The magic of matching – speech production and perception in language acquisition

Lena F. Renner

Academic dissertation for the Degree of Doctor of Philosophy in Phonetics at Stockholm University to be publicly defended on Friday 1 December 2017 at 13.00 in William-Olssonsalen, Geovetenskapens hus, Svante Arrhenius väg 14.

Abstract
This thesis investigates the relationship between speech production and speech perception in the early stages of phonological and lexical acquisition. Previous studies have mainly focused on independent investigations of speech production and perception abilities in language acquisition. This thesis connects the individual speech production capacities to the child's perception and is organized around three major studies: Study I explores methodological alternatives such as the combination of EEG and eye-tracking in different Swedish participant groups: adults, 17-month-olds, and 24-month-olds. Visual and auditory stimuli, as well as the connection between word production and word perception are explored. Study II investigates phonological capacities in terms of consonant inventory, percentage of correctly pronounced words, segmental errors, as well as phonological templates in relation to vocabulary size in a group of Swedish 18-month-olds. Study III studies the influence of the children's individual phonological and lexical capacities in speech production on their word recognition in a group of Swedish toddlers with a productive vocabulary size above 100 words.

The general results show that children accept mispronounced word forms as appropriate word candidates when the word forms are related to their individual word production. The occurrence of segmental errors increases with vocabulary size, and phonological templates are more likely to be observed in children with a productive vocabulary size above 100 words. The results thus indicate an influence of the individual child's production on word recognition, and a relationship between phonological capacities and lexical knowledge. These insights contribute to theoretical debates in linguistics regarding the abstractness of phonological word form representations and reveal a closer relationship between production and perceptual abilities in toddlers than what has previously been shown.

Keywords: First language acquisition, phonology, lexicon, vocabulary, speech production and perception.

Stockholm 2017
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-147908
THE MAGIC OF MATCHING – SPEECH PRODUCTION AND PERCEPTION IN LANGUAGE ACQUISITION

Lena F. Renner
The magic of matching – speech production and perception in language acquisition

Lena F. Renner
Contents

List of Figures v

List of Tables vii

Publications by the author ix

1. Introduction 1
   1.1. Research questions ............................................. 2
   1.2. Outline of the thesis ........................................... 2

2. The framework of phonology and lexicon 5
   2.1. Phonology and the lexicon as linguistic domains .......... 5
       2.1.1. Phonology .............................................. 5
       2.1.2. Lexicon .............................................. 9
   2.2. Methods in the study of phonology and the lexicon ...... 11
       2.2.1. Methods for investigating children’s speech perception ... 11
       2.2.2. Methods for investigating children’s speech production ... 16
   2.3. General early prerequisites to language acquisition ... 17
       2.3.1. Statistical learning .................................. 21
       2.3.2. Word-learning constraints ............................... 22
   2.4. Phonology and lexicon in language acquisition ........ 22
       2.4.1. Perception ............................................ 22
       2.4.2. Production ............................................ 27
       2.4.3. The link between production and perception ........... 33
   2.5. Summary ...................................................... 34

3. Theoretical approaches to language acquisition 35
   3.1. General approaches to language acquisition .............. 35
   3.2. Approaches to phonological and lexical acquisition ...... 37
   3.3. Concluding thoughts .......................................... 45
Contents

4. Study I: Three experiments on methodological explorations  49
   4.1. Background ........................................... 49
   4.2. Pilot eye-tracking experiment with adults .......... 52
       4.2.1. Methods ........................................ 54
       4.2.2. Results ....................................... 56
       4.2.3. Discussion .................................... 57
   4.3. Methodological explorations of EEG and eye-tracking 58
       4.3.1. Experiment with 17-month-old children ...... 59
           4.3.1.1. Methods .................................. 59
           4.3.1.2. Results ................................... 61
       4.3.2. Experiment with 24-month-old children ...... 64
           4.3.2.1. Methods .................................. 65
           4.3.2.2. Results ................................... 69
       4.3.3. Discussion .................................... 73

5. Study II: Phonological capacities at 18 months in relation to vocabulary size  77
   5.1. Background ........................................... 77
   5.2. Methods ............................................. 81
   5.3. Results .............................................. 83
   5.4. Discussion .......................................... 87

6. Study III: MAGIC Analyses Gaze In Children  91
   6.1. Background ........................................... 91
   6.2. Methods ............................................. 96
   6.3. Analysis ............................................ 103
   6.4. Results .............................................. 105
   6.5. Discussion .......................................... 108

7. General discussion and conclusion  113
   7.1. Limitations .......................................... 116
   7.2. Future directions ................................... 117
   7.3. Conclusion .......................................... 119

Summary in Swedish – Sammanfattning på svenska  121

A. Appendix study I  125

B. Appendix study III  129
List of Figures

2.1. Initial three-consonant clusters in Swedish ............. 8
2.2. Functional infrastructure of language (Bates, 2004) ........ 20
4.1. Example of a visual stimuli pair during one trial ........ 55
4.2. Looking time differences between familiar and novel objects ... 57
4.3. Number of children related to their understanding and producing of the eight target words ...................... 62
4.4. Proportional looking times for each subject .............. 63
4.5. Example of a visual stimuli pair .......................... 66
4.6. Two-step principal component analysis .................... 71
4.7. Proportional looking time for each condition ............ 72
5.1. Percentage of correct pronunciations in relation to vocabulary size 85
6.1. Examples of odd objects .................................. 100
6.2. Timeline of a trial in the correct condition ............. 102
6.3. Mean looking times towards familiar and unfamiliar object . 107
6.4. Mean looking times towards familiar and unfamiliar object for every participant .......................... 108
## List of Tables

1. Overview of the three studies included in the thesis ........................................... 3
2. Central standard Swedish consonant system ......................................................... 7
2. Swedish 9-vowel system ......................................................................................... 8
3. Auditory stimuli and the mispronunciations in the four different conditions (pilot study) ................................................................. 54
4. Auditory stimuli and the mispronunciations in the four different conditions ................................................................. 65
4. Overview of the demographic data of the participants ............................................. 68
4. Elicited speech production data for the four target words for each participant ........ 69
5. Demographic data and vocabulary scores from SECDI ........................................... 82
5. Number of types and tokens .................................................................................... 84
5. Consonant inventory in onset and coda position in the three production groups ................................................................. 84
5. Overview of the segmental errors for 10 participants ............................................. 86
5. Phonological templates adapted to target words and their shape ................................ 87
6. Stimuli in the Novel condition .............................................................................. 101
6. Overview of the 10 participants included in the final analysis .............................. 104
6. Examples of auditory stimuli .................................................................................. 105
A. Overview of the elicited speech production data and perception and production of the included words in the experiment reported by the parents ................................................................. 125
B. Objects used in the elicitation task ......................................................................... 129
B. Phonological capacities of the 10 participants ...................................................... 130
Publications by the author

Publications related to the thesis


Renner, L. F., Schwarz, I.-C., and Sundberg, U. (2012). *Combining EEG signals and eye-tracking data to investigate the relationship between phonological...*
Publications by the author


Other publications


1. Introduction

First language acquisition is a process that includes both speech production and speech perceptual abilities. These abilities have usually been investigated independently of the linguistic domains of phonology and the lexicon, although the acquisition of language requires the child to integrate many different aspects such as motor control and perceptual knowledge, as well as influences from the environment. This thesis aims to bridge the gap between speech production and perception in the early stages of phonological and lexical acquisition. Research on the interplay between perception and production in young children may contribute to different scientific fields, among them linguistic theory, as well as pedagogical and clinical implications.

The infant develops language skills by integrating linguistic characteristics of the ambient language with her own production and perceptual capacities. During this process, the physiological prerequisites, motor control, and the ability to process perceptual information undergo tremendous changes. The problem as to how the child continuously adjusts his or her speech production to match perceptual information is far from trivial, as the child’s own production leaves perceptual traces in terms of sensory clues: tactual and kinesthetic as well as acoustic. This knowledge needs to be integrated with visual and auditory information from the caregivers’ speech and other information in the environment. The characteristics of these integrating processes are truly magic: Out of the multi-faceted information emerges knowledge about ambient linguistic domains such as phonology and the lexicon. In the formation of these domains, it is unclear how perceptual and production capacities interact. The issue of if and how the child’s own production influences his or her perceptual knowledge has been largely neglected in previous research on the acquisition of phonology and lexicon. Thus, the present thesis aims to contribute to filling the research gap by examining the influence of the child’s own production on her perception of early word forms in Swedish toddlers using different methodological approaches.
1. Introduction

1.1. Research questions

The work described in this thesis explores the following research questions:

1. How, if at all, do a child’s phonological and lexical capacities in production influence her word recognition?

2. Which setting and combination of research methods might be appropriate for investigating phonological and lexical knowledge in speech production and perception at different ages during the early word-learning period?

3. Which phonological capacities can be observed in the individual word production of Swedish toddlers?

4. How, if at all, is vocabulary size related to phonological capacities in Swedish toddlers?

1.2. Outline of the thesis

The thesis has seven chapters, starting with a general background provided in chapters 2 and 3. Specifically, chapter 2 introduces the framework of phonology and lexicon in terms of a general description of the linguistic domains as well as studies related to language acquisition on those domains. Chapter 3 presents theoretical approaches. The two chapters together build the foundation for the three experimental studies to be presented in chapters 4, 5 and 6 (see table 1.1).

Study I (ch. 4) comprises three experiments exploring the combination of EEG and eye-tracking as well as the stimuli used. The first experiment tests the stimuli with adult speakers using eye-tracking. The second experiment uses the same stimuli as with the adults in a pilot on the combination of EEG and eye-tracking in 17-month-olds. The third experiment again includes both eye-tracking and EEG, but revises the previous experiment with 17-month-olds by including fewer test words, changing picture material and testing older children (24-month-olds). As a result of the study I explorations, a production study follows in study II (ch. 5). In this study the phonological capacities of 12 18-month-old children were analyzed in relation to their vocabulary size, as reported by their parents (SECDI: Swedish Early Communicative Developmental Inventory, Berglund & Eriksson, 2000). Subsequently, in chapter 6 the child’s own word production capacities are tested in a word recognition task. The thesis ends with a general discussion and conclusion in chapter 7.
1.2. Outline of the thesis

Table 1.1. Overview of the three studies included in the thesis.

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Participants</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (ch. 4)</td>
<td>Explorations of stimuli material and methods in 3 experiments</td>
<td>16 adults, 26 children at 17 months and 15 children at 24 months</td>
<td>EEG, eye-tracking, SECDI, speech recordings</td>
</tr>
<tr>
<td>II (ch. 5)</td>
<td>Phonological capacities in relation to vocabulary size</td>
<td>12 children at 18 months</td>
<td>SECDI, speech recordings</td>
</tr>
<tr>
<td>III (ch. 6)</td>
<td>Word recognition in relation to the child’s own word production</td>
<td>10 children with a vocabulary size above 100 words</td>
<td>SECDI, speech recordings, eye-tracking</td>
</tr>
</tbody>
</table>

Ethics
The methods of all studies were approved by the Regional Ethics Committee (Dnr. 2015/63-31) as part of the research plan for the Phonetics Lab at Stockholm University. This includes, among others, the anonymization of all participant names and the voluntary participation of the families. No financial compensation was provided; instead all subjects received a certificate of participation.
2. The framework of phonology and lexicon

The research questions proposed in the previous chapter involve different aspects related to language acquisition such as phonology and the lexicon, production and perception capacities as well as methods to capture these. The following chapter gives an introduction to these aspects, beginning with phonology and the lexicon as linguistic domains. Subsequently, various methods used in research on children’s speech perception and production are presented. Furthermore, phonological as well as lexical acquisition in both production and perception during the first two years of life are described.

2.1. Phonology and the lexicon as linguistic domains

2.1.1. Phonology

The infant’s early vocalizations contain articulatory sequences of openings and closures of the vocal tract, detectable with phonetic analyses during various stages of babbling. The alternations of open-closed vocal tract develop over time and become more and more regular and sophisticated. Usually, openings are most characteristic of vowels, and closures are characteristic of consonants. Together, they build the basis for the development of the phonological domain.

Generally speaking, phonology is the study of sound structure within a language and, according to Odden (2013), includes four aspects: a) the phoneme inventory of the ambient language, b) phonotactical characteristics, c) phonological variations (rules) and d) prosody. What follows is a brief summary of these aspects:

a) Every language has its own phoneme inventory, which, in broad terms, includes vowels and consonants. A phoneme is the smallest unit within a word which affects meaning. One way of testing whether a specific sound is a phoneme
of a language or not is that minimal pairs may be consulted. A minimal pair consists of two words that differ from each other in one speech sound: For instance "buss" ‘bus’ and "puss" ‘kiss’ form a minimal pair, which indicates that /b/ and /p/ are different phonemes of Swedish. The phonetic variation in the way speakers articulate a specific phoneme is considerable and is affected by numerous factors: specific language, individual speaker’s physiology, dialect, speech rate, situation, and emotional state, to name a few.

b) The phonotactic characteristics of a language describe which combinations of vowels and consonants can occur in a given position in a word. For instance, Swedish, like several other languages, including English, does not allow /N/ in onset position. The syllable structure of a language defines how vowels and consonants are grouped together, for instance, if the language tolerates consonant clusters in specific positions of the word (Odden, 2013).

c) Phonological rules are formalizations of systematic processes observed in continuous speech. In adult language, a prominent rule in some languages is the final-obstruent devoicing which occurs, for instance, in Polish or Dutch (for an example of a phonological rule in Swedish such as retroflexion, see section 2.1.1). In the study of child phonology, phonological rules describe a systemic relationship between the phonological system of the adult and the child such as substitutions, weak syllable deletions or cluster reductions (Ingram, 1974).

d) The prosodic structure of a language includes prominence patterns within words (i.e. word stress) as well as at the utterance level (i.e. accents). Common word stress patterns are sequences of strong-weak syllables (trochaic, e.g. happy) or weak-strong syllables (iambic, e.g. amuse). Additionally, the description of prosodic characteristics involves phrase and sentence intonation and the description of word tones. Tones can be realized as pitch variations, which are syllable independent, or as word accents in terms of a fixed pitch variation within a certain syllable. This can be seen in Swedish words, for instance, änden ‘the spirit’ and ånden ‘the duck’.

An alternative way of describing the phonological system of a language includes paradigmatic contrasts and syntagmatic relationships. Paradigmatic contrasts are segments that are in opposition to one another and thus belong to different categories (i.e. phonemes) (Mannell, Cox, & Harrington, 2014). Syntagmatic

---

1 Conventionally, slashes // are used in phonemic transcriptions to represent contrastive sounds in the language in question. Square brackets [] are used to represent phonetic transcriptions, which indicates how the sounds are realized in spoken language.
2.1. Phonology and the lexicon as linguistic domains

Table 2.1. Central standard Swedish consonant system (adapted from Engstrand, 1999, p. 140).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p, b</td>
<td>t, d</td>
<td>k, g</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nas.</td>
<td>m, n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fric.</td>
<td>f, v</td>
<td>s, c, j,</td>
<td>Ë, h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approx.</td>
<td>l, ŋ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The fricative Ë is articulated as voiceless alveolo-palatal; Ë as voiceless dorso-palatal/velar.

structure in phonology describes the relationship between phonemes, such as the grouping of phonemes into a higher units such as syllables and which phonemes can co-occur in a syllable.

Phonology of Swedish

The present thesis investigates the acquisition of phonology and lexicon and their influences on perception and production in toddlers. In this section, we investigate the acquisition of phonology in Swedish children and start with an overview of the phonology of Swedish to get a clearer picture of what children in this language environment acquire.

The phoneme inventory of central standard Swedish as it is spoken in the area around Stockholm consists of 18 consonants and 9 vowels (Engstrand, 1999, 2004; Riad, 2013). Among the consonants are three nasal (/m, n, nj/), six stops (/t, d, p, b, k, q/), seven fricatives (/l, v, s, c, j, Ë, h/) and two approximants (/l, ŋ/) (Table 2.1). Coronal consonants (/d, t, n, s, l/) are conventionally articulated as dentals in central Swedish (Engstrand, 1999). Five of the vowels are front (/i, y, e, ë, ø/), two are central (/0, a/) and two are back (both rounded /u, o/), (Table 2.2). All vowels, except /ë/ and /e/, can be short or long. However, the short counterparts are not always identical in terms of quality; for instance the short counterpart of /y/ is /i/ (Engstrand, 2004). The change in quality is particularly strong for the vowel /a/, whose short version in central standard Swedish is usually fronted, whereas the long counterpart is usually realized as an open back rounded vowel [u]: (Engstrand, 2004).

Swedish phonotactics include, for example, syllable onsets with one, two and three consonants with certain regularities: all three-consonant clusters start with /sl/; the second consonant is a voiceless stop (Engstrand, 2004; Garlén, 1988). The

---

2 Much of the research done in this area focuses on English children (e.g. Stoel-Gammon, 2011).
3 The Swedish fricatives are later used in this thesis for word manipulations (see chs. 4 and 6).
2. The framework of phonology and lexicon

Table 2.2. Swedish 9-vowel system (adapted from Engstrand, 2004, p. 115).

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrounded</td>
<td>Rounded</td>
<td>Rounded</td>
</tr>
<tr>
<td>Close</td>
<td>i</td>
<td>y</td>
<td>u</td>
</tr>
<tr>
<td>Close-mid</td>
<td>e</td>
<td>ø</td>
<td>o</td>
</tr>
<tr>
<td>Open-mid</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>Open</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
</tbody>
</table>

Example

/s/ /p/ /t/ /k/ /j/ /s/ /tk/ /j/ /spjut ‘javelin’

Figure 2.1. Initial three-consonant clusters in Swedish (adapted from Garlén, 1988, p. 102). Third consonant is either /v/, /j/, /l/ or /r/ (Figure 2.1). Syllable (and word) offsets may be even more complex, as word-final clusters are related to morphemes, since Swedish has several consonantal suffixes. Very rarely, codas consist of as many as six consonants4 (Engstrand, 2004; Riad, 2013).

A widespread phonological rule in central and northern Swedish dialects with apical /r/ involves retroflexion (Riad, 2013, or supradentalization Engstrand, 2004). If the apical /r/ is followed by a coronal consonant, the two merge to [ù, ú, ã, ï, í ]; however, dialectal variations occur. Another phonological rule in central standard Swedish involves the vowel lowering preceding /r/ and retroflex consonants (Engstrand, 2004; Riad, 2013). For instance, [E] is realized as [æ] and [ø] as [œfl ], if they are followed by /r/ or a retroflex consonant.

Swedish words have no fixed stress pattern; instead, the stress pattern of the words depends on lexical and morphological characteristics (Bruce, 2012). In a two-syllable word, stress can be either on the first or the second syllable. However, some predictions can be made as to which syllable is stressed. One regularity, for instance, is based on the syllable’s weight: if the last syllable within a word stem is

4Such six-consonant clusters are morphologically complex, for instance, skälmsk-t-s [ʃjɛlmskts] ‘roguish, gen.’.
closed or includes a long vowel, it tends to carry stress as in [baˈnɑːnə] banan ‘banana’ (Bruce, 2012). Additionally, central Swedish has two lexical accents: Accent I (acute) and Accent II (grave). Each of the accents is characterized by a distinct tonal curve. For central standard Swedish, accent I is characterized by one F0 peak; in accent II, two peaks can be observed (Bruce, 2012; Engstrand, 2004). Accent II only occurs in words with a poststress syllable; accent I does not have this limitation. (Riad, 2013). The realization of the tonal accents differs across Swedish dialects in both timing and pitch peak (Bruce, 2012). The above-mentioned characteristics are related to lexical word stress and words pronounced in isolation. They might change in running speech due to various phrase and sentence conditions such as speech rate.

2.1.2. Lexicon

The focus of the present thesis is not only on phonology, but also on the lexicon. A definition of the term lexicon must consider two characteristics (Jarema & Libben, 2007): On the one hand, it might be referred to as a dictionary encoded in the mind. On the other hand, psycholinguistic research on the mental lexicon focuses on lexical processes, for instance, word comprehension as a lexical activity. Additionally, a theory of the mental lexicon might depend on other linguistic domain theories, for instance, those of phonology or morphology. Thus, in a broader sense the mental lexicon can be seen as part of cognitive system, including a “degree of functional integrity that [we] know to be characteristic of lexical activity” (Jarema & Libben, 2007, p. 2).

Nevertheless, in research on language acquisition, the term lexicon is often defined as a dictionary or a ‘stock of established words’ used in human communication (Clark, 1995, p. 2). In general, during communication the listener needs to be able to identify words by searching for them in memory, and the speaker retrieves the words as appropriate forms that carry meanings. The word is thus the smallest part of a sentence that carries meaning and that can move around in the utterance (in contrast to morphemes and phonemes) (Clark, 1995). In the mental lexicon, which is part of long-term memory, words are stored as lexical entries.

Another description of the phonological representation of Swedish word accents includes a distinction between different prominence levels such as focus accent and word accent (Heldner, 2001; Riad, 2013). Taking the prominence levels into account, different accent ‘melodies’ can be identified: At the prominence level focus accent, accent I is realized as L^H, and accent II is realized as H^LH (H=high tone, L=low tone; the star (*) marks the tone that will occur on the stressed syllable). For the prominence level of word accent the distinction between accent I and II consists of an accent melody of H^L* (accent I) and H^L* (accent II) (Riad, 2013).
According to Clark (1995), a lexical entry consists of different kinds of information that are related: a) the meaning, b) the syntactic form, c) the morphology, and d) the phonological shape (for a similar description, see Aitchison, 2003, describing the syntactic form and morphological shape as the word’s “role in a sentence”, p. 36). Similarly, Allan (2006) proposes that a lexical entry needs to be accessible from three different directions: form, morphosyntax, and meaning. Meaning involves the conceptual knowledge of what the word refers to (Clark, 1995). Syntax defines the word class and function, and carries information needed to embed the word in a possible syntactic environment. For instance, the syntactic portion of a lexical entry referring to a verb includes the number of arguments: a transitive verb requires two arguments; an intransitive only one. In general, the lexical entry for a verb thus carries more information than an entry for a noun, since it includes more syntactic information. Morphology involves the building blocks from which the word is derived in terms of word root and affixes. Following Clark (1995), the morphological part of the lexical entry involves all variants of the word (e.g. plural forms), and the phonological part carries information about the order of sounds and prosodic information.

Lexical entries can be related to each other in terms of their meaning. In this way, they thus form semantic fields and/or hierarchical taxonomies. A semantic field groups together terms across word classes that are related to a specific conceptual domain. For instance, the semantic field of *schooling* includes terms such as ‘teacher’, ‘pupil’, ‘learn’, ‘reading’, ‘class’ or ‘break’. Hierarchical taxonomies describe the relation of words to each other in terms of different levels. Words on a lower level are always included in higher levels. For instance, a *poodle* (lower) is a kind of a *dog* (higher), which in turn is a kind of an *animal* (higher still).

The organization and structure of the mental lexicon and its lexical entries is the subject of ongoing debate and questions such as if and how information about morphemic and phonological units is stored and in which way (e.g. Boudelaa & Marslen-Wilson, 2015; Hanna & Pulvermüller, 2014; Marslen-Wilson, Tyler, Waksler, & Older, 1994). It is nonetheless clear that lexical storage changes also in adults in terms of losing and adding lexical entries (Jarema & Libben, 2007). The semantics of already existing entries can also change. The productive vocabulary of an adult speaking English comprises between 20,000 and 50,000 word forms (Aitchison, 2003; Clark, 1995).

However, the mental lexicon of a child is different from that of an adult, since it is under ongoing construction (Aitchison, 2003). The challenges that the child is faced with when building up a lexicon can be described from three different perspectives (Gleitman, 1994): (1) the knowledge of the concept related to the
2.2. Methods in the study of phonology and the lexicon

In order to study the research questions proposed in section 1.1, those parts of the mental lexicon concerning the meaning and phonological shape of a word (Clark, 1995) and in consequence all three aspects proposed by Gleitman (1994) should be borne in mind. These are all considered to be relevant for the process of word recognition.

2.2. Methods in the study of phonology and the lexicon

In order to capture children’s perceptual and production capacities in language acquisition, different methods must be applied. In the following section some methods available for investigating speech perception and speech production are presented in order to give an overview of the methods that might be suitable for investigation phonological and lexical acquisition.

2.2.1. Methods for investigating children’s speech perception

Researchers who investigate perceptual abilities relevant to the acquisition of phonology and the lexicon can choose from several different methods. In recent decades, in particular, new techniques have created new possibilities for gauging the perceptual abilities of children. Researchers distinguish between behavioral and neurophysiological methods. The former operationalizes different various kinds of behaviors, whereas the latter measures brain responses related to different cognitive processes.

One early technique from the 1970s is high-amplitude sucking (HAS): a well-documented behavioral method used in the investigation of speech perception. Specifically it can be used to uncover the discrimination and categorization capacities of infants from birth up to four months (Byers-Heinlein, 2014). A typical procedure of a study investigating categorical discrimination using HAS goes as follows: During a habituation phase, every time the infant provides a high amplitude suck, a sound is played. The infant learns during that phase that she can control the sound by her sucking behavior (i.e. sucking amplitude). Once children have established that they can control the speech sound, they begin to suck
on the pacifier more and more often and the sounds are played with increased fre-
quency, presupposing that the infant experiences the sound as reinforcing, and the 
high-amplitude suck reflects the infant’s interest (Byers-Heinlein, 2014). After ha-
bituation, shown as a decline in sucking rate to a predetermined criterion, the test 
phase starts with a sound stimulus change for infants in the experimental group. 
Infants included in the control group hear the same stimuli as during the habitua-
tion phase. If the infant in the experimental group perceives the sound change in 
the test phase, an increase in sucking rate is observable. The differences in sucking 
behavior between the experimental and control group in terms of increased sucking 
in the experimental group indicates, for example, the capacity of categorical dis-
crimination (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The HAS method 
was used, for instance, to study the discrimination of speech sounds in terms of 
vowel quality (Lacerda, 1992a, 1992b; Lacerda & Sundberg, 1996). In using this 
method, several considerations have to be taken into account, such as the duration 
of the experiment. According to Jusczyk (1985), the testing of one contrast might 
take approximately 15 minutes. However, others had shorter times (less than five 
minutes) for habituation, resulting in shorter experiment durations in general (Lac-
erda, 1992a, 1992b; Lacerda & Sundberg, 1996). Another aspect to consider, put 
forward by Kuhl (1985), is the information given only at group level; information 
about individual infants might be hard to extract. This is partly due to deriving 
evidence for discrimination by comparing two groups of children: Children in the 
experimental group are exposed to a different type of sound during the test phase 
than the control group. However, it is possible to use each subject as her own con-
trol in a repeated measure design (Lacerda, 1992a, 1992b; Lacerda & Sundberg, 1996). Another aspect to consider in HAS, as mentioned by Kuhl (1985), is the 
interpretation of negative results: If no change in the sucking behavior is observed 
after a change of stimulus, it is uncertain whether the child cannot discriminate the 
sound change or just does not experience it as reinforcing. Other method-specific 
challenges involve the infant’s willingness to suck and wakefulness; for instance, 
newborns and very young infants tend to sleep much of the time, and they are at 
risk of falling asleep during the experiment (Byers-Heinlein, 2014).

Another behavioral method used to measure speech perception in infants is 
the head-turn preference procedure (HPP), in which the child turns his or her head 
towards a sound source behind a flashing light (Fernald, 1985; Kemler Nelson et 
al., 1995; Kuhl, 1985). In this procedure the infant is seated between two loud-
speakers and two lights, and another, third, light source is installed in front of 
them. A camera is installed underneath the central light to record the head turns. 
In the beginning of a trial the light in the center flashes to attract the infant’s atten-
tion. As soon as the child focuses on the flashing light in the center, it turns off, and one of the lights to the right or left starts to flash. Once the infant’s attention is on the flashing light to the left or right, the sound stimulus is presented from that side and continues until the infant loses interest and turns his or her head away from the light and the sound for a predetermined criterion time, typically set to more than 2 seconds. The amount of time the child is orientated towards the speech stimulus is recorded by the camera and thereafter annotated for each trial by a researcher blind to the hypothesis. Finally, the averaged orientation time over the trials is calculated for each stimulus type. The HPP is suitable for infants who can control their head movements, which means not before four months. The best results have been obtained in 6- and 10-month-old infants (Kuhl, 1985; Werker, Shi, et al., 1998), but older children have also been tested successfully (e.g. Santelmann & Jusczyk, 1998). The method has been used to investigate whether infants prefer one kind of sound over another – for instance, infant-directed speech over speech directed to adults (Fernald, 1985), or a native language over another language (Moon, Cooper, & Fifer, 1993), or to investigate perception of speech sounds in terms of vowel prototypes (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Furthermore, it has been used to study infant word form recognition (e.g. Swingley, 2005; Vihman, Nakai, DePaolis, & Hallé, 2004).

Another variant of the HPP, also called Modified HPP, has been used to investigate word segmentation (Jusczyk & Aslin, 1995). Studies using the modified HPP include a training phase and a test phase. During the training phase infants are familiarized with a specific stimulus. During the test phase, the child listens to the familiar stimulus as well as a new stimulus. The length of time the child listens to these stimuli is compared. The modified HPP has, for instance, been used to investigate artificial language learning (Saffran, Aslin, & Newport, 1996), language discrimination (Nazzi, Jusczyk, & Johnson, 2000), or sensitivity to prosodic structure (Johnson & Seidl, 2008).

In the progression of methods used to uncover infants’ discrimination capacities, the switch task was developed (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). In this task the infant is habituated in the training phase to two word-object pairings. The test phase examines the infant’s ability to detect a switch in the learned pairing. For instance, visual stimulus A and auditory stimulus A are presented together during the training (or habituation) phase. The same procedure follows for visual stimulus B and auditory stimulus B. In the test phase, the pairing is switched; that is, visual stimulus A is presented together with auditory stimulus B. If the infant perceives such a switch, it should reinforce a longer looking time towards the screen.
2. The framework of phonology and lexicon

Another procedure based on the child’s looking behavior is called looking-while-listening, which – like the switch task – combines visual and auditory stimuli. In this procedure two (or more) pictures or scenes are shown on the screen and an auditory stimulus is presented that refers to one of the pictures/scenes. The fixation time towards the named picture or scene (target) is compared to the fixation time towards the unnamed picture or scene (distractor), presupposing that the participant will look longer at the best visual match for the auditory stimulus. The looking-while-listening procedure is suitable for children 14 months and older (Swingley, 2012) and is used, for example, to investigate how children respond to mispronounced words, that is, in word recognition tasks (e.g. Swingley, 2016; Swingley & Aslin, 2002). That means that this procedure is suitable for investigating word forms in relation to their meaning (cf. research questions 1 and 2 as well as chs. 4 and 6).

In its early days gaze behavior was analyzed off-line: The child’s gaze was recorded by a camera, and after the experiment the researcher coded the looking behavior frame by frame. Nowadays, several techniques are available to do the coding on-line, and in a smoother and easier way. For instance, eye-trackers usually use near infrared light directed to the eye and record the eye movement with a camera. Eye-trackers take advantage of the corneal reflection of the near infrared light and calculate the gaze position in real time through the camera eye. In order to reliably map the gaze position to a specific point on the screen, a calibration procedure is carried out, usually prior to the experiment. During the calibration the eye-tracker measures essential characteristics of the participant’s eyes according to position and movement in order to normalize the gaze behavior in the experiment. Eye-tracking is used both for children and adults in many different research areas, for example, in reading experiments (for a review, see Rayner, 1998), word recognition tasks or the understanding of words or sentences (e.g. Bergelson & Swingley, 2012; White & Morgan, 2008, or the MAGIC study in ch. 6). Eye-tracking glasses also make it possible to examine the gaze behavior in communication outside of a laboratory. Additionally, researchers take advantage of the pupil size measurement to unveil cognitive load and cognitive processes (Jackson & Sirois, 2009; Laeng, Sirois, & Gredebäck, 2012; Zekveld, Heslenfeld, Johnsruide, Versfeld, & Kramer, 2014). Recently, pupillometry has also been used to measure the sensitivity to mispronunciations in toddlers (Tamási, McKean, Gafos, Fritzche, & Höhle, 2017) or information mismatch (Renner & Włodarczak, 2017).

