Turn-taking and early phonology
Contingency in parent-child interaction and assessment of early speech production

Ulrika Marklund

Academic dissertation for the Degree of Doctor of Philosophy in Phonetics at Stockholm University to be publicly defended on Friday 14 December 2018 at 09.00 in hörsal 12, hus F, Universitetsvägen 10 F, plan 2.

Abstract
This thesis focuses on contingency in parent-child interaction, investigating it in the light of the linguistic capacity of the child and the status of the caregiver. Further, the thesis covers the development of two tools to assess the developmental maturity level of expressive phonology. A functional emergentist perspective on language acquisition is taken, which includes a phonetic perspective on phonological development. Both infant language development and factors that influence parent responsiveness are explored.

The thesis contains four studies. In the first study, durations of parents’ utterances and pauses in interaction with their 18-month-old infants were related to the infant’s vocabulary size. Recordings of interactions of fifteen children and their parents were made at home in daily life situations. The children were divided into three groups according to their vocabulary size: large, typical or small. The main finding is that parents in the large vocabulary size group responded faster to their children compared to the parents in the typical size vocabulary group, who in turn responded faster than the parents in the small vocabulary size group.

In study two, duration in vocal turn-taking between 6-month old infants and their caregivers was investigated, in terms of the status of the caregiver and the sex of the infant. Caregivers’ pauses were measured in 10-minute caregiver-infant interactions recorded at home. It was found that primary caregivers responded faster to their infants compared to secondary caregivers, and that in turn, infants responded faster to the primary caregiver than to the secondary caregiver.

Study three introduces the Word Complexity Measure for Swedish (WCM-SE), a tool for calculating phonological complexity in words or utterances. Calculations are based on ten parameters describing speech structures that are considered phonetically complex to produce. In the development of the WCM-SE, both language-specific and language-general descriptions of speech development were considered, as well as universal acoustic and aerodynamic principles.

Study four documents the selection of Swedish words for the word lists in the test Profiles of Early Expressive Phonological Skills for Swedish (PEEPS-SE). The selection was based on criteria of age of acquisition and word complexity, as measured by the WCM-SE.

The findings presented in this thesis contribute to our knowledge of early interaction and parents’ potential impact on the child’s early language and communication development. Further, the tools developed for the assessment of Swedish are valuable contributions both to the research field of early phonology and to clinical work in Sweden.

Keywords: parent-child interaction, turn-taking, parental responsiveness, phonological development, phonological complexity, assessment of speech production.

Stockholm 2018
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-61853

ISBN 978-91-7797-499-4
ISSN 0282-6690

Department of Linguistics
Stockholm University, 106 91 Stockholm
Turn-taking and early phonology
Contingency in parent-child interaction and assessment of early speech production

Ulrika Marklund
## Contents

Abstract ..................................................................................................................... 1

Sammanfattning ........................................................................................................ 3

Kapitel 1: Bakgrund ......................................................................................... 3
Kapitel 2: Studier ............................................................................................... 5
Kapitel 3: Diskussion .......................................................................................... 6

Acknowledgements ............................................................................................... 7

Overview .............................................................................................................. 9

Papers and manuscripts included in the thesis ....................................................... 10

1. Background ...................................................................................................... 11
   1.1 Mechanisms underlying language development ....................................... 11
       1.1.1 Introduction .................................................................................. 11
   1.2 Theoretical perspectives on language development ............................... 16
       1.2.1 Introduction ................................................................................ 16
   1.3 Parental responsiveness ........................................................................... 26
       1.3.1 Introduction ................................................................................ 26
       1.3.2 Fixed factors of parental responsiveness ................................. 28
       1.3.3 Dynamic factors of parental responsiveness ............................ 33
       1.3.4 Responsiveness related to infant communicative developmental outcomes ............................................................................. 37
   1.4 Development of communication and language ....................................... 38
       1.4.1 Introduction ................................................................................ 39
       1.4.2 Early development of communicative skills .................................. 40
       1.4.3 Phonological development .......................................................... 45
   1.5 Assessment of early phonology ................................................................. 52
       1.5.1 Introduction ................................................................................ 52
       1.5.2 Collecting and analysing speech data ......................................... 55
       1.5.3 Measures and tests to assess expressive phonology .................. 59

2. Studies .............................................................................................................. 66
   2.1 Study 1 ....................................................................................................... 67
       Pause and utterance duration in child-directed speech in relation to vocabulary size .............................................................................. 67
   2.2 Study 2 ....................................................................................................... 69
Contingent turn-taking between parents and 6-month-olds: Primary caregivers respond faster than secondary caregivers

2.3 Study 3
Introducing WCM-SE: The Word Complexity Measure phonetically justified and adapted to Swedish

2.4 Study 4
The development of a vocabulary for PEEPS-SE — profiles of early expressive phonological skills for Swedish

3. Discussion

4. References

Appendix: General definitions of terms
Abstract

This thesis focuses on contingency in parent-child interaction, investigating it in the light of the linguistic capacity of the child and the status of the caregiver. Further, the thesis covers the development of two tools to assess the developmental maturity level of expressive phonology. A functional emergentist perspective on language acquisition is taken, which includes a phonetic perspective on phonological development. Both infant language development and factors that influence parent responsiveness are explored.

The thesis contains four studies. In the first study, durations of parents’ utterances and pauses in interaction with their 18-month-old infants were related to the infant’s vocabulary size. Recordings of interactions of fifteen children and their parents were made at home in daily life situations. The children were divided into three groups according to their vocabulary size: large, typical or small. The main finding is that parents in the large vocabulary size group responded faster to their children compared to the parents in the typical size vocabulary group, who in turn responded faster than the parents in the small vocabulary size group.

In study two, duration in vocal turn-taking between 6-month old infants and their caregivers was investigated, in terms of the status of the caregiver and the sex of the infant. Caregivers’ pauses were measured in 10-minute caregiver-infant interactions recorded at home. It was found that primary caregivers responded faster to their infants compared to secondary caregivers, and that in turn, infants responded faster to the primary caregiver than to the secondary caregiver.

Study three introduces the Word Complexity Measure for Swedish (WCM-SE), a tool for calculating phonological complexity in words or utterances. Calculations are based on ten parameters describing speech structures that are considered phonetically complex to produce. In the development of the WCM-SE, both language-specific and language-general descriptions of speech development were considered, as well as universal acoustic and aerodynamic principles.

Study four documents the selection of Swedish words for the word lists in the test Profiles of Early Expressive Phonological Skills for Swedish (PEEPS-SE). The selection was based on criteria of age of acquisition and word complexity, as measured by the WCM-SE.

The findings presented in this thesis contribute to our knowledge of early interaction and parents’ potential impact on the child’s early language and
communication development. Further, the tools developed for the assessment of Swedish are valuable contributions both to the research field of early phonology and to clinical work in Sweden.

Bakgrunden, kapitel 1, är indelad i fem avsnitt: i) grundläggande mekanismer för språkinlärning, ii) teoretiska perspektiv på fonologisk utveckling, iii) föräldra-barn-interaktion och föräldrars responsivitet, iv) tidig språkutveckling med fokus på fonologisk utveckling, samt v) bedömning av expressiv fonologi.

I kapitel 2 beskrivs de fyra studierna. Studierna följer två spår: två studier rör föräldra-barn-interaktion och föräldrars responsivitet, och två studier rör utveckling av instrument för bedömning av tidig fonologi. Studierna har resulterat i tre publicerade tidskriftsartiklar och ett manuskript.

I kapitel 3 sammanfattas och diskuteras resultat och implikationer av studierna.

Appendix innehåller allmänna definitioner av termer.

Kapitel 1: Bakgrund.

I första avsnittet beskrivs grundläggande mekanismer för tal- och språkinlärning: perceptuell, motorisk, social och kognitiv utveckling. Det beskrivs hur dessa biologiska förutsättningar ligger till grund för de mekanismer som är involverade i utvecklingen av kommunikativt språk. Barnet har tidigt, redan innan födseln, förmåga att uppfatta sinnesintryck och den sensoriska och motoriska utvecklingen gör att barnet med tiden kan styra sin motorik och sina vokalisationer. Denna utveckling, tillsammans med ökad minneskapacitet och en social drivkraft innebär att barnet med tiden kan utveckla talat språk att använda i kommunikation med andra. I avsnittet beskrivs också statistisk inlärning, vilket innebär att barnet uppfattar och minns återkommande mönstert i det omgivande talet. Genom återkommande språklig input i en social miljö har barnet möjligheter att lära sig tal och språk.

Avsnitt 2 består av en översikt av tre teoretiska perspektiv på fonologisk utveckling: formalism, perceptionsmodeller och funktionalism/emergentism.
Exakt vilka mekanismer som betraktas som medfödda för inlärning av tal och språk, och på vilket sätt exponering för talspråk påverkar språkinlärmningen varierar mellan dessa perspektiv. Formalistiska teorier utgår från lingvistisk teori och bygger på antaganden om medfödda språkliga principer för barnet att upptäcka och utveckla. Perceptionsmodeller bygger på idén om att existerande språkliga kategorier lärs in genom att barnet uppfattar och minns återkommande mönster i det omgivande språket. Funktionalistiska modeller menar att inga språkspecifika medfödda förutsättningar finns för att utveckla språk. Istället är det grundläggande perceptuella, motoriska och socio-kognitiva förmågor som gör det möjligt för ett barn att uppleva, utveckla och använda tal och språk.


A snitt 5 handlar om bedömning av expressiv fonologi. Vikten av att upptäcka talavvikelser tidigt betonas och rutiner som används för att identifiera avvikande tal hos barn i Sverige beskrivs kortfattat. Följande avsnitt tar upp hur olika typer av data och olika analysmetoder kan påverka resultatet av vid bedömningar av expressiv fonologi. Det sista avsnittet handlar om olika instrument för bedömning av expressiv fonologi.
Kapitel 2: Studier


I Studie 2 undersöktes vokalturtagnings i interaktion mellan 6 månaders gamla barn och deras vårdnadshavare. Specifikt undersöktes om skillnader fanns beroende på vårdnadshavarens status (primär eller sekundär) och beroende på barnets kön (flicka eller pojke). Primär vårdnadshavare definierades som den vårdnadshavare som hade varit och var föräldralig med barnet och därmed spenderat mest tid med barnet vid tidpunkten för studien. InspeLningar av samspel med barnet gjordes i hemmen, separat för de olika vårdnadshavarna. Resultaten visar att primära vårdnadshavare svarade snabbare på sina barns vokaliseringar jämfört med sekundära vårdnadshavare, och barnen i sin tur svarade snabbare på den primära vårdnadshavarens yttranden jämfört med hur snabbt de svarade på den sekundära vårdnadshavarens yttranden.


som testet är skapat för, ökar sannolikheten att ordet kan eliciteras i testsituat-
ionen och att uttalet är representativt för barnet. De två ordlistorna innehåller ord med olika grad av fonologisk komplexitet.

Kapitel 3: Diskussion

Kapitel tre ägnas åt diskussion av resultaten från avhandlingens fyra studier. I diskussionen beskrivs också möjliga inriktningar för fortsatt forskning.

Andra metoder, som till exempel filminspelningar samt andra inspelnings-
situationer skulle kunna bidra till nya intressanta upptäckter om föräldra-barn-
interaktion och hur det kan påverka barns språkutveckling. Mer kunskap om föräldras föreställningar om barns språkliga utveckling kan också bidra till detta. Det vore också intressant att göra studier av föräldra-barn-interaktion under det skede då den sekundära vårdnadshuvaren har börjat sin föräldral-
dighet. Även andra mått på föräldras tal, som till exempel artikulatorisk tyd-
lighet, skulle kunna undersökas och relateras till barns fonologiska utveckling.

De olika materialen som utvecklats för bedömning av expressiv fonologisk förmåga har flera fördelar: WCM-SE är ett instrument som är fonetiskt grun-
dat, kan anpassas till olika dialekter och går att använda i såväl oberoende som relationell analys. PEEPS-SE är lexikalt och komplexitetsmässigt (fonolo-
giskt) åldersanpassat till små barn. Det går dock alltid, på grund av kontextu-
ella faktorer, att diskutera komplexitetsnivåer i fonologi och hur detta ska po-
ängsättas. Arbete med ett språkuniversellt komplexitetsmått för fonologi ba-
serat på fonetiska principer är påbörjat.

De resultat som presenteras i denna avhandling bidrar till kunskaperna om tidig interaktion och föräldrars potentiella inverkan på barnets tidiga språk-
och kommunikationsutveckling. De instrument som utvecklats för bedömning av svensk fonologisk komplexitet är värdefulla bidrag både till forskningen om tidig fonologisk utveckling och till det kliniska, logopediska arbetet i Sve-
rige.
Acknowledgements

There are many I would like to thank for supporting me in this work, and during my time as a PhD student. Special thanks go to:

My supervisor Professor Francisco Lacerda, for leading me into phonetics, for great academic discussions, for sharing such a positive and inspiring view on research, and for helping me understand the value of “going slalom”, yet describing things clearly.

My supervisor Associate Professor Iris-Corinna Schwarz, for all of your great support, for showing me a pragmatic and concrete perspective on research work, for telling me to keep on track when I get “carried away”, and for all of the positive energy that you spread.

Dr. Ulla Sundberg for being my supervisor during a period when I was the only clinician in the department. It was great to share research thoughts in phonetics with you, a special teacher and speech therapist that had made this journey before me.

My former PhD student colleague, and now present colleague, Dr. Ellen Marklund, for really being my “unacknowledged supervisor”, for making things so much easier by planning and structuring, always with my perspective in mind, for all of your great feedback, and for all fun!

My colleagues at the Department of Linguistics at Stockholm University, for academic discussions and for technical and administrative support.

My co-authors at Karolinska Institutet, PhD student Anna Persson and Professor Anette Lohmander, for fruitful collaboration.

My boss at Talklinken, Danderyds Sjukhus, Speech-Language Pathologist and Head Manager of Talkliniken at Danderyds Sjukhus, Helén Norström. During my years at Talkliniken I had the opportunity to do my Hanen certification and to work with parent programs, the true start of this journey. Helén, you have been a valuable support throughout my years at Stockholm University.
My boss at Waldorflärarhögskolan, President Caroline Bratt, for welcoming me to the Steiner School academics at Waldorflärarhögskolan and for giving me the opportunity to teach child language development for future Steiner teachers. It has been nourishing work for this thesis.

My family, that really has supported me during these years, all of you such wonderful personalities and role models in different ways. Very special thanks to my closest family:

My daughter Alma, for being a person true to herself and always sensitive towards others, for interesting discussions about life (and work in kindergarten!), and for helping me remember that art is one of the most important parts of life.

My son Arvid, for being constantly curious about most things in life, for interesting discussions, for your music (especially your singing and guitar playing), for making me laugh, and for always being kind and helpful.

My son Axel, for being a person that puts effort into whatever needs to be done and showing me that this is part of what makes it more fun, for great discussions, for wonderful recorder music and for cooking when I came home late.

My sister Bitta, for always being there, willing to discuss whatever the topic is: research, house renovation, garbage room treasures, ... for all fun, and for showing me a pragmatic and positive way of handling ups and downs in life.

And last but not least, my wonderful love and husband-to-be, Luca, for the total support you have shown me throughout these years in so many ways, both practical and emotional. Thank you for the discussions, for listening, comforting, cooking, planning, for (nearly) always playing the right music, for holding my bad foot, for making me laugh, and for (nearly) always being positive to my sometimes crazy ideas. I look forward to what comes next!

I also send grateful thoughts to my parents, no longer in this dimension. I think you got me into research very, very early, I just hadn’t realised that until recently.
Overview

This thesis aims to describe the early speech and communication development of typically developing hearing infants in relation to parental responsiveness, and to describe development and assessment of early expressive phonology.

The background, chapter 1, is divided into five sections: basic underlying mechanisms for language learning, theoretical perspectives on phonological development, parental responsiveness and parent-child interaction, early language development with focus on phonological development, and assessment of phonology.

