a number of factors which are likely to play a role in which sound patterns are learned first. The factors are illustrated by means of Table 6.1, which shows the ten most frequent words in the AI condition, their frequency, and the frequency with which they occurred in isolation.

Table 6.1: Ten most frequent words in the AI condition. Percentages are given between parentheses.

Word, gloss	Frequency	Frequency in isolation
ja, 'yes'	2601	1684 (64.74)
zo, 'so'	1234	730 (59.16)
je, 'you'	1098	0 (0.00)
hm, 'hm'	880	451 (51.25)
Josje, 'name'	861	157 (18.23)
he, 'eh' 'huh'	750	174 (23.20)
hee, 'hey'	675	246 (36.44)
hai, 'hi'	665	247 (37.14)
lekker, 'nice' 'good'	598	26 (4.35)
is, 'is'	535	0 (0.00)

Frequency in the Input

The first factor that is likely to play a role in detecting sound patterns in the input is the frequency with which these sound patterns occur. Simply stated, more frequent patterns are learned earlier than patterns that occur only rarely. From the figures in Table 6.1, it is possible to estimate that the average frequency per day ranges from about 30 to 145 presentations of each word. In

a period of nine months, say 275 days, this means that the most frequent words had been presented 8250 to almost 40000 times.

Evidence that infants are sensitive to word frequency in the input comes from a study by Hallé and de Boysson-Bardies (1994). In this study, it was shown that eleven-month-olds and twelve-month-olds preferred to listen to words that were likely to occur frequently in the input rather than to words that were not likely to occur frequently. Further discussion of the influence of frequency on children's early segmentations can be found in Batchelder (1997) and Horvath (1980).

Frequency of Occurrence in Isolation

As becomes clear by looking at Table 6.1, there is considerable variation in how often words occurred in isolation. While some of them occurred more often in isolation than in context (*ja*, *zo*) two of them (*je* and *is*) never occurred in isolation. The last two are function words that cannot be used as individual utterances. It does not seem likely that the sound patterns of these two words are stored in this early stage of language acquisition for two reasons. First, it is obviously more difficult to identify a unit when the boundaries of that unit are not clearly demarcated. Second, words in isolation are prosodically more highlighted than words in context. For instance, the speech rate of one-word utterances was shorter than the speech rate of longer utterances, as was shown in Chapter 4 of this dissertation.

There is evidence that words that are presented relatively often in isolation are learned earlier than words that are not likely to be presented in isolation. For instance, Mandel, Jusczyk and Pisoni (1995) showed that infants responded to the sound pattern of their own name when they were four and a half months old.

Acoustic Salience

The sound patterns of the two words that never occurred in isolation (*je* and *is*) are not likely to be stored at this early stage for a second reason. Both words are not acoustically salient because they are both very short and they contain short vowels. Although some of the other ten most frequent words were short too, they are more acoustically salient because they contained long vowels, that can more easily be lengthened or prosodically highlighted otherwise. Similarly, it can be argued that sound strings that were produced in an infant-directed speech style are more likely to be stored than sound strings that were not produced that way because these sound strings are produced with larger pitch variation or with a slower speech rate.

Location in the Input

The location of the strings in the input is likely to play a role in whether the string will be stored at an early stage or not. The context can be of influence in two ways: the variety of the context and the position in the utterance. Consider the word *lekker* in Table 6.1. This word occurred relatively rarely in isolation compared to the other words that occurred in isolation. However, it occurred relatively often (67 times) in the expression *slaap lekker* (sleep well). It might well be that these frequently occurring combinations are not segmented but stored as separate units. Other examples of these frequently occurring combinations were: *hou vast* /hau vɑst/ (hold on), *kom maar* /kɔm mar/ (come on), *kijk 's* /kɛɪk əs/ (look), etc.

The position of the words in the utterance might also play a role in whether or not a string may be stored in memory. Generally, patterns that occur at the boundaries of utterances are more likely to be stored than patterns that occur utterance internally. There are at least two reasons. First, because of the position of the string at the beginning or at the end of the utterance, one of the two boundaries of the string is marked. Second, because of its position the string might be more salient (recency or primacy effect). For instance, Newport, Gleitman and Gleitman (1977) observed that children who heard relatively many yes-no questions learned auxiliaries faster than children who heard less of these questions. The explanation for this difference was that the auxiliaries are 'highlighted' because they are placed in sentence-initial position.

Word Embedding

A fifth factor that is likely to play a role is word embedding. The string *ja* from Table 6.1 occurs not only as an individual word, but also as a substring of longer words, e.g., *pyjama*, *jarig*, etc. Generally, word embedding will be confusing for the infant, and words that have other (frequent) internally embedded words are not likely to be stored at an early stage.

6.5 Future Research

Assuming that infants store a number of sound patterns in memory that are present in the input, it has to be investigated how many of these patterns are stored, and which of the factors described above predicts best which patterns are stored.

A related question that needs to be addressed is how precise the representations of these sound patterns are. On the one hand, there is evidence that infants store specific instances of sound patterns rather than abstract representations. Jusczyk, Hohne, Jusczyk and Redanz (1993), for instance, showed that infant of approximately eight months old stored information about the speaker's voice because the infants, approximately eight months of age, showed signs of recognition of the voice of a speaker that they had been exposed to daily for a period of ten days. Similarly, Houston, Jusczyk and Tager (1998) showed that infants, eight months of age, had difficulty recognizing words in texts read by a male speaker, when they were familiarized with these words by a female speaker. On the other hand, Hallé and de Boysson-Bardies (1996) showed that the representations of words that infants have are still quite global when they are eleven months old. In this study, infants responded to words that sounded like familiar words but contained phonemic alterations.