Neurophysiological methods can assess brain responses related to language and speech perception. One of those methods is electroencephalography (EEG),

6Preferential looking task is another term used for this kind of procedure.
which measures electrical properties of neurons using electrodes placed on the scalp. The neural responses occurring at different latencies with respect to a stimulus onset (called event-related potentials, or ERPs) have been found to correspond to specific cognitive processes, several of which are linked to speech and language processing (Kaan, 2007; Swaab, Ledoux, Camblin, & Boudewyn, 2012). For instance, a negative reaction with a 400ms latency (N400) has been linked to lexical retrieval, and a positive reaction 600ms (P6) after the stimulus has been found to correspond to syntactic processing (see Luck, 2005, for an overview). A typical procedure in studies on lexical retrieval and the elicitation of a N400 response is an unexpected ending of a sentence, for example, ‘The pizza is too hot to comb’ or the mismatch of two competing pieces of information, such as visual and auditory information. In linguistic research on young infants, a picture and an auditory stimulus are usually presented (e.g. Friedrich & Friederici, 2005a). The auditory stimulus is either the picture name (match) or another mismatched word. The response to a mismatched word 400ms after word onset is usually larger in amplitude than the response to a match. If the difference between the reactions of a match or a mismatch is significant, it is called an N400 effect.

Another ERP component is connected to the oddball paradigm. The so-called mismatch negativity (MMN) is a reaction to changes in a series of events and is often used in studies of phoneme discrimination. For instance, the child hears a sequence of ‘ba-ba-ba-ba-da-ba-ba’. In the example, the fifth stimulus ‘da’ deviates from the surrounding stimuli. If the child detects this deviation, it is reflected in an MMN response (for a review, see Kuhl & Rivera-Gaxiola, 2008).

Another neurophysiological method, Magnetoencephalography (MEG) measures the magnetic field that occurs during neural electrical activity for measuring ERPs and locates the spatial origin of the ERP. The advantage of MEG over EEG is its spatial localization. So far, there are not many studies on language acquisition using MEG due to the challenges associated with this method, such as the lack of a brain model to map the measurements onto at different ages and the high cost of running a MEG machine. Nevertheless, it has, for example, been used to detect the discrimination abilities of bilingual and monolingual infants (Ferjan Ramírez, Ramirez, Clarke, Taulu, & Kuhl, 2017), and the number of publications continues to increase.

Studies using eye-tracking and methods assessing brain responses such as EEG investigate different aspects of language acquisition such as phonology as well as lexical retrieval. A combination of these has so far not been reported in the research of language acquisition but could add valuable data concerning the relationship between phonology and lexicon. A combination of the two methods
was used in two studies with 17- and 24-month-old children described in chapter 4 addressing research question 2.

2.2.2. Methods for investigating children’s speech production

Looking at the wide range of methods available for research in infant speech perception, the number of methods available for examining children’s speech production seems limited. Researchers who investigate speech production in language acquisition can rely on parental reports and diaries, audio recordings and their transcription and acoustical as well as auditory analysis of speech samples. Older children can also participate in elicitation tasks. Audio recordings can be supported by video recordings in order to reconstruct, for instance, what the child might have been referring to when the speech data is transcribed afterwards. An essential implement in the transcription of speech is the use of phonetic symbols, such as those provided by the International Phonetic Association (IPA). A narrow transcription usually includes the identification of certain details in the speech signal and marks these with diacritics (e.g. the degree of voicing), while a broad transcription omits these markings. Nevertheless, phonetic transcriptions might be influenced by the experience of the transcriber in terms of his or her knowledge of phonetics and phonology, his or her native language as well as experience with child speech, and with phonetic transcriptions in general (Edwards & Beckman, 2008). Transcriptions rely on the subjective perception of the transcriber and are thus filtered through adult’s perceptual norms. Phonetic transcription can be supported by acoustic analysis, which can be considered a tool for more objectively describing child’s production (Li, Edwards, & Beckman, 2009). Additionally, acoustical analysis provides more information to determine further characteristics of speech production, for instance, stress (Kehoe, Stoel-Gammon, & Buder, 1995). However, the transcription of speech data usually needs an evaluation by a second transcriber to ensure reliability. The comparison of transcripts as well as discussion about utterances that do not match between two transcribers is crucial in the analysis of recorded speech data (Rowe, 2012). In the present thesis, phonetic transcriptions have been used to capture the child’s own pronunciation of words (chs. 4, 5 and 6 addressing research questions 1, 2, 3 and 4).

Diary studies have the advantage of ecological validity, since the observations of the child are usually set in a naturalistic situation. Furthermore, they are age-independent, and data can be obtained in various settings (Rowe, 2012). However, some training is usually required to document a child’s utterances accurately.
2.3. General early prerequisites to language acquisition

An important tool for supporting parental reports is the McArthur Communicative Development Inventory (CDI) (Fenson et al., 1994). The CDI checklists include questions on early gesture use, vocabulary comprehension and production as well as grammatical abilities. Adaptations exist in many different languages, enabling cross-linguistic comparisons. It also exists for Swedish (Eriksson & Berglund, 2002). The tool is often used to estimate vocabulary size in both typically developing children and children with language disorders and difficulties of various kinds, such as autism spectrum disorders (Charman, Drew, Baird, & Baird, 2003) or Down syndrome (Berglund, Eriksson, & Johansson, 2001). The CDI exists for different age groups. The CDI checklist for children aged 8 to 18 months (CDI: Words & Gestures) includes sections about the comprehension and production of gestures, words, sentences and phrases. The vocabulary section asks for information on which words the child understands, says or both understands and says. The CDI checklist for children aged 16 to 30 months (CDI: Words & Sentences) includes sections about productive vocabulary as well as grammatical skills such as the use of plural markers or the use of different tenses. The Swedish adaptation of the CDI Words & Sentences and its vocabulary section is used in studies II and III (chs. 5 and 6), addressing research questions 1 and 4.

2.3. General early prerequisites to language acquisition

Acquisition of language relies on biological prerequisites, different experiences, skills related to the child’s environment, and developmental changes. For instance, an important experience is interaction with the environment, and an essential skill is the detection of a variety of regularities in the surrounding speech. What follows is a brief summary of the general early prerequisites a typically developing child goes through, with focus on perception and word learning. This section lays the groundwork for more detailed studies on phonological and lexical acquisition in section 2.4.

Perception starts in the uterus, as the auditory system develops early and is regarded as rather sophisticated approximately three months before full-term birth (Hepper & Shahidullah, 1994). However, fetal auditory experience is characterized by a low-pass filter imposed by the surrounding liquids and body tissues of

---

The open database ‘Wordbank’ (Frank, Braginsky, Yurovsky, & Marchman, 2016), accessed through http://wordbank.stanford.edu/, provides norm data of CDI adaptations in several languages. The database allows an exploration of growth curves for individual words in different languages.
the mother. Prosodic information such as intonation, stress and rhythm of the signal can pass through this filter and thereby be available to the fetus, but segmental information cannot. Several studies have shown that the heart rate and/or movements of a fetus change in reaction to music or speech (for an example see Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011; Kisilevsky et al., 2003), indicating the processing of prosodic information. The influence of prenatal exposure to speech seems to be measurable shortly after birth (May, Byers-Heinlein, Gervain, & Werker, 2011; Moon, Lagercrantz, & Kuhl, 2013). Three-day-old infants react differently to a familiar language than to an unfamiliar or artificial language. This indicates that the fetal exposure to speech through the mother and the environment leaves traces in the developing brain (May et al., 2011).

After birth, the quality of the surrounding language differs from language exposure in the womb: New acoustic filters influence the perception of the surrounding language. Still, prosodic information plays a crucial role in the speech directed to infants (infant-directed-speech: IDS). The adult typically talks to the infant in a lively and encouraging way, which carries linguistic information in what, from an adult perspective, is an exaggerated manner in terms of larger variations in pitch, word repetitions, longer pauses, shorter utterances and wider pitch excursions (Fernald & Simon, 1984; Kuhl et al., 1997; Papoušek & Papoušek, 1989, for a review, see Soderstrom, 2007). Infants at various ages show a preference for IDS (Cooper & Aslin, 1990; Fernald, 1985; J. Segal & Newman, 2015). The characteristics of IDS change over time as mothers adjust pitch and intonation (Kitamura & Burnham, 2003) as well as consonantal specifications in terms of voice onset time (Sundberg, 2004; Sundberg & Lacerda, 1999) according to the child’s developmental stage. For instance, speech directed to Swedish three-month-old infants differs from adult directed speech in terms of tonal and temporal aspects of words with tonal accents II (Sundberg, 1994). Infant-directed speech is assumed to facilitate the detection of words from fluent speech (Thiessen, Hill, & Saffran, 2005) as well as to support statistical learning (Bosseler, Teinonen, Tervaniemi, & Huotilainen, 2016), an ability to extract information based on statistical distribution in the environment. Additionally, several studies have shown that the ability of infants to associate words with objects is improved when words are presented in infant-directed speech relative to speech directed to adults (Graf Estes & Hurley, 2013; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Song, Demuth, & Morgan, 2010).

Besides the computation of distributional regularities to segment the speech stream, infants can rely on prosodic structure, which is highly modified in infant-directed speech by including wider pitch excursions, for instance (Sundberg, 1994).
2.3. General early prerequisites to language acquisition

One cue is the stress pattern of the ambient language. For instance, all Finnish words are stressed on the first syllable, resulting in a reliable trochaic (strong-weak) stress pattern. Other languages, such as Hebrew, tend to have more words with stress on the last syllable (weak-strong), that is, an iambic stress pattern. Swedish has both trochaic and iambic stress patterns and is therefore more complicated to categorize in terms of being an iambic or trochaic dominant language (see section 2.1.1 above).

In addition to the ability to extract information from the environment, some general prerequisites are assumed to be crucial for language acquisition. Based on studies involving infants with language deficits, Bates (2004) proposes a functional infrastructure for language with five different skills that together build the “starter set” for acquiring language (Bates, 2004, p. 250). Those five abilities, which are not restricted to language, converge across the first three years of life in later developmental milestones correlated with the emergence of language (see, Fig. 2.2). Firstly (1), object and social orientation are the basis for joint reference and later secondary reasoning. In addition, social orientation (2) serves as an essential skill for imitation, which is influenced by cross-modal perception (3). Both social orientation and cross-modal perception lead to later observational learning. In order to map sound to meaning, cross-modal perception as well as sensory motor precision (4) are necessary. Sensory motor precision and computational power (5) support rapid induction. In turn, rapid induction as well as the mapping of sound to meaning form the basis for the ability to do fast mapping. Altogether, they form a foundation for the successful pathway to the acquisition of language and its subdomains, not least those that are the focus of the present thesis: phonology and the lexicon.

Although not explicitly mentioned in Bates’ model, the acquisition of language and its domains of phonology and the lexicon rely on several cognitive functions such as attention and memory. In the beginning, the infant’s attention is attracted by both voices and faces (Bushnell, Sai, & Mullin, 1989; Cooper & Aslin, 1990; Ruff & Rothbart, 2001). Around two months of age attention is extended to objects (Posner & Rothbart, 1980). The onset of object recognition is observed around five months (DeLoache, Strauss, & Maynard, 1979). Thereafter, at around eight months, joint attention skills are observed (Beuker, Rommelse, Donders, & Buitelaar, 2013). Joint attention is assumed to play a causal and critical role in early word learning (Tomasello & Farrar, 1986), especially in the earliest stages of vocabulary development (Morales et al., 2000; Tomasello, 1999, 2003). In general, attention is necessary for word learning and the ability to relate new forms to new referents (Werker & Stager, 2000).
2. The framework of phonology and lexicon

Figure 2.2. Functional infrastructure of language. Reprinted from "Explaining and interpreting deficits in language development across clinical groups: Where do we go from here?" by Elizabeth A. Bates, 2004, Brain and Language, p. 251. Copyright by Elsevier. Reprinted with permission.
2.3. General early prerequisites to language acquisition

During the first year of life, the child increases his or her memory retention considerably (Hartshorn, Rovee-Collier, Gerhardstein, Bhatt, Klein, et al., 1998; Hartshorn, Rovee-Collier, Gerhardstein, Bhatt, Wondoloski, et al., 1998). Both explicit as well as implicit memory develop from early infancy (Rovee-Collier & Barr, 2010; Rovee-Collier & Cuevas, 2009). Memory functions in terms of encoding, consolidation and retention are the basis for word learning (for a review of those memory functions in relation to word learning see, Wojcik, 2013).

2.3.1. Statistical learning

Early experiences and various prerequisites facilitate the processing of the speech stream and the disentangling of language characteristics and entities such as words and phonemes. One important aspect of the infant learning mechanism is statistical learning, which facilitates the segmentation of the speech stream based on distributional frequencies of phonemic segments as well as the estimation of word boundaries (Saffran, Newport, & Aslin, 1996). In language acquisition, statistical learning has been tested by tracking sequential statistics in word segmentation or grammar learning (for a review see, Gómez & Gerken, 2000), revealing that children from eight months are able to extract distributional information from the environment (Saffran, Newport, & Aslin, 1996) and learn which units follow one another. In a study of adults learning an artificial language, CV syllables were merged to trisyllabic word forms (Saffran, Newport, & Aslin, 1996). In this study with adults, some syllables were more frequent than others and thus occurred in more word forms. The participants were exposed to those words in a continuous speech stream as the word boundaries were removed. Afterwards they were asked to identify word forms from the artificial language and were able to successfully reject those that did not occur during speech exposure, indicating that adults can extract information about word boundaries from an unsegmented artificial language based on distributional cues alone. Another study has tested the same paradigm in eight-month-old infants (Saffran, Aslin, & Newport, 1996). After exposure to the artificial language, the infants showed a novel response, namely, they listened longer to nonwords compared to words included in the artificial language, indicating that eight-month-old infants are able to extract distributional information from a speech stream. Statistical learning has been shown not only for the distribution of syllables in nonwords, but also for speech sounds (Maye, Werker, & Gerken, 2002). Additionally, statistical learning is not limited to speech, but has been observed in a study investigating distributional learning with sequences of pictures (Kirkham, Slemmer, & Johnson, 2002).
2. The framework of phonology and lexicon

2.3.2. Word-learning constraints

While learning new words, young children have been taken to rely on a set of word-learning constraints such as whole-object, taxonomy and mutual exclusivity, (Markman, 1990, 1994; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). The whole-object constraint describes the tendency to map words to the whole object rather than to their parts or characteristics such as color or other properties (Markman, 1994). The taxonomic constraint refers to the extension of words to objects of a similar kind. Mutual exclusivity refers to the tendency of infants to map a novel word onto a novel object (Markman & Wachtel, 1988). The earliest use of mutual exclusivity was reported in ten-month-old infants (Mather & Plunkett, 2010). However, another study reports that 17-month-olds map novel labels to novel objects, but 16-month-olds do not (Halberda, 2003). These diverging results could be explained by another study: Kalashnikova, Mattock, and Monaghan (2016) have shown that the use of mutual exclusivity of 17- to 19-month old children is connected to their receptive vocabulary size: children with a larger vocabulary size were more consistent in mapping a novel label to a novel object compared to children with a smaller vocabulary size, indicating that mutual exclusivity is a developing skill rather than a necessary precursor of word learning. More recently, computational approaches have shown that mutual exclusivity can guide generalizations in word learning (for a review, see Regier, 2003). The assumption of mutual exclusivity has been tested in the present thesis in all perceptual experiments (see chs. 4 and 6).

2.4. Phonology and lexicon in language acquisition

Phonology and lexicon are developing linguistic domains during language acquisition in both production and perception, particularly in the first two years of life. The child’s phonological and lexical capacities are described below from a developmental perspective, from two different angles: perception and production.

2.4.1. Perception

Speech perception in language acquisition has traditionally been investigated by experiments addressing auditory perception, speech segmentation as well as the semantic processing of words. In studies of infants and toddlers, auditory perception
2.4. Phonology and lexicon in language acquisition

includes discrimination as well categorization of phonemes. The discrimination of speech sounds includes the ability to perceive two sounds as different, whereas categorization includes the ability to treat some variations of two sounds as the same sounds (for a review of speech discrimination and categorization in infants, see Galle & McMurray, 2014). Speech segmentation involves the ability to extract words and sounds from the speech input. Studies related to semantics investigate whether children understand specific words or word forms and/or how words are retrieved lexically and reflected in brain responses.

Several studies of phoneme discrimination have shown that infants younger than six months are able to discriminate between phoneme contrasts of their mother tongue as well as from a foreign language (Polka & Werker, 1994; Werker & Tees, 1984). For instance, Werker and Tees (1984) tested American and Indian children on the discrimination of Hindi and American English phonemes. Young infants discriminated between all phonemes tested. Nevertheless, children older than six months seemed to focus in their perceptual ability to phonemes of the familiar language, discriminating native phonemic contrasts but not nonnative contrasts. In particular, as long as the infant is exposed to the unfamiliar language in the form of personal interaction two times a week, the ability to discriminate unfamiliar phonemes may be retained (Kuhl, Tsao, & Liu, 2003). Another recent study replicated the results with infants learning Arabic and Hebrew (O. Segal, Hejli-Assi, & Kishon-Rabin, 2016). These findings indicate an adjustment from global to more language-specific speech perception during the first year of life (Kuhl, 2004), which has been called perceptual narrowing (for a review of perceptual narrowing more generally, see Maurer & Werker, 2014). However, some studies have reported exceptions to perceptual narrowing related to finer acoustic contrasts: For instance, the contrast [na] [ŋa] was not discriminated by English- and Filipino-hearing infants at six to eight months, but the Filipino-hearing infants discriminated this native contrast at the end of the first year of life (Narayan, Werker, & Beddor, 2010; see also Sato, Sogabe, & Mazuka, 2010, and Mugitani et al., 2009, for a failure in vowel length discrimination in Japanese-learning infants at 6 months, but not at 10 months). Another study testing the contrast /d-ð/ showed that ten-month-old French-hearing children did not differ in discrimination performance to English-hearing infants, although /d-ð/ is a native contrast in English but not in French (Polka, Colantonio, & Sundara, 2001). Additionally, for English-hearing infants

8Auditory perception can also include phoneme identification, which can be regarded as a process closely related to phoneme categorization. Traditionally, though, phoneme identification is investigated in relation to phonological awareness in older children (for a review of methods and tools connected to phonological awareness, see Sodoro, Allinder, & Rankin-Erickson, 2002).
2. The framework of phonology and lexicon

the discrimination ability increased after 12 months of age. Given the different findings on phoneme discrimination, Maurer and Werker (2014) have suggested using the term “attunement” rather than “narrowing”.

One of the first studies on speech categorization showed in a high-amplitude sucking experiment that one-month-old infants perceived speech sounds along a voicing continuum between /b/ and /p/ categorically and similarly to adults (Eimas et al., 1971). Categorical perception could also be shown in two-month-old infants for nonspeech sounds such as saw waves differing in their onsets, which were perceived by adults as “plucks” and “bows” (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). Other studies have shown that six-month-old infants were able to categorize nasals, consonants and vowels, despite variations in speaker and context (Hillenbrand, 1984; Kuhl, 1979).

Studies of speech segmentation investigate, as mentioned above, the extraction of words from a continuous speech stream and showed, for instance that six-month-old infants are able to segment the speech stream Jusczyk and Aslin (1995). However, in a series of 13 different experiments aiming to replicate the findings of Jusczyk and Aslin with British infants, it was shown that these children segmented words only at 10.5 months and only when the stimuli were presented in an exaggerated infant-directed speech style (Floccia et al., 2016; for a similar result in nine-month-old children hearing German, see Schreiner & Mani, 2017). Nevertheless, Spanish- and Catalan-hearing infants seemed to be able to segment the speech stream already around six months of age (Bosch, Figueras, Teixidó, & Ramon-Casas, 2013). In a meta-analysis of studies on word segmentation, Bergmann and Cristia (2016) found that age did not correlate with an increased ability to segment words and suggested a revision of the general assumption that word segmentation skills improve from a specific age.

The auditory processing of phonemes and the segmentation of units from the speech stream into words can be considered as prerequisites for mapping meaning to these units. Bergelson and Swingley (2012) showed that six- to nine-month-old children recognized familiar words such as apple or milk in a looking-while-listening task. Another study reported that children as young as 4.5 months preferred their own name over other names, even when the latter shared the same stress pattern (Mandel, Jusczyk, & Pisoni, 1995). In the recognition of names, vowels seemed to play a larger role than consonants. In a study with French infants, five-month-olds were shown to recognize their own name; manipulations of the vowels led to failure in the recognition of the own name in contrast to the manipulation of consonants (Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015). At the same age, British English infants showed a similar impact on both vowel and
2.4. Phonology and lexicon in language acquisition

consonant manipulations of their own name: both led to a failure in the recognition of the own name (Delle Luche, Floccia, Granjon, & Nazzi, 2017). At the end of the first year, at 11 months, children started to recognize familiar words in isolation (Hallé & de Boysson-Bardies, 1994, 1996; Swingley, 2005; Vihman, Thierry, Lum, Keren-Portnoy, & Martin, 2007). The recognition of familiar words within sentences was observed at 12 months (DePaolis, Vihman, & Keren-Portnoy, 2014; Vihman et al., 2004).

Similarly, at 12 months children begin to process words semantically in a way similar to adults, yet this kind of semantic processing measured in an EEG experiment was observable only in children with a productive vocabulary size of more than four words, based on parental reports (Friedrich & Friederici, 2010). However, those children with a maximum of four words in their productive vocabulary understood the tested words, according to their parents. The authors acknowledge that the memory representations of word forms might be weaker in those children with a smaller productive vocabulary. The lexical-semantic representations of children two months older (14 months) seemed to be stronger, since the children’s brain responses at that age demonstrated an N400 effect on a group level (Friedrich & Friederici, 2005a). However, there is no information about the vocabulary sizes of the children in the latter study.

Of special interest in the research on phonological and lexical acquisition in the second year of life are experiments investigating word recognition and word learning. In such experiments, words are usually manipulated through substitutions of one phoneme by another, leading to a novel word. For instance, in an EEG experiment 14-month-olds were exposed to mispronunciations. The results showed that children accepted mispronounced words as word candidates when they were manipulated in one position, such as *dook* instead of *book* (Mills et al., 2004). At the same time, studies measuring visual fixations showed that 14-month-olds recognize a correctly pronounced word better than a mispronounced word (Swingley & Aslin, 2000, 2002), such as *baby* in contrast to *vaby*. These results on mispronunciations gain support from a study testing sensitivity to vowel changes in newly acquired words, where similar results were obtained (Mani & Plunkett, 2008); in a switch task it was shown that children at the same age (14 months) successfully pair novel words with novel objects (Werker, Cohen, et al., 1998). Slightly older children, at 17 months, successfully associated a novel object with a novel word even when the novel word differed in only one phoneme such as *bih* in contrast to *dih* (Werker, Fennell, Corcoran, & Stager, 2002). Another study with Dutch children reported the encoding of phonetic detail in initial and medial position (Swingley, 2003), as in the study with 14-month-olds (Swingley & Aslin, 2000, 2002), in
which correctly pronounced words seemed to be easier to recognize. White and Morgan (2008) conducted a similar study on word recognition showing two pictures, one familiar and one unfamiliar object, while presenting an auditory stimulus which either labels the familiar object with the correct pronunciation or consists of different degrees of mispronunciation. The results showed that the looking times of 19-month-old children differed significantly depending on the degree of mispronunciation, leading to shorter looking times at the familiar object when more phonetic features had been changed (see also a similar study with 24-month-olds, Bailey & Plunkett, 2002). Another recent study showed that the degree of mispronunciation was reflected in pupil size in 30-month-old children (Tamási et al., 2017).

Swingley (2016) investigated in a word learning study whether two-year-olds associate a novel word with a novel object even when the novel word was a phonological neighbor of a familiar word. The results showed that children with a larger vocabulary tended to make novel-word interpretations more often, but overall the children in this study failed to associate novel words (as a phonological neighbor of a familiar word) with novel objects. This indicates that two-year-old children do not necessarily match novel words with novel objects when the novel word is derived from a phonologically similar familiar word (e.g. book - pook - buwk). In sum, there exist quite a few studies showing that mispronunciations affect word recognition in children between 14 and 24 months. However, a few studies showed that novel words resulting from word form manipulations were treated similarly to familiar words (e.g. Mills et al., 2004; Swingley, 2016), whereas others showed that word recognition differed between manipulated and correct word forms (Bailey & Plunkett, 2002; Mani, Coleman, & Plunkett, 2008; Swingley, 2003; Swingley & Aslin, 2000, 2002; Tamási et al., 2017; White & Morgan, 2008). A discussion of those observations follows in chapter 6.

Another study examined the semantic organization of words and tested 24-months-old children with semantically related and unrelated word pairs (Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007). A related word pair consisted of two words from the same semantic field, for example, dog and horse. An unrelated word pair consisted of two words from different semantic fields, like car and apple. The results showed an N400 effect between related and unrelated word pairs, suggesting that the N400 not only indicates congruency between two pieces of information but even semantic relatedness, such as belonging to the same semantic field. Additionally, associative and taxonomic relationships between words

9Phonetic features in the sense used by White and Morgan (2008) describe specific properties of a sound: Voicing as well as place and manner of articulation.
2.4. Phonology and lexicon in language acquisition

in the infant lexical memory could be identified as early as 21–24 months of age (Arias-Trejo & Plunkett, 2013).

2.4.2. Production

The physiological premises for speech perception and speech production are, of course, very different: For instance, the child’s speech production is measurable, in contrast to speech perception, only after birth. The infant vocal tract at birth looks different than an adult vocal tract not only in size but also in shape, and with respect to a higher position of the larynx, a shorter pharyngeal cavity, a large tongue in relation to the size of the oral cavity and a shorter distance between the velopharynx and the epiglottis (Stark, 1980). The infant vocal track goes through a considerable change over the first years of life and also in adolescence and beyond (Vorperian et al., 2009). On the one hand, physiological changes enable independent movements of the tongue and the jaw, which opens the door to finer articulations involving the tongue. On the other hand, constant adjustment in speech production is required to match the articulation of target sounds to the changing premises of the vocal tract and the related increase of degrees of freedom in articulation (Lindblom, 2000). In order to limit the number of degrees of freedom, mandibular oscillation is one of the first movements related to speech and, according to a low-energy assumption in motor emergents, is retained so as to bootstrap vocal behavior and development (Lindblom, 1999, 2000).

Looking at phonological production capacities, the child passes through different vocalization stages during the first year of life (e.g. Oller, 1980 or Roug, Landberg, & Lundberg, 1989, who propose five different stages, or Stark, 1980, who proposes six). All vocalization stages overlap heavily. Starting from sounds which are based on reflexes like sucking, speech productions become more language specific as the first glottal sounds occur. Stark (1980) called the next stage vocal play, which is characterized by the exploration of the vocal capacities. Kent and Murray (1982) describe mid-front or central vowels during the same time of vocal play (3-9 months). Vocal play is followed by canonical babbling (or reduplicated babbling as in Roug et al., 1989 and Stark, 1980) between six and ten months. Canonical babbling is defined by the combination of a consonant (C) and a vowel (V) and its repetition, for instance, *mamamama*. The emergence of canonical babbling can be seen as a consequence of least-effort jaw movements together with phonation, enabling the child to explore many of the articulatory patterns of the ambient language (Lindblom, 1999, 2000). This exploration can also be seen as practice for word production (Fagan, 2009; Vihman, Macken, Miller, Simmons,
2. The framework of phonology and lexicon

Canonical babbling is regarded as the first milestone in language production. A lack of or delay in producing canonical babbling is considered a possible sign of a language delay and connected to a higher risk of a later language disorder (Oller, Eilers, Neal, & Cobo-Lewis, 1998). The use of gestures such as reaching or pointing emerges in parallel to canonical babbling (Bates & Dick, 2002; Iverson & Thelen, 1999). Canonical babbling transitions to variegated babbling, where different CV combinations are joined together.

Around the first birthday, the child usually utters his or her first word. The identification of words can be guided by criteria based on context (e.g. multiple use, or identification by the parent), on vocalization shape (e.g. complex match in terms of the child form matches the adult form in more than two segments), and on the relation of the word form to other vocalizations (e.g. all instances of the word show the same phonological shape) (Vihman & McCune, 1994). The first words are not necessarily words for objects; they may be expressions related to social interactions, for example, goodbye or peekaboo (Bloom, Tinker, & Margulis, 1993). The emergence of words relating to social interactions might not be coincidental, as these situations are usually highly interactive and encouraging for the child to respond to both in terms of gestures as well as vocalizations.

Kauschke and Hofmeister (2002) observed relational words like up, social words and onomatopoeia in the early vocabulary of German children. Some researchers have described ‘proto-words’ to refer to relatively stable word forms with a consistent use but with no similarity to any target form with the same meaning (Vihman, 1996). The phonological structure of the first words is usually related to the structure of the most common babble (MacNeilage, Davis, & Matyeyar, 1997) and varies individually and across languages (Vihman, 1992). Still, the first word usually consists of one or two syllables (Menn & Vihman, 2011). The child’s vocalizations during the transition between babbling and first words may be classified in terms of vocal motor schemes. Vocal motor schemes are consistent phonetic forms based on a mastery of a particular sound pattern (McCune & Vihman, 2001). McCune and Vihman (2001) showed that the use of vocal motor schemes was related to the production of first words: those children who had vocal motor schemes at nine months produced their first words earlier than those children who did not have a vocal motor scheme at that age. Furthermore, McCune and Vihman observed the use of context-limited words at 11 months on average (context-limited words were defined as words produced in the absence of any evidence of generalized reference).

Other researchers included phonological profiles in the description of phonological development (Grunwell, 1982). Such a profile includes mainly information
2.4. Phonology and lexicon in language acquisition

about syllable structures and phonological rules (or ‘simplification processes’) (for a detailed background of the phonological profile suggested by Grunwell, see ch. 5). Examples of phonological rules are cluster reduction or fronting of velars, which can be observed in English children up to 3;11 (Dodd, Holm, Crosbie, & Hua, 2013). For Swedish, only a few studies have addressed the phonological development of typically developing children. For instance, Nettelbladt (1983) showed phonological processes similar to those of English children in a study of one typically developing Swedish child.

Ferguson and Farwell (1975) offered a detailed description of the phonological development of three children starting with transcriptions around the first birthday until they reached the point of 50 words. They observed a considerable variation in word forms and additionally many accurate word forms at the onset of first word production. The observation of accurate forms, followed by generalization and less accuracy, is usually referred to as a ‘U-shaped curve’ in, for example, phonology and inflectional morphology. Ferguson and Farwell (1975) observed some relatively stable word forms, while others varied a great deal. Two of the children occasionally whispered words. Stoel-Gammon (1985) observed labials and alveolar stops in initial position as the first language-specific consonants, followed by nasals and glides. Similarly, Ferguson and Farwell (1975) observed labials and alveolar stops as first consonants, followed by nasals and glides and finally fricatives. The acquisition of some fricatives can take up to five years (Ingram, Christensen, Veach, & Webster, 1980). The general observation of the relatively early presence of stop consonant and glides and the later acquisition of fricatives and affricates may be explained by phonetic production and perception constraints (cf. Kent, 1992). Fricatives require a high degree of motor control in terms of synchronization between fine positioning of the tongue and high air velocity. Fricative production during babbling is often contextually constrained reflecting greater difficulty in the production of these sounds (Gildersleeve-Neumann, Davis, & MacNeilage, 2000). For instance, the production of the fricative /s/ requires the right positioning of the tongue, lips and mandible. Additionally, the central passageway is considered to be difficult to establish and to maintain. In contrast, stops, for example, require only a more ballistic movement. Given the later acquisition of fricatives in general, in the present thesis, fricatives are used for word manipulations which are not related to the child’s own production (see chs. 4 and 6). In general, the phoneme inventory increases with age and is mostly richer in initial

10From a typological point of view, the absence of fricatives is more common than the absence of, for example, nasals (Maddieson, 2013). Out of 567 studied languages, about 9% lack fricatives; mostly languages spoken in Australia.
2. *The framework of phonology and lexicon*

position (Stoel-Gammon, 1985). At 18 months, consonants (alveolar nasals and voiceless stops) were observed in final position. Vihman, Ferguson, and Elbert (1986) observed labial and dental stops as the most common consonants in first word production.