In chapter 2, the studies are described. The studies follow two tiers: two studies concern contingency in parent-child interaction and two studies are descriptions of development of measures for assessing early expressive phonology. The studies have resulted in three papers and one manuscript included in the thesis.

The findings and implications of the studies are summarised and discussed in chapter 3.

For general definitions of terms, see Appendix.
Papers and manuscripts included in the thesis


Reprinted with permission from Cambridge University Press.

**Marklund, U., Schwarz, I.-C. & Marklund E.** Contingent turn-taking between parents and 6-month-olds: Primary caregivers respond faster than secondary caregivers.

Manuscript in progress.


Reprinted with permission from Taylor & Francis Group.


Reprinted with permission from Taylor & Francis Group.
1. Background

The first section of the background describes underlying mechanisms for language learning, and section 2 describes theoretical perspectives on phonological development. Section 3 addresses how parents respond to infants’ non-vocal and vocal behaviour, depending on different factors, and the relationship between parental responsiveness and infant development. Section 4 details early language and communication development in terms of non-vocal and vocal turn-taking. It also describes early phonological development, both in general and in Swedish-learning children. Section 5 addresses the task of measuring early expressive phonology. It lists tools and tests to assess expressive phonology, and discusses how administration, type of speech data and analysis type may affect assessment.

1.1 Mechanisms underlying language development

This section begins with an introduction to underlying mechanisms and theories of language learning. This is followed by a more detailed description of the role of these mechanisms – the infants’ biological and social affordances necessary for development of communicative spoken language. It concerns perceptual learning and development of speech production, memory processes and statistical learning mechanisms, and development of social communication.

1.1.1 Introduction

During the first two years in a child’s life, basic perceptual-motor and socio-cognitive abilities set the stage for language learning. Neurophysiological changes take place, such as the integration of neuromuscular control and attention (e.g., Elman, 1993; 2003).

The infant is sensitive to regularities in the input (e.g, Saffran, Aslin & Newport, 1996; Kuhl, 2004), but not only the ambient language is important for language learning, as the infant’s own sound production also contributes sensory information. Speech perception and production work together as a dynamic system (e.g., Thelen & Smith, 1998; Yeung & Werker, 2013). Attention and memory development enable the infant to focus on and memorise events, objects and people (e.g., Squire & Zola, 1996; Baddeley, 2001; 2003), and to
link the linguistic input to other sensory information in order to create meaning (e.g., Tomasello & Farrar, 1986; Werker & Stager, 2000). Besides linking relevant context-dependent sensory information from different modalities, the infant needs to be able to understand representing symbols (Werner & Kaplan, 2014) and to gain the pragmatic insight that speech can be used for communication (McCune, Vihman, Roug-Hellichius, Delery & Gogate, 1996). The infant also need to have the ability to produce speech sounds that gradually become syllables, words and phrases (e.g., Vihman, 2013).

Thus, perceptual-motor and socio-cognitive abilities pave the way for the basic mechanisms involved in the development of communicative language, by means of recurrent linguistic input in a social setting. Infants seem to come into this world well-equipped to develop functional speech language and communication through their abilities to perceive and memorise incoming information, their potential to learn about the environment and their social interactional urges.

1.1.1.1 Perception and production prerequisites for language development

Attentional control – focus – is a skill that develops during the last months of the infant’s first year (e.g., Carpenter, Nagel, Tomasello, Butterworth & Moore, 1998). The new-born therefore experiences speech and language implicitly without this attentional control. Young infants can discriminate phoneme contrasts in most languages that they have been tested on (Kuhl, 2004). The information in the speech signal has at first no linguistic relevance to the infant, but exposure to repetitive input patterns fosters implicit learning (e.g., Kuhl, 2004; Lacerda & Sundberg, 2006). This learning is not exclusive to language but is applied to all sensory stimuli (Kirkham, Slemmer & Johnson, 2002). Regarding speech, the infant implicitly tracks regularities in the input without the purpose to learn and without attention specifically oriented toward the input (e.g., Kuhl, 2004; Saffran, Aslin & Newport, 1996). During the first year of life, infants’ ability to discriminate between non-native speech sounds decreases (Kuhl, 2004). Speech sound perception is therefore presumed to be affected by the ambient native language (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Lacerda, 1992).

At six months, infants have developed increased sensitivity to native language phonology (Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993), for example language-specific perception of vowels (Kuhl, 2004; Kuhl et al., 1992). At eight months, the infant is able to discover typical stress patterns in words (Jusczyk, Cutler & Redanz, 1993) and detects word boundaries in the speech signal through transitional probabilities (Saffran, Aslin & Newport, 1996). At nine months, infants can distinguish between sound combinations that are phonotactically legal and those that are not (e.g., Friederichi & Wessels, 1993; Sebastián-Gallés & Bosch 2002). At nine months they also can discriminate between different phonotactically legal words depending on how frequent the phonemes and phoneme combinations are (Jusczyk, Friederici,
Wessels, Svenkerud & Jusczyk, 1993; Nazzi, Bertoncini & Bijeljac-Babic, 2009). Their ability to discriminate between non-native speech sound contrasts is gradually reduced (Kuhl, 2004; Best, 1994; Best, McRoberts & Goodell, 2001; Werker, Gilbert, Humphrey & Tees, 1981; Werker & Tees, 1983; 1984). By 11 months, the infant has an increased perceptual ability of native consonant contrasts and at the same time a decrease in foreign consonant perception (Kuhl, 2004). This perceptual reorganisation means that, by the end of the first year, the infant has not only acquired a basic phonological system of the native language (or languages if the infant is exposed to more than one language), but is also specialised in listening to it. The infant’s perception of speech and listening to language has during this first year transitioned from being acoustic to linguistic in nature.

Gaining control over speech production is another prerequisite for language development. Physiological maturation of the breathing system, anatomical growth and laryngeal migration rapidly change the acoustic-articulatory settings during the first year of life (Ménard, Schwartz & Boë, 2004), enabling the infant to produce more and more sophisticated vocalisations as he or she grows older (MacNeilage & Davis, 2000). Compared to adults, the larynx of a young infant is placed more cranially, the pharynx is shorter and there is little space for the back of the tongue in the oral cavity. A new-born’s tongue is large relative to the size of the oral cavity, which creates a restricted space for vertical tongue movements. A new-born’s pharynx is much shorter than that of an adult, as well as having a more gradual slope (Paulsen & Waschke, 2013). In combination with the short distance between velo-pharynx and epiglottis in infants, this results in restricted breathing opportunities through the mouth, hence the infant typically breathes through the nose. The infant grows rapidly during the first year, and already at six months, the proportions of the anatomy involved in speech production have become more like those of an adult. At 12 months, the anatomy has almost reached adult proportions, with a larger space in the oral cavity (Ménard et al., 2004). These physiological changes and improved muscle control over larynx and articulators influence the development of speech production. Teething and gross motor development milestones that enable upright positions like sitting, standing, and walking also play a role for speech production development (e.g., Iverson, 2010).

Developmental advances in perceptual-motor abilities are one important base for the interactional loop between perception and production of spoken language. Studies that look at the link between speech production and perception in infants indicate that even infants’ own speech sound production affects their perception. For example, infants at three to six months prefer to listen to productions with infant-like properties over productions with adult-like production properties (Masapollo, Polka & Ménard, 2016). Further, 6-month-old infants could match auditory and visual stimuli of speech production better, if their own production contained more distinct sound categories (Altvater-Mackensen, Mani & Grossmann, 2016). Proprioception, that is, tactile and
kinaesthetic information of the infant’s own speech production is suggested to contribute to the linkage between production and perception along with auditory information (e.g., Vihman, 1993; Schwartz, Basirat, Ménard & Sato, 2012). Thus, speech and language develop through sensory perception of both surrounding speech and the infant’s own production, together with maturing physiology and anatomy.

1.1.1.2 Memory systems and implicit learning
Children presumably use language input and “scan” the speech stream both to find patterns and to learn words from form-meaning associations by the use of a “dual memory system” (Squire & Zola, 1996). This memory system is located in the hippocampus and in the frontal lobes in humans. The hippocampus is particularly important for memory of details and for multimodal and episodic memory while the frontal lobe for example registers and accesses the arbitrary form-meaning pairs that make up words (Squire & Zola, 1996; Baddeley, 2001; 2003). The memory system supports both implicit and explicit learning (Ellis, 2002).

Implicit learning can refer to changes in perception induced by recurrent sensory information (Aslin, 2017; Saffran & Kirkham, 2018). Recurrent patterns of linguistic information are implicitly learned, both on the supra-segmental (e.g., prosody) and segmental level (e.g., speech sound properties). Explicit – or declarative – learning can refer to attention-driven learning and this is typically not seen before the first half of the second year (Ellis, 2002).

Memory processes – in the sense of changes induced by sensory input – are active in the fetus even before birth, for example shown by new-borns’ preference for their mothers’ voice over other female voices (DeCasper & Fifer, 1980). Maternal speech that the infant has been exposed to prenatally has been found to influence the new-born infants’ speech perception: a passage from a children’s story frequently read during the last trimester of pregnancy was shown more reinforcing to the new-born infant compared to another passage that the fetus had not been exposed to (DeCasper & Spence, 1986). New-borns also react to and prefer infant-directed speech (Cooper & Aslin, 1990), and two-day-old infants have been shown to prefer the native language before a non-native language (Moon, Panneton Cooper & Fifer, 1993).

Eight-month-olds exhibit implicit learning of three-syllable nonsense-words when all other information such as prosody is controlled for (Saffran, Aslin & Newport, 1996). The infants were exposed to a two-minute string of four syllabically different tri-syllabic non-words, without prosodic cues as to word boundaries, in random order. When comparing these non-words to novel non-words consisting of the same syllable units, but in different order – that is with other transitional probabilities – the infants preferred to listen to the new “words”, following the novelty effect (Saffran, Aslin & Newport, 1996), suggesting that the infants were sensitive to “word”-boundaries in the artificial language. This type of implicit, statistical learning is a process without any
stated intention or focus on learning. The same type of learning has also been found in adults (Saffran, Newport & Aslin, 1996).

Word meaning, on the other hand, has been suggested to be explicitly learned in infants. As the child explicitly registers arbitrary form-meaning pairs, he or she builds a lexicon. Further in the process of language learning, the child uses patterns in the vocabulary – basing the information on types (categories) instead of tokens (occurrences) – in a secondary implicit learning phase, to construct abstract knowledge about syntax and morphology (Lacerda & Sundberg, 2006). However, the phases of implicit and explicit learning do not necessarily follow each other consecutively, but presumably overlap, in a process that eventually leads to more abstract grammatical representations (Lacerda & Sundberg, 2006).

1.1.1.3 Social and cognitive development
Speech development is a social process during which the infant learns in interaction with others (Tsao, Liu, & Kuhl, 2004; Kuhl, 2007). In parent-child interaction, both attention to language and the interaction itself are promoted by the adults’ infant-directed speech, and at the same time specific linguistic properties of the native language are highlighted (Golinkoff, Can, Soderstrom & Hirsh-Pasek, 2015). Social interaction with the environment starts at birth and is closely related to language development. New-borns show an interest in faces, especially faces in motion (Cooper & Aslin 1990) and eyes (Lewkowitz & Hansen-Tift, 2012). New-borns prefer their mother’s voice over another female voice (DeCasper & Fifer, 1980). Simple turn-taking, like in a reciprocal call-response behaviour, is found in new-borns as early as 15 minutes after birth, where the new-borns’ sounds are suggested to be enhanced by skin-to-skin contact with a parent and also by the parents’ vocalisations (Velandia, Matthisen, Uvnäs-Moberg & Nissen, 2010). Twelve-month-old infants attend to information relevant to communication, namely face and gaze following, and this social interest predicts productive vocabulary at 18 and 24 months (Tenenbaum, Sobel, Sheinkopf, Shah, Malle & Morgan, 2015). Further, face processing in 2-month-olds activates the same neural structures as language processing does in adults, indicating a link between interest in faces and language (Grossmann & Johnson, 2007). Early social-communicative behaviour such as reciprocal interaction and (possibly) imitation is supported when parents pretend the infant to be a fully developed communicative partner that they imitate and respond to (Meltzoff & Moore, 1977; Jones, 2009). Further demonstrating the importance of social interaction for language development, 9-month-olds maintain sensitivity to non-native contrasts only when exposed to the non-native language in a social interaction setting, not when exposed to the language as recorded speech or from a television (Kuhl, Tsao & Liu, 2003).
With memory development and the growing ability to form mental representations, the infant becomes more capable as a communicative partner, using language for communication (e.g., Kuchirko, Tafuro & Tamis-LeMonda, 2018). To engage in communication the ability to refer to an object or event is crucial, not only to have needs satisfied but also to share thoughts with others. The ability to share focus on an object – joint attention – develops as the infant acts on the environment and interacts with others (Carpenter et al., 1998).

Word learning involves both learning of the form (phonetics) and the meaning (semantics). With social and cognitive development, the child starts connecting sound shape (word phonology) and object representations. The onset of this lexical (symbolic) learning of words occurs prior to the onset of word or even proto-word production. Infants at six months have already quite a few common words in their receptive lexicon (Bergelson & Swingley, 2012).

In summary, a number of anatomical, physiological, social and cognitive prerequisites combine in order for the child to develop language for communicative use.

1.2 Theoretical perspectives on language development

In this section, an overview of different theoretical perspectives of language development with focus on phonological development is presented. It begins with the **formalist perspective**, followed by descriptions of **perception models** and **functionalist models**.

1.2.1 Introduction

Various theories attempt to explain the observed processes behind child language development. In this thesis, theories and models of phonological development are summarised under three main theoretical frameworks: **formalist theories**, **perception models** and **functionalist models**. All of these theories are in one way or another based on the idea that human beings have biological prerequisites for language, and that language learning is in one way or another input-dependent (e.g., Jakobson, 1968; Werker & Curtin, 2005; Lacerda & Sundberg, 2006). However, what mechanisms are considered innate and exactly how input and exposure affect language learning differ greatly, depending on the theoretical perspective.

**Formalist theories** are based on assumptions of innate linguistic principles for the infant to detect and develop language (e.g., Jakobson, 1968; Chomsky, 1957).

**Perception models** are based on the idea that linguistic categories are acquired, emphasising the infant’s perception of recurrent patterns in ambient language and the infant’s memory capacities (e.g., Jusczyk et al., 1993; Kuhl, 1993; Werker & Curtin, 2005).
**Functionalist models** argue that no language-specific innate prerequisites exist to develop language (e.g., MacNeilage & Davis, 1990; Lindblom, 1992; Locke, 1997; Lacerda & Sundberg, 2006). Instead, basic perceptual-motor and socio-cognitive abilities enable humans to experience, develop and use language. Language and linguistic properties are considered to emerge through the interaction between perception and production, and phonological and lexical learning are intertwined processes.

1.2.1.1 Formalist theories

Formalist theories assume that the infant is born with innate linguistic knowledge. The linguistic input is not in focus, but rather the linguistic target form. This approach to language development evolved to counter the behaviourist view (Skinner, 1957; 1966) on development. Skinner argued that child language development mainly depends on the child’s own language experience. Behaviour, language development included, was seen as the result of positive or negative reinforcement (Skinner, 1957).

Jakobson’s *structuralist*, formalist view on phonological development (Jakobson, 1968) came from studies of adult phonology and did not acknowledge that the child’s perception, production possibilities, word recognition or vocabulary learning might be different than that of adults. Jakobson focused on the order in which children acquire different phonemic contrasts and how child output is realised compared to the adult form. The phoneme as a “pre-existing” linguistic unit was a central concept. Since open-closed and front-back phoneme contrasts are available in all languages, they are considered to be universal features. According to Jakobson, babbling has nothing to do with development of actual speech, since babbling can include any speech sound of the world’s languages (Jakobson, 1968). However, when starting to speak, infants gradually specialise in their native languages and become less prone to produce non-native speech sounds. Jakobson recognised that input and social interaction is important for language development, yet he refers to babbling as “… restricted to solitary play and of waking and of going to sleep…” (Jakobson, 1968, p. 29). Jakobson describes a chronological order of phoneme acquisition based on phonemic contrasts, in which children first learn maximum contrasts. The subtler the speech sound differences are, the later they are learned (and they are also typically less common in the world’s languages). However, Jakobson’s notion that infant babbling has nothing to do with speech development is no longer current. Since the 1970s, research has shown that infants’ babbling indeed is connected to the development of speech production (e.g., Oller, Wieman, Doyle & Ross, 1976; de Boysson-Bardies, Sagart & Durand, 1984; de Boysson-Bardies, Halle, Sagart & Durand, 1989).