A further issue is whether strings that are stored in memory at this early stage of language development are related to those that the child attaches meaning to at a later stage, around nine months when receptive vocabulary starts to develop. It has been argued that the infant might use the sound patterns in memory to make inferences about other words. Therefore, it is not unlikely that the sound structure of the first words that are stored in the lexicon are phonologically similar to that of the strings that were stored in memory at an earlier age.

It remains to be investigated exactly when the infant actively starts using the sound strings that are stored in memory to isolate novel strings in utterances (as has been proposed by Brent and Cartwright, 1996). This could be tested with the preferential looking paradigm. In this case the infant would have to be familiarized with at least two, but preferably more words, presented several times in the context of another word. After the familiarization phase, the items that were in the context of the familiar items would be presented, and items that were not in the contexts. A preference for the items that occurred in the context would indicate that the child recognized these items.

One last issue is the following. In this dissertation it was shown that the total linguistic input consisted for the largest part of language that was not addressed to the infant, but that there were a number of clear differences between the language to the infant and the other language. The question *'how important is infant-directed speech?'* is still in need of further investigation. The underlying issue of how children learn language by exploiting their language environment is a basic one in the study of human cognitive capacity, and will remain an intriguing question for many years.

References

Aslin, R. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In: B. Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage and J. Morton (eds.). *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 305-315). Dordrecht: Kluwer Academic Publishers.

Baayen, H. and Schreuder, R. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, 33, 357-375.

Barnes, S., Gutfreund, M., Satterly, D. and Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of Child Language*, 10, 65-84.

Batchelder, E. (1997). *Computational Evidence for the Use of Frequency Information in Discovery of the Infant's First Lexicon.* Doctoral dissertation, The City University of New York.

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language*, 6, 183-200.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, 14, 303-309.

Bernstein Ratner, N. and Rooney, B. (1993). Segmentation of the input signal: a problem with potential solutions. *ASHA*, Anaheim.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. and Mehler. J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustic Society of America*, 82, 31-37.

Best, C., McRoberts, G. and Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.

Bloom, P. (1994). Recent controversies in the study of language acquisition. In: M. Gernsbacher (ed.). *Handbook of Psycholinguistics* (pp. 741-779). San Diego: Academic Press.

Booij, G. (1995). *The Phonology of Dutch.* Oxford: Clarendon Press.

Brent, M. (1995). Mechanisms of speech segmentation and word discovery. *Grant Proposal submitted to the National Institutes of Health.*

Brent, M. and Cartwright, T. (1994). The informational structure of word learning. In: P. Broeder and J. Murre (eds.). *Proceedings of: Cognitive Models of Language Acquisition.* Institute for Language Technology and Artificial Intelligence. Tilburg University, The Netherlands.

Brent, M. and Cartwright, T. (1996). Distributional regularity and phonotactics are useful for segmentation. *Cognition*, 61, 93-125.

Brent, M. and Weijer, J. van de (1996). Segmentation and word discovery: The interaction of lexical and phonotactic knowledge in Dutch and English. The 28th Annual Child Language Research Forum. Stanford University.

Broen, P. (1972). The verbal environment of the language-learning child. *ASHA Monographs Number 17.*

Brown, P. (1996). The conversational context for language acquisition: a Tzeltal (Mayan) case study. Plenary lecture, 5th International Pragmatics Conference, Mexico City.

Burnage, G. (1990). *CELEX, a Guide for Users.* CELEX - Centre for Lexical Information.

CELEX Dutch database (release N31), [online]. (1990). Available: Nijmegen: Centre for Lexical Information.

CELEX English database (release E25), [online]. (1993). Available: Nijmegen: Centre for Lexical Information.

CELEX German database (release D25), [online]. (1995). Available: Nijmegen: Centre for Lexical Information.

Chitashvili, R. and Baayen, R. (1993). Word frequency distributions. In: G. Altmann and L. Hřebíček (eds.). *Quantitative Text Analysis* (pp. 54-153). Trier: Wissenschaftlicher Verlag.

Christiansen, M., Allen, J. and Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221-268.

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25, 53-69.

Cole, R. and Jakimik, J. (1980). A model of speech perception. In: R. Cole (ed.). *Perception and Production of Fluent Speech* (pp. 133-164). Hillsdale, NJ: Erlbaum.

Cutler, A. (1990). Linguistic rhythm and speech segmentation. In: J. Sundberg, L. Nord and R. Carlson (eds.). *Music, Language, Speech and Brain. Proceedings of an International Symposium at the Wenner-Gren Center* (pp. 157-166). MacMillan Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, 92, 81-104.

Cutler, A. (1995). Spoken word recognition and production. In: J. Miller and P. Eimas (eds.). *Speech Language, and Communication* (pp. 97-136). San Diego: Academic Press.

Cutler, A. and Carter, D. (1987). The predominance of strong syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.

Cutler, A., McQueen, M., Baayen, H. and Drexler, H. (1994). Words within words in a real-speech corpus. In: *Proceedings of the Fifth Australian International Conference on Speech Science and Technology* (pp. 362-367).

Cutler, A. and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.

Dutoit, T. and Leich, H (1993). MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication*, 13, 435-440.