Beside the general results on group data, several researchers have mentioned individual preference and strategies (see Ferguson & Farwell, 1975, for a description of a child with extensive use of reduplications or Stoel-Gammon & Cooper, 1984, presenting a child with preference for words with a velar stop in final position, and see chapter 5). Based on individual preferences for speech sounds and word structures, the approach of phonological templates has been developed (e.g. Velleman & Vihman, 2002; Vihman, 2014). A phonological template develops out of the production experience and can either be more general or more specific. An example of a general phonological template is consonant harmony, which has been reported in many different languages (e.g. English (Keren-Portnoy, Vihman, DePaolis, Whitaker, & Williams, 2010; Sowers-Wills, 2016; Vihman et al., 1985), Italian (Keren-Portnoy, Majorano, & Vihman, 2009), or Finnish (Vihman, 2010; Vihman & Velleman, 2000)). A more specific template can be found in the description of a child by Priestly (1977) with medial [j], resulting in word forms such as [bajak] ‘blanket’ or [nr jan] ‘melon’. In general, the identification of a template is based on its overuse (Vihman, 2016); specifically the template should be present in at least 10% of the sampled data (Kehoe, 2015; Vihman, 2016).

Phonological templates can be used in two different ways: (1) for the selection of word forms from the speech input for adult target forms that fit the phonological template, in which the child chooses to produce those adult target word forms that match the phonological template, and (2) for the adaptation of adult word forms in the phonological template, in which the template is applied/overgeneralized to adult target forms that do not match the template. In the first case, the child’s practiced motor routine helps to procure new words from the input, which are similar to the practiced routine. For instance, the child presented by Ferguson and Farwell (1975) procured from the input those words that had medial sibilant fricatives. Another child acquiring German selected words with reduplication (Vihman & Croft, 2007), for example, [dr dr] Teddy ‘teddybear’. An example of adaptation to a phonological template can be observed in the above mentioned case described by Priestly (1977): the template involved CVjVC, and the child adapted adult word forms to it, for instance, [bajak] ‘blanket’ or [nr jan] ‘melon’. Evidence for the use of templates was observed in Arabic-speaking children (Khattab & Al-Tamimi, 2013), Brazilian Portuguese (Oliveira-Guimarães, 2013), Polish (Szreder, 2013), French (Wauquier & Yamaguchi, 2013) as well as bilingual children (Kehoe, 2015;
2.4. Phonology and lexicon in language acquisition

Vihman, 2016; Vihman & Vihman, 2011). More on phonological templates follows in chapters 5 and 6 addressing the research questions 1, 3 and 4.

In contrast to phonological processes, a template affects the whole word form and not only specific segments or phonemes. It is an individual routine, which means that it might not be observable in other children (Keren-Portnoy et al., 2009), and as they might change quickly it might be hard to observe them at all in some children. Moreover, the use of templates predicts an influence of production on perception.

First words in the lexicon

The link between phonology and the lexicon is particularly evident in the child’s first words, since these are embedded in the child’s babbling sequence. In the beginning of first word production, new words are added rather slowly to the productive vocabulary at 1-5 words per week (Clark, 1995). However, between the first and second birthday the vocabulary grows for the majority of the children from a few words to 50 words or more. The semantic fields of the first words often consist of people (e.g. *mummy*, *daddy*, *baby*), animals (e.g. *doggy*) or space and motion (e.g. *up*, *off*, *open*) (Clark, 1995). In a study of children between 1;5 and 2;0 years of age, general nominals (names, substance, animals) constituted around half of the lexical types, verbs and adjectives, with each around 12% (Leonard, Camarata, Rowan, & Chapman, 1982). The overextension of early words to other entities with similar properties is often reported. After passing the 50-word point, the children were usually faster to acquire new words (e.g. Goldfield & Reznick, 1990), resulting in vocabulary sizes around 500 words at 30 months (Fenson et al., 1994). However, the variation in lexical size is usually large with the largest variability between 16 and 24 months (Bates et al., 1994).

The growth of vocabulary can be described in three different phases (Bates et al., 1994): the referential, prediction and grammar phases. These phases describe which word classes (nouns, verbs or function words) dominate the word learning process during the acquisition of the lexicon (for a more detailed description of these phases, see ch. 5).

The composition of the early vocabulary is also related to environmental influences such as the target language and to whether it is considered to be noun-friendly like Swedish, German, English or French, or verb-friendly like Mandarin, Korean or Japanese, for example. Thus children acquiring Mandarin had more verbs in their early vocabulary than nouns (Tardif, 1996). The distinction between noun-friendly and verb-friendly languages was also reflected in the speech
2. The framework of phonology and lexicon

directed to children: Mandarin-speaking mothers used more verbs in comparison to English-speaking mothers when they talked to their children (Tardif, Gelman, & Xu, 1999). Thus, the early vocabulary of children acquiring Mandarin contained more verbs than nouns (Tardif, 1996). Additionally, language-specific acquisition was observed in Danish children, who seemed to be slower in their acquisition of a lexicon compared to those acquiring other languages (Bleses et al., 2008).

The differences in word types notwithstanding, the bias for nouns in the first words can explained in terms of their concreteness of reference in comparison with verbs (e.g. Gentner, 2006; Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011). Nevertheless, not all nouns among the first words produced are object nouns (Nelson, Hampson, & Shaw, 1993). Verbs often describe a relationship among nouns, suggesting that the acquisition of nouns is necessary to acquire verbs (Waxman et al., 2013).

Another study reported the acquisition of words being influenced by word length, frequency and phonological neighborhood (Storkel, 2004): Low-frequency words are acquired earlier if they have many phonological neighbors. The acquisition of high frequency words is related to word length. The acquisition of short words is related to neighborhood density and word frequency (Storkel, 2004). However, frequency is also influenced by the word’s lexical category: closed-class words such as pronouns, articles or prepositions are the most frequent in the environment, but are, nevertheless, not acquired early (Goodman, Dale, & Li, 2008). Word learning is a complex developmental trajectory, which involves separate, changing and interactive influences of different aspects such as word frequency, neighborhood density and phonotactic probability (McKean, Letts, & Howard, 2013).

Relationship between phonological and lexical development

In addition to the traditional linguistic distinction between perception and production in language acquisition, the domains of phonology and lexicon are often treated separately. However, from the perspective of a child’s development these research categories are presumably artificial, but are necessary from a research perspective. The areas of overlap of phonology and lexicon indicate an interrelationship between those two domains (for a review, see Stoel-Gammon, 2011). For instance, children who have babbled more in their first year, perform better in tasks related to the lexicon later in life (Menn & Matthei, 1992a). The phonological form of the child’s first words might be considered as another evident relationship between phonology and the lexicon, as the words added to the lexicon often follow certain preferences for specific phonological word forms (e.g. Ferguson & Farwell,
2.4. Phonology and lexicon in language acquisition

For instance, an over-representation of words starting with /b/ was observed in the early lexicon in infants learning several languages (de Boysson-Bardies & Vihman, 1991; Stoel-Gammon, 1998), despite their low frequency in infant-directed speech (Stoel-Gammon & Peter, 2008). Additionally, vocabulary size seemed to be a predictor of advances in the phonological system; that is, children with a small vocabulary tended to have a limited phonological system, while the opposite could be observed for children with a large vocabulary size (for a review on the relationship between phonological and lexical development, see Stoel-Gammon, 2011).

2.4.3. The link between production and perception

Some researchers have suggested a link between the occurrence of perceptual narrowing and the onset of canonical babbling (i.e., Vihman, 1996, 2017b; Kuhl, 2000; Kuhl et al., 2008; Westermann & Miranda, 2004). The results of a recent paper (Masapollo, Polka, & Ménard, 2016) support that suggestion: Infants’ own production abilities influenced the perceptual preference in pre-babbling infants aged three to six months, as they showed a preference for speech which was produced by an infant vocal tract. Another study by Altvater-Mackensen, Mani, and Grossmann (2016) showed that production abilities influenced sensitivity to audiovisual matches. In this study, six-month-old children looked at short clips of a woman who articulated vowels. The simultaneously presented auditory stimulus either matched the visual cue of the vowel or not. The authors were able to show that children who had more distinct sound categories in their babbling were better able to detect mismatches between the vowels seen and heard. DePaolis, Keren-Portnoy, and Vihman (2016) tested older children, at 10 months, in an auditory head-turn preference procedure on familiar and rare words, and compared their performance in this task to their production capacities as measured by vocal motor schemes. Children who were early in the production of vocal motor schemes showed either a strong novelty or strong familiarity response, whereas children who were not at that production level yet showed no preferential response to either word set.

What the very recent results of these studies have in common is that they relate perceptual abilities indirectly to production capacities. So far there appears to be no study which connects the individual child’s phonological capacities directly to his or her perceptual knowledge of words. Additionally, the link between production and perception has not yet been investigated in older children in the very dynamic period of the development of phonology and the lexicon during the second year of life.
2. The framework of phonology and lexicon

2.5. Summary

The present chapter has summarized earlier research on phonology and the lexicon in language acquisition, including methods capturing both perceptual and production capacities. The phonological system and the mental lexicon involve several aspects for the child to acquire, such as the phoneme inventory of ambient language, phonotactic characteristics, phonological variations (rules or processes) and prosody, as well as word meaning, syntactic form, morphology and phonological shape of a word. Different methods, for instance, eye-tracking and methods that assess brain responses, capture different aspects of development. While speech production capacities start to develop after birth, speech perception starts earlier, in the prenatal period. Statistical learning is a powerful tool for extracting constituent boundaries; word learning principles such as mutual exclusivity are assumed to be facilitative by some researchers. From early on, perceptual abilities include the categorization and discrimination of speech sounds. At around six months, children recognize familiar words; at twelve months children start to process words in a way that is semantically similar to adults in terms of N400 responses. At around twelve months, children produce their first words, after passing through several stages of vocalizations. During the second year of life considerable changes in word production as well as perception are observed. Some studies have shown that children recognize minor mispronunciations of word forms within the first and second year of life. At the same time the vocabulary increases considerably and speech production capacities improve, indicating a relationship between phonological and lexical acquisition. Several researchers have assumed a link between production and perception; however, studies investigating that link are rare and include children younger than one year. Thus, to the best knowledge of the author, there is so far no experimental study linking perceptual and production capacities to the acquisition of phonology and the lexicon during the second year of life. However, the next chapter presents some theoretical approaches suggesting a link between production and perceptual abilities, after introducing general approaches and specific theories on phonological acquisition.
3. Theoretical approaches to language acquisition

Several attempts have been made to develop theories, models and approaches to explain how children acquire language in general terms or in more specific terms of phonology and the lexicon. However, only a few assume a possible link between perception and production, and even fewer put this link in concrete terms. Theoretical approaches addressing phonological and lexical acquisition are presented below, after a very brief introduction to general approaches to language acquisition.

3.1. General approaches to language acquisition

In the 1960s, Noam Chomsky made his first attempt to account for language acquisition as a reaction to behaviorist theory. In his *generative* account language proceeds from a module or device which is specific to language and innate (Chomsky, 1965). The innateness of language explains the fact, as put forward by Chomsky, that children learn language in a short time and without effort. The ‘chomskian’ way of explaining language and its acquisition focuses on grammar and examines universal principles as the basis for underlying structure. The theory postulates that the child adjusts these principles through parameter settings to irregular and/or target language forms on the basis of a minimum of exposure to the ambient language. In other words, the child acquires language by fine-tuning already existing, innate knowledge of language.

In contrast, ‘usage-based’ approaches assume that language emerges through interaction with the environment and experience of language use and that the linguistic system is built up gradually with increasing linguistic experience. ‘Usage-based’ approaches suggest general cognitive capabilities, which are used to categorize and detect similarities and differences in the environment and particularly in the speech surrounding the child (Bybee, 2006). These cognitive processes include
categorization, memory and the cognitive capacity to connect form and meaning (Bybee, 2010), or, on the functional dimension, ‘intention-reading’, and, on the grammatical dimension, ‘pattern finding’ (Tomasello, 2003). The experiences that occur in the interaction with the environment are stored in memory. ‘Usage-based’ approaches presuppose powerful generalization mechanisms that make it possible to go beyond linguistic experience (Behrens, 2009).

A general biological-cultural perspective on language acquisition takes into account the emergence of speech units such as syllables and phonemes (Lindblom, 2011). Lindblom suggests that the acquisition of phonemes is connected to the “communicative need to develop large lexicons” (Lindblom, 1992, p. 158). On their path to phonemic representations, infants are guided by articulatory, perceptual and memorial processes. Lindblom argues, based on the contributions of MacNeilage (2008), that the syllable is rooted in mechanisms developed earlier than the first appearance of human language (Lindblom, 2011). Concretely, the open-closed jaw movement is used in human language as a serial organizer of speech and results in the syllable, which builds the basis for canonical babbling. In order to support the importance of jaw movements to language acquisition, several clues were proposed, among them neurobiological clues and clues from motor control. The importance of jaw movements is also reflected in canonical babbling which is characterized by the fortuitous production of pseudo-consonant and pseudo-vowels (Lindblom, 2017). As soon as the child realizes that she has some control of the articulators, she realizes that she can make other sounds, which leads to position control, which is not specific to language, and intentional variegated babbling. According to Lindblom, this is the onset of a phonological system (Lindblom, 2017).

Inspired by the ideas of Lindblom on the emergence of language acquisition, Lacerda and Sundberg (2006) proposed a framework incorporating the evolution of language. The ‘Ecological Theory of Language Acquisition (ETLA)’ suggests that the evolution of language and its acquisition follow similar processes: The combination of cultural and genetic evolutionary processes leads the human to language and the infant to language communication. The authors claim that language acquisition is a consequence of the child’s interaction with the environment and the multi-sensory information in the frame of physiological prerequisites. Some basic presuppositions are crucial for language development: firstly, the availability of multi-sensory information; secondly, the general capacity to store incoming sensory information, and also the child’s capacity for social interaction. Infant-directed speech helps the infant to perceive language as an efficient and lively instrument and as a key to interaction. Furthermore, language does not appear on its own but rather is embedded in situations including olfactory, visual, tactile, kines-
3.2. Approaches to phonological and lexical acquisition

Besides these general approaches to how children acquire language, several researchers provide domain-specific models to explain different aspects of phonological and lexical development, for instance, speech stream segmentation or word learning. Along with the assumptions about how infants and toddlers segment the speech stream or recognize a word, different suggestions have been made about what shape representations of lexical entries have and how they develop. What follows is a description of several different models connected to speech perception and speech production, and their assumptions on how items are represented in the lexicon.

One of the earliest attempts at a general model on phonological development was provided by Jakobson (Jakobson, 1941/1968). He based his model on diary studies and assumed a pre-linguistic period in the form of babbling and a linguistic period when the child acquires phonology. The role of babbling is assumed to be totally unrelated to phonological development due to its seemingly unstructured shape. By contrast, phonological development follows universal structures whereby, for example, those sounds which are most common cross-linguistically are acquired first. Jakobson’s work was very important in the comparatively short history of child phonology and served as a significant catalyst to later research of child language acquisition and development.

WRAPSA

Peter Jusczyk’s model of ‘Word Recognition and Phonetic Structure Acquisition’ (WRAPSA, Jusczyk, 1997) relies mainly on speech segmentation. According to this model, speech segmentation is conducted in several steps. First, the speech stream undergoes preliminary auditory analysis, which provides temporal and spectral information. This information is weighted in the second step by properties of the sounds as well as their distribution in the speech input. The weighting scheme is language-specific and develops over time. The third step in word recognition involves a pattern-extraction process, which analyzes the weighted output. All three
steps (auditory analysis, weighting and pattern extraction) result in successful segmentation of the speech stream and extraction of sound patterns. In the fourth and last step, Jusczyk assumes that those sound patterns are stored as exemplars in the mental lexicon and might take the form of a syllable or a larger unit. If the pattern is a close match to one of the stored exemplars, the word is recognized and access to meaning is possible. If the detected pattern does not match any of the stored exemplars, it will be stored as a new entry. Furthermore, Jusczyk proposes a rather global representation in terms of syllables incorporating salient features, especially stress and tone. These two features, among other prosodic characteristics, are the first pieces of information that are accessible to the child, followed by syllabic and phonetic information. Jusczyk concludes that “the most efficient representation is the one that embodies only enough detail to accurately recognize the word when it appears”, especially under the presupposition that perception and production capacities might influence the derivation of representations (Jusczyk, 1997, p. 226, for further reading on WRAPSA, see Houston & Bergeson, 2014).

Since WRAPSA is based on speech segmentation, it does not address the question of the child’s learning of words in any detail, that is, how meaning is mapped onto the phonological representations. WRAPSA acknowledges that production might have an impact on phonological representations, but assumes that, in general, perception comes first and speech production capacities might only refine speech perception capacities in the form of auditory processing. Hence, it remains unclear in which way production capacities influence word form representations.

PRIMIR

Werker and Curtin (2005) proposed a framework for ‘Processing Rich Information from Multi-dimensional Interactive Representations’: PRIMIR. According to this view, speech processing is influenced by three ‘dynamic filters’: initial biases, the developmental level of the child, and the nature of the task. The filters enable the child to focus attention on different types of information. The speech signal is rich in multidimensional information and is mapped onto multiple dimensions of representations. The model’s representations group the information on three multidimensional planes: the General Perceptual plane (1), the Word Form plane (2) and the Phonemic plane (3). The planes do not necessarily form a hierarchy, but are organized in parallel. Each of the planes encapsulates a different kind of regularity: The General Perceptual plane captures all information included in the signal, in particular those of phonetic and indexical characteristics. Phonetic and indexical categories are based on general perceptual features and consist of ‘clusters of
exemplar-like distributions (Werker & Curtin, 2005, p. 214). Furthermore, on this plane, language-specific categories emerge and influence the Word Form plane through sequences that constitute larger units. The Word Form plane includes ‘extracted units without meaning attached’ (Werker & Curtin, 2005, p. 216). Similar to WRAPSA, PRIMIR assumes storage of exemplars, whereby exemplars which share phonetic or indexical characteristics are grouped into clusters, resulting in neighborhoods of greater or lesser density but still lacking meaning. The Phonemic plane combines the information from the General Perceptual plane and the Word Form plane, and emerges once a criterial number of links between word forms and meaning is established.11 After establishing phoneme categories, phonemes facilitate word learning in the sense of leading the child’s attention to the “information in the word form that is essential in learning a new word–object link” (p. 218). In other words, the emergence of phoneme representations affects speech processing and word learning. Thus, the Phonemic plane is prioritized in terms of focal attention to those features that support the emergence of a phonemic system of a specific language. PRIMIR predicts that once information about segments is available, “it should be accessible at every subsequent point in development under at least one task condition” (p. 224). In contrast to WRAPSA, which assumes that phonemes are established before the extraction of word forms through the weighting scheme, PRIMIR presumes that phoneme categories emerge from the link between word forms and concepts. In addition, PRIMIR presupposes that statistical learning is a crucial mechanism at all levels of analysis. In sum, PRIMIR suggests that at the end of the first year, language-specific phonetic categories are established that incorporate sub-phonetic and indexical detail. The level of detail is influenced by perceptual bias, developmental level and task demands.

PRIMIR relies, as does WRAPSA, on the results of perceptual experiments and does not take into account the individual differences and variability observed in speech production studies and apparent in word learning. Despite suggestions about word recognition in both PRIMIR and WRAPSA, neither of them explicitly address the role that the child’s phonological and lexical capacities in production may have in word recognition (cf. research question 1).

11This assumption is reminiscent of other models, such as that of ‘lexical restructuring’ by Metsala and Walley (1998), in which the lexicon needs to be restructured under the pressure of increasing vocabulary size and neighborhood density.
3. Theoretical approaches to language acquisition

NLM-e

A more recent theory, proposed by Kuhl et al. (2008), accounts for the perceptual narrowing\(^{12}\) in the first year: ‘Native language magnet theory expanded’ (NLM-e). The ideas of NLM-e arose from earlier studies investigating vowel categories in infants (Kuhl, 1987, 1991; Kuhl et al., 1992). NLM-e assumes that some areas in the vowel space serve as prototypes based on a greater perceptual stability. These stimuli, which acoustically surround this prototype, are perceptually drawn towards the prototype on analogy to a magnetic effect. As a result, the perceived distance between the prototype and the stimuli is smaller than the acoustic distance suggests. The model consists of four different phases: Phase 1 describes the ability of the infant to discriminate all phonemes of the world’s languages, as suggested by Kuhl (2004). During phase 2 (‘neural commitment’) distributional learning together with the characteristics of infant-directed speech result in phonetic learning. Phonetic learning of vowels occurs earlier than of consonants. Social interaction strengthens phonetic learning, and as the child increases her skills in social understanding, she advances in phonetic learning. NLM-e suggests a link between speech perception and production during that phase in terms of a bi-directional effect occurring during the same time that phonetic learning is established. During phase 3, three independent skills are established: detection of phonotactic patterns, detection of transitional probabilities between segments and syllables, and mapping between words and meaning. During phase 4 neural representations become relatively stable.

Despite the concrete acknowledgment of a link between speech perception and production proposed in the model, it is unclear what the nature of this link might be and how it might change over time, in particular, when it comes to word learning.

Phonological specificity

Another approach (Swingley, 2003; Swingley & Aslin, 2000, 2002) has assumed that representations of a word are stored in an adult-like form with full phonological detail. The assumption is based on a series of studies examining word recognition and word learning in children between 12 and 24 months. For instance, studies of 14-month-old children have shown that mispronounced word forms impede word recognition (Swingley & Aslin, 2000, 2002). Sensitivity to phonological details of lexical forms was tested in word learning tasks with initial conso-

\(^{12}\)An earlier model on perceptual narrowing is provided by the ‘Perceptual Assimilation Model’ (PAM) (Best, 1994).
3.2. Approaches to phonological and lexical acquisition

nant changes (Werker et al., 2002) or vowel changes (Mani et al., 2008; Mani & Plunkett, 2008). Encoding of phonetic detail was also observed in medial position (Swingley, 2003) and in a study with increasing distance to the target word form in terms of phonological features (White & Morgan, 2008). In addition, sensitivity to mispronunciations was found to be independent of the number of phonological neighbors, which means that neighboring sound forms do not necessarily force the child to adopt to a finer word form representation (Swingley, 2003; Swingley & Aslin, 2002). This empirical evidence led to the hypothesis that “young children store words the same way adults do – not as vague or ‘holistic’ soundscapes containing only a few phonetic features, but as precise, well-specified phonetic forms” (Swingley, 2003, p. 285).

However, other empirical studies have shown that not all mispronunciations block word recognition (for instance, Hallé & de Boysson-Bardies, 1996). These results are interpreted as a question of memory, specifically:

(...) under some circumstances, this phonetic information that was accurately perceived is forgotten; that is, at the moment of hearing a word, its sounds may be accurately categorized, but nevertheless not stored in memory in a format that can then help guide the process of word recognition. Young children’s problem is not perception per se; it is encoding in memory. (Swingley, 2003, p. 288)

In sum, the approach of phonological specificity assumes that the infant learns and stores phonologically detailed, adult-like representations in the lexicon. Failures in the detection of the phonological details are explained by memory constraints. With respect to the first research question of the present thesis (“How, if at all, do a child’s phonological and lexical capacities in production influence her word recognition?”, cf. ch.1, p. 2), the phonological specificity hypothesis does not assume an influence of production capacities on word recognition. Presumably, as the word form representations are stored in phonological detail and in adult-like forms, phonological specificity would exclude an influence of speech production capacities on word recognition (more detailed reasoning on this assumption follows in ch. 6).

Whole-word approach

Other theoretical approaches link perception and production more explicitly, for example, in terms of a ‘holistic’ representations of early words (Vihman, 2014, 2017b; Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013a): A “‘holistic’
3. Theoretical approaches to language acquisition

representation means memory for the word form as a whole, with retention of salient syllables or segments, particularly those that are within the child’s output repertoire (...)” (Vihman, 2017b, p. 11). Assuming representation of a word as a whole, some parts of the word form might be better represented than others, indicating that words are learned as whole units (Vihman & Keren-Portnoy, 2013b). Vihman, DePaolis, and Keren-Portnoy (2009) suggest an ‘articulatory filter’, which unconsciously directs the attention of the child towards those segments in the input that are familiar to the child from her own production experience. Thus, the phonological representation of a word is influenced “not only by the percept of the target word itself but also by coexisting (‘whole word’) production patterns in the child’s repertoire” (Vihman et al., 2009, p. 173). Furthermore, a ‘holistic representation’ of word forms “implies that elements in one part of a word may affect memory for other parts” (Vihman, 2017b, p. 11). Evidence for the distraction of the attention from one part of the word to another is taken from a study with Italian children (Vihman & Majorano, 2016). Mispronunciations in onset position were treated differently depending on what follows later in the word: Word-medial geminates drew the child’s attention away from the stressed word-initial syllable, leading to an acceptance of word forms despite the onset mispronunciation. In contrast, onset mispronunciations in words with medial singletons blocked word recognition.

According to a whole-word approach, word learning consists of three stages (Vihman & Keren-Portnoy, 2013b): 1. Item-based word learning, 2. Distributional learning and 3. Generalizations. The first stage includes the learning of words as items (exemplars), which implies that children acquire words as wholes (Velleman & Vihman, 2002; Vihman & Croft, 2007). The unit of a word is the ground for phonological representations (Vihman & Croft, 2007; cf. Bybee, 2001). The second stage, distributional learning, includes learning of sound patterns affected by the distributional properties of the sounds, that is, the frequency of the sound and its position within the word as well as how many neighbors the sound pattern has (see, for instance, Zamuner, Gerken, & Hammond, 2004, or Storkel, 2001). From the regularities encoded by means of distributional learning the last sequence emerges, generalization, which Vihman refers to as ‘secondary distributional learning’ (2014). Generalizations over the phonological structure of words are called ‘templates’ (see section 2.4.2). Templates are dynamic and change over time with an increasing use during a limited time period which can consist of a couple of days, weeks or months. When the child starts to attain adult-like artic-
3.2. Approaches to phonological and lexical acquisition

ulation and speech planning as well as the required memory capacities, the use of templates recedes. For some children, the templates are related to the patterns that were practiced during babbling (Vihman, Velleman, & McCune, 1994). A template can either be selected or adapted or both (see section 2.4.2). The selection of words is seen as “an implicit response to what is familiar, based on input frequencies” (Vihman, 2017a, p. 40). The adaptation of target word forms to the templates allows the child to increase his or her productive vocabulary size. In terms of a dynamic system, those production patterns are assumed to “self-organize into networks based on similarity” (Vihman, 2017b, p. 17).

Several arguments have been put forward to support the assumption of a ‘whole-word’ representation (Vihman & Croft, 2007; Vihman et al., 2009). Firstly, the child’s production is variable. Varying pronunciations of a word indicate that the child has learned an item, but no precise articulation pattern to match this item is available yet. Furthermore, it indicates that no abstract sound categories have been established yet. Secondly, child word forms often match adult word forms ‘holistically’ through rhythm or word structure, for example, indicating that children aim to articulate the target form as a whole rather than as specific segments. Thirdly, the child’s own words are similar to each other, indicating that the child relies on a production pattern and uses it across words. Fourthly, the child patterns are assumed to constitute a response to challenges provided by the segmental sequences of the target word forms.

The whole-word phonology model focuses on speech production, suggesting an influence of speech production on perception. However, the term speech production in this model covers many different descriptions of vocal gestures, for instance, babbling, vocal motor schemes, or first words; yet, “there is no clear evidence of a unitary relation between these behaviors and phonological processing” (Zamuner, Yeung, & Ducos, 2017, p. 37). Additionally, the nature of the influence of speech production on perception might be related to different aspects, such as task or attentional effects, and not all of them need be positive (Zamuner et al., 2017).

In sum, according to whole-word phonology words are learned as items. Item learning, together with distributional learning, is necessary to build a phonological system. Generalizations by means of templates reflect the influence of production on perception in terms of selection and adaptation of target words. This approach predicts that speech production capacities influence word recognition (see ch. 6).
3. Theoretical approaches to language acquisition

Linked-Attractor model

The ‘Linked-Attractor’ model proposed by Menn, Schmidt, and Nicholas (2013) assumes “that our brain builds up auditory representations of input forms, articulatory/kinesthetic/sensory and auditory representations of output forms, and inter-modal representations of the mappings between them” (Menn et al., 2013, p. 484). The ‘Linked-Attractor’ model follows the ‘Two-Lexicon’ model of Menn and Matthei (1992b) and includes the following four elements:

1. **Templates** ‘attract’ output forms, which implies that ‘output templates’ are those sound patterns that are likely to be produced by the child as a consequence of motoric capacity, intensive practice and thus familiarity. ‘Input templates’ draw the attention of the child to specific sound patterns in the input. Input and output templates are continuously modified and adjusted along with increasing experience.

2. **Output constraints** describe what sounds and sound combinations can be produced. Output constraints stimulate rules, but also explain input-output mappings, which are not covered by rules.

3. **Stored output forms** include those persisting word forms that are maintained although the child has mastered the production of similar words.

4. **Rules** map input into output forms based on regularities.

Input and output representations, as well as the mapping from input to output, are seen as ‘attractors’. Those attractors emerge through listening, understanding and speaking; thus the ‘Linked-Attractor’ model is a gradual learning model. The dominance of any one attractor is dynamic and varying through development, and “each attractor has the power to drag other representations towards itself (in a landscape metaphor, it creates a basin or groove): the more frequent the form, the bigger the attractor basin or groove, other things being equal” (Menn et al., 2013, p. 485).

The input representations consist of complex connections of usage information (i.e. in which situations the word is used, speaker identity, and emotional coloring). Those representations with the least variation and strongest emotional associations are most strongly represented. Output representations consist of an accumulation of word production experience: All attempts to produce a word strengthen the representational link between form and meaning, including imitations as well as spontaneous word productions. The attractor of mapping input to output links auditory to articulatory information, incorporating rules on how input...
3.3. Concluding thoughts

The ‘Linked-Attractor’ model predicts, on the one hand, that frequent occurrences of a sound in the environment (input) will lead to a stronger auditory representation. On the other hand, frequent production patterns (output) will strengthen the articulatory representation of a word or sound pattern. In other words, if a word is infrequent in the child’s environment but frequent in the child’s production, the child will maintain his or her way of saying it as the articulatory representation is strengthened by frequent production. However, as the authors themselves acknowledge, this prediction is hard to test as input and output frequency (at least for content words) seem to correlate. Testing this prediction would require a large sample with enough phonological data to separate words that are more frequent in input than in output and vice versa. Nevertheless, the representations of a familiar word are strongly influenced not only by frequency but also by sensory, motor, social and emotional aspects. Such representations, comprising multi-modal information, might be difficult to test, but take the possibility of an influence of production on perception into account.

In sum, the approaches presented account for different aspects of speech processing and word learning. WRAPSA focuses on speech segmentation; PRIMIR and NLM-e take phoneme learning and categorization into account. The approach of phonological specificity is based on word recognition studies, whereas the whole-word approach is grounded in production studies. The ‘Linked-Attractor’ model ties in with the templates from the ‘whole-word’ approach and embeds input and output representations as well as mappings between them into a psycholinguistic model. Not all of the approaches make concrete assumptions related to the research questions in this thesis, but of special interest to the first research question are the models of whole-word phonology as well as of phonological specificity, which we will return to in chapter 6.

3.3. Concluding thoughts

The present chapter presented different theoretical approaches to understanding language acquisition and in particular the acquisition and representation of word forms. Some of the theoretical approaches acknowledge a link between perception and production, such as the whole-word approach as well as the Linked-attractor model (and to some extent NLM-e), although some predictions made by the Linked-attractor model are difficult to test. However, from the perspective of
3. Theoretical approaches to language acquisition

A model that assumes a multi-modal acquisition of language, the child’s own production is a crucial piece of the puzzle, as production provides not only auditory information, but also proprioceptive and other sensory information that supports memory. Accordingly, word form representations are likely to carry more information than what is provided by the speech of the environment. In particular, the phonological word form representation might not only include information about how a specific word sounds, but also how the word sounds or feels when the child produces it herself. The relation between production and perception might be seen as an intertwined pathway to language for each child individually, where the capacities of speech perception and production alternate in disentangling information about language in general and in the concrete case, to build word form representations. Thus, in order to understand the emergence of perceptual abilities, speech production capacities have to be taken into account.

Additionally, there is an interrelationship between phonological and lexical development, indicating that these linguistic domains are not acquired independently. Considering these interrelationships, sounds are not acquired independently or before the child acquires a lexicon. Rather, as word form representations include both types of information about the meaning and the form at the very least, phonological and lexical acquisition must occur in parallel (consider, for instance, the earliest evidence of infants recognizing familiar words at 6-9 months, Bergelson & Swingley, 2012, which is the typical age for the onset of canonical babbling, Stark, 1980; Oller, 1980; Roug et al., 1989). This implies that measures relating to the size of the expressive lexicon, which are related to the child’s phonological capacities, would give us a more detailed picture of the development of these. A possible influence of speech production on perception may contribute, not only to linguistic theory and how word forms are represented in the mental lexicon, but also to clinical implications. For instance, a possible influence might be an indication to emphasize the child’s own production of word forms in clinical interventions in addition to adult word forms.