Beginning in the late 1950s, Chomsky introduced his *generative* theory (Chomsky, 1957), a nativist view that criticised Skinner’s ideas, in particular the role of linguistic input and reinforcement for infant language development.
Chomsky proposed that as language is so complex and yet acquired so fast and with such ease, children must have access to innate foreknowledge of language. This innate linguistic knowledge is at the heart of the nativist view. An additional argument for the nativist view is that the relative scarcity and poor structure of linguistic input does not suffice to scaffold language by imitation and association only. Chomsky proposed that infants need to be born with certain knowledge of linguistic principles – a Language Acquisition Device (Chomsky, 1965), or, as it was later called, *Universal Grammar* (UG; Chomsky, 1988). In the generative sub-theory specifically concerning generative phonology, Chomsky and Halle (1965) described phonology and the relationship between abstract and surface forms. It is assumed that the child has innate knowledge of the phonological target (“adult”) form – an abstract underlying representation, and that the child’s production is the surface form, attempting to approach the target form. To derive the underlying representation, phonological rules are applied. Within phonological rules, the concept of markedness is central. Marked distinctive features in the language are those that are less common and more complex, and therefore less likely to be learned by children. Chomsky’s concepts of underlying representations, phonological rules and distinctive features, have been widely used in clinical speech therapy.

UG has been and remains very influential, and several formalist theories of phonology based on the nativist view have been suggested: *natural phonology*, *non-linear phonology*, *optimality theory* and *prosodic phonology*. All the suggested theories describe the child’s phonological development in relation to linguistic structures, but in different ways.

*Natural Phonology* has its background in speech pathology and focuses on phonological processes (Stampe, 1979; Donegan & Stampe, 1979; Ingram, 1974). The basic idea is that the child possesses natural innate phonological processes that facilitate language production, and that these are gradually suppressed with development, as the child tunes into the native language. Natural phonological processes are based on segment classes (*i.e.*, phoneme classes), but they also take into account word form and syllable structure. Examples of such natural processes are fronting, deletion of unstressed word initial syllables, metathesis and assimilation. Grunwell (1981) uses both natural processes and phonological productions independent of the target when describing stages in phonological development. These stages start when the child has reached 50 lexicon entries. In this approach, the descriptions are analysed independent of the adult target forms.

In the 1970s, a shift occurred from linear phonology, in which output description was segment-based, to *non-linear/auto-segmental phonology*, which assumes different phonological levels including phrase, prosodic word, foot, syllable structure and features (Goldsmith, 1976; Bernhardt & Stoel-Gammon, 1994). This was a new way of looking at early word forms, also taking vocal harmony, stress patterns and tone into account. Rather than describing
word forms only on the segmental level, non-linear phonology focuses on the relationships between phonological units that can act and interact more or less independently of each other on different levels.

In optimality theory (OT; Prince & Smolensky, 2004), phonological structure and markedness have central roles, but not in the same way as in generative phonology. In OT, it is presumed that the phonological structure of a language is indirectly expressed by the ranking of universal, but often conflicting constraints. This is so because humans prefer unmarked structures, but at the same time they prefer “faithfulness” to the underlying representation. The structure of the output language is seen as a result of the conflict between these two types of constraints: markedness and faithfulness.

Prosodic (metric) phonology focuses on the internal structure of syllables and words, in which these patterns are assumed to be hierarchical. This is assumed to reflect innate knowledge based on UG. A template defined by basic prosodic parameters, such as stress placement and internal structure of the word, functions as the basis for the expressive phonology. The child maps the input to this template and then adjusts the production accordingly (Fikkert, 2007).

1.2.1.2 Perception models

Perception models of phonological acquisition are based on the idea that the phonological system of a language and its relevant linguistic categories are detected by perception. This approach argues that humans have the ability to detect patterns and to use memory to recognise recurrent patterns. These patterns are neither exclusive to language, nor are they modality-specific – the child perceives them with all senses.

The main focus of the Word Recognition and Phonological Structure Acquisition model (WRAPSA; Jusczyk, 1993) is concerned with speech perception and segmentation of the speech flow. The model assumes that the preliminary representations develop based on single occurrences of words, so-called exemplars. The preliminary representations are based on the integration of various components of the acoustic signal within a certain time window, distinguishing between speech sounds depending on prosodic and durational cues. By integrating important features of the acoustic signal, relevant information within the speech signal emerges from the speech flow, forming syllables first. Then, word candidates are matched to existing preliminary representations in the lexicon, refining those or forming new representations. At early ages, this process is not language-specific, but becomes so during the first year of life (Jusczyk, 1993).

In the Processing Rich Information from Multidimensional Interactive Representations model (PRIMIR; Werker & Curtin, 2005), the rich speech signal is represented by three-dimensional “representation planes” that develop with the child's ability to perceive and match similarities in the input. Processing speech input and forming representations are interdependent, and
segments are simultaneously grouped by similarity, co-occurrence or other regularities within these three planes. There is a general perceptual plane on which any type of information in the speech input is perceived, including phonetic information with descriptive information such as emotion and speaker identity. Categories can be created on this first general plane or move into other categories on the second plane, on which word forms are processed. On the second plane, word forms are individually perceived as exemplars and not associated with meaning – just as phonetically based forms. On the third plane, the phoneme plane, representations of speech sound categories are stored, and word learning is supported further by refinement of phonemes.

According to the Native Language Magnet theory (NLM; Kuhl, 1993; Kuhl, 2004; Kuhl, Conboy, Coffrey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008), infants have an innate ability to differentiate the sound stream into gross categories that are separated by natural boundaries found by perception studies. This classification of sounds into gross categories is referred to as human phylogeny and described as a domain-general auditory processing mechanism focusing on acoustic cues. With exposure to language, the perception of phonetic differences in a particular language is altered in such a way that the distance between a prototypical speech sound exemplar (a good exemplar, or “magnet” in the specific native language) and its within-category neighbours narrows, while the perceived between-category distances (between prototypes) in the particular language are stretched out. By six months of age, infants show evidence of specific processing of prototypes (Kuhl et al., 1992), possibly as a result of exposure to the ambient language (Kuhl, 1991; 2004; Lacerda & Sundberg, 2006) Lacerda, 1995)

According to the NLM theory, speech perception processing is first limited to auditory information, but becomes multimodal when the infant also includes other information types such as articulatory feedback (Kuhl, 1993; Kuhl & Meltzoff, 1982; 1996). As they develop, the infant’s acquired skills will further influence language development through perceptual processing of production attempts. The infant does not only hear the ambient language produced by others, but also hears and feels his/her own production attempts. This feedback loop between perception and production is likely to have a stabilisation effect on learning (Kuhl & Meltzoff, 1982; 1996).

In summary, all perception models emphasise perception as the basis for language learning and present phonological development as a result of perceptual learning or construction of some kind of entities that can be derived from the linguistic input. The NLM theory focuses primarily on acoustic cues in early learning of speech sounds and does not discuss the development of word recognition. WRAPSA and PRIMIR both focus on the development of phonetic categories as well as on word recognition and word learning, and suggest that word forms are at first exemplar-based preliminary representations. The main difference between WRAPSA and PRIMIR lies in how word form representations are created in terms of the integration of speech signal
information. WRAPSA is restricted to auditory processing, whereas PRIMIR also recognises other types of sensory input, for example articulatory muscle feedback. PRIMIR argues for a distinction between indexical and phonetic information in the auditory input, even though the two are often inseparable. Indexical information refers to speaker identity and expressions of emotion and is intertwined with phonetic information. In WRAPSA, the child has first access to prosodic information, then to syllable-level information and finally to phonetic information. WRAPSA can therefore be considered as a hierarchical model. In PRIMIR, speech processing works on all levels simultaneously.

As perception models are based on experimental studies in psychology, the focus shifts from linguistic theory incorporating descriptions of language to descriptive models of perceptual processing in human behaviour. The perception models focus on basic developmental mechanisms and how these processes work rather than to explain young children’s speech deviations in relation to target language expressions.

1.2.1.3 Functionalist/emergentist models

Functionalist/emergentist models of language learning do not hold that linguistic knowledge is innate. Rather, the development of basic human biological affordances such as socialising, perceiving, vocalising and memorising abilities are important prerequisites for language learning. These basic abilities develop in interaction with the environment by general learning principles – implicit and explicit learning. The infant detects recurrent patterns in the environment, using all senses. Like the perception models, functionalists assume that in order to learn language, the infant needs to be able to detect recurrent patterns in the speech signal, and to assign meaning to these linguistic patterns by linking speech to experiences. Statistical learning mechanisms are the processes behind early learning of speech sounds, syllables and prosody (e.g., Aslin, 2017; Saffran & Kirkham, 2018). Later in development, explicit, attention-based learning takes place, in particular when learning form-meaning associations. With a basic vocabulary in place, implicit and explicit learning scaffold each other in later language development, for example when acquiring grammar.

According to Locke’s biological model (1997), language acquisition is based on neurocognitive mechanisms as well as sensitive periods for language learning. Basic mechanisms such as perception, social cognition and analytic abilities are the active processes as the infant is exposed to language. Locke postulates four phases in language learning: During the first phase, starting during the last trimester of pregnancy, the infant becomes familiar with the prosody and sound segments of the native language in the ambient speech. During the second phase, starting around five to seven months, the child recognises stereotyped utterances. During the third phase, starting from 20 months, the child learns and uses morphology, syntax and phonology. During
the fourth phase from three years on, the lexicon is expanded and language operations become automatised.

Variants of the biological perspective include *self-organising models* and *usage-based models*. In self-organising models, the focus lies on the development of motor control and how this affects the child’s speech production. Examples of self-organising models are the *Frame/Content Model* (MacNeilage & Davis, 1990), *Dynamic Systems Theory* (Thelen & Smith, 1998), *Lindblom’s emergentist model* (1992) and *Ecological Theory of Language Acquisition* (ETLA, Lacerda & Sundberg, 2006).

Self-organising models are concerned with general child development and not with phonological development in particular. Central is that individual development equals increased complexity of the system, with the consequence that the child acquires new opportunities to manage him/herself in relation to experiences in the outside world. That the child gets varied experiences is important for his/her change and adaptation. Self-organisation of a natural system can only be achieved if the system is in itself complex, while at the same time open to interact with, and to be influenced by, the outside world. Thus, perception and action are two independent, but intertwined factors, and, embedded in cognition, interaction between the two occurs simultaneously on multiple levels.

The *Frame/Content model* (MacNeilage & Davis, 1990) is a self-organising model that describes speech motor development. The model suggests that representations are based on oscillations of the lower jaw. The concept of “Frames” refers to syllable frames, and the model suggests that syllables have evolved through the original motor actions for chewing, sucking and licking. An open jaw refers to vowels and a closed jaw to consonants, just as in canonical babbling. This suggests that the early speech motor development results in a limited number of prototypical consonant-vowel (CV) combinations in babbling, determined by more passive articulators such as tongue and lips that depend on movements of the lower jaw (mandibula). Some CV combinations are given, such as dentals and alveolars are followed by front vowels, labials are followed by central vowels and velars are followed by back vowels. Thus, the front-back dimension of the vowel is determined by the articulation of the adjacent consonant. The child’s first word productions are firstly achieved from the frame and from the content, which means more possible variations in the manner of articulation of consonants and more possible variations of height in terms of vowels. The model is based on the development of motor skills, and assumes that children's early productions are initially not influenced by the ambient language.

*Lindblom’s emergentist model* also emphasises the interaction between biology and the environment (Lindblom, 1992). In this model, phonetic forms have their roots in universal human constraints of perception and production – the so-called “performance factors”. These are individual factors in the child, but increasingly also language-specific factors. Lindblom poses that
phonological units are emergent consequences of lexical development. The child starts with a few word forms, which set off certain motor activity in different articulators. There is a somato-topic (i.e. neural) storage of these motor activities, to which perception of the ambient language will then be linked. By using statistical regularities in the speech signal, the child derives phonological structures of the language. Lindblom also highlights the tendency to save energy in any motor activity. The clearer the articulation, the more energy is recruited to speech motor activity. In the Hyper-Hypo Theory (1990), Lindblom describes how the speech signal of the speaker can be adjusted depending on the listener’s information need.

From Lindblom’s theories, the Ecological Theory of Language Acquisition (ETLA, Lacerda & Sundberg, 2006) has developed. It is an emergentist theory that has the view that the ability to communicate by means of language has evolved with humanity’s needs for communication, and that language on all levels emerges as a consequence of interacting biological and ecological components. The infant is seen as a biological system with production, perception and memory abilities. In interaction with the environment, multimodal sensory information is processed and phonology emerges as the infant produces vocalisations and the adult listener interprets them according to expectations related to mature language – and also responds thereafter (Lacerda & Sundberg, 2006).

Usage-based models are similar to self-organising models. According to Pierrehumbert, the implicit bottom-up learning is not sufficient to develop phonological knowledge: Instead, the phonological system is built while it is in use (Pierrehumbert, 2003). The author describes a combination of bottom-up processing of the speech signal, lexical coding and generalising of patterns found in the lexicon. The model consists of several levels. First, acoustic and articulatory information both from perception and production is mapped, and then spectral information is phonetically encoded, which in turn enables word form representations, and phonology can be refined. Thus, phonology continues to develop through ongoing language use and word learning takes place regardless of age and maturation. This has been shown experimentally: Infants with larger vocabularies recognise familiar words faster compared to those with smaller vocabularies, when presented with initial word parts only (Fernald, Swingley & Pinto, 2001). Beckman and colleagues refer to usage-based theory and propose that vocabulary size, not age, is a critical factor in phonological processing (Beckman, Munson & Edwards, 2007).

Whole-word Phonology took a emergentist, usage-based perspective, in the 1970s (Waterson, 1971). This theory is based on child production data from babbling and word productions and the idea that phonology develops through representations of whole words. Waterson (1971) described child phonology as a system separate from adult phonology, and argued that patterns in child phonology should be seen as holistic rather than as segmental substitutions of
adult realisations, as a child’s perception is schematic and incomplete in comparison with that of an adult.

Vihman and Croft (2007) take the idea of whole-word based early phonological structures further with three main arguments: 1. There is variability in the production – the child can produce the same speech sound differently in different words, which indicates that the child has some word knowledge, but not the abstract categories for phonemes. 2. The relationship between child words and adult words is difficult to find on a segmental level. This is often characterised by assimilations of vowels or consonants (i.e. harmony) or fixed patterns of speech sound in the word form (i.e. melody). 3. Relations between a child’s word forms can be more obvious than their relationship to the adult word forms, as they often depend on the use of phonological favourites in selected word patterns (templates) (Vihman & Croft, 2007).

1.2.1.4 Perspectives on phonological development – summary

The various theories regarding phonological development have changed focus over time, and recent models emphasise a systemic perspective of development with the human infant and its environment as the base for learning, rather than the idea of a pre-determined development as a process for reaching a specific goal. With an increased focus on perception – and functionalist models, input has become important again. In contrast to formalism, functionalist models emphasise the role of the development of memory and learning processes based on study findings that were not available a couple of decades ago.

The importance of the reciprocal interaction between perception and production is stressed, thus emphasising the notion that linguistic units are not fix or innate, but are rather a practical way of categorising different linguistic elements bearing different meaning (e.g., Edwards Munson & Beckman, 2011). The phoneme, as a concept, is constructed based on an “adult” understanding of phonological systems, thus, it has little to do with the infant’s maturing phonological system and forthcoming phonological, lexical and morphological awareness.