Eimas, P., Siqueland, E., Jusczyk, P. and Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.

Everitt, B. (1977). *The Analysis of Contingency Tables*. Bristol: J.W. Arrowsmith.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181-195.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development*, 60, 1497-1510.

Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64, 657-674.

Fernald, A. and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 209-221.

Fernald, A. and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27, 209-221.

Fernald, A. and McRoberts, G. (1996). Prosodic bootstrapping: a critical analysis of the argument and the evidence. In: J. Morgan and K. Demuth (eds.). *Signal to Syntax* (pp. 365-389). New Jersey: Lawrence Erlbaum Associates.

Fernald, A. and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B. and Furui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477-501.

Friederici, A. and Wessels, J. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics*, 43, 287-295.

Gallaway, C. and Richards, B., eds. (1994). *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press.

Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children. In C. Snow and C. Ferguson (eds.). *Talking to Children:*

Language Input and Acquisition (pp. 63-88). Cambridge: Cambridge University Press.

Gleitman, L. and Wanner, E. (1982). Language acquisition: the state of the state of the art. In: E. Wanner and L. Gleitman (eds.). *Language Acquisition: The State of the Art* (pp. 3-48). Cambridge: Cambridge University Press.

Goodsitt, J., Morgan, J. and Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, 20, 229-252.

Grieser, D. and Kuhl, P. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24, 14-20.

Grieser, D. and Kuhl, P. (1989). Categorization of speech by infants: support for speech sound prototypes. *Developmental Psychology*, 25, 577-588.

Hallé, P. and de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, 17, 119-129.

Hallé, P. and de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463-481.

Haselager, G., Slis, I. and Rietveld, A. (1991). An alternative method of studying the development of speech rate. *Clinical Linguistics and Phonetics*, 5, 53-63.

Hayes, J. and Clark, H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. Hayes (ed.). *Cognition and Development of Language* (pp. 221-234). New York: Wiley and Sons, INC.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., Wright Cassidy, K., Druss, B. and Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.

Hoff-Ginsberg, E. and Shatz, M. (1982). Linguistic input and the child's acquisition of language. *Psychological Bulletin*, 92, 3-26.

Hohne, E. and Jusczyk, P. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, 56, 613-623.

Horvath, R. (1980). *Frequencies as an Aid to Segmentation in Language Acquisition*. Doctoral dissertation, The University of Michigan.

Houston, D., Jusczyk, P., Kuijpers, C., Coolen, R. and Cutler, A. (submitted). Both Dutch- and English-learning 9-month-olds segment Dutch words from fluent speech.

Houston, D., Jusczyk, P. and Tager, J. (1998). Talker-specificity and persistence of infants' word representations. In A. Greenhill, M. Hughes, H. Littlefield and H. Walsh (eds.). *Proceedings of the 22nd Annual Boston University*

Conference on Language Development (pp. 385-396). Somerville: Cascadilla Press.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. and Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236-248.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge MA: The MIT Press.

Jusczyk, P. and Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1-23.

Jusczyk, P., Cutler, A. and Redanz, N. (1993a). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V. and Jusczyk, A. (1993b). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402-420.

Jusczyk, P., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A. and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252-293.

Jusczyk, P. and Hohne, E. (1997). Infants' memory for spoken words. *Science*, 277, 1984-1986.

Jusczyk, P., Hohne, E., Jusczyk, A. and Redanz, N. (1993). Do infants remember voices? *Journal of the Acoustical Society of America*, 93, 2373.

Jusczyk, P., Luce, P. and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.

Kelly, M. and Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua*, 92, 105-140.

Kemler Nelson, D., Hirsh-Pasek, K., Jusczyk, P. and Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 1989.

Kemler Nelson, D., Jusczyk, P., Mandel, D., Myers, J., Turk, A. and Gerken, A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111-116.

Kitamura, C. (1994). Infant preferences for age-related infant-directed speech: the salience of vocal affect. In: *Proceedings of the 5th Australian International Conference of Speech Science and Technology* (pp. 70-75).

Kuhl, P. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1668-1679.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93-107.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U. and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684-686.

Kuhl, P., Williams, K. Lacerda, F. Stevens, K. and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, 255, 606-608.

Kuijpers, C. and Coolen, R. (1996). The role of the syllable in infants' speech segmentation. In: *Abstracts of the Seventh International Congress for the Study of Child Language* (pp. 207-208). İstanbul: Boğaziçi University.

Kuijpers, C., Coolen, R., Houston, D. and Cutler, A. (1998). Using the head-turning technique to explore cross-linguistic performance differences. *Advances in Infancy Research*, 12, 205-220.

Lively, S. (1993). An examination of the perceptual magnet effect. *Journal of the Acoustical Society of America*, 93, 2423.

MacWhinney, B. (1995). *The Childes Project: Tools for Analyzing Talk.* New Jersey: Lawrence Erlbaum Associates.

Mandel, D., Jusczyk, P. and Pisoni, D. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, 6, 314-317.

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21-46.

McQueen, J., Cutler, A., Briscoe, T. and Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, 10, 309-331.

McQueen, J., Norris, D. and Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 621-638.

Mehler, J., Dommergues, J., Frauenfelder, U. and Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.

Mehler, J., Jusczyk, P. Lambertz, G. Halsted, N., Bertoncini, J. and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 145-178.

Miller, J. and Eimas, P. (1994). Observations on speech perception, its development, and the search for a mechanism. In: J. Goodman and H. Nusbaum (eds.). *The Development of Speech Perception: The Transition*

from Speech Sounds to Spoken Words (pp. 37-57). Cambridge, MA: The MIT Press.