The present thesis aims to contribute to bringing studies of speech production and perception together, filling the research gap relating the child’s own production capacities to her perceptual abilities. The studies presented in the thesis seek to find evidence for the links between perception and production as well as lexicon and phonology in terms of exploring the related research questions, which address the influence of phonological and lexical capacities in speech production on word recognition, the relationship between lexical and phonological acquisition, as well as the evaluation of methods capturing phonological and lexical performance. Additionally, the research questions relate to the lack of studies examining
3.3. Concluding thoughts

phonological capacities in Swedish toddlers in general, for example, with respect to phonological characteristics related to vocabulary size. Furthermore, the questions address the under-investigation of a possible relationship between production and perception during the acquisition of phonology and the lexicon. Hence, the link between perception and production related to phonological and lexical capacities in Swedish toddlers is investigated in the following chapters.
4. **Study I: Three experiments on methodological explorations**

This chapter is related to research questions 1 and 2, namely the question of how, if at all, do a child’s phonological and lexical capacities in production influence word recognition and which setting and combination of research methods might be appropriate for investigating phonological and lexical knowledge in speech production and perception at different ages during the early word-learning period. In order to investigate these questions a series of experiments was conducted that explored methodological options as well as stimulus materials, taking typical mispronunciations of children into account. Two out of the three experiments used a combination of EEG and eye-tracking, one only eye-tracking. The three experiments differ in the choice of picture material and in the ages of participants: The first experiment includes adults and evaluates stimulus material in an eye-tracking experiment. The second experiment pilots the combination of EEG and eye-tracking in 17-month-old children. The third experiment constitutes a revision of the second experiment and includes 24-month-old children. The purpose of conducting the three experiments was to explore and evaluate methods and stimuli for investigating the relationship between speech perception and production.

4.1. **Background**

First language acquisition is a broad research field, which is mainly investigated through three different aspects: the language environment, and the perceptual and the productive capacities of the infant. Many studies investigate perception and production separately (for a review, see Kuhl, 2004). Although the relationship between the child’s lexical representation and phonological ability is of special interest in the research on early language acquisition, the acquisition of lexicon and phonology have mostly been investigated independently. For instance, in preferential looking tasks, 14-month-old children showed sensitivity to mispronunciations in terms of vowel change for newly acquired novel words (Mani et al., 2008; Mani...
Similar effects were found for 15-month-old children (Swingley & Aslin, 2002). Children at this age adjusted their looking behavior depending on the pronunciation (close as well as distant mispronunciations and correct pronunciations) of the target word. Closely mispronounced words were created by changing a sound to a similar one, e.g. dog to tog; distant mispronounced words were created by changing one sound to a phonetically more dissimilar sound, e.g. dog to mog. The results showed that mispronunciations were treated differently in terms of looking time than correct pronunciations. There was no difference between the two mispronounced conditions (Swingley & Aslin, 2002). At 19 months of age, children also showed sensitivity to the degree of phonological mismatch (White & Morgan, 2008). In this eye-tracking study the children looked at two pictures, consisting of one familiar and one unfamiliar object. The auditory stimuli consisted of correct word forms as well as different degrees of mispronunciations: change of place of articulation, as in bear / gare, or changes of the place of articulation as well as the manner of articulation and voicing, as in book / sook. The results showed that the more the pronunciation deviated from the correct one, the less the children looked towards the familiar object. Changes in one feature led to increased looking towards the familiar object, but not to the same extent as when the word was correctly pronounced. These results suggest that familiar words could be represented in detail regarding the phonological specification of the stored representations (cf. Bailey & Plunkett, 2002; Mani & Plunkett, 2007; Swingley & Aslin, 2000).

Another type of perceptual study uses event-related potentials (ERPs) to measure brain responses to test, among other things, processes related to lexical retrieval. These brain responses are measured by electrodes positioned on the participant’s scalp that detect electrical activity occurring in the brain (Luck, 2005). ERPs typically reflect postsynaptic potentials, which means the voltages that occur when neurotransmitters connect to ‘receptors on the membrane of the postsynaptic cell, causing ion channels to open or close and leading to a graded change in the potential across the cell membrane’ (Luck, 2005, p. 28). The electrodes on the scalp record the summed voltages as a result of many neurons firing together. A typical procedure in event-related studies examining lexical retrieval is the following: infants listen to auditory stimuli, and look simultaneously at a picture. This picture is semantically either a match or a mismatch to the auditory stimulus (for a review, see Friederici, 2006). A mismatch evokes a so-called N400 effect in neural activation. One study indicates that even 14-month-old children have lexical representations and show signs of understanding basic-level words (Friedrich & Friederici, 2005a). A different version of this type of study uses nonsense words as a mismatch instead of existing words. The relationship between brain responses in
ERP and mismatched stimuli with nonsense words showed that even 12-month-old infants showed an N400 effect (Friedrich & Friederici, 2010). Nevertheless, measuring the word production of the children in this study, the effect occurred only for children with high early vocabulary scores in parental reports. The authors postulate that the neural mechanisms underlying N400 are related to the infant’s early word-learning skills, which points to the relevance of the N400 neural mechanism in the process of early word acquisition. Similar results in the brain activity of 14- and 20-month-old children have been observed using phonetically dissimilar nonsense words compared to mismatched known words (Mills et al., 2004), with auditory stimuli only. Phonometically dissimilar nonsense words were, for example, neem, blick, zav, which were dissimilar to the known words used, such as dog, cat or milk. While presenting phonemically similar words, e.g. bog, gat or nilk, the N400 amplitude was smaller compared to mismatching known words for the 20-month-old children. Children at 24 months showed the same N400 component with semantically related and unrelated word pairs (related word pair: dog-horse, unrelated word pair: car-apple) just as is observed in adults (Torkildsen et al., 2007).

Neither the studies using preferential looking tasks to capture reactions to mispronunciations nor the studies using EEG to unveil processes of lexical retrieval involved productive phonological capacities of the children. While the children’s production capacities are in some sense predictable in terms of typically occurring mispronunciations during language development (cf. Grunwell, 1982; for Swedish Nettelbladt, 1983; see also chapter 2), it is unclear whether and how those mispronunciations might influence the child’s perceptual processing. Phonological processes or phonological deviants from adult word forms occur in particular when the lexicon emerges during the second year of life and can usually be observed in Swedish children up until the time when they reorganize their phonological system around the age of four years (Nettelbladt, 2007b). Evidence for a relationship between phonological acquisition and the lexicon has been found in several different studies (for a review, see Stoel-Gammon, 2011). For instance, the structure of the first words emerges from the vocal patterns practiced during babbling (Stoel-Gammon & Cooper, 1984; Vihman et al., 1985). However, this relationship has not been reflected in much of the existing work on perceptual phonological knowledge. On the one hand, perceptual phonological knowledge was mainly investigated in preferential looking tasks; on the other hand, perceptive lexical knowledge was investigated with ERPs. A combination of those two methods with stimuli based on typically occurring developmental mispronunciations might shed light on the influence of phonology on the lexicon and vice versa, as well as the relationship between speech production and perception.
The object of the following three experiments was to examine the relationship between perceptual responses using EEG and eye-tracking and productive knowledge measured by recordings of the words tested. The two methods have been used independently to investigate perceptual phonological knowledge of word forms and lexical retrieval in young infants and toddlers (e.g. Friedrich & Friederici, 2004; White & Morgan, 2008). The combination of the two methods will give us a deeper insight into the relationship between the acquisition of the lexicon and phonology. In order to tackle the methodological challenge in the combination of EEG and eye-tracking and additionally to evaluate the stimulus material, three different studies have been conducted. First, an eye-tracking experiment with adults was carried out to evaluate the stimuli (section 4.2). Second, a pilot experiment with 17-month-old children and the combination of EEG and eye-tracking was carried out (section 4.3.1, p. 59); and finally a third experiment with 24-month-old children was conducted (section 4.3.2, p. 64). In all three experiments the auditory stimuli consisted of word form changes according to developmentally typical as well as atypical phonological deviants.

4.2. Pilot eye-tracking experiment with adults

The purpose of this pilot study with adults was to evaluate the stimuli to be used in later studies with infants on word recognition. Earlier studies on the detection of mispronunciations in adults showed that word-initial stops play an important role in speech processing (Cole, Jakimik, & Cooper, 1978). Mispronunciations in the form of devoicing affected word recognition to a larger extent when a stop was devoiced than if it affected a fricative. Another study examined the detection of mispronunciations in fluent speech by preschool children and adults (Bernthal, Greenlee, Eblen, & Marking, 1987). The manipulations consisted of substitutions that could be observed in early child language such as a stop substituting for a fricative and substitutions that are rarely observed in early child language like a fricative substituting for a stop. The participants listened twice to the story including these mispronunciations. During the second telling, the story was presented sentence by sentence, and the participants were asked to indicate whether they detected a mispronunciation by writing down the correct word form (children were asked to raise their hand and say the correct pronunciation of the word). The results show that adults were in general better in the detection of mispronounced words than children. The detection of manipulations that are commonly observable in child language was for both groups in general easier than those manipulations which are
rarely observed. However, this early study on the detection of mispronunciations does not take into account how word recognition measured by newer techniques might be affected by the manipulations. The use of a newer technique like the measurement of looking behavior provides a highly sensitive measure of lexical activation to potential referents (Allopenna, Magnuson, & Tanenhaus, 1998). Swingley (2009) tested in a visual fixation task the knowledge of phonological forms of familiar words by different manipulations of those words in children and adults. The manipulations included mispronunciations of onset or coda consonants, for instance, boat-poot (onset mispronunciation) or boad (coda mispronunciation). In every trial two pictures with familiar objects were presented, and one of the objects was labeled by playing the sound. Depending on the condition, the object labeling was either mispronounced in the coda or onset consonant or correctly pronounced. The participants gaze was recorded and coded off-line to determine the visual preference in relation to condition. The results show that adults detect both onset and coda mispronunciations, and look to a lesser extent at the target pictures compared to correctly pronounced labels (Swingley, 2009).

In the present pilot study, the sensitivity of adults’ phonological representations to different manipulations in terms of mispronunciations usually occurring in typically developing children is investigated. This is a preparatory step towards a study of infants investigating sensitivity to mispronunciations which are related to typical production patterns in first language acquisition. The experimental setup was similar to studies with infants (e.g. White & Morgan, 2008): The participants looked at two pictures at the same time, one picture with a familiar object and one picture with a novel object, while listening to an auditory stimulus, which named the familiar object with its correct pronunciation or one of two different kinds of mispronunciation. Another condition consisted of an auditory stimulus in form of a novel word paired with two pictures: one of the two pictures depicting a novel object, and the other a familiar object, which was also included in the conditions with correct and mispronounced word forms. We expected a longer look at the familiar object when the word is pronounced correctly and a decrease in looking time in relation to the degree of mispronunciation. For novel words we predicted an increased looking time towards the novel object, which should be comparable to the looking time in the Correct condition.
## 4. Study I: Three experiments on methodological explorations

### Table 4.1. Auditory stimuli and the mispronunciations in the four different conditions (pilot study).

<table>
<thead>
<tr>
<th>Target word</th>
<th>Correct</th>
<th>M1</th>
<th>M2</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>lampa</em> 'lamp’</td>
<td>[lamp:a]</td>
<td>[bamp:a]</td>
<td>[fampa]</td>
<td>[’omga]</td>
</tr>
<tr>
<td><em>boll</em> 'boll’</td>
<td>[bol]</td>
<td>[bo]</td>
<td>[sal]</td>
<td>[nuck]</td>
</tr>
<tr>
<td><em>mössa</em> ‘cap’</td>
<td>[møs:a]</td>
<td>[møtt:a]</td>
<td>[fosa]</td>
<td>[søla]</td>
</tr>
<tr>
<td><em>sko</em> ‘shoe’</td>
<td>[sku:]</td>
<td>[ku:]</td>
<td>[fu:]</td>
<td>[ru:]</td>
</tr>
<tr>
<td><em>buss</em> ‘buss’</td>
<td>[bos]</td>
<td>[bo]</td>
<td>[sos]</td>
<td>[mul]</td>
</tr>
<tr>
<td><em>napp</em> ‘pacifier’</td>
<td>[nap:]</td>
<td>[bap:]</td>
<td>[jap:]</td>
<td>[mir]</td>
</tr>
<tr>
<td><em>strumpa</em> ‘sock’</td>
<td>[strampa]</td>
<td>[fømpa]</td>
<td>[fømpa]</td>
<td>[ty:ra]</td>
</tr>
<tr>
<td><em>klocka</em> ‘watch’</td>
<td>[klɔ:kə]</td>
<td>[’okə]</td>
<td>[vɔ:kə]</td>
<td>[byɡa]</td>
</tr>
</tbody>
</table>

### 4.2.1. Methods

#### Stimuli

The familiar auditory stimuli consisted of eight concrete nouns, which were selected from Swedish data from the CLex Database (Jørgensen, Dale, Bleses, & Fenson, 2010). According to this norming data at least 40% of the children produced the relevant word at 18 months and at least 25% could comprehend the words by 14 months of age. The eight auditory stimuli for novel words consisted of nonsense words made up in accordance with Swedish phonotactics (see table 4.1). The syllable structure was identical to the familiar words (four with one syllable and four with two syllables).

Aside from the correct word forms (Correct), each of the words was manipulated in two different ways, resulting in two additional conditions: Mispronunciation 1 (*M1*) and Mispronunciation 2 (*M2*). We used information about mispronunciations from a larger project, in which 647 parents of 17-month-old children reported their children’s words and typical pronunciations (Renner, Sundberg, & Schwarz, 2011). The mispronunciations most frequently reported for the selected words were used in the *M1* condition. The changes in this condition included initial, medial as well as word-final position. The words in *M2* consisted of initial position changes to fricatives, which are atypical mispronunciations in Swedish first language acquisition presupposing that the word forms in condition *M1* are closer to the correct pronunciation than the word forms included in condition *M2*.

---

14SPRINT, conducted by the Department of Linguistic and Department of Special Education at Stockholm University, 2009-2012, and funded by the Swedish Research Council: VR2008-5094.
4.2. Pilot eye-tracking experiment with adults

![Image](image_url)

*Figure 4.1.* Example of a visual stimuli pair during one trial. On the left hand side is the familiar object depicted (‘sock’); on the right hand side the novel object.

The words in the conditions were spoken by a female native speaker of Swedish in three natural variants in a child-friendly manner and recorded in an anechoic chamber.

The auditory stimuli were presented together with visual stimuli, a familiar and a novel object. The pictures with a novel object were abstract and did not depict any existing object (for an example, see Figure 4.1). All of the pictures were child-friendly drawings in modest colors on a gray background. The novel objects were similar in visual complexity to the pictures of the familiar words. For each item we used three different colors, resulting in three different pictures per item.

**Procedure**

The experiment was conducted in a sound attenuated and softly lit room with subjects seated in front of a Tobii T120 eye-tracking monitor at a distance of approximately 60 cm. After a successful 9-point calibration, the experiment started. The experiment consisted of 96 trials, the sum of labeling each item three times as one of the three natural variants: 24 trials involved correct labeling of one object, 24 trials involved an M1 mispronunciation, 24 trials involved an M2 mispronunciation and 24 trials involved a novel word. Each picture pair was introduced 2.5 s before the auditory stimulus was presented in order to obtain a baseline for looking behavior. The auditory stimuli were presented via a pair of external computer speakers, and the sound level of the stimuli was approximately 60 to 80 dB SPL. After the auditory stimulus presentation, the pictures remained on the screen for another two

---

15The data for this study was collected by SLP students from Karolinska Institutet, Stockholm, Sweden, during spring term 2012.

55
4. Study I: Three experiments on methodological explorations

seconds. The participants were advised to sit comfortably and as still as possible and were asked to listen attentively to the audio and look at the pictures shown on the screen. The experiment was run using E-prime 2.0, and gaze was recorded and analyzed with Tobii Studio 3.0.2.

Participants

Seventeen adult subjects participated in the experiment (nine male, mean age = 24.5 years; range 19-40 years). They were all native speakers of Swedish who had lived in Sweden all their lives; four of the participants had a second mother tongue (Cantonese, Greek or Spanish). All subjects had normal or corrected-to-normal vision and normal hearing. None of them reported any language difficulties or dyslexia.

4.2.2. Results

The analysis is based on the data of 16 subjects, as one of the participants was excluded from the analysis due to insufficient eye-tracking data. Two areas of interest (AOIs) were defined, each involving 15% of the total screen area, covering the area in which the pictures were located. Looking times were summarized for each trial and condition per participant, resulting in an accumulated sum of looking times for each subject in each condition within each AOI. A paired samples t-test was done on the baseline preceding each auditory stimulus presentation in order to check for any potential bias towards one of the objects. No significant differences ($t(15) = 1.76, p = 0.099$) between the novel and familiar object were found during the baseline period. After presentation of the speech stimulus, a significant difference in looking times between the familiar objects and unfamiliar objects was found ($t(15) = 3.87, p = 0.001$), showing an increased amount of looking time for the familiar objects in total. Following this, the gaze data was divided into groups for each of the conditions. Within these conditions, the total looking time differences (TLD) between target and non-target after auditory stimulus onset were calculated. Subsequently, an ANOVA test was performed, showing significant differences ($F(3, 60) = 13.7, p < 0.001$) between the conditions. Post-hoc comparisons using the Bonferroni correction indicated a decrease in looking time towards the familiar object according to condition (Figure 4.2). In particular, the mean score of the looking time difference in the Correct condition ($\bar{x} = 8.16$, $s = 5.10$) differed significantly from the mean looking time differences in M2 ($\bar{x} = 2.75$, $s = 5.26$, $p = 0.016$) and Novel ($\bar{x} = -2.23$, $s = 4.48$, $p < 0.001$).
4.2. Pilot eye-tracking experiment with adults

Figure 4.2. Looking time differences between familiar and novel objects during test phase. The error bars depict the 95% CI of the mean. Y-axis: Looking time differences (target – nontarget) in seconds, X-axis: Conditions (Correct = correct words, M1 = words of mispronunciations type 1, M2 = words of mispronunciations type 2, Novel = Novel words)

The mean scores of the looking time difference in the M1 condition ($\bar{x} = 6.01$, $s = 4.63$) as well as the M2 condition differed significantly from those of in the Novel condition ($M1 - Novel: p < 0.001, M2 - Novel: p = 0.033$).

Taken together, the participants looked more towards the familiar object when the target word was correctly pronounced compared to the mispronounced conditions. The difference in looking time between the familiar and novel object decreased as the similarity to the stimulus differed more phonologically. Additionally, the participants did not look at the novel picture during the novel condition to the same extent as they looked at the familiar object during the correct condition, which indicates that the participants not only look towards the unfamiliar object in the Novel condition.

4.2.3. Discussion

The results showed that adults looked at the familiar object less when they heard a mispronunciation of the word than when the word was pronounced correctly. The looking time to the familiar object decreased in relation to the degree of mispronunciations connected to the conditions. Earlier studies with adults showed that listeners find mispronunciations occurring in typically developing children easier
4. Study I: Three experiments on methodological explorations

...to detect than unusual sound shifts (Bernthal et al., 1987) or a lesser looking time towards the familiar object when the words were mispronounced (Swingley, 2009), which could not be confirmed in the present study as there was no significant difference between the looking times in Correct and M1 conditions. However, the stimuli material between the latter and the present study differed significantly: In the present study mispronunciations were in line with expected typical mispronunciations in child language; in Swingley’s (2009) study the mispronunciations were systematically varied in terms of changes of places of articulation as well as voicing. The significant difference between looking times in the Correct and M2 conditions could be observed, which indicates that the word representation is sensitive to some mispronunciations. M1 mispronunciations might be more acceptable for a target word than M2 mispronunciations, which could be explained by the activation of the target word despite the mispronunciation (Bailey & Plunkett, 2002; Swingley & Aslin, 2000).

For novel words there was less looking towards the unfamiliar object than expected. The prediction that the looking behavior to correct labels and novel labels should be comparable could not be fully confirmed; instead, it was found that the participants look towards the familiar object in the novel word condition as well.

In the current study the mispronunciations were adapted to typical language acquisition. Adults seem to process mispronunciations differently, depending on the phonological changes. In general, it was observed that they adjust their looking behavior to the degree of mispronunciation. This indicates that M1 mispronunciations are closer to the correct pronunciations than M2 mispronunciations. Further studies are needed to evaluate if this also holds for infants or if the infant maps the mispronounced word to a greater extent onto the familiar object if it corresponds to her own production pattern. The next studies with 17-month-old and 24-month-old children investigate this question.

4.3. Methodological explorations of EEG and eye-tracking on children at 17 and 24 months

The following two studies with 17- and 24-month-old children aim to investigate whether the processing of a child’s own mispronounced word is different than the processing of the target form of the same word. Data on eye movements as well as...
4.3. Methodological explorations of EEG and eye-tracking

on brain responses were collected. The idea behind the combination of EEG and eye-tracking draws from, on the one hand, studies on phonological representation using preferential looking tasks (Bailey & Plunkett, 2002; Swingley & Aslin, 2000, 2002; White & Morgan, 2008) and, on the other hand, studies on lexical retrieval using EEG (Friedrich & Friederici, 2005a, 2005b, 2010; Mills et al., 2004). Using these two methods, we hoped to bring together insights on the acquisition of phonological representations and on lexical retrieval aiming for a better understanding of the relationship between lexicon and phonology. Specifically, we aimed at investigating if and when mispronunciations are ignored both in terms of lexical retrieval as well as word recognition in connection with the child’s production of the word in question. Whereas the looking time window usually consists of 2 seconds, an N400 response would be expected 400ms after word onset. Thus the neurophysiological response should be observable earlier than the behavioral reaction measured in terms of looking time. The combination of these two methods should thus give us insights into phonological and lexical processing of mispronounced words and its temporal course. The first experiment on 17-month-old children pilots the combination of EEG and eye-tracking as well as the experimental design. The second experiment with 2-year-old children includes major improvements in the experimental design of the study with 17-month-old children.

4.3.1. Experiment with 17-month-old children

The first experiment with 17-month-olds is based on earlier studies from the literature and aims to pilot technological possibilities in the combination of EEG and eye-tracking in relation to production abilities using the same stimuli material as in the previous study with adults. Our predictions consisted of a longer looking time towards the familiar object during the correct condition and a decreasing looking time in the other conditions, \( M1, M2 \) and Novel, with the shortest looking towards the familiar object in the Novel condition based on earlier studies (e.g. White & Morgan, 2008) and the results from the adult study. In EEG we expected a similar pattern in the form of the largest negativity in the novel condition and decreasing values in the other three conditions.

4.3.1.1. Methods

Participants

The participants were recruited through the address register of the Swedish tax agency (Skatteverket). Families in the metropolitan area of Stockholm with 17-
4. Study I: Three experiments on methodological explorations

Month-old children were sent a letter with information about this study and a registration form, which they were asked to send back to the department if they were interested in participating. The registration form included questions about language background and whether there were any complications (e.g., regarding their child’s hearing, or premature birth). After registration, an appointment was made for a visit to the lab. A total of 28 children visited the lab, two of those refused to wear the EEG cap, resulting in 26 children available for further EEG analysis (mean age = 17.6, 12 boys, 14 girls). Eye-tracking data could only be collected if the combination of the two methods worked out technically, which was the case in 13 children (mean=17.5, 4 boys, 9 girls).

Stimuli

The same auditory and visual stimuli in the different conditions (Correct, M1, M2, Novel) were used as in the study with adults, see section 4.2.1 on page 54. Condition Correct refers to the correct pronunciation of the target word; M1 refers to mispronunciations that typically occur during child development; M2 consists of mispronunciations that are atypical during child development and Novel consists of nonsense words made up in accordance with Swedish phonotactics.

Procedure

Prior to the experiment, the parents were asked to fill in a questionnaire concerning their child’s word understanding and production of the eight target words. Additionally, if they reported that their child said the word, they were asked to write down what the production sounds like. After that, the child and her parents were guided to the experiment room. The parents were instructed to look in a big bag filled with the objects to be used in the experiment, together with the child, in order to elicit the child’s production of each of the target words; if necessary the experimenter helped the child to get the object out of the bag and tried to elicit the word form from the child. The speech production during the elicitation task was audio recorded. After the elicitation, the parents were asked to sit in front of a Tobii eye-tracker T120 with the child on their lap at 60 cm distance from the screen. The EEG-cap was shown to the child, and the child was told what was going to happen. After cap application and impedance measurement of the electrodes, the experiment started after a successful 5-point calibration of the eyes. The parents wore headphones and listened to music during the whole experiment. The experiment consisted of alternating different types of trials: EEG trials with one picture on the screen, and eye-tracking trials with two pictures on the screen. These two type
4.3. Methodological explorations of EEG and eye-tracking

of trials were conducted for a specific analysis of different measurements (looking time versus neurophysiological response approximately 400 ms after word onset) in the two methods.

In an EEG trial, a picture of a familiar object was presented in the middle of the screen. After one second the auditory stimulus in one of the conditions was played. The auditory stimulus could either be a match in form to the correct pronunciation (Correct) or a mismatch in terms of one of the three other conditions M1, M2 or Novel. The picture stayed on the screen for another three seconds after word onset (following Friedrich & Friederici, 2004).

In an eye-tracking trial, two pictures were presented, one on the left and the other on the right side of the screen. One picture depicted a familiar object, the other one an abstract object. After 2.5 seconds the auditory stimulus in one of the four conditions was played. The pictures stayed on the screen for another three seconds. The order of EEG and eye-tracking trials was randomized. The whole experiment consisted of 96 trials (48 EEG trials and 48 eye-tracking trials). After every second trial a short reward film (3–11 seconds long) with child cartoons was shown. The duration of the whole experiment was 20 minutes, but the procedure was interrupted if the child started crying or became restless.

4.3.1.2. Results

Out of the 28 children, 9 understood all of the 8 words tested in the experiment according to their parents (Figure 4.3, for a detailed overview over the individual performances, see Table A.1 in Appendix A); one child understood only the word boll of the eight words in question. The word boll was understood by most of the children (27), the word klocka by the least of the children (17). The collected information from the parents on speech production showed that none of the children could say all of the words yet, one child could say six of the words and nine children did not say any of the words in question yet. During the elicitation task, half of the participant did not name any of the objects; in the other half of the group five named one of the objects. The information from the parents and from the elicitation task did not overlap in many cases: Either the child said fewer words than the parents reported, more words than the parents reported, or the pronunciation differed between the parental report and the transcribed production from the elicitation task.

The analysis from the perception experiment is divided into two parts according to the method. The EEG data processing included typical steps in terms of 1-40 Hz bandpass filtering, segmentation, bad channel replacement and artifact detec-
4. Study I: Three experiments on methodological explorations

Figure 4.3. Number of children related to their understanding and production of the eight target words. The size of the circle indicates the number of children.

Segments were marked as bad if they contained more than 10 bad channels. The threshold for a bad channel was +/-200 μV. Subjects with less than 30% good trials in each condition were excluded, leading to 15 participants being included in the final analysis. The data from electrodes in the parietal-occipital brain region was averaged for each participant in the different conditions in the time window of 400 to 700ms after auditory stimulus onset. An analysis of variance showed that the effect of CONDITION on mean amplitude was not significant ($F(3, 56) = .846$, n.s.).

The eye-tracking analysis is based on the looking time towards each of the objects for each subject. The proportional looking time towards the familiar object was computed for the baseline and the test phase (cf. White & Morgan, 2008). After that, for each trial the difference between the proportional looking time towards the familiar object during the test phase and baseline was calculated. Thus, test-minus-baseline difference scores below zero indicate a longer looking time towards the unfamiliar object during the test phase, whereas positive test-minus-baseline difference scores indicate a longer looking time towards the familiar object during the test phase. The individual results of the children showed a great deal of variation (Figure 4.4). Zooming in to the perceptual reactions, three children (RV, SA...
4.3. Methodological explorations of EEG and eye-tracking

![Graph showing proportional looking times for each subject. The y-axis represents the difference between the proportion looking at the familiar object in the test phase and the proportion looking at the familiar object in the baseline phase. (Correct = correct words, M1 = words of mispronunciation type 1, M2 = words of mispronunciation type 2, Novel = Novel words)](image)

*Figure 4.4. Proportional looking times for each subject. Participants and conditions are on the x-axis. The y-axis represents the difference between the proportion looking at the familiar object in the test phase and the proportion looking at the familiar object in the baseline phase. (Correct = correct words, M1 = words of mispronunciation type 1, M2 = words of mispronunciation type 2, Novel = Novel words)*

and IM) showed the predicted pattern of a longer looking time towards the familiar object during the correct condition and a shorter looking time in the other conditions. Looking at the production of these children, participant IM said three of the tested target words, only one close to the predicted mispronunciation ([gi:] sko), which was tested in condition M1. Additionally, the parents reported that he understood seven out of the eight tested target words. RV understood all of the tested target words according to her parents, but did not say them yet, which was evident in the elicitation task with no target words being produced. SA understood five of the tested target words and neither parents reported any pronunciation of the target words nor did any of the target words appear in the elicitation task. Another child (QT) showed an unpredicted pattern with a longer looking time in all conditions except Correct. She produced one target word in the elicitation task, which was different from the pronunciation that the parents reported for this word.
4. Study I: Three experiments on methodological explorations

No clear tendencies in the relationship between the perceptual reaction and the expected child’s production in M1 could be detected in any of the measurements. In general, the results are characterized by variability in both perception and production measurements. This could be explained by the following factors:

1. The experiment was too long, since none of the children completed the whole experiment. Interruption of the experiment resulted in fewer trials available for analysis than expected.

2. Some children did not understand the tested words, according to the parents, nor was the speech production of the words elicitable, which indicates that neither the selected target words nor the data collection method was appropriate to this age group.

3. The drawings of the objects used as visual stimuli included different variances (i.e. three different color variants). The feedback that we obtained from the parents after the experiment revealed that some of the drawings were hard to recognize. For instance, one parent reported that the picture of a pacifier reminded the child of a candy car.\textsuperscript{16}

4. It turned out that the parental report on how the words were produced was not as helpful as expected, since it did not always overlap with the child’s productions from the elicitation task.

In the following study, these reasons were taken into account to make major improvements to the experiment. In order to reduce the length of the experiment in general, the reward movies were shortened. The number of trials was adjusted to obtain fewer eye-tracking trials and more EEG trials. Older children, at 24 months, were tested to increase the probability of eliciting production data. Additionally, only four target words were retained. The visual stimuli consisted of photographs instead of drawings. An estimate of vocabulary size (Swedish CDI, Berglund & Eriksson, 2000) was used instead of parental reports on the target words and how the words were produced according to the parents. The following section describes this study with 24-month-old children in more detail.

4.3.2. Experiment with 24-month-old children

This study is a consequence of the preceding one in terms of major improvements of the experimental setup as well as the age of the participant group. The meth-

\textsuperscript{16}In Sweden, candy in the shape of a car is quite common, and the manufacturer advertises the candy as ‘the best-selling car in Sweden’.
4.3. Methodological explorations of EEG and eye-tracking

ods, EEG and eye-tracking, as well as the four conditions as in the two previous experiments were retained (Correct, M1, M2 and Novel). The aim of the study was to investigate the child’s perceptual reactions to their own production reflected in condition M1. The hypotheses were the same as the previous study with younger children: We predicted a longer looking time toward the familiar object when the word was pronounced correctly and a decreasing looking time in the other conditions M1, M2 and Novel. For novel words we predicted an increased looking time towards the novel object. However, if the child produced the word in the way that we predicted and tested in M1, the looking time toward the familiar object might increase in this condition.

In EEG we expected the largest negativity in the novel condition and decreasing values in the other three conditions. However, if the child produced the word in the way that we predicted and tested in M1, the mean amplitude in M1 might be similar to the one in Correct.

4.3.2.1. Methods

Stimuli

The stimuli consisted of both auditory and visual material as in the previous studies. However, of the eight different target words evaluated in the previous studies (section 4.2 and 4.3.1), four were retained for use in the current experiment (lamp ’lamp’, ball ’ball’, cap ’cap’, shoe ’shoe’; see table 4.2).

Table 4.2. Auditory stimuli and the mispronunciations in the four different conditions.