Jakobson’s argument that babbling is unrelated to speech seems outdated, and has been criticised, for example in Locke’s biological model (Locke, 1997). Locke argues that content in babbling is found in adult speech and that babbling actually is what the infant within developmental constraints is able to produce. Babbling provides the opportunity for word production practice on the way to the target form. This is backed up by phonological analyses of babbling and of first words which demonstrate similarities and a likely continuity between the two (e.g., Stoel-Gammon, 1998; 2011).

Describing child production in relation to target forms can be practical, for example in speech pathology, but the formalist perspective gives little room to incorporate findings on how perceptual learning underlies phonological learning. In addition, focusing on phonemes as Jakobson does, or on distinctive features as Chomsky does, does not take into account suprasegmental
structures. Linear descriptions of output with segments that are weighted equally are problematic, both when describing expressive phonology and when designing intervention, since different speech sounds vary according to complexity characteristics and because suprasegmental structures have an impact on production.

As far as our present knowledge on speech perception development is concerned, we know that infants acquire language structures on several linguistic levels simultaneously, and that language learning on all levels is input-dependent (e.g., Locke, 1997; Lacerda & Sundberg, 2006; Stoel-Gammon, 2011). Chomsky does recognise to some extent the impact of input, but more as “triggers” for innate linguistic structures to develop. The development of suprasegmental structures is overlooked in generative phonology. The idea of underlying phonological representations and a surface output level itself can be questioned, since children are not known to have knowledge of the underlying forms.

Natural Phonology is widely used in clinical settings to describe deviations in expressive phonology, and offers practical tools with which to analyse patterns of substitutions and reductions with natural processes. However, with knowledge about perceptual learning (e.g., Werker & Tees, 1984; Jusczyk, 1993; Kuhl, 2004) the basic idea that the developing infant gradually suppresses innate phonological processes that function as simplifications is problematic, as infants presumably cannot be expected to have prior knowledge of target phonology or any underlying representation.

Non-linear phonology is also used increasingly in clinical testing and intervention, for example in the English test Profiles of Early Expressive Phonological Skills (PEEPS; Stoel-Gammon & Williams, 2013) and the Swedish Profiles of Early Expressive Phonological Skills (Marklund, Lacerda, Persson & Lohmander, 2018) in which word forms, syllable structures and sound classes are assessed. Even though the theoretical base of non-linear phonology comes from formalistic theory, non-linear phonology has developed a way of describing output phonology further by taking all phonological levels into account, thereby offering a practical way to describe expressive phonology in children.

There are obvious similarities between perception models and functionalism/emergentism in that both emphasise perceptual learning, but there are differences according the view on linguistic categories. In perception models it is assumed that linguistic categories are to be detected by perception of phonetic information, while the emergentist approach see linguistic categories as emerging on the basis of speech motor activity and speech perception during interaction with the environment. Emergentist theories also have a phylogenetic, evolutionary perspective on language development.

Studying the development of spoken language cannot be reduced to how surrounding speech affects the child’s vocalisations, as the speech stream
needs to be linked to experiences (other types of sensory input) to bear meaning. The child in the ecological setting sets the ground for this development.

1.3 Parental responsiveness

This section addresses the notion of parental responsiveness and how it is affected by “fixed” long-term factors and more “dynamic” short-term factors. Both fixed and dynamic factors are suggested to impact parental responsiveness.

Fixed factors are cultural and socioeconomic contexts, caregiver status and the sex of the child – factors that are usually stable over a long time.

Dynamic factors are short time factors, like parental awareness and understanding of the child’s developmental state in a given moment, non-verbal and verbal behaviour, and conversational settings.

1.3.1 Introduction

The child develops language in interaction with speakers of the ambient language, present in the child’s immediate ecological environment (Elman, 1993; 2003; Lacerda & Sundberg, 2006). In these settings, parents are typically the most salient and active interlocutors for the very young child, who has not yet gained sufficient motor ability to move around and interact with other speakers. It is reasonable to assume that new-borns produce sounds and actions that lack obvious initial communicative intent. The infant’s first sounds and actions are essentially reflexive, but their parents respond to them as if they actually had communicative intent. For example, mothers interpret burps and smiles as intended and intentional – not as random (Snow, 1977). The mother uses “turn-passing rather than turn-keeping or turn-grabbing devices” (Snow 1977, p. 20), thus helping the infant to take the turn in a conversation-like fashion. Soon, with increasing experience in interaction with speakers of the ambient language, the child’s sounds and actions become signals that represent states of being that can actually be interpreted by the parents to have particular meanings. In tune with the parents’ behaviour, the child starts to infer causality, forms associative representations and explores these in vocal interaction and communication with the parents. Learning is facilitated by social interaction – recurrent, contingent parental responses in different modes help the infant to learn about the environment and to learn language (Kuhl, 2007).

Parental responsiveness is defined as the parents’ tendency to respond to child actions. It is suggested that it is part of an intuitive universal parenting style that has developed with evolution and is important for attachment development (Papoušek & Papoušek, 2002). Despite its universal aspects, parental responsiveness also reflects the parents’ socialisation goals, child rearing principles, and knowledge and understanding of the child’s developmental state and behaviour. Besides immediate natural responses to expressions of pain or
distress, parents in particular respond to children’s communicative and explorative actions. Parents respond to about 70% of infant vocalisations (e.g., Dominguez, Devouche, Apter & Gratier, 2016; Gros-Louis, West, Goldstein & King, 2006), and mothers are much more likely to talk following infant vocalisations than talk when infants are silent (Tamis-LeMonda, Kuchirko & Tafuro, 2013).

Parental responses can be contingent in two ways: conceptually linked to infant action, or temporally connected. Temporally appropriate connection is sometimes called “contiguous” (e.g., Tamis-LeMonda, Kuchirko & Song, 2014). In this thesis, both conceptually and temporally linked responses are referred to as contingent. Parental responses occur usually within two to three seconds from the offset of infant action/vocalisation (e.g., Messinger & Fogel, 2007; Gros-Louis et al., 2006). Parents also typically adjust their responsive behaviour to match the developmental level of the child – referred to as scaffolding (Tamis-LeMonda, Kuchirko & Song, 2014). Yet not all vocalisations are responded to for a number of reasons, as parents may not be paying attention to the child, may be speaking on the phone, talking to someone else or otherwise occupied. Depressed mothers are found to show atypical response behaviours such as being either more withdrawn or intrusive in interaction with their infants, who in turn are affected by their mothers’ behaviour (e.g., Field, Diego & Hernandez-Reif, 2009). Under more typical conditions though, an intentional reason for not responding may be that the parent provides the infant with an opportunity to be more active in the interaction, since parents are inclined to see their infant as a conversational partner (e.g., Snow, 1977; Dominguez et al., 2016).

Parental responses come in different modes, such as sounds or vocalisations, words, utterances, facial expressions, touch, gestures and/or movements (e.g., Van Egeren, Barratt & Roach, 2001). Parental responses may be dyadic or extra-dyadic (Tamis-LeMonda, Kuchirko & Song, 2014). Dyadic responses, which are most common in interaction with infants up to about four to five months, are caregiver-directed attention, “face-to-face” interaction. Extradyadic responses direct attention towards the environment, for example towards objects. Parents’ contingent vocal responses to infants can be labels, imitations, affirmations, questions, descriptions or redirectives, depending on context and situation (e.g., Paavola, Kunnari, Moilanen & Lehtihalmes, 2005). Although parents often respond verbally to infant behaviour, parental input in general is multimodal (e.g., Tamis-LeMonda, Song, Leavell, Kahana-Kalman & Yoshikawa, 2012; Tamis-LeMonda, Kuchirko & Song, 2014). Parent responses are embodied in combinations of visual cues, touch, gestures, pointing, and movement with verbal language (Tamis-LeMonda, et al., 2014).

Socio-cultural context (Tamis-LeMonda, Kuchirko & Song, 2014), and socio-economic factors such as educational level and income (e.g., Hart & Risley, 1995), and the sex of the child (Johnson, Caskey, Rand, Tucker & Vohr, 2014), affect parental responsiveness. These are considered fixed factors,
since they are usually rather stable over longer periods of time. Dynamic factors, on the other hand, vary over shorter time scales, and include the level of infant speech and language development, vocalisations and motor control, as well as the activity or conversational setting in which parent and child are engaged (e.g. Tamis-LeMonda, Kuchirko & Song, 2014).

1.3.2 Fixed factors of parental responsiveness

The sections below summarise how parental responsiveness varies with socio-cultural factors, socio-economic status, as well as with caregiver status and the sex of the infant. The majority of studies of parental responsiveness rely on data from mothers, but the text refers to parents in general, unless a specific study involving only mothers is described.

1.3.2.1 Socio-cultural contexts

Socio-cultural contexts may influence parental responsiveness (e.g., Ochs & Schieffelin, 2011; Tamis-LeMonda et al., 2012), because of differences in parenting practice and/or differences in parental knowledge or parental beliefs on child development (e.g., Harkness & Super, 1996; Rowe, Denmark, Harden & Stapleton, 2016). Studies on Western, educated, industrialised, rich and democratic (WEIRD) societies dominate research on parental responsiveness, but some recent studies on non-WEIRD contexts have contributed to the field of infant social interaction and language learning (e.g., Kärtner, Keller, Lamm, Abels, Yovsi, Chaudhary & Su, 2008; Tamis-LeMonda et al., 2012). Even though there are components of parental responsiveness that are thought to be universal across cultures, for example the inclination of mothers to vocally respond to infant vocalisations (e.g., Bornstein, Putnick, Cote, Haynes & Suwalsky, 2015), there seem to be some long-term context-dependent circumstances, for example child-rearing principles that affect the way parents respond to infants. Socialisation goals may differ between cultures according to the degree to which the individual depends on the society (Tamis-LeMonda et al., 2014; Keller, Lamm, Abels, Yovsi, Borke, Jensen ... & Su, 2006). Most WEIRD cultures typically emphasise independence or autonomy of the individual, while in many rural cultures, and in East Asia, interdependency or relatedness is more important than individualism. This causes variation in parental response patterns. In cultures that emphasise autonomy, parental responses are more often reactions to infants’ positive signals. They also occur mostly in face-to-face mode and often concern object stimulation (Keller et al., 2006; Kärtner, et al., 2008). In cultures in which relatedness is central, parents promptly satisfy the physical needs of the child, and respond predominantly with body contact and motor stimulation (Keller et al., 2006; Kärtner, et al., 2008).
Different modes of responding are thought to reflect parenting principles described as proximal versus distal parenting (Kärtner, et al., 2008). For example, parental responses differ between mothers in the USA, France and Japan. In their responses to 5-month-olds, US American mothers were found to be more extra-dyadic than French mothers who in turn were more extra-dyadic than Japanese mothers were. US American mothers emphasised events, objects and actions outside the parent-child dyad in their responses, more than French and Japanese mothers did, compared to social-interactional behaviour such as smiles (Bornstein, Tamis-LeMonda, Tal, Ludemann, Toda, Rahn, ... & Vardi, 1992).

Also, contingency in vocal parent-infant interaction seems to vary across communities. Contingent vocalisations in mother-infant interaction were studied in mothers and their 5½-month-old infants in 11 countries, (Argentina, Belgium, Brazil, Cameroon, France, Israel, Italy, Japan, Kenya, South Korea, and the United States (Bornstein, et al., 2015). Rates of contingent vocalisations in interactions for both mothers and infants varied across communities. Maternal responses were contingent on infant vocalisations overall, except in Cameroon and Kenya, but these countries also had very small samples. Mothers were more responsive than their infants in Belgium, USA and Italy, while mothers and infants were equally responsive in the other communities.

Another study found that US American and Japanese mothers play and use language somewhat differently in interaction with 13-month-olds despite similar socio-economic backgrounds. US American mothers responded more to object play and used more labelling and verbal descriptions compared to Japanese mothers, who in turn gave more responses to social play and demonstrated symbolic play more than US mothers did (Tamis-LeMonda, Bornstein, Cyphers, Toda & Ogino, 1992).

Kärtner and colleagues (2008) studied variations in mothers’ contingent responses to 3-month-old infants in relation to socio-cultural context. They defined three types of socio-cultural contexts according to their socialisation goals. Besides independent and interdependent socio-cultural contexts as described above, they described a third, autonomy-related socio-cultural context. This involves socialisation goals located in between independent and interdependent contexts. Independent contexts were found among mothers in Berlin and Los Angeles. Autonomous-related contexts were found among mothers in Beijing, Delhi and urban Nso in Cameroon and interdependent contexts in rural Nso in Cameroon. Overall, mothers in all contexts gave contingent responses mostly to infants’ non-distressed vocalisations and non-verbal behaviour (in this case, gaze and smiles) by vocalising or using other sounds. However, mothers in independent contexts responded more using infants’ visual modality by using facial expressions and showing objects, compared to mothers in the other two contexts, while mothers in rural contexts responded more using the tactile modality (Kärtner et al., 2008). Luo and Tamis-LeMonda (2016) compared types of responses to infants from mothers with different
cultural backgrounds. Interaction between US American mothers with Mexican, African and Dominican immigrant backgrounds and their infants were studied in a bead stringing activity at infant ages 14, 24 and 36 months. Mothers with Mexican background were more inclined to respond non-verbally and offer non-verbal strategies such as gestures and hands-on guidance for supporting the infant in the activity compared to the mothers with Dominican and African background. Mothers with Dominican and African background, on the other hand, used more verbal strategies such as attention-getting and encouraging language to support their infants in the activity (Luo & Tamis-LeMonda, 2016).

However, parental responsiveness in different socio-cultural contexts is a difficult area to investigate, since random samples can be hard to find and representativeness may be skewed also because of uncontrolled personal biases. Despite some differences across countries, vocal contingency seems to be a universal behaviour that exists in parent-infant interaction in different cultures and communities.

1.3.2.2 Socio-economic status
Parental responses towards infants seem to be affected by socio-economic status (SES) (e.g., Hoff-Ginsberg, 1991; Hart & Risley, 1995, which in turn is linked to knowledge about child development (e.g., Rowe, 2000; Rowe, et al., 2016). In a study by Vanormelingen & Gillis (2016), mothers’ communication with infants aged 6-24 months was compared. High-SES mothers had more utterances per hour, longer durations of speech and more turns per hour than low-SES mothers. Also, high-SES mothers responded significantly more often to their children’s utterances and used more expansions compared to what low-SES mothers did (Vanormelingen & Gillis, 2016).
In another study mothers with higher education talked more and pointed more in joint interaction with their 14-month old infants and they were also more inclined to view the infant as understanding compared to mothers with lower education (Rowe, 2000).

This is in contrast with results from Hoff-Ginsberg (1991) where no differences between high SES and low SES mothers were found according to their reports about infant understanding. However, according to Hoff-Ginsberg (1991), low-SES mothers had fewer contingent verbal responses – defined as topic-continuing replies – to their children aged 18 to 29 months, compared to the mothers with higher SES during the contexts of meal time, dressing, book reading and play.

1.3.2.3 Caregiver gender, caregiver status and sex of the child
Traditionally, mothers have been primary caregivers, in that they carry the main responsibility for the infant, spend the majority of their time with the infant and care for the infant’s needs. As a result, research on infant language development and parental responsiveness has predominantly focused on
mothers. However, some studies focus also on fathers’ impact on infant language development and both differences and similarities between fathers and mothers in their communication with infants have been described.

Just like mothers, fathers adapt their way of speaking with infants (Fernald, Taeschner, Dunn, Papoušek, de Boysson-Bardies & Fukui, 1989; McRoberts & Best, 1997; Rondal, 1980), even though mothers seem to exaggerate the characteristics of infant-directed speech more than fathers (e.g., Fernald et al., 1989). There are other differences in speech characteristics as well as other aspects of language in infant-directed speech that may differ as a function of caregiver status.