Morgan, J. (1986). *From Simple Input to Complex Grammar*. Cambridge, MA: The MIT Press.

Morgan, J. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development*, 17, 389-403.

Morgan, J. and Demuth, K. (1996). Signal to syntax: An overview. In: J. Morgan and K. Demuth (eds.). *Signal to Syntax* (pp. 1-25). New Jersey: Lawrence Erlbaum Associates.

Morgan, J. and Saffran, J. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66, 911-936.

Newport, E. (1977). Motherese: The speech of mothers to young children. In: N. Castellan, D. Pisoni, and G. Potts (eds.). *Cognitive Theory, Vol. 2*, (pp. 177-215). Hillsdale NJ: Lawrence Erlbaum Associates.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11-28.

Newport, E., Gleitman, H. and Gleitman, R. (1977). Mother I'd rather do it myself: some effects and non-effects of maternal speech style. In: C. Snow and C. Ferguson (eds.). *Talking to Children* (pp. 31-49). Cambridge: Cambridge University Press.

Newsome, M. and Jusczyk, P. (1995). Do infants use stress as a cue in segmenting fluent speech? In: D. MacLaughlin and S. McEwen (eds.). *Proceedings of the 19th Boston University Conference on Language Development*. Boston, MA: Cascadilla Press.

Norris, D., McQueen, J., Cutler, A. and Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191-243.

Papoušek, M., Papoušek, H. and Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their three-month-old infants. *Journal of Psycholinguistic Research*, 16, 491-516.

Phillips, J. (1973). Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Development*, 44, 182-185.

Plunkett, K. (1998). Language acquisition and connectionism. *Language and Cognitive Processes*, 13, 97-104.

Polka, L. and Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, 100, 577-592.

Polka, L. and Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.

Ratner, N. and Pye, C. (1984). Higher pitch in BT is not universal: acoustic evidence from Quiche Mayan. *Journal of Child Language*, 2, 515-522.

Richards, B. (1987). Type/token ratios: what do they really tell us? *Journal of Child Language*, 14, 201-209.

Richards, B. and Malvern, D. (1997). *Quantifying Lexical Diversity in the Study of Language Development.* Reading: Faculty of Education and Community Studies.

Saffran, J., Newport, E. and Aslin, R. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

Schieffelin, B. (1985). The acquisition of Kaluli. In: D. Slobin (ed.). *The Crosslinguistic Study of Language Acquisition* (pp. 525-594). Hillsdale, NJ: Erlbaum.

Shi, R., Morgan, J. and Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, 25, 169-201.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In: C. Ferguson and D. Slobin (eds). *Studies of Child Language Development* (pp. 175-208). New York: Holt, Wineheart and Winston.

Snow, C. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549-565.

Snow, C. (1977a). Mothers' speech research: from input to interaction. In: C. Snow and C. Ferguson (eds.). *Talking to Children* (pp. 31-49). Cambridge: Cambridge University Press.

Snow, C. (1977b). The development of conversation between mothers and babies. *Journal of Child Language*, 4, 1-22.

Snow, C. (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In: P. Fletcher and B. MacWhinney (eds.). *The Handbook of Child Language* (pp. 180-194). Oxford: Basil Blackwell Ltd.

Snow, C., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J. and Voster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research*, 5, 1-20.

Stern, D., Spieker, S., Barnett, K. and MacKain, K. (1983). The prosody of maternal speech: infant age and context-related changes. *Journal of Child Language*, 10, 1-15.

Stern, D., Spieker, S. and MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, 18, 727-735.

Sussman, J. and Lauckner-Morano, V. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. *Journal of the Acoustical Society of America*, 97, 539-552.

Swingley, D., Pinto, J. and Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. *Advances in Infancy Research*, 12, 257-277.

Trehub, S. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47, 466-472.

Vroomen, J. and Gelder, B. de (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 98-108.

Vroomen, J., Zon, M. van and Gelder, B. de (1996). Cues to speech segmentations: Evidence from juncture misperceptions and word spotting. *Memory and Cognition*, 24, 744-755.

Wagner, K. (1985). How much do children say in a day? *Journal of Child Language*, 12, 473-487.

Werker, J. (1994). Cross-language speech perception: Development change does not involve loss. In: J. Goodman and H. Nusbaum (eds.). *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 93-121). Cambridge MA: The MIT Press.

Werker, J. and Lalonde, C. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology*, 24, 672-683.

Werker, J. and Polka, L. (1993). Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics*, 21, 83-101.

Werker, J. and Tees, R. (1984a). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

Werker, J. and Tees, R. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-1878.

Wijk, C. van and Kempen, G. (1980). Funktiewoorden - Een inventarisatie voor het Nederlands (An inventory of Dutch function words). *Review of Applied Linguistics*, 47, 53-68.