<table>
<thead>
<tr>
<th>Target word</th>
<th>Correct</th>
<th>M1</th>
<th>M2</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>lamp ’lamp’</td>
<td>[lampa]</td>
<td>[bampa]</td>
<td>[fampa]</td>
<td>[nuga]</td>
</tr>
<tr>
<td>ball ’ball’</td>
<td>[bol]</td>
<td>[bo]</td>
<td>[sol]</td>
<td>[nuk]</td>
</tr>
<tr>
<td>cap ’cap’</td>
<td>[mosa]</td>
<td>[mot]</td>
<td>[fosa]</td>
<td>[sola]</td>
</tr>
<tr>
<td>shoe ’shoe’</td>
<td>[sku]</td>
<td>[ku]</td>
<td>[fu]</td>
<td>[ru]</td>
</tr>
</tbody>
</table>

These four words were understood by at least 85% of 16-month-old children according to the CLex norms (Jørgensen et al., 2010). Thus, most of the 24-month-old children could be expected to be familiar with these words. The word manipulations of these four target words in the conditions M1 and M2 were the same as in the previous studies: M1 consisted of changes that reflect common errors in typical language acquisition, such as consonant cluster reduction (e.g. [ku:] for [sku:]).
4. Study I: Three experiments on methodological explorations

Figure 4.5. Example of a visual stimuli pair. On the left hand side the familiar object (lamp) is depicted; on the right hand side the unfamiliar object.

sko ‘shoe’). According to typical phonological changes in children’s production, the mispronunciations in M1 involved initial, medial and word-final-positions. M2 consisted of manipulations of the correct word form that are rare or non-occurrent in typical language acquisition (e.g. fricativization) and involved only initial-word position (e.g. [fampa] for lampa ‘lamp’). Similar to the previous studies, four different, phonotactically legal, novel Swedish words (Novel) were included that matched the general syllable structure of the correct word forms. Each word form was produced by a female native Swedish speaker in a child-directed manner and was recorded in an anechoic chamber in the phonetics lab at Stockholm University.

The visual stimuli consisted of photos of familiar as well as abstract objects (Figure 4.5). The abstract objects had an unusual shape, e.g. a design version of an office punch.

Procedure

The procedure was in general the same as in the pilot study with 17-month-old children (cf. section 4.3.1, p. 59). Still, a few adjustments were made: The parents were asked to fill in the Swedish CDI (Berglund & Eriksson, 2000) to estimate the productive vocabulary size prior to the experiment. As in the study with 17-month-olds, prior to the perceptual experiment a speech elicitation task was conducted: The child unpacked together with her parent toys from the bag, and the child was encouraged to talk about the objects. The perceptual experiment consisted of 36 eye-tracking trials and 96 EEG trials. The order of EEG and eye-tracking trials was randomized. After every second trial, a short reward movie lasting four seconds was presented. The movies consisted of a chicken crossing the screen accompanied by instrumental music.
4.3. Methodological explorations of EEG and eye-tracking

Eye-tracking
Each trial in the eye-tracking condition consisted of a baseline phase, a fixation phase and a test phase. In the baseline phase two pictures – one familiar and one abstract object – were presented on the screen for 2.5 seconds. During the fixation phase the two pictures from the baseline remained on the screen, but a smiley face occurred between the two pictures after 2.5 seconds. The smiley face remained on the screen between the pictures for a maximum of four seconds; as soon as the child had fixated on it for 500ms it disappeared and the test phase – playing the auditory stimulus – began. The test phase with the two pictures lasted for three seconds. If the child did not fixate on the smiley face, the experiment continued after five seconds with the presentation of the auditory stimulus. The 36 eye-tracking trials consisted of twelve target words presented in the Correct condition and eight trials each of the remaining conditions (M1, M2 and the Novel condition).

EEG
In the EEG trials, one familiar object was presented in the middle of the screen. The auditory stimulus was presented, when the child fixated on the picture of the familiar object for one second. After word onset, the picture remained on the screen for another two seconds. As in the eye-tracking trials, if the child did not fixate on the picture within five seconds, the auditory stimulus was presented and the experiment continued.

The total duration of the experiment varied by participant, because the experiment only continued when the child fixated on the one picture in EEG trials or the smiley face in eye-tracking trials, which occurred in the middle of the experiment, prior to the test phase. Given that the children looked at the screen most of the time, the overall duration was about 15 minutes.

Participants
Recruitment followed the same procedure as in the study with 17-month-old children, except for the age of the children (24-month-olds). Of the 35 children visiting the laboratory, some had to be excluded from either the EEG or the eye-tracking analysis, and some from both. Of the 35 children, 11 were excluded from both, EEG and eye-tracking, due to technical issues (6) or being restless during the experiment (5). Of the remaining 24 children, 9 refused to wear the EEG cap and were thus excluded from the EEG analysis, resulting in 15 children available for EEG preprocessing (mean age = 24;11, range = 23;26–25;19, eight boys, seven girls) (table 4.3).
4. Study I: Three experiments on methodological explorations

Table 4.3. Overview of the demographic data of the participants.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Age</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>M</td>
<td>24:11</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>AT</td>
<td>F</td>
<td>24:7</td>
<td>ET</td>
</tr>
<tr>
<td>ED</td>
<td>F</td>
<td>24:15</td>
<td>EEG</td>
</tr>
<tr>
<td>EF</td>
<td>F</td>
<td>24:3</td>
<td>ET</td>
</tr>
<tr>
<td>FJ</td>
<td>F</td>
<td>24:3</td>
<td>ET</td>
</tr>
<tr>
<td>FK</td>
<td>F</td>
<td>24:8</td>
<td>EEG</td>
</tr>
<tr>
<td>GE</td>
<td>M</td>
<td>25:17</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>ID</td>
<td>F</td>
<td>24:1</td>
<td>EEG</td>
</tr>
<tr>
<td>IG</td>
<td>M</td>
<td>24:6</td>
<td>EEG</td>
</tr>
<tr>
<td>IQ</td>
<td>M</td>
<td>24:4</td>
<td>ET</td>
</tr>
<tr>
<td>IV</td>
<td>M</td>
<td>24:23</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>LD</td>
<td>M</td>
<td>24:1</td>
<td>ET</td>
</tr>
<tr>
<td>LF</td>
<td>M</td>
<td>24:6</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>LV</td>
<td>M</td>
<td>24:14</td>
<td>ET</td>
</tr>
<tr>
<td>ON</td>
<td>M</td>
<td>23:30</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>PO</td>
<td>F</td>
<td>24:1</td>
<td>ET</td>
</tr>
<tr>
<td>SL</td>
<td>F</td>
<td>25:19</td>
<td>EEG</td>
</tr>
<tr>
<td>SN</td>
<td>M</td>
<td>24:2</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>SR</td>
<td>F</td>
<td>23:29</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>TA</td>
<td>F</td>
<td>23:26</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>TH</td>
<td>M</td>
<td>25:10</td>
<td>ET, EEG</td>
</tr>
<tr>
<td>VG</td>
<td>F</td>
<td>24:11</td>
<td>ET</td>
</tr>
<tr>
<td>WJ</td>
<td>M</td>
<td>24:5</td>
<td>ET</td>
</tr>
<tr>
<td>WO</td>
<td>F</td>
<td>24:22</td>
<td>EEG</td>
</tr>
</tbody>
</table>

If the child refused to wear the EEG cap, eye-tracking data was still collected. Eight children were excluded from the eye-tracking analysis due to difficulties in calibration (5), technical issues limited to the eye-tracker (2) or below 20% of the samples with non-missing data (1), indicating that the eye-tracker was not able to detect the child’s gaze in 80% of the trials. In the remaining group of 16 children included in the eye-tracking analysis, were 9 boys and 7 girls (mean age = 737 days, range = 726–771). With respect to language background, seven had one parent with a mother tongue other than Swedish; five included in the analysis of EEG data (German [1], French [2] and Spanish [2] and five included in the eye-tracking analysis (Amharic [1], French [1], German [1], Spanish [1], Greek [1]). As these children had been in a Swedish kindergarten or nursery care for at least six months at the time of the experiment, Swedish was still considered the dominant language in their environment.
4.3. Methodological explorations of EEG and eye-tracking

4.3.2. Results

From the parental report on vocabulary size (SECDI, Berglund & Eriksson, 2000), information was extracted as to whether the parents reported the production of the target words for the child (table 4.4). Three parents reported that their child did not produce all of the target words yet (ED, EF and IG). Some parents did not hand in the SECDI (5), thus resulting in missing data.

For each child the recorded productions were phonetically transcribed and compared to the predicted mispronunciations in the auditory stimuli condition M1. Not all of the target words could be elicited for every child. Still, in four children the elicitation of all of the target words was successful (AT, TH, WJ, GE). Three children produced none of the target words (ED, EF and IV). Most of those children who produced the word *boll* did so correctly, whereas the realizations of the word *sko* ‘shoe’ for the most part showed the predicted initial cluster reduction. Some children (for instance AT, PO or TA, SL) produced the target words several times in different ways, showing variable production.

<table>
<thead>
<tr>
<th>Subject</th>
<th>SECDI (words prod.)</th>
<th>SECDI (no. prod. target words)</th>
<th>Production target word</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>-</td>
<td>-</td>
<td><em>boll</em> [bol]; <em>lampa</em> [lampa]; <em>sko</em> [gono]</td>
</tr>
<tr>
<td>AT</td>
<td>412</td>
<td>4</td>
<td><em>boll</em> [bol]; <em>lampa</em> [lampa]; <em>sko</em> [guen]; [gu]; <em>mössa</em> [moxa]</td>
</tr>
<tr>
<td>ED</td>
<td>91</td>
<td>1 (boll)</td>
<td>-</td>
</tr>
<tr>
<td>EF</td>
<td>89</td>
<td>1 (lampa)</td>
<td>-</td>
</tr>
<tr>
<td>FJ</td>
<td>540</td>
<td>4</td>
<td><em>boll</em> [bol]; <em>lampa</em> [lampa]</td>
</tr>
<tr>
<td>FK</td>
<td>444</td>
<td>4</td>
<td><em>mössa</em> [mösa]</td>
</tr>
<tr>
<td>GE</td>
<td>402</td>
<td>4</td>
<td><em>lampa</em> [lampa]; <em>sko</em> [dæ]; <em>boll</em> [bol]; <em>mössa</em> [moxa]</td>
</tr>
<tr>
<td>ID</td>
<td>143</td>
<td>3 (boll, mössa sko)</td>
<td><em>lampa</em> [lampa]; <em>jampa</em> [lampa] [lampa]</td>
</tr>
<tr>
<td>IG</td>
<td>13</td>
<td>0</td>
<td><em>boll</em> [bol]; <em>sko</em> [gø]</td>
</tr>
<tr>
<td>IQ</td>
<td>184</td>
<td>4</td>
<td><em>boll</em> [bol]; <em>lampa</em> [abac]; <em>mössan</em> [can]</td>
</tr>
<tr>
<td>IV</td>
<td>273</td>
<td>4</td>
<td>-</td>
</tr>
</tbody>
</table>

*continued on next page*
4. Study I: Three experiments on methodological explorations

Table 4.4 – continued from previous page

<table>
<thead>
<tr>
<th>Subject</th>
<th>SECDI (words prod.)</th>
<th>SECDI (no. prod. target words)</th>
<th>Production target word</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD</td>
<td>-</td>
<td>-</td>
<td>missa [mossa]</td>
</tr>
<tr>
<td>LF</td>
<td>296</td>
<td>4</td>
<td>lampa [lampa]; sko [guc]; missa [moican]</td>
</tr>
<tr>
<td>LV</td>
<td>401</td>
<td>4</td>
<td>missa [bosaa]; sko [guc]</td>
</tr>
<tr>
<td>ON</td>
<td>116</td>
<td>4</td>
<td>lampa [blampan]</td>
</tr>
<tr>
<td>PO</td>
<td>390</td>
<td>4</td>
<td>missa [mossa] [menca]; skorna [guna]</td>
</tr>
<tr>
<td>SL</td>
<td>-</td>
<td>-</td>
<td>lampa [lampa]; sko [skuc]; bol [bol]; missa [mossa]</td>
</tr>
<tr>
<td>SN</td>
<td>300</td>
<td>4</td>
<td>bol [bolc]</td>
</tr>
<tr>
<td>SR</td>
<td>136</td>
<td>4</td>
<td>lampa [wampa]; sko [gona] [sko] [gunc]; missa [boca] [boca] [moena]</td>
</tr>
<tr>
<td>TA</td>
<td>325</td>
<td>4</td>
<td>bol [bol]; sko [unc] [gunc]; missa [moena] [moena]</td>
</tr>
<tr>
<td>TH</td>
<td>-</td>
<td>-</td>
<td>bol [bol]; lampa [lampa]; sko [kot] [skuc]; missa [mossa]</td>
</tr>
<tr>
<td>VG</td>
<td>-</td>
<td>-</td>
<td>lampa [labla]; missa [moena]; sko [gunc]</td>
</tr>
<tr>
<td>WJ</td>
<td>590</td>
<td>4</td>
<td>lampa [lampan] [lampa] [lama]; sko [skoca] [guc]; bol [bol]; missa [mossa]</td>
</tr>
<tr>
<td>WO</td>
<td>344</td>
<td>4</td>
<td>lampa [lampan]; sko [kucn]; missa [moica]</td>
</tr>
</tbody>
</table>

The analysis from the perception experiment is divided into two parts according to method. First, the EEG analysis\(^\text{17}\) was conducted in terms of filtering, artifact detection, bad channel rejection and baseline correction, as in the pilot study with 17-month-olds (section 4.3.1, p. 59). Segments were marked as bad if they contained more than 10 bad channels. The threshold for a bad channel was +/- 200 $\mu$V. Subjects with less than 30% good trials in each condition were excluded, which means that those children who moved a great deal were excluded. After these preparatory steps, eight participants were included in the statistical analy-

\(^{17}\) The statistical analysis of the EEG data was carried out in collaboration with Petter Kallioinen.
4.3. Methodological explorations of EEG and eye-tracking

Figure 4.6. Two-step principal component analysis. The first temporal component has as its maximum 388ms after word onset. This component is decomposed into spatial components, the two first shown here with amplitude maximum at channel 9c and 2c, respectively.

ysis, which showed no significant results in the time window 300-700ms after word onset in the four different conditions ($F(3, 407) = .382$, n.s.).

Subsequently, in order to explore the data and lack of significant results all of the participants (even those previously excluded) were included in a data-driven summary of a two-step principal component analysis (PCA) (see Figure 4.6). The first temporal component and its first spatial component reasonably match the N400 component criteria: central or parietal negativity maximum at 400ms and sensitivity to semantic manipulations. Visual inspection of the component patterns showed more posterior negative responses to semantic anomalies. The first spatial component showed a greater negativity in Novel compared to the other conditions. The second spatial component displayed greater amplitudes in the conditions Novel and M2 compared to M1 and Correct. In sum, some aspects of the N400 are unique for novel words, in others M1 mispronunciations are processed in a similar way as correct word forms, and M2 mispronunciations and Novel words are processed alike.

The analysis of the eye-tracking data is based on average on 31 eye-tracking trials (median=25) per child. The proportional looking time towards the familiar object was computed for the baseline and the test phase (cf. White & Morgan, 2008). After that, for each trial the difference between the proportional looking time towards the familiar object during the test phase and baseline was calculated. Thus, test-minus-baseline difference scores below zero indicate a longer looking time towards the unfamiliar object during the test phase, whereas positive test-minus-baseline difference scores indicate a longer looking time towards the famil-
4. Study I: Three experiments on methodological explorations

The familiar object during the test phase. To establish whether there is a general effect of mispronunciations on responses, a one-way ANOVA was conducted, showing a main-effect on CONDITION ($F(3, 310) = 3.246, p = 0.022$). A planned post-hoc comparison revealed significant differences in reactions to the condition with novel words (LSD: Correct – Novel $p = .006$, M1 – Novel $p = .008$, M2 – Novel ns.). Since the aim of the study was to investigate the child’s perceptual reactions to their own pronunciation of the tested words in relation to condition M1, the children’s productions for each of the tested word were placed in three different groups:

1. The child could produce the tested word correctly.
2. The child pronounced the tested word as predicted in condition M1.
3. The child showed a different mispronunciation from what was predicted, e.g. /pal/ for /lamp/, but still different from M2.

These production groups for each trial were related to the perceptual responses on each trial. The perceptual responses for each production group are depicted in Figure 4.7.

![Figure 4.7](Image)

**Figure 4.7.** Proportional looking times and 95% confidence intervals for each condition. Production group is represented on the x-axis. The y-axis represents the difference between the proportional looking time towards the familiar object during the baseline and the test-phase.
4.3. Methodological explorations of EEG and eye-tracking

A two-way ANOVA conducted to examine the effects of CONDITION (four levels: Correct, M1, M2 and Novel) and PRODUCTION GROUP (3 levels: Correct pronunciation, Mispronunciation like M1, Unpredicted mispronunciation) on LOOKING TIME (dependent variable) showed no significant differences (Condition: $F(3, 7.241) = 4.276, p = .050$) Production group: $F(2, 6.089) = 2.127, p = 0.199$, Interaction Condition*Production group ns.). However, visual inspection of Figure 4.7 shows different patterns for each condition and production group. When the child could produce the word correctly, looking time towards the familiar object decreased for conditions M1, M2 and Novel accordingly. A mispronunciation in the child’s production as predicted in M1 resulted in increasing looking time towards the familiar object in conditions M1 and M2, whereas, in general, other mispronunciations in the child’s own production led to shorter looking times towards the familiar object. Furthermore, the children in all production groups showed a great deal of variability in all conditions.

4.3.3. Discussion

The aim of the study was to investigate the children’s perceptual reactions to their own pronunciation reflected in the condition M1. We predicted a longer looking time toward the familiar object when the word was pronounced correctly and a decreasing looking time in the other conditions M1, M2 and Novel. For novel words we predicted an increased looking time towards the novel object.

Our results from the study with 24-month-old children showed a significant difference between looking behavior in response to Novel and Correct as well as Novel and M1, indicating that the looking times towards the familiar object differ in those conditions. This to some degree supports our predictions, which were based on White and Morgan’s (2008) study with 19-month-old children, albeit to a lesser extent than expected and not for all conditions. In particular, the data is characterized by considerable variability, which might account for the lack of significance in all conditions.

Another prediction concerned the child’s own production and how it might influence the perceptual reaction if it had the shape that we predicted and tested in M1. The grouping of the trials of the participants according to whether they mispronounced the words according to our prediction showed no significant differences between the conditions. However, children who mispronounced the words as we predicted showed, on visual inspection, longer looking time toward the familiar object in the conditions, including mispronunciations as in M1 and M2. Notably, most of the trials included in this group involved the word sko ‘shoe’ and the reduc-
4. Study I: Three experiments on methodological explorations

tion of consonant cluster in initial position. Earlier studies have shown that changes in initial position hinder word recognition (Bailey & Plunkett, 2002; Swingley & Aslin, 2002; White & Morgan, 2008). The results presented in the current study might give a tentative suggestion that this is not always the case, but depends on the child’s own production of the word, as assumed for instance by Wang and Seidl (2016).

However, the results in both studies with 17- and 24-month-old children reveal a large amount of variability in both perception and production. Additionally, the predicted mispronunciations did not appear as we expected. On the one hand, in the group of younger infants at 17 months, there were many children who neither understood nor produced all of the tested words according to their parents. On the other hand, testing older children did not solve the problem, as the variability in production remained and additionally more children of the older group refused to wear the EEG cap (nine versus two in the younger age group). In fact, what we gained in choosing older children in the understanding and production of the target words, we lost in EEG data collection. In EEG we expected the largest negativity would be in the novel condition and decreasing values in the other three conditions, but with too little or too noisy data the statistical testing of the hypothesis was unsuccessful. However, an exploration of the data indicated that our attempt to elicit an N400 response pointed to the predicted direction and gives a tentative indication that the anticipated N400 response might be found in some of the conditions. In comparison to other studies investigating mispronunciations (e.g. Mills et al., 2004), we used both visual as well as auditory stimulus material, which involves another information processing than when the target words are only auditorily presented. The number of exclusions of participants in the present study due to uninterpretable EEG is not uncommon for this age group: For instance, Carver, Meltzoff, and Dawson (2006) had to exclude 80 out of 113 18-month-old toddlers when investigating the recognition of familiar and unfamiliar objects, most of whom refused to wear the electrode cap. This indicates that in future studies a very large number of participants needs to be tested to compensate for the high exclusion rate due to uninterpretable EEG data or challenges in electrode cap application.

The combination of EEG and eye-tracking to measure perception abilities on both phonology as well as lexical retrieval was challenging, as reflected in the number of participants excluded due to technical reasons. However, future studies might consider an improvement in the experiment to take greater advantage of the simultaneous recording of behavioral and neural reactions. For instance, a study with adults observed processes of lexical retrieval in pupil size changes as well
4.3. Methodological explorations of EEG and eye-tracking

as in N400 response (Kuipers & Thierry, 2011). A recent study detected changes in pupil sizes in 30-month-old children in response to mispronunciations (Tamási et al., 2017). Furthermore, the results from an experiment with adults using a similar experimental setup as in the infant EEG studies by Friedrich and Friederici (2004, 2005a, 2005b) indicate that pupil dilation is sensitive to semantic mismatch (Renner & Włodarczak, 2016, 2017). The combination of EEG and eye-tracking and the measurements connected to the development of phonology and lexicon open the door to further investigate those domains towards a better understanding of how the lexicon and phonology are related to each other beyond the present thesis.

Looking at the children’s production, not all of the words could be elicited in either age group. One reason might be that the child did not produce the target word yet - a probable explanation for the younger age group, since not all parents reported production of the words. Another factor might be the elicitation task per se: Many children might not have felt comfortable in the lab situation and might not have produced the word for that reason. Future studies might consider another setup for speech recordings, for instance, recordings in the home environment and in a larger play situation where the child feels comfortable, as has been done in earlier production studies (e.g. Vihman et al., 1986; Vihman & McCune, 1994). Another issue concerning the speech production data addresses the variability in production which led to a mismatch between our predicted mispronunciations and the actual speech production by the individual child. Variability in production has been observed in several earlier studies (e.g. Vihman et al., 1986). Our predictions were based on parental reports and on deviance typically occurring in child development (Grunwell, 1982). However, as the results from the study with younger infants showed, the parent’s estimation of the child’s speech production did not always agree with the productions elicited. The deviance occurring in typical child development were based on parental reports and English-speaking children (Grunwell, 1982) and were assumed to be similar to Swedish-speaking children (Nettelbladt, 2007b). Thus, beyond the general assumption that mispronunciations in Swedish children are similar to those of English children, it is unclear when mispronunciations in Swedish children occur and, when they do, which form they have. In the following chapter, the production capacities of Swedish 18-month-old children are investigated.
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

This chapter examines the production capacities of Swedish toddlers at 18 months in relation to their vocabulary size and is related to research questions 3 and 4, namely which phonological capacities can be observed in the individual word production of Swedish toddlers and how, if at all, vocabulary size is related to phonological capacities in Swedish toddlers. The results from chapter 4 showed that the predicted mispronunciations based on parental reports were unapparent or unelicitable in 17- and 24-month old children. Thus, we sought to determine typical production patterns in a group of Swedish toddlers. In order to examine the child’s production abilities more closely, another strategy was considered. The acquisition of phonology is interrelated with the acquisition of the lexicon (Stoel-Gammon, 2011), which indicates a relationship between phonological production capacities and vocabulary size. Namely, possible segmental changes and/or the adaptation of target words to phonological templates as well as the consonant inventory will be related to vocabulary size. The background gives an overview of existing speech production studies focusing on the development of the lexicon and phonology during the second year of life.

5.1. Background

The phonological development of three English-speaking children has been described longitudinally by Ferguson and Farwell (1975). They started transcribing the child’s vocalization around the first birthday and continued until the child reached the point of producing 50 words. Ferguson and Farwell observed both word forms that vary as well as many accurate, stable word forms at the onset of first word production. The first consonants consisted of labials and alveolar stops, followed by nasals and glides, and finally fricatives. In a larger study examining the
5. *Study II: Phonological capacities at 18 months in relation to vocabulary size*

form of consonant inventories, 34 English-speaking children were recorded longitudinally from 15 to 24 months (Stoel-Gammon, 1985). The consonant inventory of these 34 children contained stops and nasals and, later, glides and fricatives. This inventory increases with age and tends to be richer in word-initial position. Labials and alveolars usually occurred earlier than velars. The consonant inventory in initial position contained voiced before voiceless stops. In final position this observation was reversed, with voiceless appearing prior to voiced stops. Menn and Vihman (2011) compiled several studies investigating early speech production from 11 different language backgrounds, including 57 children between the ages of 9 and 20 months. Of 250 listed target words 38% included more than one consonant; for only 18% the consonant in word-onset position differed from the consonant in a later word position. Consonant clusters were rare (three words).

In Swedish 18-month-old typically developing children, nine different consonants were found, most of them dentals, followed by labials (Lohmander, Olsson, & Flynn, 2011).

Grunwell (1982) proposes a phonological profile in four stages from 9 to 36 months, focusing on the development of phonological capacities. She suggests phonological rules (which she terms ‘simplification processes’) that can be observed until the age of five in a typically developing child. The consonant inventory in the first stage consists of nasals, plosives, fricatives and approximants. CV, CVCV and CVC word structures as well as all simplifying processes occur. Stage II lasts from 1;6 to 2;0 and comprises the production of labial and alveolar nasals as well as stops. The same word structures as in stage I occur: CV, CVCV and CVC. The following phonological processes are noted: reduplication, consonant harmony, final consonant deletion, cluster reduction, fronting of velars, stopping, gliding and context-sensitive voicing. Most of the processes are also apparent during stage III, except for fronting of velars, which starts to disappear around 24 months, at the time when velar consonants first appear. Stage IV lasts from 2;6 until 3;0, and only specific sounds undergo the phonological processes of stopping and fronting. Final consonant deletion and cluster reduction are still apparent in stage IV. Dodd et al. (2013) observed error patterns like stopping until the age of 3;5 and fronting of velars /k,g/ until the age of 3;11. Furthermore, the deletion of weak syllables was present until 3;11. The substitution of /l/ with /j/ or /l/, so-called ‘gliding’, was observable for some of the children until the age of 5;11. Nettelbladt (1983) has shown similar phonological processes for one Swedish child. A norma-

---

18Note, the term ‘voiceless’ in English stops refers to an aspirated production, whereas ‘voiced’ refers to zero voice onset time (VOT).

19Swedish /d/ and /t/ are phonologically classified as dentals, cf. section 2.1.1.
A study with Danish children aged 2;6 to 4;11 showed that children at 2;6 years master all consonants and all vowels of the Danish phonetic inventory (Clausen & Fox-Boyer, 2017). Most phonological processes were resolved at 3;11 years; the most common phonological process in the youngest age group between 2;6-2;11 was initial and final cluster reduction as well as weak syllable deletion.

In addition to the general results on group data in the studies mentioned above, several researchers have mentioned individual preferences and strategies. Ferguson and Farwell (1975) observed in one child an extensive use of reduplication and in another a persistent avoidance of one specific sound. One child seemed to have a preference for sibilant fricatives that led to an overrepresentation of words including such sounds in her early vocabulary (e.g. *ice*). Another child presented by Stoel-Gammon and Cooper (1984) showed a preference for words with a velar stop in final position, leading to a vocabulary of first words with an overrepresentation of velar-final words such as *clock, milk, talk, walk* or *frog*. This child assimilated the initial consonant to the velar in final position, resulting in CVC(V) structures where both consonants consisted of velar stops. For instance, the adult word form of *sock* was produced as [gak]. Priestly (1977) described a child who overused the word structure CVjVC and transferred this pattern to adult word forms.

These observations of individual preferences for speech sounds and word structures led to the approach of Vihman and colleagues (e.g. Keren-Portnoy et al., 2010; Velleman & Vihman, 2002; Vihman, 1981, 2014), which suggests phonological templates based on a whole-word phonology (cf. chs. 2 and 6). A template expresses “the idea of a word production pattern or routine specific to a particular child. The production pattern develops through vocal practice with both babble and first words and is shaped by its use in matching adult word targets” (Vihman, 2014, p. 176). An example of adaptation to a phonological template can be observed in the case described by Priestly (1977): the template involves CVjVC, and the child adapts it to adult word forms, for instance, [bajak] ‘blanket’ or [narjan] ‘melon’. The adaptation of phonological templates to adult word forms can be found in several languages. For instance, in Italian a VCV pattern [eta] *forchetta* ‘fork’ was found (Keren-Portnoy et al., 2009) and in a bilingual child acquiring English and Estonian a CVi pattern [bvi] *apple* ‘apple’ (Vihman, 2016). A Finnish-speaking child showed a phonological template in terms of consonant harmony ([nenä] *häänti* ‘tail’); a French-speaking child favored a medial [l] ([bolu] *bross* ‘brush’) (Vihman, 2010). Additional languages, where phonological templates were reported, include, for example, Norwegian (Garmann, Kristoffersen, & Simonsen, 2017), Arabic (Khattab & Al-Tamimi, 2013) and Welsh (Vihman, 2010, 2014).20

---

20For further examples in several languages, see, for example, Appendix III in Vihman (2014).
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

The youngest age for the use of adapted templates in the sample provided by Vihman (2014) was observed in a bilingual child (Estonian-English) at 14 months, a few weeks after production of the first words. In contrast to phonological processes which describe mispronunciations on a segmental level, a template might affect the whole word form (as was seen in the example of the Estonian-English in \[\text{[bi:] apple ‘apple’ Vihman, 2016}\]) and not only specific segments or phonemes.\(^{21}\) Additionally, it is an individual routine based on the child’s motor practice, which implies that different children learning the same language follow different developmental paths in the use of phonological templates (Keren-Portnoy et al., 2009).

Looking at lexical development, between the first and second birthday the vocabulary grows for the majority of the children from a few words to 50 words or more. After passing the 50-word point, children are usually quicker to acquire new words (e.g. Goldfield & Reznick, 1990), resulting in vocabulary sizes around 500 words by 30 months (Fenson et al., 1994). However, the variation in lexical size is considerable, reaching its maximum between 16 and 24 months (Bates et al., 1994). On the basis of parental reports, Bates et al. (1994) proposes three different phases covering a range of months to describe the growth of vocabulary. The first, ‘referential phase’, lasts until the child has acquired around 100 words. During this phase, the lexicon increases mostly by the addition of nouns. In contrast, the next phase (‘prediction phase’) is characterized by a large growth in the number of verbs. The prediction phase encompasses a vocabulary size between 100 and 400 words. The final, ‘grammar phase’, refers to a large increase of function words. During this phase the vocabulary size consists of 400 to 700 words. At 24 months, English-speaking typically developing children have a productive vocabulary size of 250 to 350 words (Stoel-Gammon & Sosa, 2008). Fenson et al. (1994) gives a tentative chronological directive of a lexical size of circa 500 words around 30 months of age for typically developing children.

Which words are acquired first is also influenced by phonology. Phonotactic probability, neighborhood density as well as word length influence the acquisition of words (Storkel, 2001, 2004, 2006; Storkel & Lee, 2011). Besides the preferences for specific sounds or word structures as described above, the relationship between lexical and phonological development is apparent in children who show a reduced use of canonical babbling and have delayed lexical development (Paul & Jennings, 1992; Rescorla & Ratner, 1996; Stoel-Gammon, 1991). Furthermore,

\(^{21}\)A similar distinction is made in other contexts, for instance, clinical linguistics, in terms of syntagmatic and paradigmatic processes. Syntagmatic processes effect the phonotactic and prosodic structure of the target word, and paradigmatic processes influence different types of segments and consist of, for instance, substitutions.
the consonant inventory mostly correlates with vocabulary size in children with a language delay: the larger the vocabulary, the more advanced the phonological system (Fasolo, Majorano, & D’Odorico, 2008; Stoel-Gammon, 1991). In addition, articulation accuracy measured in typically developing children at the age of 33 months is a good predictor for vocabulary sizes at 36 and 39 months (Schwarz, Burnham, & Bowey, 2006).

Marklund, Sundberg, Schwarz, and Lacerda (2012) related the vocabulary size of 15 Swedish 30-month-old children to their phonological capacities. The results showed that children with a large vocabulary produced more words with higher phonological complexity, e.g., consonant clusters, word-final consonants, fricatives or velar consonants, than are produced by children with a smaller vocabulary score. However, the authors relied on parental reports rather than the actual speech production of the children.

The present study connects vocabulary size to phonological performance recorded in parent-child interactions. The phonological performance will be described by consonant inventory and mispronunciations, aiming for a better understanding of phonological capacities in a group of Swedish toddlers. Specifically, the occurrence of possible segmental changes and/or the adaptation of target words to phonological templates will be described as well as the consonant inventory. The expectations were based on the results of studies in other languages: Grunwell’s phonological processes as well as the most common template forms, such as consonant harmony, were expected. The production capacities are related to vocabulary size.