In a review by Leaper and colleagues (1998), fathers, in comparison to mothers, were found to speak less often and use less supportive language, but also less negative language when talking with their infants. Fathers also used more directive and informing language than mothers did (Leaper, Anderson & Sanders, 1998). Johnson and colleagues (2014) found that mothers responded to vocalisations from infants more often than fathers did during the infants’ first seven months. Taking into account the amount of time each parent spent with the infant, this means that infants heard about three times as much speech from mothers than they did from fathers (Johnson, et al., 2014). When taking turns with their 24-month-old children, fathers tended to participate in fewer turns and use shorter utterances compared to mothers (Pancsofar & Vernon-Feagans, 2006. Golinkoff and Ames (1979) as well as Hladik and Edwards (1984) investigated mothers’ and fathers’ language in both dyadic (parent-infant) and triadic situations (both parents present). The two studies differ in their findings. Golinkoff and Ames (1979) found that mothers and fathers produced about the same number of utterances in dyadic situations with their 19-month-old infant, but fathers produced fewer utterances than mothers than in dyadic situations. Hladik and Edwards (1984) found that mothers talked more than fathers did in both dyadic and triadic situations with their child between 24 and 42 months. Mothers also produced more complex utterances, in contrast to utterances with simplified grammar that were more common among fathers. However, fathers talked more and used more complete utterances in triadic situations than in dyadic situations. The different results may be due to age differences of the children or to the fact that Golinkoff & Ames’ observations were made in a laboratory and Hladik while Edwards’ in the families’ homes. Leaper and colleagues (1998) in their review found no differences in the number of general parental questions across studies, but fathers used more wh-questions than mothers did, and fathers also asked more questions in dyadic situations than in triadic situations (Leaper et al., 1998). Mothers, on the other hand, asked more questions in triadic situations compared to dyadic situations, and tended to use more yes/no-questions than fathers (Leaper, et al., 1998).

In another study (Wu & Gros-Louis, 2015), mothers and fathers were studied separately in a play situation with their 12-month-old infant. Both mothers
and fathers responded more often with labels to infants’ pointing than to infants’ object-directed vocalisations, that is they used the adequate content word in the situation, for example, “oh, I see the cow is here”. However, mothers used naming, for example, “this is a cow”, more than fathers. Fathers, on the other hand, used more wh-questions and yes/no-questions compared to mothers (Wu & Gros-Louis, 2015).

Socio-cultural factors that are linked to caregiver status may also affect parental communication patterns in interaction with infants. Tamis-LeMonda and colleagues (2004) studied US American low-income mothers’ and fathers’ as they played with their 2- and 3-year-old infants (Tamis-LeMonda, Shannon & Cabrera, 2004). They found that fathers’ support in terms of sensitivity, positive regard, cognitive stimulation, but also intrusiveness (defined as being overcontrolling and overinvolved) at 24 months predicted mothers’ sensitivity, cognitive regard and intrusiveness towards their 36-month-olds. Only mothers’ intrusiveness predicted fathers’ intrusiveness (Tamis-LeMonda et al., 2004). Thus, mothers’ and fathers’ parenting styles seem to be connected to each other.

While mothers have traditionally been the primary caregiver of infants in most cultures, but since the last couple of decades in some WEIRD countries, fathers have been more engaged in early infant care. In Sweden, for example, fathers are encouraged to go on parental leave during the first year, and authorities support this by securing 60 days of the paid parental leave period exclusively to the father (Försäkringskassan, 2018). There are not many studies on fathers as primary caregivers, but in a study by Field (1978), primary caregiver fathers’ interactional behaviour with their child was investigated and compared to primary caregiver mothers’ and secondary caregiver fathers’ interactional behaviour. Primary caregivers, both fathers and mothers, were found to smile more, make more imitative grimaces and also produce more high-pitched imitative vocalisations compared to secondary caregiver fathers. These results indicate that the caregiver status (thus, time spent with the infant) affect interactional behaviour rather than the gender of the parent.

Parent responses also seem to vary according to the sex of the child. For example, mothers of infant boys touch their infants more than do mothers of infant girls when the infants are between three and six months of age, but at ages 7 to 12 months no difference was found (Fausto-Sterling, Crews, Sung, García-Coll & Seifer, 2015). Mothers of 3- to 6-month-old infants match their own facial expression to that of the infant more often if the infant is a boy (Malatesta & Haviland, 1982). In another study, both mothers and fathers of boys used more explanations than mothers and fathers of girls (Kruper & Uzigiris, 1987).

Parents have been found to spend on average more time vocalising to daughters than to sons (Leaper, et al., 1998). This has also been supported by Johnson and colleagues who found that mothers respond more to new-born girls and when their girls are ten months old, while fathers showed a trend
towards preferring to respond to boys at ten months (Johnson et al., 2014). Further, mothers of boys used longer utterances in interaction with their child from two to 24 months, compared to mothers of girls (Ahl, Fausto-Sterling, Garcia-Coll & Seifer, 2013; Gratier, Devouche, Guellai, Infanti, Yilmaz & Parlato-Oliveira, 2015). An earlier study showed that fathers of 18-month-old infants talked more to boys than to girls, while mothers talked more to girls than to boys (Weinraub & Frankel, 1977), but even more recent research shows that mothers’ vocal responses to new-borns were more frequent to girls than to boys (Johnson, et al., 2014).

Differences were also found in Australian mothers’ speech registers and prosodic modifications when speaking to their infant boys versus girls at ages 3, 6, 9, and 12 months old: Mothers used a greater pitch range when talking to girls than to boys, and these pitch range differences increased with infant development (Kitamura & Burnham, 2003).

1.3.3 Dynamic factors of parental responsiveness

Besides the fixed factors of parental responsiveness, parental responses also seem to be affected by dynamic factors, that is, factors that change over a relatively short time period. These include the developmental state and the non-verbal and verbal communication of the child, the particular contextual setting or activity, and the health status of the parent.

1.3.3.1 Non-verbal behaviour of the child

Children’s use of gestures, gross and fine motor development, facial expressions and gaze have been shown to influence parental responsiveness, including response type. Pointing gestures typically emerge around 12 months of age (e.g., Carpenter, et al., 1998) and there seems to be a correlation between pointing gestures and language (e.g., Brooks & Meltzoff, 2008). In a meta-analysis by Colonnesi, Stams, Coster & Noom (2010) possible relations between pointing gestures and language outcomes are described.

Both mothers and fathers label objects more often when infants point to them than when infants make object-directed vocalisations (Wu & Gros-Luis, 2015). Infant gestures are also found to elicit gestural responses from mothers (LeBarton, Goldin-Meadow & Raudenbush, 2015). There are differences found in how mothers respond to infants depending on their gross motor development. Mothers respond to children’s communicative initiatives differently if the child has learned to walk or not; responses to sitting or crawling infants were often labels, such as “book”, while responses to infants who can walk were more often predicate phrases such as “want to read?”, even when infant communicative initiative was controlled for (Karasik, Tamis-LeMonda & Adolph, 2014). The results suggest that verb use in parent responses is related to infant mobility.
With infant development, mothers’ referential language tends to increase and directive language to decrease when they respond to infant actions, and between 14 to 24 months specifically, mothers tend to decrease responses to gestures (Tamis-LeMonda, Kuchirko & Tafuro, 2013). When mothers and their 14-month-old infants were engaged in book reading and a bead-on-a-string activity, mothers were more likely to use contingent verbal responses (within three seconds) to task-relevant infant object exploration compared to when the infant shows off-task behaviour (Tamis-LeMonda, Kuchirko & Tafuro, 2013).

1.3.3.2 Vocalisations, speech and language of the child

As described above, parents respond to various infant behaviours, and several studies have specifically shown the relationship between infant vocal behaviour and parental responses. Parents can identify mature syllables in babbling without training (Oller, Eilers & Basinger, 2001), and they also rate infants’ characteristics in terms of communicative intent, interest and social favourability when listening to their babbling (Papoušek, 1989; Beaumont & Bloom, 1993). However, not all types of infant vocalisations have communicative intent, especially not the early ones.

Infants’ grunts – composed of a laryngeal closure that is followed by a vowel-like release – are among the first infant vocalisations. At first, grunts are expressed as accompanying physical effort, then as an expression for attention, after which they develop further towards having communicative intent. Further, once the infant has started to use grunts for communicative purposes, the likelihood of joint infant-parent attention increases (McCune et al., 1996).

Parents have been found more vocally responsive to infants’ early sounds and vocalisations (ages 2 to 14 weeks) if the infant’s vocalisation is interpreted as positive rather than negative (Keller & Schölmerich, 1987). In this study, positive vocalisations were described as

“... uttered with low to moderate intensity and are formed in the process of gentle exhalation. They are commonly thought to express positive emotions and include ‘a’-sounds, happiness sounds, babbling, cooing, blurtling, consonant-vowel groups, laughing, ‘r’-strings, and repetition sounds.” (Keller & Schölmerich, 1987, p 63).

Negative vocalisations were sounds interpreted as sounds of discomfort, such as whining, fussing, crying and sighing. Positive vocalisations were predominantly verbally or vocally responded to by the parent, with eye contact. Negative vocalisations, as well as vocalisations expressing effort and physiological vocalisations (respiratory or throat-closing sounds) were to a greater extent responded to by changes in movements, or tempo of movements by the parent (Keller & Schölmerich, 1987).
Babbling has been thought to function as a tool to organise social interaction with caregivers in order for infants to access the information in parents’ contingent responses (Albert, Schwade & Goldstein 2017). Hsu and Fogel (2003) found that the occurrence of non-distress vocalisations from infants aged two weeks to six months, was the main factor regulating vocal maternal responses, not the speech quality of the infant’s vocalisations. However, other studies indicate that parental responses depend on the quality of the infant’s vocalisations. In a study by Beaumont and Bloom (1993), infants from three to five months vocalising with full voice, syllabic and with melodic complexity were considered as having more communicative intent compared to the infants using mostly vocalic sounds. Parents’ responses to babbling are found to be related to the quality and the direction of babbling of 9-month-olds (Albert et al., 2017). Less mature vocalisations, for example pre-canonical babbling, elicit fewer responses while more mature pre-linguistic vocalisations, such as canonical syllables, tend to elicit more parental responses, and parental interpretations suggesting that the infant is “wanting” something (Goldstein & West, 1999). Mothers tend to respond predominantly with play vocalisations to less mature vocalisations and with imitations and expansions to more mature vocalisations (Gros-Louis, West & King, 2016). Mothers show significantly more vocal responses to infant vocalisations compared to interactive responses such as gaze, smile and touch in 8- to 9-month-olds (Gros-Louis, et al., 2006).

Parents use many simple labels and descriptions in speech to infants. They use fewer words, less complex grammar and less diverse language with less linguistically developed infants, compared to more linguistically advanced infants (e.g., Soderstrom, 2007; Tamis-LeMonda, et al., 2012) Mothers tend to respond to children’s exploratory and communicative actions with didactic language that has informative and pedagogic attempts, rather than regulatory language that aims to control behaviour (Tamis-LeMonda, Kurchirko & Tafuro, 2013). Mothers are more likely respond to their 24-month-olds when they use new words (Masur, 1997). Additionally, the number of responsive questions from parents increases over the second year as the child becomes more skilled in language (Bornstein, Tamis-LeMonda, Hahn, Haynes, 2008). Thus, parental responses seem to vary according to the quality and direction of the infant’s vocalisation, and the infant’s maturity in terms of speech and language use.

1.3.3.3 Conversational settings
Parents usually respond to infants’ initiatives (e.g., Dominguez, et al., 2016; Gros-Louis, et al., 2006). Studies of parental responsiveness and its effect on child language development may use goal-oriented situations such as play or book reading, but also free activities in both laboratory settings and in families’ homes. Hoff-Ginsberg (1991) has showed differences in mothers’ speech to 18- to 29-month-olds in four different daily situations: book reading, meal
time, getting dressed and toy play. These four situations differed in lexical diversity, grammatical complexity, overall speech rate and contingency, the latter defined as topic-continuing replies (Hoff-Ginsberg, 1991). Mothers’ contingency was greatest during the book reading situation, while all four parameters showed complex patterns in the other situations. Mothers’ speech during toy play had the highest number of directives and the lowest number of conversation-eliciting utterances of all situations. During mealtime, mothers’ numbers of utterances were the lowest and had, together with the maternal utterances during the dressing situation, the lowest rate of conversational utterances as well as the lowest lexical diversity (Hoff-Ginsberg, 1991).

In another study, book reading elicited more parental responses to infant vocalisations compared to for example toy playing situations (Gros-Louis et al., 2016). Parents tended to interpret the infants’ vocalisations during book reading as having meaning, presumably supported by the book that restricted the possible meaning of the vocalisations. Labelling, the use of contextually relevant content words, and acknowledgements, the use of confirming responses like “mm-hmm”, “uh-huh” were the most common responses in the book situation (Gros-Louis, et al., 2016).

Further, mothers use more referential language during a book reading session and more directive language during a bead stringing task (Tamis-LeMonda, Kuchirko & Tafuro, 2013).

Parents’ responses vary according to different conversational settings or activities, and book reading seems to be one activity where parent responses are frequent. However, different situations and activities are of course differently appreciated by parents, which may affect interaction with the infant. Book reading, for example, may be a listening activity or have a more conversational focus. In a study by Bus, Leseman and Keultjes (2000) parents with less interest in reading themselves were less inclined to deviate from the text to discuss or negotiate with the infant, compared to parents more interested in reading.

1.3.3.4 Parental emotional health status
Parental responsiveness also varies according to parental health status. For example, in interactions with 4-month-olds, depressed mothers showed less frequent touch responses than non-depressed mothers (Mantis, Mercuri, Stack & Field, in press), and depressed fathers are less responsive in play with their 3-month-old infants compared to non-depressed fathers (Parfitt, Pike & Ayers, 2013). Both mothers’ and fathers’ depressive symptoms are negatively associated with positive activities such as reading, singing and telling stories to the infant (Paulson, Dauber & Leiferman, 2006).
Responsive behaviour to a child’s needs and thoughts is necessary for any child growing up, and associations have been found between responsiveness and infant development in several studies. Parental responsiveness to infant behaviour influences infant attention (e.g., Bornstein & Tamis-LeMonda, 1997; Miller, Ables, King & West, 2009), exploratory and social communicative behaviour (e.g., Kuchirko, Tafuro & Tamis LeMonda, 2018), language comprehension and language use (e.g., Hart & Risley, 1995; Baumwell, Tamis-LeMonda & Bornstein, 1997; Rollins, 2003).

Even though there are some differences found in mothers’ and fathers’ responses to infants, both mothers’ and fathers’ responsiveness are found to predict later child communicative skills (e.g., Shannon, Tamis-LeMonda, London & Cabrera, 2002; Tamis-LeMonda, Shannon & Cabrera, 2004). Responsiveness of adopting parents is found to be related to language outcomes also in adopted children (Stams, Juffer & van IJzendoorn, 2002), and responsive techniques used in intervention enhance children’s language learning (e.g., Girolametto, Pearce & Weitzmann, 1996; Landry, Smith, Miller-Loncar & Swank, 2008).

Contingent parent vocal responses have positive effects on prelinguistic infant vocalisations. It has been shown that 8-month-old infants (Goldstein, King & West, 2003), and 9.5-month-old infants (Goldstein & Schwade, 2008) who received contingent vocal responses from their mothers changed their vocalisations to be more mature, while infants receiving non-contingent responses did not. Gros-Louis and colleagues (2014) found a positive association between mothers’ responsiveness to infants aged 6-14 months and an increase in infants’ advanced consonant-vowel vocalisations, and that mothers’ responsiveness correlate with an increase of mother-directed infant vocalisations (Gros-Louis, West & King, 2014).

The use of didactic, referential language supports child vocabulary growth, as labelling and descriptions will give opportunities for the child to make connections between phonetic forms (words) and what they represent. For example, parents’ labelling responses to infants’ object-directed vocalisations at nine months have been found to correlate to vocabulary size at age 15 months (Goldstein, Schwade, Briesch & Syal, 2010), while labelling of non-present object is negatively correlated with vocabulary size (e.g., Goldstein et al., 2010). Number of responses conceptually linked to the infant’s focus and action are positively related to language learning, while re-directive responses and responses not related to joint focus correlate negatively with later language development (Tomasello & Farrar, 1986; Akhtar, Dunham & Dunham, 1991; Paavola et al., 2005). Maternal vocal responses to their infants at 9 and 13 months predict the timing of infants’ 50 first words (Tamis-LeMonda,
Bornstein & Baumwell, 2001), and infants who receive more maternal responses to their vocalisations score higher both for vocabulary production and gestures on the CDI at 15 months (Gros-Louis, et al., 2014).