Appendix A
Codes used in the transcriptions

*MOD:	Mother, Dutch utterance
*MOG:	Mother, German utterance
*MOT:	Mother, Dutch/German utterance
*FAT:	Father
*BBS:	Baby sitter
*MEN:	Baby sitter's husband
*OTH:	Other speaker
@date:	Date of recording
@tape:	Sound file
[\]	Restart in the utterance
+...	End of interrupted utterance
+,	Beginning of interrupted utterance
xxx	not understandable fragment
[9]	Utterance length in words
[DEC]	Declarative utterance
[QUE]	Interrogative utterance
[IMP]	Imperative utterance
[FIL]	Filler
[BBT]	Baby talk
[ROU]	Routine
[VOC]	Vocative
[SOC]	Social expression
[TAG]	Tag
[X]	Not clear
word@e	incidental English word
word@g	incidental German word
word@f	incidental French word
word@d	incidental Dutch word
word@x	incidental nonword

Appendix B

Syllables that were used to create the stimuli in the experiments, described in Chapter 5.

syllable	pitch (Hz)	duration (*s*)	Exp[1]	
ba	129.70	0.199	1	1
bi	129.10	0.192	1	1
bo	129.60	0.192	1	1
bu	129.30	0.192	0	1
fa	129.40	0.214	0	1
fe	130.20	0.214	1	1
fu	130.40	0.214	1	1
ka	128.50	0.199	1	1
ke	132.60	0.199	1	1
ki	130.10	0.199	0	1
ku	130.00	0.199	0	1
la	130.70	0.192	0	1
lo	130.70	0.192	1	1
lu	130.80	0.192	1	1
mi	130.80	0.215	1	1
pe	133.90	0.199	1	1
pi	129.00	0.199	1	1
pu	130.50	0.199	1	1
ri	128.50	0.192	1	1
ro	126.10	0.192	1	1
ru	127.80	0.199	1	1
sa	130.10	0.214	1	1
si	130.10	0.214	1	1
so	129.90	0.214	1	1
su	129.80	0.214	1	1
Average:	129.90	0.203		
Std.dev:	1.47	0.010		

[1] A 1 in the first column indicates that this syllable was used in Experiment 1; A 1 in the second column indicates it was used in Experiment 2.

Appendix C
Word Onsets and Word Offsets

(DISC notation; the onsets and offsets are listed in order of frequency of occurrence, the most frequent first; 0 denotes a vowel).

AI Onsets

j 0 h m z d n w k l b x v t sl kl sx p xr f sp st bl br kr dr s vr fr r kn fl tr zw str kw sn pl pr vl sm tw sxr S spr mw wr xl

AI Offsets

0 t r s n j x k m p l nt S st xt ns rt f pt ks mt rst w ft kt lk N ls Nkt lt lf rs ts Nk jt ps rf rk rp rx rts js Nt nS tst

AC Onsets

0 j m d h n z b w k x v l p t st sx kl br xr r sl kr bl sp dr str pl f s kn tr zw sn vl spr pr fl tw fr S vr sm sxr xl kw Z tS pf dw px sk wr spl

AC Offsets

0 n t r m s k x p l nt j f xt st ft rt rst ls w pt ks rs mt lk kt lf lt ts N ns Nk rm jt rk S jk rx ps mp Nkt tst Ns lpt js lp Nt rp rkt ms nts wt Nks xts rf rn rxt lm jNk lfs mpt xst

AA Onsets

0 d j m h w n z v x b k t l p st r tw s xr pr vr f kl sx kr dr sl kw br bl str tr sp pl sn kn fr sm spr fl vl sxr zw Z ps xl S tS dw sk sw _ ks sf tj vw xw skr spl

AA Offsets

0 n t r k s x l p nt m f xt rt ft ls st N ts lf ns kt w lt j ks Nk rs pt mt rst lk jt Nkt rx nts rm rts rk rkt S ms xs tst Ns wt Nks xts mp Nt nst ps lft lm lp lfs rn mst rp lpt jk rf ws mpt rft rmt rxt xst jl nS St tS fst jmt lxt

Appendix D
Local Depfile

```
@Activities: *
@Age of child: *
@Bck: *
@Begin
@BG
@BG: * =
@Birth of: @d<dd-lll-yyyy>
@Blank
@Coding: *
@Coder: *
@Comment: * "* -
@Date: *
@DSS
@DSS: Start
@Education of: *
@EG
@EG: * =
@End
@Filename: *
@G:    * =
@Group of: *
@ID: *
@Indent
@Language: *
@Location: *
@New Episode
@Participants:    Target_child Child Mother Father Brother Sister Sibling
      Narrator German Dutch BabySitter Menno MotherGerman MotherDutch
@Room Layout: *
@SES of: *
@Sex of: *
@Situation: * "* -
@Stim: * =
@Tape Location: *
@Tape: *
@Time Duration: @t<hh:mm-hh:mm> @t<hh:mm:ss-hh:mm:ss>
@Time Start: @t<hh:mm:ss> @t<mm:ss>
@Transcriber: *
@Warning: * [*]
*:     * *-0* *-able *-al *-ary *-er *-ed *-en *-er *-es *-est
       *-ful *-ic *-ie *-ify *-ity *-ize *-ing *-less *-ly
       *-ment *-ness *-s *-'s *-s' *-' *-tion *-y
       *un# *re# *dis# *mis# *pro# *anti# *ex# *pre# *over#
       *under# *mono# *quasi#
       *+achtigs *+achtige anti+* 's*+* T+* mega+* *@*+*
       - -? -. -^ -' -'' -'' -` -, -_ -: #*
       + ++ +" +". +... +/. +//. +, +^ +"/. +/? ,,
       [=! *] [= *] ["] [: *] [:=*] [=? *] [\%*] [#*]
       [?] [>*] [<*] [<>] [+ *] [$*] [0 *] [0 * *]
       [/] [//] [///] [/?] [\*]i [\\] [!] [!!] [c]
       $* ^* &* *&ed 0* */* *//* *///*
       @*@u 6*@u 3*@u
       *@b *@c *@d *@f *@i *@l *@n *@o *@p *@s *@sl *@sas *@t *@e *@w *@ *@x
       '*@* *@g *@dg
       'n 's 'ns 'ne 't 'k da's di's d'r 'm 'r 'raus 'rein 'tzelfde
       [IMP] [QUE] [DEC] [ROU] [VOC] [SOC] [TAG] [FIL] [X] [*]
%flo: * '*
%pho: *
```