5.2. Methods

Participants

The recorded data of the participants comes from a larger study investigating parent-child interactions.22 Out of 60 18-month-old children, 12 (average age: 18;13; range 17;1-18;7) were selected based on the children’s productive vocabulary size according to the Swedish version of the MacArthur-Bates Communicative Development Inventory (Berglund & Eriksson, 2000). Four of the selected children were considered to have a high productive vocabulary based on percentiles in the Swedish CDI of 80 or above. Six of the children fell between the 50th and 75th percentiles and so were considered to have a medium vocabulary size. The remaining

22SPRINT, conducted by the Department of Linguistic and Department of Special Education at Stockholm University, 2009-2012, and funded by the Swedish Research Council: VR2008-5094.
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

Table 5.1. Demographic data and vocabulary scores (words produced) from the Swedish Early Communicative Development Inventory (SECDI) (Berglund & Eriksson, 2000) for the 12 participants. F = female; M = male. Age is given in months and days.

<table>
<thead>
<tr>
<th>Prod. Group</th>
<th>Subject</th>
<th>Gender</th>
<th>Age</th>
<th>SECDI score</th>
<th>SECDI %tile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Simon</td>
<td>M</td>
<td>18:1</td>
<td>4</td>
<td>&lt;10</td>
</tr>
<tr>
<td></td>
<td>Eva</td>
<td>F</td>
<td>18:10</td>
<td>15</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Medium</td>
<td>Albin</td>
<td>M</td>
<td>17:19</td>
<td>36</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Henry</td>
<td>M</td>
<td>18:3</td>
<td>37</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Astrid</td>
<td>F</td>
<td>18:20</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>August</td>
<td>M</td>
<td>18:3</td>
<td>51</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Fredrik</td>
<td>M</td>
<td>18:7</td>
<td>53</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Ivan</td>
<td>M</td>
<td>18:1</td>
<td>59</td>
<td>65</td>
</tr>
<tr>
<td>High</td>
<td>Anders</td>
<td>M</td>
<td>18:0</td>
<td>134</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Celine</td>
<td>F</td>
<td>18:25</td>
<td>161</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Cecilia</td>
<td>F</td>
<td>18:10</td>
<td>262</td>
<td>&gt;90</td>
</tr>
<tr>
<td></td>
<td>Emma</td>
<td>F</td>
<td>18:22</td>
<td>376</td>
<td>&gt;90</td>
</tr>
</tbody>
</table>

Two children fell below the 25th percentile and made up the small vocabulary group. Table 5.1 gives an overview of the demographic data of the 12 participants. The names are changed for anonymity. All 12 children were born full-term with normal birth weight, and no family reported speech or language difficulties for any family member.

Data analysis

The speech material consisted of home recordings of spontaneous parent child-interactions during a reading situation. The book was individually chosen by the families. The recordings lasted on average 15 minutes and were phonetically transcribed in Praat (Boersma, 2001). Productions with very poor sound quality, e.g. overlapping speech with an adult or background noise from toys, were excluded from the transcriptions. All remaining utterances were transcribed by a phonetically trained annotator.

A second phonetically trained annotator made an independent transcription of the samples from three of the twelve participants. The reliability between transcribers was calculated as the number of matching consonant labels divided by the
5.3. Results

The results are based on the types and tokens produced by the children during the recording session. The number of types ranged from 2 to 49 (table 5.2). The number of produced types was largest in the high-production group; the maximum of tokens was found in the medium-production group (Albin). Astrid from the medium-production group had the smallest number of types and tokens in this production group. The low-production group produced the fewest types and tokens.

The consonant inventory of the different production groups shows an increase in the number of consonants in relation to vocabulary size (see table 5.3). The children in the low production group had two consonants (plosives) in onset position and one consonant in coda position. The medium-production group produced mainly nasals in coda position and additionally the approximant /j/. In the onset position stops (including a velar stop), nasals and a bilabial fricative were observed. The high-production group had the most consonants in both onset and coda position, including stops (both velar, labial and dental), nasals, fricatives and one affricate ɗʒ.

The children from the low-production group produced fewer words in comparison to the other production groups (table 5.2). Eva produced two words; both were correctly pronounced. Simon produced nine words; one of those was mispronounced: [ɛːɑ]ārtə ‘pea’.

The proportion of correct pronunciations tends to slightly decrease with a larger vocabulary: The group with a high SECDI score (Anders, Celine, Cecilia, Emma) achieves around 50% correct pronunciations (figure 5.1) and has the lowest
### Table 5.2. Number of types and tokens.

<table>
<thead>
<tr>
<th>Prod. Group</th>
<th>Subject</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Simon</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Eva</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Albin</td>
<td>23</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Henry</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Astrid</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Medium</td>
<td>August</td>
<td>22</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Fredrik</td>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Ivan</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>High</td>
<td>Anders</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Celine</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Cecilia</td>
<td>39</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Emma</td>
<td>49</td>
<td>87</td>
</tr>
</tbody>
</table>

### Table 5.3. Consonant inventory in onset and coda position in the three production groups.

<table>
<thead>
<tr>
<th>Prod. Group</th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>b d</td>
<td>n</td>
</tr>
<tr>
<td>Medium</td>
<td>p b l m h n t v g</td>
<td>j m n η</td>
</tr>
<tr>
<td>High</td>
<td>l b t p k d h m c d j</td>
<td>n f t k l η η c</td>
</tr>
</tbody>
</table>
percentage correct at 27% (Anders). The proportion of correct pronunciations in the medium-production group ranges from 67% to 93% and for the low-production group between 89% and 100%.

All children in the medium- and high-production groups showed phonological processes affecting segments (table 5.4). The most frequent segmental error was the reduction of consonant clusters (seen in eight children). The omission of a final consonant was observed in five children; the substitution of /t/ to /j/ or /l/, gliding, was found in three children. The same number of children omitted an initial consonant or showed a stopping process. Two children fronted velar consonants. One child (Emma) showed a substitution of initial fricatives to /h/ (here called ‘H-zation’) and a substitution in the form of labialization. The highest number of segmental errors were observed in the two children with the largest vocabulary, Cecilia (5) and Emma (6). The other children showed one to a maximum of three phonological processes.

Mispronunciations affecting the whole word occurred in seven children and included assimilations, syllable reductions and metatheses. Nevertheless, the number of occurrence was in some children too small to determine the existence of a template. For instance, Astrid and August from the medium-production group showed two assimilations each. Emma from the high-production group had one assimilation and three metatheses. The metatheses had different shapes, for instance,
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

Table 5.4. Overview of the segmental errors for 10 participants. Only those segmental errors are included that occurred at least twice in the child’s production, leading to the exclusion of the low-production group (Eva and Simon), due to little production data. RCC=Reduction of consonant cluster. OIC=Omission of initial consonant. OFC=Omission of final consonant. G=Gliding. F=Fronting. L=Labialization. H=H-zation. S=Stopping.

<table>
<thead>
<tr>
<th>Subject</th>
<th>RCC</th>
<th>OIC</th>
<th>OFC</th>
<th>G</th>
<th>F</th>
<th>L</th>
<th>H</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albin</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Henry</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astrid</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fredrik</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ivan</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anders</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Celine</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cecilia</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emma</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

86
5.4. Discussion

Table 5.5. Phonological templates adapted to target words and their shape.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Phonological template</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anders</td>
<td>CH</td>
<td>[lumlå] blomma ‘flower’</td>
</tr>
<tr>
<td></td>
<td>CV(C)CV</td>
<td>[drlär] istället ‘instead’</td>
</tr>
<tr>
<td></td>
<td>C_{velar}VCV(C)</td>
<td>[gţjan] tröjan ‘pullover’</td>
</tr>
<tr>
<td></td>
<td>C_{velar},V(C)</td>
<td>[bulan] baljan ‘tub’</td>
</tr>
<tr>
<td>Celine</td>
<td>CH</td>
<td>[dadi] vatten ‘water’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[crecz] mössa ‘hat’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[gâka] kossa ‘cow’</td>
</tr>
<tr>
<td>Cecilia</td>
<td>(C)VC_{fric}.V</td>
<td>[tacz-400ms,e:] trasig ‘broken’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[barcz-396ms,t] plask ‘splash’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[cz-241ms,å:] läsa ‘to read’</td>
</tr>
</tbody>
</table>

Table 5.5. Phonological templates adapted to target words and their shape.

[kêni] ka’nin ‘rabbit’ or [jæbOl] hjälpa ‘to help’. Ivan had three syllable reductions in two words: [da] blåddra ‘to flip pages’ and ballong ‘balloon’, realized as [buŋ] and [bʊ]. The number of affected target words were too small to determine whether these mispronunciations reflected an adaptation of target words to phonological templates. Still, the adaptation of target words to phonological templates was identified in three children from the high-production group (Anders, Celine and Cecilia), with a vocabulary size between 100 and 300 words. The shape of the phonological templates included consonant harmony in two of the children and the word structures CVC_{velar}V(C), C_{velar},VCV(C), CV(C)C_{velar}V(C) and (C)VC_{fric}.V (table 5.5).

The phonological template of Cecilia is salient, with an occurrence of 20 cases, and stands out because it involves a break in medial position, for instance, [tacz-400ms,e:] trasig ‘broken’. A further analysis revealed an average duration of the break for 279 ms before she proceeded to the next syllable or consonant. In comparison, the longest duration of a break after a plosive was 123 ms.

5.4. Discussion

The current study investigates the phonological status of 18-month-old Swedish children in terms of consonant inventory, accuracy in word production, the occurrence of segmental errors as well as phonological templates in relation to their vocabulary size as reported by their parents. In general, the vocabulary size reported by the parents with SECDI (Berglund & Eriksson, 2000) was reflected in
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

the number of types and tokens produced in the recording session: The children in the high-production group according to parental reports had the largest number of produced types. The results of the consonant inventory showed that the consonant inventory increases with the size of vocabulary. A similar developmental path was observed in earlier studies with English speaking children (Stoel-Gammon, 1985, 1991). The accuracy in word production showed that children with a small vocabulary size have a higher accuracy than those with a larger vocabulary, which indicates that the first words are more accurate. At the same time a variability was observed, for instance, for Anders, who made several attempts to pronounce the word *inte* 'not’, several times successfully, but sometimes resulting in different mispronunciations. This is in line with earlier studies where first words were reported as accurate and stable, but later there was variation in the words in the child’s performance (Ferguson & Farwell, 1975; Vihman, 1996).

In the current study, phonological processes occurred in all children. Still, the number of segmental errors was in general higher in children with larger vocabularies than in those with smaller ones. The most common mispronunciation consisted of the reduction of consonant clusters: a process which was expected according to Grunwell (1982) for this age group and also is reported as the most common segmental process in older Danish children (Clausen & Fox-Boyer, 2017). However, in one child a few more processes were observed that were not mentioned by Grunwell (1982), that is, metatheses, H-zation and labialization. Still, metatheses are, for instance, reported in Italian children (Keren-Portnoy et al., 2009), and Nettelbladt (1983) reports H-zation as well as labialization in one typically developing Swedish child. Thus the results from the present study confirm Nettelbladt’s single-case study in the observation of H-zation and labialization.

The occurrence of phonological templates adapted to target word forms was observed in children with a vocabulary size above 100 words and below 300 words. Several earlier studies found phonological templates earlier (e.g. Vihman, 2010, 2014, 2015), although the productive vocabulary was mostly determined by words produced within a recording session, for instance, a ‘25-word point’ (Vihman, 2015) corresponding to a parental-report level of 50-70 words (Vihman & Miller, 1988). In longitudinal studies with Italian children (Keren-Portnoy et al., 2009; Majorano, Vihman, & DePaolis, 2014; Vihman & Majorano, 2016), phonological templates were found in children with a smaller vocabulary size than 100 words. Garmann et al. (2017) reports the use of phonological templates in four Norwegian children aged 1;7-1;9 in a 30-minute video-recorded session. Those children produced around 50 different words in the recorded playing situation (50-word point), more words produced than the children in the medium- and low-production groups
5.4. Discussion

in the present study. The lack of templates in the medium- and low-production groups might be due to the low number of words produced overall. Additionally, the recordings in the current study differed in terms of duration, situation (reading) and the use of audio instead of video recordings, which might help to identify more word candidates. The combination of vocabulary size estimates provided by parental reports and the measurements of speech production capacities allowed us to connect phonological capacities to vocabulary size and showed that CDI provides a reasonable complement to the research of phonological capacities.

At an individual level, the adaptation of target words to phonological templates in the form of consonant harmony (Celine and Anders) has been reported for several other languages, for instance, English (Keren-Portnoy et al., 2010; Sowers-Wills, 2016; Vihman et al., 1985) and Italian (Keren-Portnoy et al., 2009). In the current study, one child (Anders) showed some fixed patterns, two involving velars in different positions and one involving [l] word medial. Similarly, Priestly (1977) reported a pattern including medial [j].

The phonological pattern of another child (Cecilia) involved a break after the production of a fricative, which is seldom described in the literature. Still, Vihman (1976) observed in a diary study one child leaving a break after a fricative, but without a measurement of the duration of the break. The acquisition of fricatives usually takes a longer time than other phonological categories (Ingram et al., 1980; Nettelbladt, 2007b), and the observed break might reflect this phenomenon. Further studies with an acoustical analysis of fricatives are needed to investigate whether breaks after fricatives can be observed in several children and if this is a more common developmental pattern. An articulatory study could highlight the motor coordination involved in producing the break.

In general, the formation of templates could be regarded as an attempt to harness phonetic variability on behalf of the child. The child regulates both the acoustic as well as articulatory characteristics of sounds by adapting word forms into a well-practiced productive routine. Templates are phonological, as they reflect an abstraction that the child uses when producing target words that do not fit yet into already acquired productive sequences of sounds. They emerge through speech production and, consequently, the practice of articulatory sequences. For instance, Anders attempted to produce the Swedish target word istället ‘instead’ and fell back on the template CVC\V [dr1æ].

In sum, the results from this study expand the evidence from other languages and describe which phonological capacities can be observed in the individual word production of Swedish toddlers (cf. research question 3). in terms of the use of segmental errors, phonological templates, percentage correct, and consonant in-
5. Study II: Phonological capacities at 18 months in relation to vocabulary size

Ventory. Additionally, individual differences in the shape of phonological templates were described. The results on the use of phonological templates suggest a vocabulary size between 100 and 300 words, where phonological templates are more likely to occur. This gives an indication for the experimental study to follow, in which the influence of production on perception was to be tested.
6. Study III: MAGIC Analyzes Gaze In Children: A perception study based on individual children’s speech production

The previous chapter focused on the speech production of 18-month-old Swedish toddlers. The present chapter integrates the findings into a study investigating the influence of the individual child’s production on perception and is thus related to research question 1: How, if at all, do a child’s phonological and lexical capacities in production influence word recognition.

6.1. Background

Some researchers, focusing on phonology and perception, have assumed very detailed phonological representations from 12 months on (cf. Swingley, 2003; Swingley & Aslin, 2000, 2002) (cf. ch. 3). Based on tasks on word form recognition or word learning of minimal pairs, it has been shown that infants and toddlers detect mispronunciations of tiny details. Hallé and de Boysson-Bardies (1996) tested word form recognition in several experiments with 11-month-old French infants. The omission of word-initial consonants blocked word form recognition. In contrast, changes in voicing and manner of the word-initial consonant did not affect word form recognition. In a study with Dutch infants of the same age, Swingley (2005) shared that changes to the onset or final consonant led to a failure in word form recognition. Slightly older children, at 14 months, showed a sensitive reaction to mispronunciations of newly learned words in a preferential looking task (Mani et al., 2008; Mani & Plunkett, 2008). Specifically, the word manipulations in these experiments consisted of vowel changes. Similar results were found at 15 months.
6. Study III: MAGIC Analyses Gaze In Children

(Swingley & Aslin, 2002). In this study, children were exposed to closely mispronounced, distantly mispronounced and correctly pronounced words. Closely mispronounced words were created by changing one sound to a similar sound, e.g. *dog* to *tog*; distantly mispronounced words were created by changing one sound to a dissimilar sound, e.g. *dog* to *mog*. The reactions to the correct pronunciations of the words differed significantly from the reactions to the two conditions of mispronunciations, indicating that the lexical representation of the tested words is stored in detail. This study gets further support from White and Morgan (2008), who showed in a similar experiment different reactions to phonetic feature changes of word-initial consonants. In this study, the looking times of 19-month-old children differed significantly depending on the degree of mispronunciations: The more the phonetic features were changed, the shorter the looking time to the familiar object was. Additionally, if the novel object was named, the children presented a longer looking time towards the novel object, indicating the presence of ‘mutual exclusivity’ (Markman & Wachtel, 1988). Bailey and Plunkett (2002) tested 18- and 24-month-old children on different mispronunciations of the word-initial consonant and showed that in both ages children differentiated between accurate and inaccurate pronunciations. Concretely, different one- or two-feature changes led to shorter times looking towards the target object, independently of age, receptive vocabulary size or phonological neighborhood of the tested words. Another study on 24-month-old children showed that the sensitivity to mispronunciations in word-medial onset and coda positions in a word learning task differs: Mispronunciations in word-medial onset positions hindered word recognition but mispronunciations in coda positions did not (Wang & Seidl, 2016). The importance of initial phonetic information was also shown in a study with 18- and 21-month-old children who recognized partial words (first 300 ms of a word) as quickly and reliably as whole words (Fernald, Swingley, & Pinto, 2001). Moreover, children with a productive vocabulary larger than 100 words were faster and more accurate in word recognition.

In sum, most of the studies supporting the approach of phonological detailed representations have investigated perception of word initial consonants (Bailey & Plunkett, 2002; Hallé & de Boysson-Bardies, 1996; Swingley & Aslin, 2002; White & Morgan, 2008); fewer studies have explored vowels (Mani et al., 2008; Mani & Plunkett, 2008), word medial (Wang & Seidl, 2016) and coda consonants (Swingley, 2005; Wang & Seidl, 2016).

In contrast to the phonological specificity approach, studies based on production have proposed more ‘holistic’ representation of early words based on a whole-word phonology (Ferguson & Farwell, 1975; Vihman, 2017b; Vihman &
6.1. Background

Keren-Portnoy, 2013a; see also chapters 2 and 5). According to this approach, children learn a language through the learning of patterns and acquire words as wholes (Velleman & Vihman, 2002), presupposing a close interrelationship between the acquisition of the lexicon and phonology. In the acquisition of phonology, patterns in the form of phonological templates have been proposed (cf. section 5.1 and ch. 3). In this view, “The [phonological] template is a simplification pattern, usually based on earlier word preferences (selection), which is now applied to a variety of adult words through a process of adaptation” (Velleman & Vihman, 2002, p. 20), which implies two possible ways of using a phonological template: (1) for the selection of word forms from the speech input for adult target forms that fit the phonological template, in which the child chooses to produce those adult target word forms that match the phonological template, and (2) the adaptation of adult word forms in the phonological template, in which the template is applied/overgeneralized to adult target forms that do not match the template. Given the two different types of application, this approach predicts an influence of production abilities on perception. The child’s practiced motor routine helps to procure new words from the input, albeit ones that are similar to the practiced routine, and facilitates their transfer to production. The use of phonological templates has even been termed ‘secondary distributional learning’ (Vihman, 2014) in terms of abstraction or generalization of schemas or templates from word forms familiar from production, resulting in the child attempting more challenging word forms, but with a loss of accuracy. Thus, the occurrence of template [...] can be seen as expressing the child’s growing ability to generalize [...] from what is known [...] to what is unknown. This seems to be a plausible effect of the integration of implicit and explicit memory functions. (Vihman, 2017b, p. 16-17)

The connection to memory functions as well as the parallels to statistical learning indicate a possible psychological reality for the use of templates. De-Paolis, Vihman, and Keren-Portnoy (2011) showed that a one-year-old’s own consonant production experience influences perception of nonwords in a head-turn experiment. Children who had multiple vocal motor schemes showed longer looking times in response to consonants that they were not producing, whereas children with a single vocal motor scheme seemed to show a preference for consonants they knew.

The influence of production on perception beyond phonological templates has been shown in still younger children in two recent studies. The results of a study by
Masapollo et al. (2016) indicate that pre-babbling infants’ own production abilities influence their perceptual preference, since infants tested 3-6 months of age show a preference for speech produced by an infant vocal tract over speech produced by an adult vocal tract. Another study by Altvater-Mackensen et al. (2016) suggests that production abilities influence the sensitivity to audiovisual matches. In this study, six-month-old children looked at short clips of a woman who articulated vowels. The simultaneously presented auditory stimulus either matched the visual cue of the vowel or not. The authors showed that children who had more distinct sound categories in their babbling were better able to detect mismatches between the seen and heard vowel. DePaolis et al. (2016) used an auditory headturn preference procedure to test ten-month-old children on familiar and rare words; the children’s performance in this task was compared to their productive capacities indexed by vocal motor schemes. Those children who were early in the production of vocal motor schemes showed either a strong novelty or strong familiarity response, whereas those children who had not yet reached that production level showed no preference for either familiar or rare words.

Studies claiming phonological detailed representations of early word forms are based on perceptual studies. In contrast, the assumption of holistically presented word forms is based on production data assuming an interrelationship between production and perception. However, a link between production and perception is also, although less commonly, hinted at in the phonological detail literature. For instance, the results of the study by Swingley (2016) indicate that vocabulary size—words produced—has an impact on the perceptual performance, pointing to a link between lexicon and phonology. The results of Wang and Seidl (2016) and the absence of influence on word recognition in changes to the coda are discussed by the authors in connection to word production: Early words often lack the coda consonants, so children do not focus on final consonants, resulting in a mapping of the mispronounced word onto the familiar object. Nevertheless, the approach of phonologically detailed representations does not predict a clear influence of production onto perception other than that perception precedes production. In other words, the perceptual abilities start to develop earlier than the capacities of speech production. On the other hand, the assumption of an effect of production on perception interpreted in terms of whole-word phonology predicts that the child’s own production capacities influence the perception of word forms. Accordingly, word form changes which are similar to the child’s own production should not hinder word recognition. Following the assumption of phonological detailed representations, manipulations would block word recognition, especially if the manipulations occur in a salient position of the word and if they include a distant mispronuncia-
6.1. Background

Within the holistic approach, the use of phonological templates plays an important role. The motor routines practiced in the form of phonological templates guide the child through the jungle of the surrounding speech stream to select word forms that match the motor routine. If and how this influences the recognition of words under experimental conditions is yet unclear. The present study aims to find evidence for an influence of the child’s own production on perception in terms of word recognition. Thus, the following research questions were evaluated:

1. How, if at all, do a child’s phonological and lexical capacities in production influence her word recognition? (cf. research question 1 in ch. 1) And more concretely:
   a) Does the individual child’s production influence the word recognition of manipulated words, if the manipulations are based on the child’s own pronunciation?

2. If the child uses phonological templates, does this influence the recognition of words?

In order to answer these questions, children were tested in four different conditions tailored to the individual child in a preferential looking task. Words and their references were chosen based on the child’s production. Besides the correct pronunciation of the words, the children were exposed to three other conditions: Condition CM (Child’s Mispronunciation) includes pronunciations that the child uttered herself which were deviant from the adult target form. Condition AM (Atypical Mispronunciation) included atypical mispronunciations of the adult target form. As in White and Morgan (2008), a condition Novel containing rare, infrequent words was tested.

The hypotheses for the different conditions were that (1) the children will recognize the correct word forms easily, which will be reflected in a longer looking time towards the familiar object. According to whole-word phonology and the expectation of an influence of production onto perception, a similar reaction is expected for the word forms that are close to their own pronunciation (CM) (2). For the AM condition we expect either a longer looking time towards the unfamiliar object, or at least no significant difference in looking times (3). For the Novel condition a longer looking time towards the unfamiliar object is expected based on the concept of mutual exclusivity (4) (see chapter 2). Following the results from the production study presented in chapter 5, only children with a vocabulary size over 100 words were included to increase the probability of phonological template
6. Study III: MAGIC Analyses Gaze In Children

presence and to capture a sufficient amount of words showing phonological characteristics in the early child’s adult-like word form production. How the use of templates influences the results in the perception experiment will be explored.

6.2. Methods

Participants
The participants were recruited through the Swedish Tax Agency (skatteverket) address register. Families with children between 18 and 26 months living in Stockholm were sent a letter with instructions on how to enroll in this study on a website. In total 200 letters were sent out in two different rounds. Additionally, adverts were put up at a kindergarten close by the university and on the university campus. Twenty-six families registered for the study through the website. In the registration form the parents were asked about their language background as well as about any complications concerning their child’s hearing or premature birth and whether there are younger or older siblings. None of the registered children had any reported health issues and all were at 37 weeks gestation. Written informed consent was obtained from all participating families.

Procedure
This study consists of four different steps, which were carried out sequentially.

1. Examining the child’s vocabulary size
2. Examining the child’s speech production capacities
3. Stimuli design in the four conditions
4. Testing the child’s perception in an eye-tracking experiment

The first step covers the examination of the child’s vocabulary size based on parental reports. The second step includes speech recordings and transcriptions for examining the child’s speech production. The third step involves the stimuli design in four different conditions, three of which are connected to the child’s speech production. During the fourth and last step the child’s perception was tested in an eye-tracking experiment in four different conditions. The design of the eye-tracking

23The online registration form was set up with the free software LimeSurvey.
6.2. Methods

experiment, with two pictures presented on the eye-tracking screen, is inspired by White and Morgan (2008) with respect to inclusion of the Novel condition and a similar presentation of stimuli. Each step is described in detail below.

Examining the child’s vocabulary size (step 1)
After registration for this study, the families were asked to fill in the Swedish version of CDI (Berglund & Eriksson, 2000) to examine the vocabulary size of the child. If the vocabulary size was 100 words or above, an appointment was made to visit the families in their home environment to record the child’s production (step 2) within one week. If the parents reported fewer than 100 words produced for the child, the parents were asked to fill in the questionnaire weekly until the vocabulary size was above 100 words.

Examining the child’s speech production capacities (step 2)
The child’s speech production was captured via audio and video recordings at home. Two researchers visited the families, recording the children wherever they felt comfortable, which was usually in the living room. The parents were asked to stay in the background as far as possible and not to interact with the children. The child’s speech production was recorded by an audio recorder and a separate camera. During the recording, the child’s speech production was transcribed on-line by one of the researchers. The other researcher interacted with the child in a semistructured conversation. As suggested in the discussion of chapter 5, the recording session included both a playing part as well as a book reading situation. Various objects in a big bag were shown to the child and the child was asked to name them. The objects were from different semantic categories, specifically: food, clothes, animals and toys. The objects were chosen based on the norm data for 16-month-old children, provided by Jørgensen et al. (2010). A complete list of the objects in the bag and information about the norm data can be found in Appendix B. Since the children in this study were older than the children in the norm data, more children were expected to be able to name the objects than the percentage from the norm data suggests. Furthermore, the toys were chosen by phonological complexity, including words with more than two syllables, consonant clusters in initial and final position as well as fricatives. To elicit enough speech material from the child, the child and the researcher played together after unpacking the objects out of the bag. If the child had a favorite toy, for instance, a doll, it was integrated in this playing session. Concretely, the clothes from the bag were put on the child’s doll or cud-

---

24The parental report was collected through an online survey conducted with the free software LimeSurvey.
25The 100 words criterion is based on the results of the production study in the previous chapter.
6. Study III: MAGIC Analyses Gaze In Children

dly toy. Furthermore, the playing session included a part in which the researcher together with the child pretended to feed the animals. After approximately 15 minutes the child was asked to pick a book that he or she would like to read for the animals. Together with the researcher the child looked through and talked about the book. At the end of the session the child was asked to help put the toys back into the bag, naming every single toy again. The child was never asked to just repeat a word produced by the test leader (e.g. ‘say ball’), but was asked to name it (i.e. ‘what is this?’). The recording session lasted around 30 minutes. After the recording sessions the speech material was transcribed phonetically in PRAAT (Boersma, 2001) and supported by the video recordings to see, for example, where the child was pointing. Productions with very poor sound quality, e.g. overlapping speech with an adult or background noise from toys, were excluded from the transcriptions. All transcriptions were compared to the on-line transcriptions of the second researcher who also took notes. In cases in which the two transcriptions were not consistent, both researchers listened again to the recording and agreed on one version. For each child, the occurrence of segmental errors and/or phonological templates was examined. As in earlier studies, a phonological template was identified if it affected at least 10% of the produced types (cf. Garmann et al., 2017; Kehoe, 2015; Sowers-Wills, 2016; Vihman, 2016). Furthermore, the consonant inventory for each child was analyzed in order to design the stimuli in condition AM (see following section).

Stimuli design in the four conditions (step 3)

For the perceptual experiment, visual and auditory material was created for four different conditions: Correct, CM, AM and Novel. The auditory stimuli in the three conditions Correct, CM, AM were tailored for each individual child according to their own production. This results in individually chosen auditory stimuli and objects for every single child with exception of the Novel condition (a detailed description of Novel follows after the description of Correct, CM, AM). Based on the phonetic transcriptions of the recordings in the previous step, six words were chosen for each child. These six words were mispronounced by the child during the recording session, either under the influence of a phonological template or in form of a segmental error. The words were chosen according to the following criteria:

Stability: No other pronunciations of the word occurred during the entire recording session, whether correctly or incorrectly pronounced.

Transfer: More than one word showed the same type of mispronunciation.

Imageability: The word is sufficiently easy to present in a photograph, since the
word was to be visually presented later in the eye-tracking experiment. In other words only nouns were candidates for inclusion in the experiment.

Based on these criteria, six objects from the elicitation task were chosen for presentation in three different conditions in the perception experiment (step 4). In the first condition the object was named correctly (Correct). In the second condition, CM, the object was named according to the child’s own production. In the third condition, (AM,) the first consonant of the correct word form was replaced by a fricative. If the word started with a vowel, a fricative was added in the initial position. The selection of fricative was in accordance to the child’s consonant inventory, avoiding those fricatives that the child herself produced. Furthermore, the resulting word form was not to refer to any existing object.

**Visual stimuli**

During the experiment (step 4) in each trial of the three conditions, two pictures were shown. One depicted the known, familiar object that the child had named in the recording session and was chosen to be included in the experiment based on the four criteria described above. All of the familiar pictures were photographs of actual objects that the child named during the recording session and was familiar with. The other picture depicted an odd, novel object. The pictures of the odd objects were mainly taken from a company that manufactures, among other things, office supplies (for an example of odd objects, see figure 6.1).\(^\text{26}\) To ensure that the objects were unfamiliar to the children in some of the photographs, a detail was removed, e.g. a handle bar. Prior to the actual experiment with children, photographs of 30 novel objects were presented to 13 Swedish adult speakers. They were asked to rate the familiarity of the objects presented on a scale from 1 (“I know exactly what this objects is”) to 7 (“I have no idea what this picture could be”). Furthermore, they were asked what each object reminded them of the most in order to get the most common associations. Out of these 30 pictures the 18 with the lowest familiarity ratings (below 2.3) and with no common associations were chosen. The 18 odd objects were presented together with the familiar objects in the three conditions Correct, CM and AM. In other words, during the experiment (step 4) in the conditions Correct, CM and AM, the familiar object was paired with a different odd object. All pictures of the familiar and odd objects were photographs presented on a gray-green background.

\(^{26}\)The German company Manufactum sells products for everyday life in high-quality materials. The pictures in the present study were mainly taken from the product category of office supplies and technical toys, and the author would like to thank the company for permission to use them in the study.
6. Study III: MAGIC Analyses Gaze In Children

Figure 6.1. Examples of odd objects presented in the conditions Correct, CM and AM together with a familiar object.27

Novel condition

In addition to the three conditions above (Correct, CM, AM), which were based on the child’s production, a fourth, novel condition was tested during the experiment (Novel), which was the same for every child. In this novel condition two pictures were presented, one depicting a presumably familiar object, e.g. a key, and one an unfamiliar object, e.g. a tweezer (see table 6.1). In the first experiments with three children, an odd, novel object was matched to a pseudoword, but since these children did not show the expected reaction (a longer looking time towards the unfamiliar object) in this condition, we decided to use low-frequency words instead, following White and Morgan (2008). The three children are not included in the final analysis. Notably, the objects included in condition Novel were unrelated to the child’s production and were thus not included in any other condition.