Temporal contingency in parents’ conversations with their 18-month-olds is also found to be correlated with vocabulary size in Swedish infants. Parents of infants with large vocabularies respond faster to their infants, compared to parents of infants with typically sized vocabularies, who in turn respond faster than parents of infants with small vocabularies (Marklund, Marklund, Lacerda & Schwarz, 2015). In addition to contingency, frequency in responsiveness is an important factor for phonological development (e.g., Kuhl, 2007) and vocabulary development (e.g., Goodman, Dale & Li, 2008).

Parental responsiveness in the second year predicts not only vocabulary size but also timing of language milestones (e.g., Tamis-LeMonda et al., 2001). Maternal contingent responses at infant age nine months are related to infant productive syntax at 30 months (Rollins, 2003), and mothers’ responsiveness to 13-month-old infants predicted grammar in terms of combinatorial speech and ability to talk about the past at 21 months (Tamis-LeMonda et al., 2001).

Infant language development is affected by parental responsiveness, that in turn varies according to different conversational settings. For example, variability in maternal language during meal times was found to be associated with the subsequent vocabulary development of children (Beals, 1997; Dickinson & Tabors, 2001; Weizman & Snow, 2001). The book reading situation has shown to promote responsive behaviour in parents and also parents’ labelling and use of diverse language (Hoff-Ginsberg, 1991; Gros-Louis et al., 2016). However, such variability within maternal interactional patterns in the book situation has not been found to be related to child language outcomes (Weizman & Snow, 2001), instead the amount of time spent in book-sharing with an adult was related to the development of the child’s vocabulary (Payne, Whitehurst & Angell, 1994; Scarborough & Dobrich, 1994).

In summary, parents are responsive to infants’ various expressions in different ways due to fixed long-term factors and more dynamic short-term factors, and parents often automatically vary mode, content and timing of their communication. Parents offer contingent responses that seemingly support infants’ social, cognitive and language development. Parents’ speech directed to infants offers various informative perceptually salient cues in the social context by being temporally contingent, and the infant can with time use these cues for word learning, and the development of social communication.

1.4 Development of communication and language

This section describes the infant’s early growing skills in communication and language learning. The section starts with a general introduction on infants’
communication and language development. With parental responsiveness as the base, the next section describes the development of communicative skills, first the non-verbal communication/socio-pragmatic development, and then the development of vocal turn-taking. The following section describes infants’ phonological development from early vocalisations and babbling to word production from an ecological phonetic perspective, and the last section describes specifically the phonological development in Swedish, primarily on a segmental level.

Although phonological learning continues to develop with more advanced vocabulary and morphology, this thesis focuses primarily on descriptions of early development of expressive phonology.

1.4.1 Introduction

Shared intentionality, joint attention and theory of mind are crucial for developing understanding and use of social language (e.g., Bruner, 1983). Joint attention is the ability to coordinate one’s focus with someone else’s. Theory of mind is the ability to attribute mental states to oneself and others, and the understanding that other people have other beliefs, thoughts and perspectives than oneself. Infant development of such social communicative behaviour and engagement in interaction are skills developing with cognitive maturation, but they are also known to be influenced by parental responses. By being conceptually and temporally contingent and also recurrent, parental responses support the development of interaction, non-verbal communication and understanding and use of verbal speech and language (e.g., Rollins, 2003).

In phonological development, universal learning processes are the basis for the use of context-dependent input to form a phonological system of the target language(s). The infant’s development towards the native language has started already before birth (DeCasper & Spence, 1986). Long before the infant has uttered the first words, perceptual learning of linguistic information in parents’ speech forms the basis for the development of the phonological system of the target language(s) (e.g., Kuhl, 2004). Infants’ vocal productions develop from immature vocalisations and babbling to more mature productions, and this takes place in interaction with experienced language users – most often the responsive parents. Thus, infants’ vocal development is embedded in a social feedback loop where infant and parent influence each other (e.g., Goldstein & Schwade, 2008; Warlaumont, Richards, Gilkerson & Oller, 2014). The development of expressive phonology is a universal process based on infant capacities and ecological settings. Taking on a language-specific perspective when describing development of expressive phonology offers descriptions of infants’ productions in relation to the target language, but at the same time, universal patterns are revealed in these descriptions. A phonetic perspective in describing the development of expressive phonology will further help to understand infant and child productions.
1.4.2 Early development of communicative skills

In the light of parental responsiveness, the development of infants’ early communicative skills is described. Non-verbal/non-vocal communication refers to gaze, smile, pointing and other gestures and movements. Vocal turn-taking concerns interaction with the use of vocalisations and/or spoken language.

1.4.2.1 Development of non-verbal/non-vocal communication

Before infants become verbal and use words, they vocalise, move and look at people and objects in the environment. Parents typically interpret and respond to non-verbal infant behaviour and expressions in various ways. This starts early. Parents respond even to very young infants’ actions as if the infant was communicating (e.g., Snow, 1977), supporting the development of interaction. The engagement in interaction – primary intersubjectivity – is thought to be innate (Trevarthen, 1979) and to be present already in the prenatal period by rhythmic turn-taking between mother and infant (Trevarthen, 2011) and rudimentary imitations (Meltzoff & Moore, 1977, but see also Suddendorf, Oostenbroek, Nielsen & Slaughter, 2013; Oostenbroek, Redshaw, Davis, Kennedy-Constantini, Nielsen, Slaughter & Suddendorf, 2018). However, the development of intersubjectivity cannot merely be a product of imitation; it is also dependent on reciprocal social exchange (Rochat & Passos-Ferreira, 2008).

Infants are sensitive to patterns, can detect contingency early and seek to match and recognise stored patterns. For example, Bertin & Striano (2006) have shown that the still-face paradigm which consists of a person looking at the infant with a non-responsive, expressionless face, increases negative affect in infants as young as 1.5 months. The first smile in infants usually appears around two months of age (Messinger & Fogel, 2007), and at that age, infants are engaged in proto-conversations with caregivers (e.g., Bateson, 1979; Csibra, 2010). Caregivers usually respond to infant smiles within two seconds (Messinger & Fogel, 2007). At five months, infants are capable of social smiling (Messinger & Fogel, 2007). Ten-month-olds who turn around to see why their chair they are sitting on is pushed forward (indicating they search for a cause of an effect) show increased capability to coordinate vocalisations in social interaction (Harding & Golinkoff, 1979), and 8- to 14-month-olds respond to maternal contingency with more caregiver-directed vocalisations (Gros-Louis et al., 2014; Miller & Gros-Louis, 2013). This shows that infants in the first year of life both affect social communication and interaction and are affected by it.

Secondary intersubjectivity is when the child acts upon the environment and the parent responds to these actions. This type of interaction offers further possibilities for the infant to develop an understanding of social interaction and intentional, referential communication and joint attention. It usually emerges around the end of the first year (e.g., Striano & Rochat, 1999; Rochat & Passos-Ferreira, 2008).
According to Tamis-LeMonda and colleagues, interaction with responsive parents facilitates child development of secondary intersubjectivity in the child, and in turn expands the possibilities of learning language (Tamis-LeMonda, Adolph, Lobo, Karasik, Ishak & Dimitropoulou, 2008). An indicator guiding infant attention to third objects in interaction and thereby initiating the development of secondary intersubjectivity is parental gaze. For example, a mother’s gaze helps an infant to redirect attention from the mother’s face to the object (Beier & Spelke, 2012; Yu & Smith, 2015). This joint attention helps the infant to interpret the mother’s intentions and leads to infant responses that promote interaction, aligned between parent and child (Déak, Krasno, Triesch, Lewis & Sepeta, 2014).

Mothers’ responses to infants’ initiatives in a joint-attention framework have been shown to be positively related to later language outcomes (Tomasello & Farrar, 1986; Tamis-LeMonda et al., 2001). Joint mother-infant activities in interaction with 15- and 21-month-old infants yield more utterances than non-joint activities (Tomasello and Farrar, 1986). However, general measures of maternal speech such as amount of speech input or mean length utterance (MLU) did not correlate with infant vocabulary, but different types of object referents did. Mothers’ use of gestures and object labels following infants’ attention focus on the object has been shown to be positively correlated with infant vocabulary size. On the other hand, mothers’ re-directive responses correlated negatively with infant vocabulary size, but not if maternal re-directives were accompanied by maternal gestures in infant focus (Tomasello and Farrar, 1986). The authors describe joint attention as a transactional process, a loop in which joint focus scaffolds infant language development, helping the infant at the same time to maintain interaction for longer – which in turn supports language learning.

Infants actively take part in maintaining interaction. Rutter and Durkin (1987) showed that infant gaze behaviour approximates the adult pattern at around 18 months. They examined what they called “the terminal look”, defined as an upward look at the conversational partner to hand over the turn. Infants used the terminal look already at 12 months, increasingly at 18 months and steadily at 24 months (Rutter & Durkin, 1987). Besides vocalisation, gaze is the dominant modality used in infant communication. Gaze is a signal in turn-taking and for joint attention. Gaze timing is crucial both to signal interactional turns and for the infant to learn language as it means opportunities for the infant to link a label to the object or action in visual focus.

With the development of joint attention at the end of the first year, the infant is more engaged in turn-taking. Both infants and mothers have been shown to speak more during joint activities compared to in other situations, and infants learn more language if the maternal references to objects inside joint activities are in focus of the infant (Tomasello & Farrar, 1986). However, language outcomes are not related to joint attention only, since infants with autism can learn language without joint attention abilities (e.g., Rollins, 2003).
Another type of non-vocal communication is the pointing gesture, that usually emerges around 12 months of age (e.g., Leung & Rheingold, 1981; Carpenter, Nagell, Tomasello, Butterworth & Moore, 1998). Pointing is a gesture that directs the extended arm, index finger or other body-parts towards an object, person or location (Liszkowski, Carpenter, Henning, Striano & Tomasello, 2004). Pointing does not necessarily include eye contact (Goldin-Meadow, Goodrich, Sauer & Iverson, 2007). Pointing may develop from the infant’s attempt to reach for something, and may therefore not from the start be an action that serves to share attention, but it can develop to a referential communicative act (e.g., Tomasello & Farrar, 1986), a precursor to labelling (Werner & Kaplan, 2014).

Tomasello and colleagues describe social and communicative intentions that can underlie the act of referential pointing (Tomasello, Carpenter & Liszkowski, 2007). A social intent can comprise a request or a declaration of emotion or information and is often expressed by facial or other emotional expressions. The infant may want an object, help, something to happen, or to inform about something. A communicative intent is often expressed with distinct eye contact and leading to the infant and the adult being aware that communication occurs in a shared context – they share “common ground”. By expressing these social and communicative intentions, pointing is also used for referentially (Tomasello et al., 2007). Just like speech, pointing can direct attention to external events and objects. It can also represent and influence others’ mental states as well as motivation to communicate (e.g., Liszkowski et al., 2004, Liszkowski, 2008; Tomasello et al., 2007).

Pointing often elicits verbal responses from others, which in turn facilitate language learning (Kishimoto, Shizawa, Yasuda, Hinobayashi & Minami, 2007; Goldin-Meadow et al., 2007). When an infant points, the parent often takes the opportunity to teach and give information (Southgate, Van Maanen & Csibra, 2007; Begus & Southgate, 2012). Rates of maternal labelling after infant pointing predict later infant object-labelling (Masur, 1982). Frequency of infant pointing predicts subsequent receptive and productive language development (e.g., Camaioni, Castelli, Longobardi & Volterra, 1991; Blake, Vitale, Osborne & Olshansky, 2005; Brooks & Meltzoff, 2008; Colonnese, et al., 2010). Pointing relates more to vocabulary development than other infant gestures (e.g., Blake et al., 2005). The onset of pointing has shown to be linked to animal sound comprehension and gesture production at 14 months (Butterworth & Morrisette, 1996), and the number of objects pointed to at 14 months predicts vocabulary comprehension at 42 months (Rowe & Goldin-Meadow, 2009).

1.4.2.2 The development of vocal turn-taking
The development of vocal turn-taking is one aspect of the development of communication in children. In vocal interaction, typically two communicative partners – a dyad – switch between vocalising and pausing: they take turns.
Turn-taking behaviour can be described in relation to different aspects of the conversation, such as timing of utterances and pauses and/or conceptual contingency in response types. A conversational turn is typically defined as one or two switches between speaker and listener: first, one speaker’s vocalisation is followed by silence and then the other speaker’s utterance (e.g., Levinson, 2016). An utterance can be described in terms of its linguistic content, its communicative intent and its duration. A pause is usually defined as silence between utterances. Pause duration can be difficult to measure as it depends on prosodic cues, on the linguistic content of the utterances and on speech style variation, for example speech rate (Heldner & Edlund, 2010). A pause between speakers in a turn-taking situation is often referred to as a gap or switching pause, while silence between utterances from the same speaker is an intrapersonal pause. If two people are speaking at the same time, there is an overlap of speech – a negative pause (Heldner & Edlund, 2010). Typically, two speakers match the duration of their intrapersonal and switching pauses during a conversation. This is often referred to as “vocal congruence” (e.g., Beebe, Alson, Jaffe, Feldstein & Crown, 1988; Pardo, Jay, Hoshino, Hasbun, Sowemimo-Coker & Krauss, 2013).

Parents typically adapt their vocal responses to infant age and developmental state, but independent of age or developmental state, parents see the infant as a communicative partner (Snow, 1977). Both conceptual and temporal contingency in vocal turn-taking between infants and parents vary according to fixed as well as dynamic factors, having an impact on infant communication and language development. The timing aspects of vocal turn-taking, such as utterance duration and pause duration are important features when adjusting speech input to the listener, and those aspects are even more important when speaking to a child with a language disorder (Vigil, Hodges & Klee, 2005). In order for vocal turn-taking to work smoothly, each interaction partner needs to be sensitive to communicative initiatives from the other partner (Duncan, 1972). Infants show early sensitivity towards timing in interaction. For example, reciprocal vocal turn-taking patterns are already found in new-borns, in that infants respond with vocalisations to mothers’ infant-directed vocalisations within five-second units (Rosenthal, 1982). These early turn-taking instances occur only within parent-infant dyads and when the new-born is in skin contact with the parent, and are found as early as 15 minutes after birth (Velandia et al., 2010). Dominguez and colleagues (2016) studied turn-taking abilities in 2- to 4-day-old infants interacting with their mothers and found that very few infant vocalisations were isolated, which means that most vocalisations were either preceded or followed by a maternal vocal response or followed by a second infant vocalisation, which indicates early turn-taking (Dominguez et al., 2016). Infants at three months can engage in vocal interaction as vocal onsets were shown more likely to occur in both infants and mothers if the other dyad member also vocalised (Anderson, Vietze & Dokecki, 1977). However, parents still take the main responsibility for the
conversational structure of the interaction with infants at this age (Snow, 1977).

Vocal turn-taking in an adult-like fashion has been observed in infants at four months. From this age, vocal overlaps decrease and more alternations take place in turn-taking (Elias, Hayes & Broerse, 1986; Elias & Broerse, 1996; Ginsburg & Kilbourne, 1988). Studies of vocal congruence in mother-infant dyads found matched switching pause duration (Beebe et al., 1988), but not matched duration of intrapersonal pauses, which naturally occurs in adult conversations of a certain length (Braarud & Stormark, 2008).

At five months, infants are found to associate their own vocalisations with responses from parents. This has been shown by a still-face experiment, in which one minute of naturalistic interaction was followed by a two-minute still-face episode, and then one minute of naturalistic interaction with responsive facial expression. During the still-face episode, infants showed first an increase and then a decrease in vocalisations, suggesting that infants have learnt by five months that vocalisations elicit responses from others (Goldstein, Schwade & Bornstein, 2009).