Samenvatting

Het herkennen van woorden in gesproken taal levert zelden problemen op tijdens communicatie. Toch is het opdelen van uitingen in woorden (segmenteren) geen vanzelfsprekendheid, en wel om de volgende redenen. 1) In gesproken taal worden woorden meestal aan elkaar uitgesproken. Ze worden niet gescheiden door pauzes of andere grensmarkeringen zoals geschreven woorden door spaties gescheiden worden. 2) Veel woorden kunnen worden opgedeeld in kortere woorden. Het woord *kandelaar* bijvoorbeeld, kan worden opgedeeld in *kan*, *de*, *la* en *aar*. 3) In gesproken taal worden klanken die bij een bepaald woord horen soms uitgesproken alsof ze bij een ander woord horen. In de uiting *hij heeft 't* bijvoorbeeld wordt de *t* van *heeft* in de uitspraak aan het woord *'t* verbonden: *hij heef t't*.

Ondanks de aanwijsbare problemen die optreden bij het segmenteren van uitingen, zijn er ook verscheidene soorten van informatie aanwezig waar de luisteraar gebruik van kan maken voor het herkennen van de woordgrenzen. Drie belangrijke soorten informatie zijn: ritmische structuur, fonotaktische structuur en statistische informatie.

De ritmische structuur van de taal wordt door verschillende factoren bepaald. Dat verschilt van taal tot taal. In het Engels en het Nederlands wordt de ritmische structuur bepaald door de afwisseling van volle en gereduceerde klinkers. Volle klinkers hebben een relatief hoge amplitude en een relatief lange duur (de eerste klinkers uit de woorden *tafel* of *table* bijvoorbeeld), terwijl gereduceerde klinkers relatief kort zijn en een relatief kleine amplitude hebben (de eerste klinkers uit de woorden *beleg* of *balloon* bijvoorbeeld). Zowel in het Engels als in het Nederlands komen lettergrepen met volle klinkers vaak aan het begin van inhoudswoorden voor, terwijl lettergrepen met gereduceerde klinkers vaak op andere plaatsen voorkomen. De distributie van volle en gereduceerde klinkers zegt dus iets over mogelijke plaatsen in het spraaksignaal waar inhoudswoorden beginnen.

De fonotaktische structuur van taal houdt in dat niet alle combinaties van spraakklanken op elke plaats geoorloofd zijn. Een Nederlands woord kan beginnen met /kn/ bijvoorbeeld, maar het kan er niet mee eindigen. De kennis van de fonotaktische structuur kan dus door de luisteraar gebruikt worden om te bepalen of op een bepaalde plaats een woordgrens is of niet. Evenals de ritmische structuur verschilt de fonotaktische structuur van taal tot taal. De combinatie /kn/ kan het begin zijn van een Nederlands woord, maar bijvoorbeeld niet van een Engels woord.

Statistische informatie tenslotte houdt in dat de frequentie waarmee eenheden (bijvoorbeeld lettergrepen) elkaar opvolgen doorgaans groter zijn *in* een woord dan *tussen* twee woorden. De lettergreep *jul* in het Nederlands, bijvoorbeeld, wordt meestal gevolgd door de lettergreep *lie* zodat het woord *jullie*

ontstaat. Na *lie* kan echter een veelvoud aan andere lettergrepen volgen doordat op die plaats een veelvoud aan woorden mogelijk is.

Experimentele studies hebben aangetoond dat luisteraars deze soorten informatie gebruiken voor het segmenteren van gesproken taal. Aangezien geen van de hierboven beschreven informatiebronnen eenduidig aangeeft waar woordgrenzen zijn, wordt aangenomen dat dit proces vooral dient ter versnelling van de herkenning van woorden die de luisteraar in zijn of haar geheugen heeft opgeslagen.

Eén vraag blijft op dit punt echter nog onbeantwoord. Kinderen die hun moedertaal nog moeten leren weten nog niet wat de woorden zijn. Hoe leren zij de woordgrenzen op de juiste plaatsen in uitingen te lokaliseren?

In een poging deze vraag te beantwoorden is onderzocht of kinderen kennis hebben van de hierboven beschreven regelmatigheden in taal die informatie bevatten over woordgrenzen. De resultaten van dit onderzoek hebben aangetoond dat in de periode tussen zes en negen maanden kinderen inderdaad deze kennis ontwikkelen. Wanneer kinderen negen maanden oud zijn vertonen zij bijvoorbeeld een voorkeur te luisteren naar woorden met de fonotaktische structuur van hun moedertaal, terwijl ze, als ze zes maanden oud zijn, deze voorkeur nog niet vertonen.

Algemeen wordt dan ook wel verondersteld dat in de periode tussen zes en negen maanden, kinderen leren hoe de structuur van woorden in hun moedertaal kan zijn en op basis daarvan uitingen leren segmenteren. Als een logisch gevolg, is geconstateerd dat de verwerving van de woordenschat zich met name vanaf de leeftijd van negen maanden begint te ontwikkelen.