The frequency of the unfamiliar objects was controlled by the Korp tool.28 The frequency of the words referring to the unfamiliar objects did not exceed more than 3,300 occurrences. In contrast, for the familiar objects in this condition, norming data was taken into account. According to the norming data from CLEX (Jørgensen et al., 2010) the familiar objects were understood by at least 50% of children at 16

27 No familiar object is depicted here, but see figure 6.2 for a presentation of an odd object together with a familiar object.

28 The Korp tool summarizes several Swedish corpora. In this study nine corpora were included (Fin- landssvenska texter, Tidningstexter, texter, LaSSbarT - Läntläst svenska och barnbokstäver, SNP 78 70, SUC3.0, Svensk-författarförskrift, Svenska wikipedia (januari 2015)) and Textbanken (Bolin, Forsberg, & Roxendal, 2012) comprising 116,797,871 sentences. Bilingual corpora as well as corpora on forensic, medicine and political parties were excluded. Furthermore, the corpus of a Swedish hymnbook from 1937 was excluded. These exclusions were designed to obtain as realistic numbers as possible on frequencies for language in the environment of the children. The corpus LaSSbarT -Läntläst svenska och barnbokstäver, which is based on texts for children had only 105,000 sentences and was considered to be too small to provide a comprehensive picture of word frequencies. For the record, only the words pokal ‘trophée’ and matter ‘but’ had any occurrences (two each) in this corpus.
6.2. Methods

Table 6.1. Stimuli in the Novel condition. Only the unfamiliar object was labeled.

<table>
<thead>
<tr>
<th>Familiar object</th>
<th>Understand</th>
<th>Unfamiliar object</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>flaska ‘bottle’</td>
<td>64.3 %</td>
<td>spole ‘coil’</td>
<td>169</td>
</tr>
<tr>
<td>potta ‘potty’</td>
<td>54.1 %</td>
<td>kavel ‘rolling pin’</td>
<td>564</td>
</tr>
<tr>
<td>soffia ‘couch’</td>
<td>65.6 %</td>
<td>pokal ‘trophee’</td>
<td>3273</td>
</tr>
<tr>
<td>borsie ‘brush’</td>
<td>65.0 %</td>
<td>mutter ‘nut’</td>
<td>1545</td>
</tr>
<tr>
<td>kudde ‘pillow’</td>
<td>64.3 %</td>
<td>tuba ‘tuba’</td>
<td>615</td>
</tr>
<tr>
<td>nyckel ‘key’</td>
<td>64.3 %</td>
<td>pincett ‘tweezer’</td>
<td>1136</td>
</tr>
</tbody>
</table>

Note. The percentage of children understanding the familiar word is based on norm data for 16-month-old children (Jørgensen et al., 2010). The frequencies of the unfamiliar words are taken from Korp (Borin et al., 2012) reflecting the number of occurrences in 116,797,871 sentences.

months. A similar procedure of familiar and rare words is used, for example, in DePaolis et al. (2016). Since the children in this study were older than the children in the norming data, the familiar objects were assumed to be known by the children. Only the novel object was named in this condition during the experiment (step 4). As in the three other conditions, the objects were presented as photographs with a gray-green background.

All words in all conditions were recorded while spoken by a female Swedish speaker in a child-directed manner in an anechoic chamber in the Phonetics Lab at Stockholm University.

Testing the child’s perception in an eye-tracking experiment (step 4)

After the recordings in the home environment (step 2) and the stimulus design (step 3), the participants were asked to come to the Phonetics Lab at Stockholm University to participate in an eye-tracking experiment. The time between the recording of the child’s speech and the eye-tracking experiment was approximately one week (median days 7, 25th percentile 6 and 75th percentile 13). As mentioned above, every child had her own experiment since the stimuli were based on the child’s own production. Nevertheless, the structure of the experiment was always the same. The child sat on the parent’s lap (except for two children who preferred to sit alone) at 60 cm distance to the eye-tracking screen (Tobii T120). Before the experiment started, the child’s gaze was calibrated at five points. During calibration, parents were asked to close their eyes or to turn their heads to the side. During the experiments, parents wore headphones and listened to music. The experiment was presented with E-Prime 2.8, and gaze was recorded with Tobii Studio 3.2. In each
trial two pictures were presented for 2.5 s at the left and at the right side of the screen. One of the pictures depicted a familiar object and the other one was odd in the conditions Correct, CM and AM. In the Novel condition one picture was familiar and the other one was unfamiliar. After 2.5 s an attention-getter in the form of a smiley face occurred in the middle of the screen in between those two pictures. The two pictures remained in place. As soon as the child directed her gaze to the attention-getter a green frame occurred around it. If the child fixated on the attention-getter for 500 ms, the auditory stimulus was presented and the attention-getter disappeared while the two pictures on the right and left remained for another three seconds. Figure 6.2 gives an example of the timeline of one trial in the correct condition with the familiar object on the left side of the screen and the odd object on the right. The time until the auditory stimulus is presented is called the baseline. Since the time point of auditory stimulus presentation is depending on the child’s fixation on the attention-getter, the duration of the baseline is variable, but lasts a minimum of three seconds. The test phase starts with word onset of the auditory stimulus and lasts for three seconds. If the child did not fixate on the attention-getter at all, the experiment continued after five seconds by presenting the auditory stimulus. The trials without a fixation on the attention-getter around onset were disregarded in the final analysis.
6.3. Analysis

A total of 24 different trials (six in each condition Correct, CM, AM and Novel) were included in the experiment. The trials were repeated three times in the Correct condition, whereas trials in the conditions CM, AM and Novel were repeated once over the course of the experiment. This results in 54 trials in total (18 trials in the Correct condition and 12 each in CM, AM and Novel (cf. White & Morgan, 2008, who also had a higher number of trials in the correct condition than in the other conditions). Before any repetition in the conditions CM, AM and Novel occurred, all the stimuli had been presented exactly once. The side on which the familiar and unfamiliar object occurred was randomly chosen by E-prime. The order of the trials was pseudo-randomized with a maximum of two repetitions of the same condition. Additionally, no images were repeated on consecutive trials.

The experiment started with a short animated clip to get the child’s attention. Subsequently, after every second trial an animated clip was shown to get the child’s attention and keep it focused while looking towards the screen. The clips lasted for four seconds and included comic scenes like a bouncing ball or a sailing ship. They included no speech sounds, but did sometimes include music or short vocalizations like snoring. The experiment lasted a minimum of 10 minutes; the actual time was dependent upon the child’s attention and looking towards the screen. If the child’s interest in the experiment declined, the experiment was interrupted after the first half of the experiment was presented, ensuring that the child had the chance to see each stimulus at least once.

6.3. Analysis

Out of the 26 families that registered, 16 had to be excluded for a variety of reasons. One of the registered families had a language in the home environment other than Swedish. Four dropped out during step 1 and did not fill in the questionnaire on vocabulary size, and one did not show up for the last step, the eye-tracking experiment. The first three participants completed the version of the test with pseudowords rather than low-frequency words in the Novel condition, as mentioned above, and were thus excluded from the final analysis. Of the remaining 20 participants, 3 were excluded because of technical problems (i.e. calibration difficulties), and 2 were excluded for continually leaping out of their parents’ laps and walking around instead of watching the screen. This led to an interruption of the experiment before half of the trials were presented. Of the remaining 14 participants, only those who had a minimum of two trials in each condition where they fixated on the attention-getter at word stimulus onset were included in the final eye-tracking anal-
6. Study III: MAGIC Analyses Gaze In Children

Table 6.2. Overview of the 10 participants included in the final analysis. F = female; M = male. Age is given in years and months. SECDI = Swedish Early Communicative Development Inventory; the raw score refers to the words produced. Types and Tokens refer to the productions during the recording session.

<table>
<thead>
<tr>
<th>Subject [Age]</th>
<th>SECDI raw score (%tile)</th>
<th>Types/ Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anton [2:0] (M)</td>
<td>513 (85th)</td>
<td>103/352</td>
</tr>
<tr>
<td>Elsa [1:10] (F)</td>
<td>304 (80th)</td>
<td>75/220</td>
</tr>
<tr>
<td>Hans [2:0] (M)</td>
<td>118 (30th)</td>
<td>38/118</td>
</tr>
<tr>
<td>Inger [2:0] (F)</td>
<td>443 (80th)</td>
<td>85/294</td>
</tr>
<tr>
<td>Isak [2:0] (M)</td>
<td>520 (85th)</td>
<td>94/220</td>
</tr>
<tr>
<td>Kerstin [2:0] (F)</td>
<td>211 (45th)</td>
<td>55/112</td>
</tr>
<tr>
<td>Linus [2:5] (M)</td>
<td>435 (-)</td>
<td>118/495</td>
</tr>
<tr>
<td>Majken [2:0] (F)</td>
<td>319 (65th)</td>
<td>80/220</td>
</tr>
<tr>
<td>Sara [1:11] (F)</td>
<td>388 (85th)</td>
<td>102/226</td>
</tr>
<tr>
<td>Tilda [1:10] (F)</td>
<td>219 (65th)</td>
<td>55/153</td>
</tr>
</tbody>
</table>

("-" no percentile is reported since norming data exists only until the age of 2;4.

ysis. This resulted in 10 participants between the ages of 1;10 and 2;5 (M = 2.2; four boys), see table 6.2. The vocabulary size examined by Swedish Early Communicative Development Inventory (SECDI) (Berglund & Eriksson, 2000) ranged from 118 to 513 words (M = 347). The number of tokens produced during the recording session ranged from 112 to 495 and the number of types from 38 to 118.

For every child six words from the recording session were chosen to be included in the eye-tracking experiment (see table 6.3 for an example of two children and table B.2 in Appendix B for a detailed overview of the phonological capacities for each of the children and the individually chosen words for the experiment in the conditions Correct, CM and AM). The target words included both tonal accents I and II. The stress pattern of the target words involved both trochaic (påron ‘pear’) and iambic (giraff ‘giraffe’) word forms, word-final stress (krokodil ‘crocodile’) as well as monosyllabic words (sko ‘shoe’). The child’s own mispronunciation, and consequently word forms included in condition CM, involved changes in all word positions, for instance, initial: [fjyplΛn] flygplan ‘airplane’; medial: [rk:ɔ:l] ekkore ‘squirrel’ or final: [mɔ:m] mona ‘carrot’. Additionally, major changes in the form of syllable reductions, e.g. [kæNgU] känguru ‘kangaroo’, syllable additions, e.g. [rkojɔ:n] ekkore ‘squirrel’ or metathesis, e.g. [mɔ:ɔ] mossa ‘hat’, were included.
Table 6.3. Examples of auditory stimuli in the three conditions based on the child’s own production.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Target word</th>
<th>Correct</th>
<th>CM</th>
<th>AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elsa</td>
<td>giraffen 'the giraffe'</td>
<td>[fjrafan]</td>
<td>[fafan]</td>
<td>[srrafan]</td>
</tr>
<tr>
<td></td>
<td>dockan 'doll'</td>
<td>[dokan]</td>
<td>[gokan]</td>
<td>[fkan]</td>
</tr>
<tr>
<td></td>
<td>flygplan ‘airplane’</td>
<td>[flygplan]</td>
<td>[blfplan]</td>
<td>[slfplan]</td>
</tr>
<tr>
<td></td>
<td>morot ‘carrot’</td>
<td>[nrut]</td>
<td>[nujut]</td>
<td>[rut]</td>
</tr>
<tr>
<td></td>
<td>päron ‘pear’</td>
<td>[parrn]</td>
<td>[pren]</td>
<td>[rdn]</td>
</tr>
<tr>
<td></td>
<td>strumpa ‘sock’</td>
<td>[strompa]</td>
<td>[dompr]</td>
<td>[compa]</td>
</tr>
<tr>
<td>Kerstin</td>
<td>krokodil ‘crocodile’</td>
<td>[krkodil]</td>
<td>[dkodil]</td>
<td>[skodil]</td>
</tr>
<tr>
<td></td>
<td>morot ‘carrot’</td>
<td>[nrut]</td>
<td>[nujut]</td>
<td>[rut]</td>
</tr>
<tr>
<td></td>
<td>mössa ‘cap’</td>
<td>[morsa]</td>
<td>[nrrna]</td>
<td>[forsa]</td>
</tr>
<tr>
<td></td>
<td>sko ‘shoe’</td>
<td>[sku:]</td>
<td>[tu:]</td>
<td>[cu:]</td>
</tr>
<tr>
<td></td>
<td>strumpa ‘sock’</td>
<td>[strompa]</td>
<td>[bompa]</td>
<td>[fompa]</td>
</tr>
<tr>
<td></td>
<td>tröja ‘pulllover’</td>
<td>[troeja]</td>
<td>[terja]</td>
<td>[tcej]</td>
</tr>
</tbody>
</table>

Six children showed use of phonological templates: Elsa, Hans, Kerstin, Majken, Sara and Tilda, and their vocabulary size ranged from 118 to 388 words. Elsa showed the use of a template in the form of consonant harmony as in [fafan] giraffen ‘the giraffe’ or [krkr] täcke ‘blanket’. Hans used a template in the form of CVCV(C), as in [dujut] morot ‘carrot’ or [didun] flygplan ‘airplane’. Kerstin used a form of consonant harmony, as in [data] äta ‘eat’. Similarly, Majken showed consonant harmony as a template (e.g. [bopa] strumpa ‘sock’) on the one hand, and additionally CVVC, as in [mucu] mössa ‘cap’. Sara had two different templates: consonant harmony, as in [fyflut] flygplan ‘airplane’, or [papa] tappa ‘loose’, and CVCV(C) [kuick] trasig ‘broken’ or [sloba] klubba ‘lollipop’. Tilda adapted word forms to a CVCV structure, e.g. [batu] elefant ‘elephant’. The remaining four children showed segmental errors in the form of gliding and reduction of consonant clusters as well as the lateralization of /l/. All word positions were affected by segmental errors. The vocabulary size ranged between 435 and 520 words.

6.4. Results

For each of the objects on the screen an area of interest was defined. The gaze behavior was calculated based on the overall looking time in seconds towards each
area of interest including the objects during baseline and test phase, respectively. The raw scores for looking time for every trial were extracted from Tobii Studio and imported to SPSS v24 for further analysis. If the child did not look at the area of interest, Tobii Studio defined it as missing data. Since it was ensured that the child looked at the screen at word onset, these data points were treated as zeros in the subsequent analysis in SPSS.

In order to control for any general bias towards objects, a t-test for independent samples was conducted on the looking times towards the familiar and unfamiliar object during baseline. No significant difference was observed. For the test phase, a factorial ANOVA was conducted to examine the influence of the two independent variables CONDITION and OBJECT and the interaction effect of CONDITION and OBJECT on looking time (dependent variable). CONDITION included four levels (Correct, CM, AM, Novel) and OBJECT consisted of two levels (familiar, unfamiliar). No main effect of CONDITION was found. However, the main effect for OBJECT yielded an F ratio of $F(1, 514) = 6.600, p = .01$ indicating a difference between looking times towards the familiar and unfamiliar object. Furthermore, the model showed a significant interaction between OBJECT and CONDITION ($F(3, 514) = 4.793, p = .003$). The planned contrasts revealed a significant difference in looking time towards the familiar object in the conditions Correct ($p < .012$) and CM ($p < .001$). No significant difference was observed between the familiar versus unfamiliar object in AM and Novel. However, in descriptive terms children seem to prefer the unfamiliar object in the Novel condition (see figure 6.3).

Zooming in to the looking behavior of the individual children, a considerable variability between the participants as well as conditions can be detected (see figure 6.4). For instance, in the Correct condition, the looking times towards the familiar and unfamiliar object for Sara and Tilda vary in terms of a smaller difference for Sara and a clear preference for the familiar object for Tilda. All children except two (Elsa and Kerstin) showed a longer average looking time towards the familiar object in the Correct condition. Eight children followed the expected pattern of a longer looking time towards the familiar object in CM (all but Linus and Tilda). In the condition AM, seven showed the expected longer looking time towards the unfamiliar object, in the Novel condition.

Those children who used a phonological template during the recording session (Elsa, Hans, Kerstin, Majken, Sara and Tilda) showed a longer looking time towards the unfamiliar object in AM, although the difference is sometimes small (i.e. Elsa, Hans and Kerstin). In the condition Correct only Elsa and Kerstin showed a longer looking time towards the unfamiliar object; in CM only Tilda
6.4. Results

Figure 6.3. Mean looking times towards familiar and unfamiliar object in the four different conditions. Error bars display the 95% confidence interval. Correct = Correctly pronounced, CM = Child’s mispronunciation, AM = Atypical mispronunciation, Novel = Novel word was named.
6. Study III: MAGIC Analyses Gaze In Children

Figure 6.4. Mean looking times towards familiar and unfamiliar object in the four different conditions for every participant. Participants with a solid frame used templates in the recording session. Participants with no frame showed only segmental errors during the recording session.

showed the unexpected pattern in terms of a longer looking time towards the unfamiliar object. In the Novel condition only Hans had a longer looking time towards the familiar object. Additionally, he had a preference for the familiar object in all conditions, except for AM, where he looked for almost the same amount of time to each of the objects. In sum, no general tendency could be determined for those children who used a phonological template during the recording session.

6.5. Discussion

This study examined whether an individual child’s word production influences the word recognition of a mispronounced word that corresponds to the child’s own pronunciation. In the perceptual experiment, which was based on individually chosen words, four different conditions were tested. The hypotheses for the different conditions were as follows: First, the child would recognize the correct word form easily, which would be reflected in longer looking time towards the familiar object. Second, a similar reaction was expected for the word forms that correspond
to their own pronunciation (CM) according to the assumption of a ‘holistic’ word representation. Third, for the AM condition, including word manipulations that are uncommon for early word production, either no preference for any of the objects or significantly longer looking time towards the unfamiliar object was expected. Fourth, in the Novel condition, where an infrequent and presumably unfamiliar object was named, a longer looking time towards the unfamiliar object was expected following the constraint of ‘mutual exclusivity’, the tendency of mapping novel words to novel objects.

The results from the present study support the hypotheses proposed for the conditions Correct and CM (1-2). The children looked significantly longer towards the familiar object in both conditions. This implies that children accept those phonological manipulations that correspond to their own pronunciations as word candidates as predicted by a holistic word representation (Vihman, 2014, 2017b; Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013a). In contrast, reactions in condition AM do not show a statistically significant difference between the looking times towards one of the objects, indicating that the word manipulation is not experienced as a clear word candidate to the subjects (hypothesis 3).

In an earlier study, the distance of the mispronunciations to the adult word form led to decreased looking time towards the familiar object (White & Morgan, 2008), which seems not to be the case in the present study, given that CM word forms were accepted but not AM word forms. This might be surprising given that the word form changes in condition CM included, for instance, metathesis as well as syllable additions, resulting in word forms that might not be clearly identifiable at first sight (see, for example, Tilda’s [bata] for elefant ‘elephant’). However, the rejection of word-initial changes as word candidates in condition AM might be due to the importance of the initial consonant. Language unfolds over time and the first syllable is an anchor in word recognition, especially if it is stressed (cf. Fernald et al., 2001). If the first syllable neither matches the adult nor the child’s own word form, word recognition fails. Earlier studies showed that changes in initial position led to a rejection of the manipulation as a word candidate (Bailey & Plunkett, 2002; Swingley & Aslin, 2002; White & Morgan, 2008). Others showed that a failure in the recognition of words depends on the position of the change within the word (Swingley, 2005) or the stress pattern of the target language (Vihman et al., 2004). In the present study, the individually chosen words for every child included

29The manipulations can only be claimed to be corresponding, since all auditory stimuli were articulated by a female adult speaker in a child-directed manner. Consequently, they reflect the researchers’ perception of what the child said, rather than the child’s perception of her own speech (cf. Strömbärgsson, 2013).
monosyllabic (e.g. *s*ko 'shoe') and multisyllabic words (e.g. *e*lefant 'elephant'). The multisyllabic words involved both trochaic (e.g. *mössa* 'cap'), iambic (e.g. *giraff* 'giraffe') as well as anapest stress patterns (e.g. *krokodil* 'crocodile'), which implies that the manipulations in the condition *AM* also affected the unstressed syllable (e.g. *sokodil* or *siraff*). French infants at 11 months failed to show a preference for familiar words when the onset consonant was omitted in bisyllabic, iambic words (Hallé & de Boysson-Bardies, 1996). The result in condition *AM* could also indicate that the child does not know how to process the information and that neither of the objects seems for the child to be a good candidate for the mispronounced word.

The rejection as a word candidate of a word form that is neither a match to the child’s own production nor to the target form is, as mentioned above, compatible with the assumption that word forms are stored ‘holistically’. The holistic representation comprises the child’s knowledge of speech sounds and auditory characteristics, sensory and kinesthetic information derived from the child’s own production as well as auditory and visual information from input speech (cf. Vihman et al., 2009, 1994). Such an assumption is not necessarily in contrast to the results supporting phonologically detailed representations (Swingley, 2003; Swingley & Aslin, 2000, 2002). Rather, it could be assumed that the target word forms and the word forms based on the child’s production are stored in parallel (cf. Vihman, 2017b). As motor control improves and phonological abilities increase, the lexical representation gravitates towards a more fully specified target form. Even though the results in general support a more holistic representation of words, it remains unclear just where the boundary for acceptance as a word candidate lies. Possibly, the word forms included in the *AM* condition lie outside the frame of a holistic representation. Even in the present study, some of the word forms tested in the *CM* condition might be at the borderline of the holistic word form frame, as suggested by the individual results that did not follow our prediction (e.g. Linus and Tilda). Future studies could further examine boundaries of lexical representations by testing word forms that fall between the child’s own production and the target word form as well as *AM* word forms.

The perceptual reactions of the individual children reveal variability between the participants as well as between conditions. This is in line with studies based on word production, which showed variation in children’s production abilities (Menn & Vihman, 2011; Vihman et al., 1986, 1985). Six children showed the use of a phonological template, and four children showed only the use of segmental errors. The results from the perceptual experiment connected to the use of phonological templates did not display a clear pattern, which emphasizes individual variance
6.5. Discussion

between the children. The question as to whether the use of templates has an impact on perception in comparison to segmental errors needs further exploration in future studies.

The reactions in the Novel condition revealed no preference for any of the objects, as in some word-learning studies where children did not respond according to a ‘mutual exclusivity’ principle in word recognition tasks (cf. Swingley, 2007; Swingley & Aslin, 2007). This is in contrast to the findings of White and Morgan (2008), where children showed a ‘mutual exclusivity’ response. Nevertheless, the present study as a whole differs from White and Morgan (2008) in some respects. First of all, the children in the present study were chosen by vocabulary size, in terms of having at least 100 words in the productive vocabulary according to their parents, rather than by age. White and Morgan used feature changes to the initial phoneme as phonological mispronunciations, while in the present study the stimuli of the other conditions was tailored to the production patterns of the individual child. The Novel condition differed from the other conditions in the present study in terms of the presentations of two completely new pictures, pictures that the children had not seen before. This might have led to a surprise response by the child. In White and Morgan (2008) both object pictures were involved in the other conditions. Thus the occurrence of the two pictures was probably not as surprising or attention-attracting as in the present study. However, despite the lack of statistical significance in the Novel condition in the present study, in descriptive terms the children tended to look longer towards the unfamiliar object. Further studies are needed to evaluate whether the constraint of ‘mutual exclusivity’ will appear in a larger participant group and whether confounding variables such as the relative novelty of the pictures affect the results.

In the present experiment, the children covered a larger age range, namely 1;10–2;5, larger than observed in typical cross-sectional studies. Language acquisition per definition describes an ever-changing individual developmental path. Usually, the individual differences are in particular apparent in production studies (Vihman, 1992). In order to increase the probability of observing phonological templates, the present study only included children with a vocabulary size above 100 words. This also ensured that sufficient amount of adult-like word forms would be captured, which in turn allowed tailoring auditory stimuli in the different conditions for the individual child. The inclusion criterion leads to a larger age range, which was considered a price worth paying in order to capture the individual productive phonological capacities of each child and to use that information in the perception experiment.
6. Study III: MAGIC Analyses Gaze In Children

In general, the results of this study and especially the children’s responses in the perceptual experiment in condition CM support the view of a holistic representation of early word forms and an influence of production on perception. However, there was no clear evidence that the use of templates influences the recognition of words in an experimental situation. The responses in AM showed that not all mispronunciations are accepted, strengthening the importance of word initial segments since this was the only manipulated position in that condition. In the Novel condition no ‘mutual exclusivity’ response was observed. Further studies are needed to explore how the child’s own phonological patterns influence the recognition of systematically varying word forms.
7. General discussion and conclusion

The relationship between speech production and perception in the early stages of phonological and lexical acquisition has so far not received a lot of attention in experimental studies. The child’s acquisition of phonological and lexical characteristics develop in an intertwined and, from the child’s perspective, inseparable way. Moreover, the child’s speech production and perceptual capacities are reflected in those developing domains.

From a theoretical perspective, models of phonological acquisition reflect differing and, to a certain extent, conflicting approaches explaining how children build up a phonological system as well as a lexicon and integrate production and perceptual abilities. The differing theoretical approaches as well as the under-investigated relationship between perception and production raised research questions explored in the present thesis.

The first research question consists of two parts, namely, (1) do the child’s phonological and lexical capacities in production influence word recognition at all and (2) how do phonological and lexical capacities in production influence the child’s perception? The results of study III (ch. 6) indicate that the children’s speech production capacities do influence their word perception, and thus the first part can be answered in the affirmative. The second part of the question, how do phonological and lexical capacities in production influence a child’s perception of words, requires a more complex answer. Our results suggest that not all mispronunciations block word recognition. Rather, mispronunciations that are based on the child’s own production are accepted as word candidates. Earlier studies with stimuli without a connection to the child’s own production have shown that mispronunciations may hinder word recognition, which led to the assumption of phonologically detailed word representations (Swingley, 2003; Swingley & Aslin, 2000, 2002). At first sight, a holistic representation contradicts the assumption of phonological specificity, but assuming that a representation comprises the child’s experiences – including information about the child’s own production as well as information about the speech input – adult target forms and the child’s production
are stored in parallel (Vihman, 2017b). Future studies considering the child’s own production are needed to evaluate the phonological boundaries of lexical representations in terms of which word forms are accepted as word candidates, and how these lexical representations change over time. The answer to the first research question is in line with earlier studies on younger children, which indicate that the child’s own production influences perception (Altvater-Mackensen et al., 2016; DePaolis et al., 2016; Masapollo et al., 2016). Moreover, the present thesis expands the influence of production on perception to children around two years, who are in a very dynamic period of phonological and lexical development. Additionally, it supports theoretical approaches that assume a link between production and perception capacities, but were based – so far – on little experimental evidence, like for instance, the Linked-Attractor model (Menn et al., 2013), whole-word phonology (Vihman, 2014, 2017b; Vihman & Croft, 2007; Vihman et al., 2009; Vihman & Keren-Portnoy, 2013a), and to some extent, NLM-e (Kuhl et al., 2008).

The second question concerning setting and combination of research methods might be appropriate for investigating phonological and lexical knowledge in speech production and perception at different ages over the early word-learning period. In this thesis, several setups for speech recordings as well as EEG and eye-tracking and the combination of these were explored (chs. 4 and 5). For speech recordings, video as well as audio recordings in the home environment in a play and reading situation resulted in the most usable speech data, since they made it easier to elicit a sufficient amount of speech. Hence those situations formed the basis for the investigation of productive phonological capacities in chapter 6. The audio recordings made it possible to analyze the speech production in terms of phonetic, phonological and lexical characteristics, as suggested by Li et al. (2009) and Kehoe et al. (1995). The value of using EEG and thereby adding information about processing capabilities needs to be balanced against the age of the participants and the effort of testing a large number of children in combination with the time-consuming production data assessment (cf. ch. 4).

Eye-tracking proved to be a valid method for examining word recognition in different age groups, as it has been used in earlier studies (e.g. Bergelson & Swingley, 2012; White & Morgan, 2008). The time spent gazing at familiar and unfamiliar objects in the different conditions can serve as an index of matching word forms to objects and word form interpretation (cf. Swingley, 2016). Eye-tracking allows a detailed description and analysis of the child’s looking behavior (e.g. fixation time, duration of longest fixation or pupil size changes), which constitutes an obvious advantage relative to a head-turn procedure, for instance. Additionally, the advantage of the automatic measurement of looking time provided by eye-tracking...
is less time-consuming and minimizes potential influence on the experimental procedure by the researcher. The combination of EEG and eye-tracking is useful if a simultaneous recording of behavioral data as well as the assessment of brain responses is desirable, considering that a neuronal response precedes a behavioral one. For instance, it could be used in studies exploring the reaction to mispronunciations reflected pupil size changes (Tamási et al., 2017) as well as brain responses, using the same paradigm for both measurements (for a study with adults see, for example, Kuipers & Thierry, 2011). An obvious advantage of combining the methods is a deeper analysis of the child’s processing capability and a validation of the observations made in either measurement. Future studies are needed to evaluate the combination of these two methods in the research of infants and toddlers.

The third and fourth research questions concerned the phonological capacities of Swedish toddlers in relation to vocabulary size. The results from studies II (ch. 5) and III (ch. 6), indicate that, in general, Swedish children show the same kind of phonological templates and segmental errors as those that have been reported in other languages. Forms adapted to templates were observed in some children when they had greater than 100 words in their vocabulary, suggesting a relationship between vocabulary size and the phonological capacities of Swedish toddlers. As reported in earlier production studies (e.g. Vihman et al., 1986), individual differences were observed, for example, a template involving sibilants and a following pause, or the phonological process of H-zation, which was described earlier mostly for older Swedish children with phonological disorders (Nettelbladt, 1983, 2007a). The results in the present thesis are based on a substantially larger sample than earlier studies and thus expand earlier studies on Swedish typically developing children (Lohmander et al., 2011; Nettelbladt, 1983) by providing a detailed description of the phonological capacities of typically developing children (study II, ch. 5, and III, ch. 6). Moreover, this thesis adds evidence of Swedish children’s use of phonological templates. The phonological capacities measured by consonant inventory, number of words produced correctly, and phonological processes and templates were related to vocabulary size, supporting earlier studies on the relationship between phonology and the lexicon (cf. Stoel-Gammon, 2011). The data in this thesis offer a valuable basis for future studies evaluating different phonological and lexical characteristics in Swedish children.

The main result of this thesis suggests an influence of the individual child’s production on perception in relation to phonological and lexical capacities. Thus lexical representations are not only phonologically specified but word forms are accepted that are related to the child’s own production. Additionally, the child’s phonological capacities are related to vocabulary size. In the early stages of vocab-
7. General discussion and conclusion

Vocabulary development the word pronunciations tended to be more precise; the children with a larger vocabulary size tended to produce words with less accuracy. The use of phonological templates was more likely to be observed in children with a vocabulary of more than 100 words.

7.1. Limitations

Some limitations should be considered when interpreting the findings reported in this thesis. The first concern relates to the sample sizes in the different studies. Study I examined methodological advances in the combination of EEG and eye-tracking. In general, EEG studies with children include a large number of participants, given the high attrition rate. Still, although the results from study I are limited, the main findings of this work can be supplemented with studies II and III. Study II examined the phonological capacities of Swedish children in relation to vocabulary size, and 12 children were included. However, production studies often include smaller sample sizes in comparison to perception studies (cf. Vihman et al., 1986; Vihman & Velleman, 2000, who had 10 and 5 participants per group), due to the time-consuming procedures required for transcriptions of speech materials. Study III follows up on studies I and II. Hence in study III a relatively complex design was chosen that included several steps to examine the interaction between production and perception. The sample consisted of 10 children due to the necessity of carefully tailoring visual and auditory stimuli to the individual child.

A further concern involves the choice of methods. Studies relying on the gaze behavior of children have previously been used in a variety of studies examining, for example, reactions to phonological manipulations. The measurements of brain responses (ERP) have been used to investigate lexical retrieval, and thus the combination of those methods was considered likely to be useful for research on the lexicon and phonology, despite the lack of precedent from earlier studies. The combination of these technically advanced methods in connection with the time-consuming analysis of the child’s production capacities was challenging; the relatively limited results could be regarded as a price to pay for the effort of exploring new combinations of techniques. In the end, the combination of those methods was set aside and eye-tracking was retained since more participants were cooperative and comfortable with that method. The results regarding the influence of production on perception are therefore based on eye-tracking data, which was, as shown in earlier studies, a reliable and established method for child language research. Still, EEG data might give further insights into lexical retrieval and a combination...
of these two methods should be considered in future research on the development of phonology and the lexicon, especially in young infants.