Pause duration has been shown to be influenced reciprocally in 9-month-olds and their mothers (Jaffe, Beebe, Feldstein, Crown, Jasnow, Rochat et al., 2001). Switching pauses following maternal utterances were longer in mother-infant dyads than switching pauses following a speaker in adult-adult conversations. This has been shown in interactions with 4-month-olds, (Beebe et al., 1988), 9-month-olds (Jasnow & Feldstein, 1986), as well as 4 to 5-year-old children (Welkowitz, Bond, Feldman & Tota, 1990). Hilbrink and colleagues (2015) studied vocal interaction longitudinally in 12 mother-infant dyads (seven boys) at the infant ages 3, 4, 5, 9, 12 and 18 months to describe turn-taking development. They examined the development of pause duration in turn-taking and the amount of overlaps, more specifically infant switching pauses, overlaps and within-turn pauses (switching pauses and overlaps were examined separately). Taking gap and overlap duration as an indicator for adult-like turn-taking based on anticipation of upcoming turn switches, the authors’ hypothesis was that infant gaps (switching pauses between parent and infant utterance) would become longer at around 12 months, around the time of the child’s first words. Their data shows, however, that this apparently occurs even earlier, at nine months (Hilbrink, Gattis & Levinson, 2015). Also, the number of infant overlaps decreased with age, while Rutter & Durkin (1987) had found earlier an increase in overlaps between 9 and 24 months after an initial decrease. However, Hilbrink and colleagues (2015) found that overlap duration remained stable over time. They argue that this duration stability indicates that infants do not yet reach for the end of the preceding turn as adults do (Hilbrink et al., 2015). On the other hand, Casillas & Frank (2013) found that infants at 12 months can anticipate their upcoming turn in conver-
sations (Casillas & Frank, 2013). This behaviour is exclusive to vocal conversations as infants will not expect a new turn to follow a non-speech utterance (Thorgrimsson, Fawcett & Liszkowski, 2014).

At 12 months, infants seem to have at least some ability to identify turn-taking features, but they need to further develop their interaction skills. By the end of the second year, children start to play an active role in maintaining interaction as they actively structure their vocal interaction and time their utterances in a more adult-like way, with fewer overlaps (Rutter & Durkin, 1987).

Besides the timing of parental responses, the content of parental speech affects the development of turn-taking. The quality of infant vocalisations in turn-taking is associated with contingent vocal responses: 3-month-old infants’ vocalisations to contingent adult responses follow a speak-listen pattern, while random adult interaction incurs infant responses in more bursts of vocalisation (Bloom, Russell & Wassenberg, 1987). Infants from one to six months were more likely to produce vocalic and syllabic sounds in symmetrical interaction, that is mother and infant are mutually engaged in interaction, compared to in asymmetrical interaction, where the mother is more active than the infant (Hsu & Fogel, 2001).

Parental switching pause duration has been shown to be related to 18-month-old infants’ vocabulary size: parents of infants with large vocabularies respond faster than parents of infants with typical vocabularies who in turn respond faster than parents of infants with small vocabularies (Marklund, et al., 2015).

Different contexts and activities influence parental behaviour (see section 1.3.3), and these differences affect infant learning. For example, 12-month-old infants vocalised more in interaction with their mothers in a book reading situation compared to in a toy playing situation (Gros-Louis et al., 2016). The infants also produced more mature vocalisations – operationalised by more consonant-vowel vocalisations compared to vowel-only vocalisations – in the book reading context, even when maternal responsiveness was controlled for. Vocal turn-taking behaviour therefore does not develop separately from the development of language skills in other domains, but rather the two are intimately intertwined.

1.4.3 Phonological development

As developmental constraints in the child set the limits for speech sound production, studies of phonological development most often describe children’s phonology in relation to the target language, and many studies are related to speech intervention. However, the difficulty in the description of phonological development does not primarily lie in describing phonology in speech production during development; it rather lies in how to define phonology and relate production to the basic processes of language development (as described in
section 1.1). Phonology reflects the conventional adult description of established language forms which makes the child’s production difficult to assess, as the infant is still learning the language.

Phonological development has been studied in many languages, and one recurrent question is what can be seen as universal and what is language-specific. From an ecological phonetic perspective, the infant’s affordances and the learning processes involved are universal, but language learning must by definition converge to the ambient language and is therefore context-dependent at the same time. As infants’ perceptual-motor and cognitive abilities provide the base for statistical learning of incoming information, this is also valid for phonological form. Infants learn the phonology of the language(s) they are exposed to, and the amount and timing of exposure will affect infant learning. Thus, parents’ contingent and frequent responses affect infants’ phonological learning.

This section describes the development of infant early vocalisations and babbling, early expressive phonology, and early words, mostly in English-learning children. In the last part of this section, the development of Swedish phonology is described, predominantly on a segmental level.

1.4.3.1 Early vocalisations and babbling

Due to their small resonating oral and nasal cavities and limited lung capacity, the new-born has different resonance capabilities compared to adults, but cries, whines and produces other types of sounds. In the earliest stages of vocal production at two to four months of age, the infant produces voluntary comfort sounds – cooing – either spontaneously or in response to talk or smiles. Adults can distinguish between different infant cries and comfort sounds (Green, Jones & Gustafson, 1987; Papoušek, 1989), and parents typically respond to these early expressions. These early vocalisations are typically short and are often perceived as pharyngeal because of the proportions of the infant vocal tract. Later on, and especially if the infant is focused on something, these sounds become longer, and when repeated, separated by glottal stops. Goldstein and colleagues (Goldstein, et al., 2010) suggest that infant grunts have some self-stimulating purpose, as if the infant needs to practice before using vocalisation for communication. Grunts associated with effort and attention are common in 9- to 12-month-olds, while grunts with communicative intent are used from around 12 months (McCune et al., 1996).

Already as early as from three months, parental responsiveness is thought not only to be important for turn-taking abilities, but also to affect the quality of infant vocalisations (Bloom et al., 1987). Responsiveness also affects the quality of vocalisations in 8-month-olds (Goldstein, King & West, 2003), 9-month-olds (Goldstein & Schwade, 2008), and 14-month-olds (Gros-Louis et al., 2014). Parents show a preference for syllabic infant vocalisations over shorter, vocalic sounds that are more nasal (Masataka & Bloom, 1994). Tem-
porally contingent responses from adults – as opposed to temporally non-con-
tingent – enhance more mature vocalisations in 3-month-old infants (Bloom et al., 1987).

At around four to seven months, the infant has gained better control over articulation, and experiments with variations in production of fundamental frequency and intensity of the voice, and different periodic and aperiodic sounds. The infant uses fricative, nasal, bilabial, and uvular sounds that can be interpreted as “consonantal” and that together with more adult-like vowel sounds form productions beginning to resemble single CV-syllables – the foundation of babbling (MacNeilage, Davis, Kinney & Matyear, 1999; Mac-
Neilage, Davis & Matyear, 1997; Oller, Eilers, Neal & Schwartz, 1999). However, the timing of the speech sound transitions is not yet adult-like. From around eight months of age, the infant starts reduplicating these CV-syllables, resulting in canonical or reduplicated babbling. Babbling contains syllable se-
quencies that are similar to adult syllables. The babbling phase usually has a sudden onset (Oller et al., 1999). The syllables consist of a complete or partial supra-glottal closure that forms a consonant-like sound that is then repeated with a vowel between repetitions. The transitional timing between consonants and vowels is now more adult-like, which together with the temporal regular-
ity of the sequences makes infant productions resemble adult speech. The syll-
able production rate during babbling also closes in on the adult speech rate (Dolata, Davis & MacNeilage, 2008).

Parents are able to identify CV-vocalisations among other types of infant vocalisations. Adults interpret infants as being more intentionally communi-
cative when they use CV-vocalisations (Beaumont & Bloom, 1993) and in-
fants’ vocalisations contain acoustical patterns that can be interpreted as emo-
tive (Papaeliou, Minadakis & Cavoura, 2015). Late onset of canonical bab-
bling is shown to correlate with later language outcomes (Oller, Eilers, Neal & Cobo-Lewis, 1998; Oller et al., 1999).

At around 11 months of age, the infant starts producing variegated babbling (Oller et al., 1976; 1999; Oller et al., 1999). This type of babbling sounds even more like adult speech as it is not built on syllable reduplication; instead the speech sounds in the syllables vary. Therefore, word-like structures are often heard. Canonical and variegated babbling may at first co-occur. With time, variegated babbling increases dramatically. Temporally contingent ma-
ternal responses have been shown to be associated with more mature vocali-
sations and learning of phonological categories also in 8- and 9-month-olds, while no such learning occurs if responses from parents are non-contingent (Goldstein, King & West, 2003; Goldstein & Schwade, 2008).

1.4.3.2 Early development of speech sounds

Studies of deaf and hard-of-hearing infants (Oller & Eilers, 1988; Stoel-Gam-
mon & Otomo, 1986) as well as infants who during a period have been unable to vocalise because of tracheostomy (Locke & Pearson, 1990) have shown that
early vocalisations and babbling are precursors to mature speech. Babbling rate at six months has been found positively related to speech accuracy at 30 months (Scherer, Williams, Proctor-Williams, 2008). Thus, babbling provides both practice of articulation as well as experience of the link between production and perception to the infant (Vihman, 1993; Schwartz et al., 2012).

With babbling as a base for further development towards functional speech, a universal perspective sheds light on general learning principles for infants’ language development, while a language-specific perspective will describe how phenomena such as speech sound inventories, frequencies and functions of different structures may affect the development of the target phonology.

The phonological inventory and the frequency of different structures of a specific language, for example phonological complexity and vowel/consonant ratio, can influence development of phonological patterns. Children who acquire languages with a lower vowel-consonant ratio are found to have a faster vocabulary growth rate compared to children acquiring languages with a high vowel-consonant ratio (Bleses, Basbøll & Vach, 2011). This result is supported by the consonant bias in lexical learning, while vowels may play a more important role in morphological and syntactic learning (e.g., Gervain & Mehler, 2010; Nazzi, Poltrock & von Holzen, 2016; Hojen & Nazzi, 2016). For example, Danish infants (8-15 months) were found to have a slower vocabulary growth compared to infants learning other languages, and this was suggested to be linked both to unclear syllable-boundaries and to the relative low consonant-vowel ratio of Danish phonology (Bleses et al., 2011). This can further be related to the influence of extensive vowel reduction in realisations of Danish speech, and also to the fact that consonants are more categorical than vowels.

In a cross-linguistic study, phonological development in 12 languages was investigated in children aged three to six years (Canadian French; Granada, Mexican and Chilean Spanish; European Portuguese; German; English; Swedish; Icelandic; Kuwaiti Arabic; Japanese; Mandarin; Bulgarian and Slovenian) (Bernhardt, Stemberger & Bérubé, 2017). Language-specific phenomena were found, but could be related to phoneme inventories, frequencies or the function of a specific element. For example, when comparing vowel sequences in Spanish and Mandarin, the most challenging vowel sequence in each language – diphthongs and triphthongs, respectively – seems to have the same level of accuracy in productions in spite of other differences between the languages. When measuring whole-word matches in typically developed children, 4-year-olds had 80% matching, irrespectively of the complexity of the language (Bernhardt et al., 2017). This suggests that within-language complexity levels may determine when structures with different complexity are developed, and not the specific complexity level per se.

In a review on the acquisition of phonological patterns in 2:0 to 6:6-year-old Mandarin speaking children, a study found that despite large variation between dialects and also in how to categorise phonological errors, phoneme
acquisition largely follows what is observed in other languages. Further, it was found that development of consonant production occurred hand in hand with speech motor development, with variations that could be related to dialect variants and lexical frequency of words with specific syllabic structures (Li & To, 2017).

Looking at specific structures or phonological patterns, some similarities are found between languages (see also the description of the development of Swedish phonology below). In American English-learning 2-year-olds, consistent phonological patterns are found such as simple word patterns and syllable shapes, and easier speech sounds such as glides, nasals and stops (Stoel-Gammon, 1998; 2011). Typically, children’s phonological development changes at 18 to 24 months from a holistic whole-word learning approach to a more analytic, phonologically rule-based strategy (e.g., Scherer, Williams, Stoel-Gammon & Kaiser, 2012). From 24 months to 36 months, the number of sound classes and syllable/word shapes increases and accuracy of speech improves, resulting in a higher level of intelligibility (Stoel-Gammon & Williams, 2013).

At age three, the phonemic system has usually developed to include consonants with all types of relevant manner, place and voicing, and a variety of syllable shapes and word forms (Stoel-Gammon, 1998; 2011). Bernhardt & Stemberger have specifically studied the development of word initial clusters containing a trill or a tap in preschool children in 12 languages. The findings indicated age effects (3-year-olds versus 6-year-olds) and group effects (typically developed children versus children with late or deviant development of phonology). Clusters were generally less accurate than singleton trills and taps, but more accurate in stressed syllables. Rhotics were less advanced than alveolar /l/ in all languages except Portuguese, where the lateral is velarised. Deletion of /r/ is common among children in young ages regardless of language, and later /r/ is often substituted. The sounds /r/ and /l/ share many articulatory and acoustic properties, but /r/ acquires finer motor control that may affect its development. Adult rhotics vary among adults according to language, both within and between word positions and contexts, but it is not known to what extent this variability affects the development of rhotics in children (Bernhardt & Stemberger, 2017; Stemberger & Bernhardt, 2018).

1.4.3.3 Development of word production in relation to phonological development

Early word production has universal characteristics as well as language-specific differences, and there are also individual differences among children both regarding vocabulary size and phonological patterns. Infants usually imitate words and sequences at around 12 months; thereafter more consistent referential word forms are used. The first words are lexically learned and highly context-bound since the infant is not yet able to generalise and therefore the words are not yet symbolic.
First words often have the same patterns as the CV-vocalisations in babbling (for example /mama/) (e.g., Stoel-Gammon, 2011). However, there are also differences between babbling and first words (MacNeilage, Davis & Matyear, 1997). In their first words, infants produce more labials, while the production of other consonantal sounds decreases. Labials are thought to be facilitated through input because they are visually salient (de Boysson-Bardies & Vihman, 1991). However, neither babbling nor mature speech are dominated by bilabial consonantal sounds (Locke, 1997). This preference for labials seems like a regression in speech motor development as labials do not involve active tongue movements in the same way as for example coronals or velars. MacNeilage and colleagues also found more variegated productions among words compared with babbling. In first words, a larger degree of variation among vowels was found – especially in final position. The authors suggest that this occurs in the last segment because production of a vowel in final position does not demand coarticulation of any following speech sounds (MacNeilage, Davis & Matyear, 1997).

The phonology of the target word also has an impact on early lexical learning. In English, early words are usually comprised of one or two syllables and contain consonants that are known to be acquired early (Stoel-Gammon, 2011). This leads Stoel-Gammon to suggest that early phonological development affects lexical learning to a greater extent than lexical learning affects phonological development (Stoel-Gammon, 2011). Vihman (2013) proposed that the infants’ own production is a part of the auditory speech input for the infant and therefore affects lexical learning through an articulatory filter. Vihman has also found that infants are more attentive to words that contain already acquired speech sounds compared to other words with not yet acquired speech sounds (Vihman, 2013). Some infants seem to have individual preferences regarding the word forms and sound classes they produce, and it is suggested that infants often select and use specific word forms to meet articulatory, perceptual and memory challenges (Vihman, 2013). These word forms, so-called templates, are thought to be universal but at the same time also affected by the ambient language. In typically developing children, templates are only used over a short period of time to scaffold phonological learning (Vihman & Wauquier, 2018). However, preferences for certain word forms and syllable structures may persist also after the first 50 words are acquired. In one study this was the case for English-learning infants, in whose lexica most acquired words were mono- or di-syllabic, with uniform stress placement and few words containing a consonant cluster (Stoel-Gammon, 1998; 2011). A positive relationship has also been found between phonological complexity in words and vocabulary size in Swedish infants aged 30 months (Marklund, Sundberg, Schwarz & Lacerda, 2012; Marklund, Marklund & Lacerda, 2014).