Het onderzoek dat in dit proefschrift beschreven wordt, bestond voor een groot deel uit het verzamelen van taalaanbod aan een Nederlands kind tussen zes en negen maanden oud. Ten behoeve van dit onderzoek was 90 procent van de tijd dat een kind in deze periode wakker was, op band vastgelegd. De taal die gesproken werd tijdens 18 van de in totaal 91 dagen werd met behulp van een spraak-editor apart opgeslagen. Verder werd alle taal die de volwassenen die bij het onderzoek betrokken waren tegen het kind produceerden (in het proefschrift aangeduid als AI conditie) getranscribeerd, evenals de taal die de volwassenen produceerden tegen andere volwassenen (aangeduid als AA conditie) en tegen andere kinderen (aangeduid als AC conditie). Met behulp van het materiaal werden een aantal aspecten van het taalaanbod onderzocht.

Ten eerste werd de vraag gesteld hoe groot het taalaanbod was. Deze vraag wordt behandeld in Hoofdstuk 2. Om een inschatting te geven van de grootte van het taalaanbod werd de *spreektijd* berekend, de duur van alle uitingen die van het bandmateriaal geïsoleerd waren. Per dag was de spreektijd gemiddeld ruim twee en half uur. De spreektijd waarin volwassenen uitingen

tegen het kind produceerden was ongeveer 14% van de totale spreektijd, wat neerkomt op ongeveer 21 minuten. Verder werd ongeveer 30 procent van de totale spreektijd in beslag genomen door uitingen die volwassenen tegen andere kinderen produceerden, eveneens ongeveer 30 procent door uitingen die andere kinderen tegen de volwassenen produceerden, en ongeveer 19 procent door uitingen die volwassenen tegen elkaar produceerden. Dus de taal rechtstreeks aan het kind gericht was maar een klein gedeelte van alle taal die het kind hoorde.

Ten tweede werden een vijftal aspecten geanalyseerd die te maken hebben met segmentatie: uitingslengte, woordenschat, ritmische structuur, fonotaktische structuur en woord-inbedding. De resultaten van de analyses worden beschreven in Hoofdstuk 3. De analyses werden in drie condities uitgevoerd: de AI, AC en AA condities.

De gemiddelde uitingslengte in de AI conditie was 2.66 woorden per uiting. Dit was significant lager dan de gemiddelde uitingslengte in de AC of de AA condities. De lage uitingslengte in de AI conditie bleek vooral veroorzaakt te worden door een relatief groot percentage uitingen bestaande uit de naam van het kind, begroetingen, en woorden die het gesprek gaande houden (*ja*, *nee*, *zo*, enz.). Wanneer deze uitingen namelijk niet meegerekend werden, steeg de gemiddelde uitingslengte naar 4.01 woorden per uiting, wat niet significant verschillend was van de gemiddelde uitingslengte in de AC conditie.

Het totale aantal verschillende woordtypen dat tijdens de 18 dagen in de AI conditie gebruikt werd was significant kleiner dan het totale aantal in de AC of de AA condities. Verder werd berekend dat het gemiddelde aantal woordtypen per 300 woordtokens significant lager was in de AI conditie dan in de AC of de AA condities. De woordenschat in de AI conditie was dus beperkt en woorden werden relatief vaak herhaald.

Vervolgens werd de metrische structuur in de drie condities geanalyseerd. Het percentage inhoudswoorden in de AI conditie met een volle klinker in de eerste lettergreep was significant hoger dan dat in de AC of de AA condities. Echter, in alle drie de condities kwamen volle klinkers ook vaak op andere plaatsen voor: in grammaticale woorden of in de tweede of latere lettergrepen van inhoudswoorden. Daarom werd geconcludeerd dat de distributie van volle klinkers feitelijk vaker tot foute segmentaties leidde dan tot goede, maar dat - als deze strategie gevolgd zou worden - het in ieder geval vaker goed zou gaan in de AI conditie dan in de andere twee condities.

Het vierde aspect dat beschreven wordt in Hoofdstuk 3 is de fonotaktische structuur. Voor de analyse hiervan werd een inventarisatie gemaakt van de consonanten en de consonantclusters die voorkwamen aan het begin en het einde van de woorden in elke conditie. Het totale aantal was het kleinste in de AI conditie en het grootst in de AA conditie. Vervolgens werden consonantclusters

rondom woordgrenzen geclassificeerd als 'duidelijk' (wanneer aangegeven kon worden waar in de cluster de woordgrens lag) of 'onduidelijk' (wanneer dat niet mogelijk was). Het percentage duidelijke woordgrenzen in de AI conditie was ongeveer 25%, wat significant hoger was dan in de AC of de AA condities.

Tenslotte werd woord-inbedding geanalyseerd. Het percentage woorden dat andere woorden bevatte werd twee keer bepaald. De eerste keer werden lettergreepgrenzen niet in de analyse betrokken. De tweede keer moesten de woordgrenzen van het ingebedde woord precies samenvallen met de lettergreepgrenzen van het woord waarin het ingebed was. In beide analyses was het percentage woorden dat ingebedde woorden bevatte significant het laagste in de AI conditie.

Uit de resultaten beschreven in Hoofdstuk 3 werd geconcludeerd dat segmentatie van spraak in de AI conditie minder vaak tot problemen leidt dan segmentatie in de andere twee condities.