The measurement of vocabulary size was based on parental reports such as SECDI (Berglund & Eriksson, 2000). Other production studies, especially those with a longitudinal design, have often defined word points, such as a 25-word point identified when the child produced 25 words spontaneously in a recording session. The studies included in the present work had a cross-sectional design. However, there are many advantages in longitudinal designs, including further control over vocabulary size changes during development, for instance by controlling words produced in one recording session. Hence future studies on the relationship between the acquisition of phonology and the lexicon might consider several recordings to define specific word points (as in earlier studies e.g. Vihman et al., 1986, 1985) to further examine developmental aspects in both phonological as well as lexical acquisition.

Another limitation involves the selection of word stimuli in all of the studies. Earlier studies showed an impact of phonological neighborhood on word learning (Storkel, 2002, 2004). Additionally, Swedish has two lexical accents, which may influence the acquisition of phonology and lexicon, much as geminates influence word learning in Italian children (Vihman & Majorano, 2016). However, the selection of stimuli was based on the child’s production (ch. 6), or what was assumed that the child was already saying according to parental reports (ch. 4). Hence, the stimuli consisted of words varying in phonological neighborhood, syllable number, stress pattern as well as word accent. As a consequence, those aspects constitute potential confounding factors. Similarly, besides phonological neighbors, other finer phonological characteristics were not varied systematically, since the aim of the present thesis was to examine the influence of production on perception in general, and the selection of naturalistic stimuli was prioritized.

### 7.2. Future directions

Even though this thesis has generated some answers, several future directions can be identified. Evidence for the influence of production on perception in relation to phonological and lexical acquisition needs replication in different developmental stages to further establish how such influence may be characterized. Further investigation is needed regarding how the child’s phonological and lexical capacities shape word form representations. In particular, future studies could investigate which manipulated word forms are accepted as word candidates for the individ-
ual child. For instance, if the child produces ‘elephant’ as [batɔ], a manipulation of that word form could result in [bafɔ] for medial changes, or [fatɔ] for initial changes. Another type of stimulus material could involve mispronunciations that are in-between the child’s own production and the target word form; for instance, if the child says [řkɔjɔnɔ] instead of [řkɔrɔnɔ] for ekorre ‘squirrel’, a manipulation could be [řkɔrɔnɔ]. Such an experiment could illustrate how robust the word form representations are and which parts of the words involve a more specific representation. Moreover, such experiments could be extended to clinical groups and would give us further insights into phonological representations in atypical language acquisition. Intervention studies that take the influence of the individual child’s speech production on her speech perception into account could contribute to pedagogical considerations.

Another future direction could focus on the use of templates and their assumed influence on perception, preferably in a longitudinal design since templates are ephemeral and occur in different time windows. A longitudinal design would provide the possibility of capturing the dynamic characteristics of phonological templates and allow testing their influence on perception in a manner closely timed to developmental patterns in individual children. Furthermore, a longitudinal study would allow an additional estimate of vocabulary size, given the possible identification of word points. Moreover, the influence of Swedish word accents on the acquisition of words needs further investigation. In addition to the investigation of production capacities, a longitudinal design might be considered to study how word form representations change over time. Such an investigation would also make it possible to answer questions on the interaction between production and perception in terms of direction of influence during phonological and lexical development.

Future studies might also reconsider the use of ERP recordings in order to explore individual perceptual knowledge in relation to production capacities. For instance, an MMN paradigm could be used to contrast word forms with different phonological characteristics, based on the child’s production. Investigating the relationship between phonology and lexicon with ERP recordings would entail further insights into how word forms are represented in the mental lexicon.
7.3. Conclusion

Language acquisition depends on a complex interplay between various factors related to speech production and perceptual capacities of the individual child as well as to a variety of influences from the ambient language and the environment as a whole. The disentangling of various factors is the topic of the present work and future studies. Concretely, the present thesis builds a first step for bridging the gap between speech production and perception in the early stages of phonological and lexical acquisition. The results showed that children accept mispronounced word forms as appropriate word candidates when the word forms are related to their own specific word production. Furthermore, phonological capacities are related to vocabulary size. The results also indicate an influence of the individual child’s production on word perception, as well as a relationship between phonological and lexical capacities. These insights contribute to linguistic theoretical approaches concerning abstract representations of phonological word forms and reveals a closer relationship between production and perceptual abilities in toddlers than has previously been demonstrated. The dynamics of language acquisition are in the process of being investigated and new insights are continuously emerging. The present thesis, with its research focus on characteristics of individual children’s language capacities as well as the exploration of new methods, has emphasized the importance of connecting production and perception in different linguistic domains in order to achieve a deeper understanding of how children acquire language.
Summary in Swedish –
Sammanfattning på svenska

I denna avhandling undersöks förhållandet mellan talproduktion och talperception i barnets tidiga fonologiska och lexikala utveckling. Syftet med avhandlingen är att närmare studera sambandet mellan den lexikala och fonologiska förmågan i form av talproduktion och talperception, och hur dessa områden ömsesidigt kan påverka varandra. Barnets individuella talproduktion relateras till dess perceptuella förmåga, något som är nytt inom barnspråksforskningen.

Följande forskningsfrågor studeras:

1. Påverkar barnets individuella fonologiska och lexikala produktionsförmågor dess ordperception i form av ordigenkänning? Om det sker en påverkan, vad kännetecknar denna?

2. Vilken uppsättning och kombination av forskningsmetoder kan vara lämpliga för att undersöka tidig fonologisk och lexikal förmåga, i talproduktion och talperception, med avseende på olika åldersgrupper?

3. Vilka fonologiska förmågor kan observeras i den individuella ordförmågan hos en grupp svenska småbarn?

4. Är ordförrådstorleken kopplad till den fonologiska förmågan hos en grupp svenska småbarn? Om så är fallet, vad känneteckna denna kopplingen?

I bakgrundskapitlen beskrivs olika metoder för att undersöka tidig talperception och talproduktion. Dessutom tas allmänna förutsättningar som påverkar barnets språkutveckling upp. Baserat på tidigare forskning om barnets orduppfattning och ordförmåga, beskrivs olika teorier och teoretiska modeller med kopplingar till den fonologiska och lexikala utvecklingen. Exempelvis har tidigare perceptionstudier visat att barns ordigenkänning kan hindras av vissa ordförändringar, även om manipulationen består av en liten förändring av ordförändringar. Detta har lett till antagandet att barnets ordförmåga omfattar detaljerad information om fonologiska egenskaper hos ordet i fråga. Å andra sidan har andra...
studier som undersökt barns talproduktion kommit fram till att barns ordrepresen-
tation kan vara mer holistisk. Detta innebär att uttalsvarianter baserade på barnets
egut uttal accepteras som ordkandidat, även om uttalsvarianten inte följer ordets
målform. Detta antagande tyder på ett samband mellan barnets egen talproduktion
och talperception. Inom antagandet om en holistisk ordrepresentation ingår även
begreppet fonologiska templates, något som i talproduktion kännetecknas av en
övergeneralisering från ett motoriskt väl övat talproduktionsmönster till nya ord.

Avhandlingen är organiserad kring tre studier. I studie I utforskas i tre exper-
iment metodologiska möjligheter så som kombination av EEG- och Eye-tracking,
något som inte genomförts tidigare inom barnspråksforskning. Försökspersoner i
de tre experimenten är vuxna samt barn vid 17 månaders och 24 månaders ålder.
Experimenonen med barn innehåller en kombination av EEG och Eye-tracking samt
inspelningar av barnets talproduktion. Experimentet med vuxna använder bara
Eye-tracking. Resultaten visar att vuxna talare upptäcker fonologiska ordforms-
manipulationer och anpassar sitt tittbeteende efter graden av manipulationerna.
Resultaten i experimenten med barn visar att barnens talproduktion varierar my-
cket, speciellt i den yngre barngruppen. Kombinationen EEG och Eye-tracking
visade sig tekniskt komplex med mycket brus i EEG-datan, speciellt i den äldre
barngruppen. Eye-tracking däremot var en metod som visade sig lämplig för att
undersöka barnets ordigenkänningsförmåga, vilket reflekteras i förhållandevis lågt
bortfall och resultat som generellt stämmer överens med tidigare forskning.

I studie II undersöks en grupp svenska småbarns fonologiska förmåga i rela-
tion till ordförrådsstorlek. Den fonologiska förmågan bedöms utifrån antal kon-
sonanter i initial och final ordposition, antal korrekt uttalade ord, uttalsvariationer
i termen av segmentella ändringar och fonologiska templates. Försökspersonerna
i denna studie bestod av tio 17-månaders gamla barn med varierande produktiv
ordförrådsstorlek mätt genom en standardiserad föräldrarenkät. Resultaten visar att
konsontinventariet ökar med ordförrådsstorlek. Barn med större ordförråd visar
mindre antal korrekt uttalade ord än barn med mindre ordförrådsstorlek. Utöver
detta visas att barn med över 100 ord i sitt produktiva ordförråd visar en större
sannolikhet att använda fonologiska templates.

I studie III studeras inflytandet av barnets egna fonologiska och lexikala för-
mågor i talproduktion i ett ordigenkännningsexperiment. Urvalet av barn i denna
studie är baserat på ordförrådsstorlek: alla barn hade en ordförrådsstorlek över 100
ord enligt en standardiserad föräldrarenkät. Innan det perceptuella experimentet
genomfördes, spelades barnets talproduktion in i barnets hemmiljö. På basis av
barnets ordproduktion utvaldes för varje barn sex individuella ord som testades i ett
Eye-tracking experiment. De sex individuella orden var underlag för auditiva stim-
7.3. Conclusion


Baserat på resultaten av de tre studierna kan forskningsfrågorna besvaras enligt följande: Den första forskningsfrågan huruvida barnets individuella fonologiska och lexikala förmågor påverkar dess ordperception i form av ordigenkänning besvaras främst i studie III. Resultaten i denna studie tyder på att barns egen talproduktionsförmåga påverkar ordigenkänning. Andra delen av forskningsfrågan, vad som kännetecknar påverkan, är komplex. Resultaten indikerar att inte alla ordmanipulationer behandlas lika: Bara de manipulationer som är baserade på barnets eget uttal accepteras som ord. Vidare studier behöver bland annat undersöka närmare var gränserna går för acceptans av olika typer av ordmanipulationer.

Den andra forskningsfrågan omfattar vilken uppsättning och kombination av forskningsmetoder som kan vara lämplig för att forska på tidig fonologisk och lexikal förmåga, i talproduktion och talperception, med avseende på olika åldersgrupper. Flera uppsättningar av undersökningsmetoder har använts och testats i studie I, II och III. När det gäller att studera barnets talproduktionsförmåga, gav kombinationen av ljud- och videoinspelningar i barnets hemmiljö samt en läs- och leksituation de bästa resultaten i form av kvantitet och kvalitet av talproduktionsdata och användes därmed i studie III. Eye-tracking visade sig vara en lämplig metod för att testa ordigenkänning i olika åldrar, eftersom resultaten bland annat föreföll rimliga jämfört med tidigare studier och många barn kunde fullfölja experimenten. Eye-tracking tillåter en detaljerad beskrivning och analys av barnets tittbeteende: Tittiden som barnen uppvisar när de tittar på bilderna på skärmen kan tolkas som index på kopplingen mellan ordformer och objekt. Kombinationen av EEG och Eye-tracking visade sig vara komplex och tekniskt utmanande men torde vara viktigt att utforska vidare, då en djupare analys av barnets bearbetningsförmåga undersöks.
7. General discussion and conclusion

Tredje och fjärde forskningsfrågorna gäller kopplingen mellan den fonologiska förmågan och ordförrädsstorleken hos svenska barn. Resultaten från studie II och III tyder på att svenska barn generellt sett visar liknande fonologiska processer och fonologiska templates som barn i andra språk. Utöver detta visar resultaten att antalet fonologiska processer och användandet av fonologiska templates är relaterade till ordförrädsstorleken, vilket i sin tur tyder på ett nära samband mellan lexikon och fonologi.

Framtida studier kan lämpligen undersöka förändringar av ordrepresentationer över tid, till exempel med en longitudinell studie-design. Studier som undersöker ordrepresentationernas gränser, det vill säga vilka olika ordformer som accepteras och vilka som inte accepteras, ger information om hur detaljerat eller holistiskt information om ord är representerat. Utöver detta behövs mer information om svenska barns tidiga talutveckling i form av användandet av fonologiska templates och om och hur, exempelvis, den svenska ordaccenten påverkar den fonologiska utvecklingen.

Barns språkutveckling undersöks i denna avhandling med avseende på fonologi och lexikon vad gäller talproduktion och talperception. Att koppla samman flera lingvistiska domäner med talproduktion och talperception är en viktig del i forskningen inom barns språkutveckling. Avhandlingen bidrar med ökad kunskap om barnets tal- och språktillägnan under det tidiga, dynamiska och betydelsefulla skedet av fonologisk och lexikal utveckling till teoretiska, lingvistiska antaganden och kan därmed bidra till pedagogisk och klinisk forskning med fokus på sambandet mellan talproduktion och talperception.
## A. Appendix study I

Table A.1. Overview of the elicited speech production data and perception and production of the included words in the experiment reported by the parents. - symbolizes missing data: the parents did not report the words and/or we were not able to elicit the words.

<table>
<thead>
<tr>
<th>Participant [Age] (Gender)</th>
<th>Parents report the child understands (number)</th>
<th>Parents report how those words sound</th>
<th>Pronunciation of the tested objects in the elicitation task</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN [17:25] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka (7)</td>
<td>mpa (lampa), uss (buss), ula (klocka)</td>
<td>[boo] buss</td>
</tr>
<tr>
<td>BU [19:28] (F)</td>
<td>lampa, boll, mössa, strumpa, sko, klocka, napp (7)</td>
<td>ampa (lampa), össa (mössa), sumpa (strumpa), klocka, app (napp)</td>
<td>[gga] klocka; [bobla] strumpa</td>
</tr>
<tr>
<td>CK [17:15] (F)</td>
<td>lampa, boll, mössa, strumpa, sko, napp (6)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DW [16:27] (M)</td>
<td>boll (1)</td>
<td>laampa (lampa), boll, ko (sko), na (napp)</td>
<td>-</td>
</tr>
<tr>
<td>EV [17:0] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>-</td>
<td>[gic] sko; [gga] strumpa</td>
</tr>
<tr>
<td>FU [17:4] (F)</td>
<td>lampa, boll, mössa, strumpa, sko (5)</td>
<td>prump (strumpa)</td>
<td>-</td>
</tr>
<tr>
<td>GT [17:8] (M)</td>
<td>lampa, boll, mössa, buss, sko, klocka (6)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HS [17:2] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>tumpa (strumpa), ko (sko)</td>
<td>[kal] bolli; [dædr] klocka; [ggi] guis; [maama] lampa; [bonpa] strumpa; [boo] buss</td>
</tr>
<tr>
<td>IM [18:0] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka (7)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table A.1 – continued from previous page

<table>
<thead>
<tr>
<th>Participant [Age] (Gender)</th>
<th>Parents report the child understands</th>
<th>Parents report how those words sound</th>
<th>Pronunciation of the tested objects in the elicitation task</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR [17;2] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, napp (8)</td>
<td>lampa, klocka, p(app) (napp)</td>
<td>-</td>
</tr>
<tr>
<td>JJ [18;14] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>ampa (lampa), buss (buss), kocka (klocka)</td>
<td>-</td>
</tr>
<tr>
<td>JQ [17;9] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (6)</td>
<td>wroom (buss)</td>
<td>-</td>
</tr>
<tr>
<td>KL [18;3] (F)</td>
<td>boll, strumpa (2)</td>
<td>[bo] buss; [buss] lampa;</td>
<td></td>
</tr>
<tr>
<td>KP [17;15] (M)</td>
<td>lampa, mössa, strumpa, sko, klocka, napp</td>
<td>bo (boll), buch (buss), o (sko)</td>
<td>[buss] [oun] boll; [kxo] skor; [maeheu] mössa; [age] boll</td>
</tr>
<tr>
<td>LF [17;29] (F)</td>
<td>lampa, boll, mössa, buss, sko, napp (6)</td>
<td>bollen, mässa (mössa), buss, kona (skorna)</td>
<td>[bompan] strumpa; [akol] boll; [g'wha] skorna; [tresin]; [abros] buss</td>
</tr>
<tr>
<td>LO [17;10] (F)</td>
<td>lampa, boll, strumpa, sko, klocka, napp (6)</td>
<td>mamma (napp)</td>
<td>-</td>
</tr>
<tr>
<td>MO [19;21] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>ampa (lampa), boll, buss, umpa (strumpa), ko (sko), napp</td>
<td>[bo'da] klocka</td>
</tr>
<tr>
<td>OH [18;8] (M)</td>
<td>lampa, boll, mössa, buss, sko, napp (6)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OL [17;29] (M)</td>
<td>lampa, boll, mössa, strumpa, sko, napp (6)</td>
<td>bapp (napp)</td>
<td>-</td>
</tr>
<tr>
<td>PK [17;26] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, napp (7)</td>
<td>mlå (boll), müttta (mössa), butt (buss), bumpa (strumpa), slola (skorna)</td>
<td>-</td>
</tr>
<tr>
<td>QT [17;14] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>kaka (klocka), tutte (napp)</td>
<td>[bo'da] klocka</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Participant [Age] (Gender)</th>
<th>Parents report the child understands</th>
<th>Parents report how those words sound</th>
<th>Pronunciation of the tested objects in the elicitation task</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV [17;14] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SA [17;27] (M)</td>
<td>lampa, boll, mössa, buss, sko (5)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SP [18;2] (M)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (8)</td>
<td>na (napp)</td>
<td>-</td>
</tr>
<tr>
<td>TE [17;27] (M)</td>
<td>lampa, boll, mössa, buss, sko (5)</td>
<td>-</td>
<td>[baʃa] napp</td>
</tr>
<tr>
<td>UL [17;18] (F)</td>
<td>lampa, boll, mössa, buss, strumpa, sko, klocka, napp (7)</td>
<td>napp</td>
<td>[dɔːʃa] klocka</td>
</tr>
<tr>
<td>VH [17;22] (F)*</td>
<td>lampa, boll, mössa, buss, strumpa, sko (6)</td>
<td>pa (lampa)</td>
<td>[duː] sko</td>
</tr>
<tr>
<td>WK [18;0] (F)</td>
<td>lampa, boll, mössa, strumpa, sko, klocka (6)</td>
<td>amp (lampa), oka (klocka)</td>
<td>[mpa] strumpa; [ga] klocka: [mpa] lampa</td>
</tr>
</tbody>
</table>
### B. Appendix study III

Table B.1. Objects used in the elicitation task in the four categories with phonetic transcriptions in standard Swedish and the percentage of children producing the Swedish word at 16 months in the norm data (Jørgensen et al., 2010). Words with (-) could not be found in the norm data, but are included because of their phonological complexity.

<table>
<thead>
<tr>
<th>Clothes</th>
<th>Toys</th>
<th>Animals</th>
<th>Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>sko [skuː] ‘shoe’ (26.8)</td>
<td>boll [boːl] ‘ball’ (30.6)</td>
<td>ekorre [ˈɛkːɔrɔ] ‘squirrel’ (3.8)</td>
<td>melon [ˈmeːlɔn] ‘melon’ (-)</td>
</tr>
<tr>
<td>strumpa [ˈstrʊmpa] ‘sock’ (-)</td>
<td>lampa [ˈlampa] ‘lamp’ (49.7)</td>
<td>giraff [ˈɡiːrɑf] ‘giraffe’ (1.3)</td>
<td>morot [ˈmʊrɔt] ‘carrot’ (3.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kyckling [ˈɛkvɛlɪŋ] ‘chicken’ (1.3)</td>
<td>äpple [ˈɛplɛ] ‘apple’ (19.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>känguru [ˈkɑŋɡʊru] ‘kangaroo’ (-)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>papegoja [ˈpɑːpɛɡɔja] ‘parrot’ (-)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pingvin [ˈpɪŋvɪn] ‘penguin’ (0)</td>
<td></td>
</tr>
</tbody>
</table>
Table B.2. Overview of the phonological capacities of the 10 participants and the word forms included in the eye-tracking experiment in the conditions Correct, CM and AM. Only word final-stress is marked.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Consonant inventory</th>
<th>Segmental errors</th>
<th>Phonological template</th>
<th>Words included in the ET-experiment in Correct/CM/AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anton</td>
<td>Onset: b d m h</td>
<td>Gliding, reduction of consonant clusters, omission of final consonants</td>
<td>no</td>
<td>[Ek:] [Or:] [Ek:] [Or:] 'squirrel'</td>
</tr>
<tr>
<td></td>
<td>Coda: n l k t f</td>
<td></td>
<td></td>
<td>[krokodil] [l] [k] krokodil 'crocodile'</td>
</tr>
<tr>
<td></td>
<td>f v t k l h</td>
<td></td>
<td></td>
<td>[murnad] [m] [murnad] morot 'carrot'</td>
</tr>
<tr>
<td></td>
<td>Coda: n t k f</td>
<td></td>
<td></td>
<td>[pæren] [næren] [cæren] pinon 'pear'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[skan] [g] [t] sko 'shoe'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[torja] [t] [t] [torja] morja 'pallorbev'</td>
</tr>
<tr>
<td>Elsa</td>
<td>Onset: b p m d</td>
<td>Gliding, reduction of consonant clusters</td>
<td></td>
<td>[Ek:] [Or:] [Ek:] [Ol:] 'dol'</td>
</tr>
<tr>
<td></td>
<td>f v t k l h</td>
<td></td>
<td></td>
<td>[flygplan] [v] [flygplan] flygplan 'airplane'</td>
</tr>
<tr>
<td></td>
<td>Coda: n t k f</td>
<td></td>
<td></td>
<td>[Tafan] [T] [Tafan] giraffen 'the giraffe'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[murnad] [m] [murnad] morot 'carrot'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[pæren] [næren] [cæren] pinon 'pear'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[stompa] [s] [stompa] stumma 'sock'</td>
</tr>
<tr>
<td>Hans</td>
<td>Onset: d j g</td>
<td>Omission of unstressed syllables, reduction of consonant clusters</td>
<td></td>
<td>[Ek:] [Or:] [Ek:] [Ol:] 'banana'</td>
</tr>
<tr>
<td></td>
<td>Coda: n t</td>
<td></td>
<td></td>
<td>[flygplan] [v] [flygplan] flygplan 'airplane'</td>
</tr>
<tr>
<td></td>
<td>f v t k l h</td>
<td></td>
<td></td>
<td>[Tafan] [T] [Tafan] giraffen 'the giraffe'</td>
</tr>
<tr>
<td></td>
<td>Coda: n t k f</td>
<td></td>
<td></td>
<td>[murnad] [m] [murnad] morot 'carrot'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[pæren] [næren] [cæren] pinon 'pear'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[stompa] [s] [stompa] stumma 'sock'</td>
</tr>
<tr>
<td>Inger</td>
<td>Onset: b d t k l m h</td>
<td>Gliding, assimilation, substitution of /s/ to /f/</td>
<td>no</td>
<td>[Ek:] [Or:] [Ek:] [Ol:] 'squirrel'</td>
</tr>
<tr>
<td></td>
<td>f v t k l h s v e s</td>
<td></td>
<td></td>
<td>[flygplan] [v] [flygplan] flygplan 'airplane'</td>
</tr>
<tr>
<td></td>
<td>Coda: n e t r f</td>
<td></td>
<td></td>
<td>[murnad] [m] [murnad] morot 'carrot'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[stompa] [s] [stompa] stumma 'sock'</td>
</tr>
</tbody>
</table>

continued on next page
<table>
<thead>
<tr>
<th>Subject</th>
<th>Consonant inventory</th>
<th>Segmental errors</th>
<th>Phonological template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isak</td>
<td>Onset: b p d t j g k l m v f h c</td>
<td>Gliding, reduction of consonant clusters, lateralization of /s/</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>Coda: n j t l k f p e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kerstin</td>
<td>Onset: b m d t j</td>
<td>Reduction of consonant clusters, Fronting of velar consonants, Omission of final consonants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coda: -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linus</td>
<td>Onset: d t b p j l m j f s h k</td>
<td>Gliding, reduction of consonant clusters, lateralization of /s/</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>Coda: n t p j k</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majken</td>
<td>Onset: b d l m j g</td>
<td>Omission of final consonants, reduction of consonant clusters, stopping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coda: n s t p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sara</td>
<td>Onset: b m j v j l g k j e</td>
<td>Omission of unstressed syllables, reduction of consonant clusters, gliding</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coda: n t f l j t k</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.2 – continued from previous page

Words included in the ET-experiment in CI/AM

- [Ek:Or@] | [Ek:Oj@] | [fEkOr@] ekorre 'squirrel'
- [flygplan] | [flygplan] flygplan 'airplane'
- [grat] | [grat] sko 'shoe'
- [lampa] | [lampa] lamp 'lamp'
- [luks] | [luks] sko 'shoe'
- [mossa] | [mossa] 'cap'
- [pappa] | [pappa] 'apple'
- [pin] | [pin] 'penguin'
- [sås] | [sås] 'apple'
- [strumpa] | [strumpa] 'sock'

Continued on next page
<table>
<thead>
<tr>
<th>Subject</th>
<th>Consonant inventory</th>
<th>Segmental errors</th>
<th>Phonological template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilda</td>
<td>Onset: b p d m</td>
<td>Reduction of consonant clusters, gliding, omission of final consonants</td>
<td>CVCV as in [baːn] \textit{elephant}</td>
</tr>
</tbody>
</table>

Words included in the ET-experiment in \textit{Correct} CM AM

<table>
<thead>
<tr>
<th></th>
<th>CM</th>
<th>AM</th>
<th>Words included in the ET-experiment in \textit{Correct}</th>
</tr>
</thead>
<tbody>
<tr>
<td>[baːn]</td>
<td>[baːn]</td>
<td>[baːn]</td>
<td>banana</td>
</tr>
<tr>
<td>[lampa]</td>
<td>[bampa]</td>
<td>[kampa]</td>
<td>lamp\textsuperscript{a}</td>
</tr>
<tr>
<td>[μrμt]</td>
<td>[μμμ]</td>
<td>[suːrμt]</td>
<td>mor\textsuperscript{a}</td>
</tr>
<tr>
<td>[pærOn]</td>
<td>[pælO]</td>
<td>[særOn]</td>
<td>pær\textsuperscript{a}</td>
</tr>
<tr>
<td>[skuː]</td>
<td>[duː]</td>
<td>[Cuː]</td>
<td>sko\textsuperscript{a}</td>
</tr>
<tr>
<td>[trœj:a]</td>
<td>[dœj:a]</td>
<td>[fœj:a]</td>
<td>tröja\textsuperscript{a}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Words in brackets indicate cases of reduced pronunciation.


Acknowledgements

This thesis would not have been possible without the support and contributions by so many people. First of all, I would like to thank all the participating families. Without a doubt, without your interest and time you spent on my research, there would not be any research. Thank you!

My supervisors Ulla Sundberg, Marilyn Vihman and Mattias Heldner. Ulla, you guided me from the very first day to the last in such a supportive and encouraging way that I have no words to describe my gratefulness. Starting from recruiting participants, driving with your car to the suburbs of Stockholm to record children, recording speech stimuli, reading and discussing texts and ideas in museums, in the green, during car-rides... Not to mention all the laughter and the tears you dried. Thanks for your optimism and for constantly reminding me of the importance and relevance of this thesis. I will carry your positive attitude with me! Marilyn, you came into the picture right on time. Thanks for generously sharing your ideas and home. The stays at your place were always so fruitful and enjoyable -discussing child phonology for breakfast and dinner and almost all the time in between was just what I needed to regain the enthusiasm for work. Your dedication to child phonology and your energy will always be something I strive for. Thanks for being a role model in so many respects! Mattias thank you for assistance in statistics, your continuous reminders of APA conventions and for your joyful attitude toward research and thanks, of course, for the PTW’s! Iris-Corinna Schwarz, you were there as a supervisor in the beginning, thanks for accompanying me at the outset.

I wish to thank, also, Björn Lindblom, the few times we met were always a source of inspiration and every time I learned something new. Christina Kauschke, thank you, for planting the seed of my research career and fruitful collaboration beyond my employment in Marburg. Your trust in me and involvement in research projects were invaluable experiences.

I got a lot of inspiration through the NorPhlex network. The network meetings allowed me to make friends and very fruitful contacts in the research group at York University. Thanks, Tamar Keren-Portnoy for your advice regarding stimuli design and other aspects of the experiments. I am indebted to Gålöstiftelsen, which generously supported three of my research travel expenses, two of them to York in

I am grateful to the Department of Linguistics at Stockholm University for all the open doors and the open-minded working environment. In particular, I would like to thank my fellow and former Ph.D. students who made my days at the department pleasurable: Andrea, Benjamin, Calle, David, Desalegn, Elísabet, Ellen, Emil, Francesca, Ghazaleh, Hatice, Johan, Josefin, Natalia, Niklas, Pernilla, Pia, Richard, Robert, Signe, Susanne, Thomas, Ulrika and Yvonne. Special thanks to those of you who passed down the LATEX root document to me and to those of you who contributed to the textseminarier and read my texts. I am also grateful for the opportunity I was given to obtain teaching experience. Thanks Henrik L. for guiding me through my very first teaching experiences, thank you Elísabet for the lab hours and Lisa, Iris, Tove, Thomas and Mattias H. for letting me be part of your courses. The administrative staff, in particular Linda, Nada and Cilla as well as Liisa and Ann deserve a big acclamation for their helpfulness and kindness in solving everyday problems (special thanks to those of you who I met at various friskvård-passes at Frescatihallen, which gave me extra motivation). The department provides incredible support for the technical stuff: Peter, Hassan and Jeffy. Thank you for your patience and help in so many respects. A special thank goes to Johan Stark for helping to fix the eye-tracking lab, as well as Rickard for assistance in PHP and for installing the prerequisites for LimeSurvey. Ulla Bjursäter for administrating the homepage connected to the MAGIC project. Petter, thank you for your EEG expertise and supportive attitude! Thanks to Anders Eriksson for apple breaks and scientific discussion walks.

I thank Ulrika Marklund and Christine Cox Eriksson for providing SPRINT material and the transcriptions of the parent’s utterances. Thanks to Andrea Strandberg, who transcribed some of the speech recordings from the production study. Johannes Bjerva for the support during my very first eye-tracking experiment. Myra Öberg and Marie Markelius for their help in data collection during the first studies with infants. In particular, thanks Marie for your commitment and engagement in running experiments at any time of the week or day! Frasse, for taking me to VBO—you and Cecilia contributed in many ways to making me feel at home from the very first day. Ellen Marklund and Ulrika Marklund for sharing the office during the first years and beyond. I have fond memories of the time with the three of us and feel privileged having shared the office with two ‘lefties’.

I was very fortunate to find close friends at the department: Elísabet, Marcin and Pernilla. Thanks for your tremendous support throughout the years, for run-
ning, for social activities and sausage discussions, laughter and so much more!
In particular, Elisabet, thanks for your hawk-eye in reading texts, the rides home, for discussions about life and research/teaching, for our Friday sessions during my parental leave, and all the fun we had teaching together. Marcin, you read my very first drafts on this thesis, your positive attitude and encouraging feedback let me continue. Thanks for your open ear and support in lab-related issues and for constantly reminding me “it will be fine”. I definitely needed it. Thanks, also, for your creativity in finding magical project names and for being a great collaborator on the Zombie project. Pernilla, thank you for all the time you spent on reading abstracts or Swedish texts, your gentle reminders of work-life balance, for lending me your first child as a guinea pig, for your distinct sense of how life and work should be and for helping me to look over the edge.

I am grateful to the friends from near and far who supported and thought of me and who reminded me of non-academic and social life as well as the good research conditions that I was privileged to have: Doro & Lisa, Anna, Caro, Sandra S., Sandra G., Julia, Francesca & Raphaël, Pernilla & Jon, Hans (thanks for all the Monday evening rides home!) and Anders H. (klopf klopf!). I appreciated a lot all the social activities, or thoughts and best wishes you sent from the distance when I moved to Stockholm or occasionally disappeared in the research world. Thanks for your never-ending support!

My parents and Uwe as well as my brothers and sisters and their families. Thanks for always being there and for preparing me to take steps that I would not have thought I would be able to do. You shaped me as a person to the better, and without it, I would not be here.

Mattias, your support and expertise on most of the problems in life made my days more fun, more love-filled, more balanced and well, more of everything. Thanks for being at my side, for your endless patience and love. I feel very blessed in having Miriam in our life, who shows us every day what it means to be patient and to challenge yourself in learning new skills. I know I am lucky 🌸