Ten derde werd onderzocht of de volwassenen die betrokken waren bij het onderzoek met een typische spreekstijl tegen het kind spraken: met een hoge stem, met veel toonhoogte-variatie en met een lage spreeksnelheid. Deze spreekstijl is vaker aangetoond in verschillende talen en tegen kinderen van verschillende leeftijden. De resultaten staan in Hoofdstuk 4.

Voor de akoestische metingen was het noodzakelijk een selectie te maken van materiaal waarin geen bijgeluiden voorkwamen. Daarom werden uitingen geselecteerd van de drie meest voorkomende sprekers in het corpus: de vader, de moeder en de babysitter. De analyses werden opnieuw uitgevoerd in de AI, AC en AA condities. De volgende aspecten werden onderzocht: gemiddelde toonhoogte, toonhoogte-variatie, intonatie en spreeksnelheid. De resultaten vertoonden het volgende patroon. De gemiddelde toonhoogte en de toonhoogte-variatie waren significant hoger in de AI en de AC condities dan in de AA conditie. De verschillen tussen de AI en de AC conditie waren niet altijd consistent. Soms was de gemiddelde toonhoogte of de toonhoogte-variatie hoger in de AI conditie, soms in de AC conditie. De intonatiepatronen werden onderverdeeld in enerzijds patronen waarvan in eerder onderzoek was vastgesteld dat ze typerend zijn voor spraak tegen kinderen en anderzijds patronen die dat niet zijn. Het percentage 'typische' intonatiepatronen was significant hoger in de AI en de AC condities dan in de AA conditie. Spreeksnelheid tenslotte was significant lager in de AI conditie dan in de AC en de AA condities.

Er waren dus duidelijke verschillen tussen de spreekstijl in de AI en de AA conditie, maar de verschillen tussen de AI en de AC conditie waren minder duidelijk. De resultaten pasten goed in het beeld van spreekstijl tegen kinderen dat geschetst was in eerder onderzoek.

Als aanvulling op de analyse van het verzamelde corpus werden twee experimenten uitgevoerd. De experimenten waren gebaseerd op een segmentatiemodel dat ervan uitgaat dat uitingen gesegmenteerd kunnen worden op basis van het principe dat klankpatronen van woorden frequent zijn en in een verscheidenheid aan contexten kunnen voorkomen. De resultaten van deze experimenten worden vermeld in Hoofdstuk 5. De experimenten waren opgezet om te testen of luisteraars dit principe ('distributional regularity' genoemd) gebruiken voor het segmenteren van spraak. Om het proces van het leren van een nieuwe taal te modelleren werden een aantal nonwoorden gesynthetiseerd die in zinnen aan volwassen proefpersonen werden aangeboden. Vervolgens kregen de proefpersonen een aantal woorden te horen die - volgens het distributional regularity principe - wel of geen goede woorden waren. Zowel de reactietijden van de proefpersonen als de percentages *ja* responsen werden gemeten. Over het algemeen reageerden de proefpersonen volgens het verwachtte patroon: ze reageerden vaker met *ja* wanneer ze een 'goed' woord hoorden dan wanneer ze een 'slecht' woord hoorden. Bovendien waren de reactietijden voor gemakkelijke testitems sneller dan voor moeilijke. Echter, de proefpersonen reageerden ook vaak met *ja* wanneer ze items hoorden die wel erg veel leken op goede woorden maar die er toch enigszins van verschilden. Gedurende het experiment hadden de proefpersonen dus niet genoeg gelegenheid gehad om exacte representaties van de woorden op te bouwen.

Uit de resultaten van de experimenten werd geconcludeerd dat distributional regularity een mogelijke bron van informatie is voor segmentatie. Of deze informatie ook door kinderen gedurende de taalverwerving wordt benut en op welke leeftijd zal uit vervolgonderzoek moeten blijken.

In Hoofdstuk 6 worden de resultaten van het proefschrift nogmaals op een rijtje gezet. Geconcludeerd wordt enerzijds dat spraak in de AI conditie door een aantal factoren gemakkelijker te segmenteren was dan spraak in de andere condities. Anderzijds, bleek de AI conditie maar een beperkt gedeelte van het totale taalaanbod. Het belang van het taalaanbod voor de taalverwerving lijdt geen twijfel. Welke factoren exact de meeste invloed hebben staat nog open voor verder onderzoek.

Curriculum Vitae

Joost van de Weijer studied French literature and linguistics and speech and language pathology at Nijmegen University. He graduated in speech and language pathology in 1990. After that, he studied speech therapy at the Hogeschool Nijmegen. After completing his studies in 1992, he worked as a speech therapist in Enschede, and, simultaneously as a research assistant at the Max Planck Institute for Psycholinguistics and at the department of Language and Speech at Nijmegen University. In December 1994 he was offered the possibility to carry out his dissertation research at the Max Planck Institute which resulted in the present dissertation.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking. Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout.*
2. The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsaggital articulography. *Niels Schiller.*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette Schmitt.*
4. The open/closed class distinction in spoken-word recognition. *Alette Haveman.*
5. Gesture and speech production. *Jan-Peter de Ruiter.*
6. The acquisition of phonetic categories in young infants: a self-organising artificial neural network approach. *Kay Behnke.*
7. Comparative intonational phonology: English and German. *Esther Grabe.*
8. Finiteness in adult and child German. *Ingeborg Lasser.*
9. Language input for word discovery. *Joost van de Weijer.*