Early lexical learning is also linked to the frequency and distribution of words in the input, and to the grammatical class of the words. Even though closed-class words are the most frequent in the ambient language, infants’ first
words are typically open-class words, such as commonly used nouns, important persons, objects, actions and sound effects. Parents’ referential language when talking to infants is relatively high in lexical diversity (i.e., number of different words), which seems to promote infant vocabulary growth (Song, Spier & Tamis-LeMonda, 2014). Lexical diversity in the parental speech is beneficial for the infant’s word learning, as long as the lexical entities in the parental speech are linked to other relevant sensory information.

Around the end of the second year, infant vocabulary grows quickly, followed by word combinations, syntax and morphological development (Clark, 2009). Parents’ speech input is also linked to the infants’ complex language use. Around 20 months, language use of possessives emerges, together with the awareness of the self (Rochat & Passos-Ferreira, 2008).

Phonology and lexicon are not separate phenomena, nor hierarchically or chronologically ordered. Instead, the infant learns about the speech sound system by incoming auditory and visual information from others, and proprioceptive information from him/herself. Yet, both incoming speech sounds and the infant’s own productions vary with language, family and infant. Although complex, speech variation is not random as it correlates with other sensory input, which enables the infant to eventually acquire the phonological system of the ambient language.

1.4.3.4 Development of expressive phonology in Swedish-learning infants
A study of early phonological development in 1- to 17-month-old infants (Roug, Landberg & Lundberg, 1989) describes five distinct vocalisation stages in Swedish: glottal, velar/uvular, vocalic, reduplicated babbling and variegated babbling. The stages mirror the above description of babbling development in section 1.4.3.1. From around ten months of age, vocalisations become more like the speech sounds of the native language. In first word productions, the infant uses speech sounds present in the target language(s), while a true, stable phonological system has not yet been achieved. That the phonological system is unstable can be noticed by the variable productions of the same word during the same developmental period. From 1:6 to 4:0 years of age, the child’s ability to analyse word structure and word segments improves, and the phonological system develops further through links to lexical and morphological development (Nettelbladt & Salameh, 2007).

There are diverse findings in regard to Swedish vowel production. Some studies suggest that Swedish vowels are acquired early (Bjar, 2003), as most vowels are acquired before age three (Blumenthal & Jakobsson, 2013). However, front, round vowels are usually acquired late (Linell & Jennische, 1980), and Lawrence and Henriksson found that 50% of a sample of Swedish 3-year-olds still had deviations in vowel production (Lawrence & Henriksson, 2014). These differences in the reports can be explained by the fact that there are not many studies on vowel acquisition in Swedish and that assessment strategies...
vary between studies. Vowel variability is typically large between individual children and therefore vowels may be difficult to assess.

There is more agreement with regard to the acquisition of Swedish consonants. Lohmander-Agerskov and Forss (2005) studied a sample of Swedish 3-year-olds and found that all the children had acquired bilabial and dental stops and nasals, while velar stops and /l/ were found in only 90% of the children. According to Lawrence and Henriksson (2014), the most difficult speech sounds to produce for typically developed 3- to 4-year-olds are /ɭ/, /r/, /ʈ/ and /ʃ/. Lundeborg Hammarström (2018) as well as Linell & Jennische (1980) found that most Swedish speech sounds including /l/ are acquired by age four and that /r/ is one of the late acquired speech sounds. This may be related to the demands /r/ puts on fine motor control, an aspect that is also relevant when considering difficulties regarding the production of fricatives. Fricatives are among the speech sound classes with the most challenging manners of articulation (Ohala, 1983; Marklund, Marklund, Schwarz & Lacerda, 2018) and can be difficult even at age 4;6 (Blumenthal & Jakobsson, 2013). The distinction in place of articulation between dental and supradental is also difficult (Linell & Jennische, 1980). Lohmander-Agerskov and Forss (2005) shows that other difficulties in articulation include the fact that around 20% of 3-year-olds delete initial consonants, 33% have occurrences of assimilations, and 45% make reductions in consonant clusters. Lundeborg Hammarström (2018) found that di-consonantal word-initial clusters, including clusters containing /l/, are generally established between three to four years. Furthermore, clusters with /r/ are particularly difficult and can remain so even at age six. In younger children, the /r/-sound in clusters is often deleted, while later often substituted by /j/. Deletion of clusters with the /l/-sound appears before age two. At age three, most typically developed children master the majority of /l/-clusters and also have a high timing unit match which means that all elements of the cluster are present, even if substituted (Lundeborg Hammarström, 2018).

1.5 Assessment of early phonology

This section describes the assessment of expressive phonology in children. The introduction stresses the importance of detecting speech disorders early, and briefly describes common routines used to identify deviant speech in young children in Sweden. The next section discusses how speech data collection and analyses affect outcomes of expressive phonology assessment. The final section concerns different assessment measures and tests.

1.5.1 Introduction

Early speech and language development is known to be associated with later language skills such as vocabulary development and combinatorial speech
(Tamis-LeMonda et al., 2001), reading (Morgan, Farkas, Hillemeier, Scheffner Hammer & Maczuga, 2015), school success McCormack et al., 2011, and cognitive achievements and academic functioning (Marchman & Fernald, 2008; Morgan et al., 2015) and behavioural adjustment. Language difficulties not only risk affecting children's school achievements (e.g., Conti-Ramsden, Durkin, Simkin & Knox, 2009), but may also have an impact on socio-emotional development (McCabe & Meller, 2004) and mental health (Snowling, Bishop, Stothard, Chipchase & Kaplan, 2006).

Phonological development, in particular expressive phonology and phonological awareness, is associated with delayed outcomes in reading and writing (e.g., Gillon, 2000; Bus & Van IJzendoorn, 1999; Castles & Coltheart, 2004; Ehri, Nunes & Willows, 2001). Early language difficulties may also have long term effects into adulthood and are also associated with negative effects such as poor adult literacy skills, poor academic functioning (Morgan et al., 2015), adult mental health problems and unemployment (e.g. Law, Rush, Schoon & Parsons, 2009), and difficulties in managing higher education, work and social life (Whitehouse, Watt, Line & Bishop, 2009). Estimates suggest that 5 to 8% of preschool children in Sweden have some form of speech or language impairment, of which 1 to 2% of these children present severe difficulties (Nettelbladt, Samuelsson, Sahlén & Ors, 2008). The prevalence declines rapidly after five years of age, which implies that a large number of children have recovered by this age (Nettelbladt & Salameh, 2007). However, the prognosis for those children whose language problems persist after this age is worse.

Speech sound disorders (SSD) are common among pre-school children, documented for example in Great Britain, (Bishop, Snowling, Thompson & Greenhalgh, … Boyle, 2017), and in Sweden (Nettelbladt & Salameh, 2007). SSD are among the most apparent types of speech problems, and may cause speech production difficulties also affecting intelligibility (Bishop et al., 2017). SSD is an umbrella term for difficulties with delayed or deviant speech production compared to peers, and SSD could appear with or without difficulties in language perception. SSD includes cases with phonological disorders and/or speech motor disorders (Bishop et al., 2017). Phonological disorder is differentiated from speech motor disorder on the basis of articulation. It is not primarily articulation-based (like a speech motor disorder), but a linguistic difficulty that often entails that the child produces extensive phonological errors and, in more severe cases, unintelligible speech. Phonological disorders may occur alone or together with problems in other linguistic domains regardless of the severity of the language problem, but usually has better prognosis if not accompanied by other language difficulties (Bishop et al., 2017).

To find the right intervention, it is important to find out the nature of an SSD. Hearing problems should always be taken into consideration, and it is important to check for possible underlying structural anomalies and physiological problems when establishing whether the difficulties are phonological or phonetic. Studies have shown that training of phonemic and phonological
knowledge supports the development of reading and writing abilities in children at risk for dyslexia (e.g., Gillon, 2000; Puolakanaho, Ahonen, Aro, Eklund, Leppänen, Poikkeus, ... Lyytinen, 2007; Puolakanaho, Ahonen, Aro, Eklund, Leppänen, Piokkeus, ... Lyytinen, 2008).

In cases of speech production difficulties, phonology should be assessed early – prior to the occurrence of reading and writing problems at school. A 4-year-old is expected to be fully understood by listeners (Coplan & Gleason, 1998 cited in Namavigayam, Pukonen, Goshulak, Vickie, Kadis Kroll ... Luc, 2013). If less than 66% of the words produced by a 4-year-old, are understood, then speech therapy is recommended (Gordon-Brannan & Hodson, 2000). Therefore, speech production difficulties should preferably be investigated whenever a child is still difficult to understand by close family members at age three, and regardless whether the cause is known or not. Since phonological difficulties are a part of expressive language problems (i.e., not just a matter of articulation), assessment of the child’s ability in speech sound identification needs to be included before starting an intervention - even though speech production training without speech error sound identification may improve the perception of a speech error (Wolfe, Presley & Mesaris, 2003). In a study on treatment of phonological disorders in British children, it was found that the majority of speech therapists used intervention approaches that target speech discrimination, minimal pairs and phonological awareness, but about half of the therapists still also used less effective methods, such as articulation training (Hegarty, Titterington, McLeod & Taggart, 2018).

In Sweden, virtually all families attend regular developmental checks at the Child Healthcare Centres, which include language screenings by a nurse (Swedish Children Medical Association, 2013). The Child Healthcare Centres play an important role in identifying language and communication problems early in infancy, and if there are concerns regarding language or communication development, the child should be referred to a speech-language pathologist. However, even though there are national guidelines regarding language screening in young children, these routines are not consistently used across all municipalities in Sweden. Most Child Healthcare Centres offer screening at age 2:6 or 3:0, but far from all have follow-up screenings at the year-four or year-five check-ups. Furthermore, multi-lingual children with language impairment tend to be identified later by the Child Healthcare Centre nurses and their referral to a speech language pathologist occurs later than for monolingual children with speech and language difficulties (Nayeb, Wallby, Westerlund, Salameh & Sarkadi, 2015).

Thus, early detection of deviant speech production is important, regardless of its origin, as the consequences of SSD can be harmful for school achievement, social development (McCormack, Harrison, McLeod & McAllister, 2011) and later life success (Whitehouse, et al., 2009). Early detection of problems with expressive phonology enables detection of potentially co-existing disorders and early intervention (Gillberg, 2010).
1.5.2 Collecting and analysing speech data

To ensure a high degree of validity when assessing expressive phonology, it is important to reflect on which type of speech data is used, as different types of speech data alter how children’s general production is represented. Independent of the type of speech data, the assessment situation and elicitation strategies will also influence assessment results. Furthermore, different types of analysis offer different ways of describing children’s expressive phonology. There are two types of analysis used to describe expressive phonology: independent and relational. Independent analysis is not related to target productions, and usually comprises descriptions of occurring word forms, syllable structures and/or phonemic segments. Independent analysis is often used for very early or deviant phonology. In relational analysis, child productions are compared to mature target productions, and often specifically selected words or utterances are used for comparisons.

1.5.2.1 Speech data

Samples with spontaneous speech show one perspective of the child’s production, elicited words and structures a second, and repetition of words and phrases a third. Spontaneous speech samples enable descriptions of a child’s typical performance and use of speech sounds, both in the context of connected speech and in isolated words. Spontaneous speech samples make it possible to analyse supra-segmental information, for example prosody. This type of speech data is usually easy to collect. Any interactive play can elicit spontaneous speech in children, and toy play or book reading can be particularly effective (Rowe, 2012). Spontaneous speech samples are usually ecologically valid in the sense that they reflect the child’s typical production. Using data from conversational speech to assess expressive phonology has been shown to be linguistically and psychometrically robust regardless of semantic content (e.g., Shriberg, 1993). However, in conversations it is difficult to ascertain that certain phonological target structures are captured, which makes spontaneous speech data difficult to use in comparisons between different children or age groups. Therefore, combining spontaneous speech samples with additional test words that include phonological structures and sounds not present in spontaneous speech may be a good strategy to get a full overview over a child’s phonological capabilities (Stoel-Gammon, 1987).

A single word test to assess the development of a young child’s phonological system should be valid and reliable, and needs to use target words that are familiar to children in the same age group. Furthermore, single word tests for expressive phonology are suitable for relational analysis where comparisons of a child’s production to target on segmental and suprasegmental levels are made. A relational analysis is suitable once the child has an established phonological system. Relational analysis also enables comparisons between children. Even if the child does not produce all target words spontaneously during an assessment session, the words may be elicited through strategies such as
forced choice or delayed imitation, as there is the possibility that the words are stored in the receptive vocabulary. However, a more structured test with elicited words and structures might not reflect functional speech, as repetitions of words can be produced differently from words in spontaneous speech.

To ensure that speech production is representable, a speech sample needs to be of a certain size. Typically, when assessing expressive phonology, a speech sound should occur in at least two different words to be assessed as included in a child’s phoneme inventory (Rowe, 2012). If the speech sample is too small, it is difficult to draw conclusions about patterns of expressive phonology. For example, inconsistent productions of the same target word may not be apparent. Variation itself is also informative and contributes to the analysis. Target words in any test material also need to be selected to represent important phonological structures on both segmental and supra-segmental levels. For example, if a specific speech sound occurs spontaneously in a child’s speech but never in word-initial position, the words used in a test need to target all word positions to capture this.

1.5.2.2 Elicitation strategies

One commonly used strategy to elicit words in speech and language tests is to show pictures of familiar objects and actions for the child to name, and using a set of toys can also be useful to map phonological target structures, especially for young children (e.g., Stoel-Gammon & Williams, 2013; Rowe, 2012). However, elicited speech data may not reflect the child’s typical performance and is therefore less ecologically valid than spontaneous speech. A child who normally talks a lot spontaneously can speak very little in a test situation in which elicitation strategies are used. An advantage with assessment materials chosen to target specific phonological structures is that they facilitate comparisons between children or age groups.

In some cases, if the child does not name the presented picture or toy or responds to questions, having the child simply repeat the word can be a useful alternative. The test leader produces the target word or utterance and waits for the child to repeat it without further prompting. Alternatively, the test leader specifically asks the child to repeat the word. Just like other non-spontaneous speech elicited by a therapist or test leader, repetitions might not be as ecologically valid as spontaneous speech, but if specific structures are targeted, this type of speech data makes it possible to collect useful data for comparisons among children in different age groups.

A more guided and structured assessment situation usually works well if both the child and the parent are comfortable with the situation. If the parent is nervous or uncertain, for example about the purpose of the assessment, the child may notice this and become uncomfortable. If the child is uncomfortable, shy or feels pressured, the result can be inaccurate because the child’s performance is not representative. The test leader should be aware of the impact of interaction style in the test situation, having in mind that speech and
language is the tool used for assessing speech and language (Bruce, Hansson & Nettelbladt, 2007).

In formal test situations, elicited speech of target words or utterances is typically collected to facilitate comparisons between children and age groups. However, it has been shown that different elicitation strategies may impact the assessment results differently, depending on the type of speech disorder. Children with cleft palate are found to perform better on single-word naming tasks compared to sentence repetition tasks that require retelling of narrative and conversational speech, while no such differences were found in children without cleft palate (Klintö, Salameh, Svensson & Lohmander, 2011). Goldstein, Fabiano & Iglesias (2004) tested children aged 3:1 – 4:9 years with mild-moderate phonological disorders, and found that spontaneous naming and imitation productions in repetition tasks were phonologically identical in 62% of the cases, and that no significant differences were found according to Percent Consonants Correct (PCC, see section 1.5.3.1) in the productions. Depending on whether the basis of speech errors is phonetic or phonological, different elicitation strategies may affect the child’s output differently. Optimal performance in the test situation might depend on which elicitation strategy is used and not only because of the underlying cause of the speech disorders. Also the child’s own feelings and preferences in the test situation may affect the results. Thus, a sensitive interaction style, and a systematic use of different elicitation strategies and materials is presumably the best way to map young children’s speech and to effectively obtain differential diagnostics.

1.5.2.3 Reliability

It is important to keep in mind that assessment of a child’s phonology is based on the test leader’s own perceptual estimations. There is therefore always a risk for the child’s performance to be over- or underestimated. Yet standardisation and consistency in test procedures and analytical methods diminishes the problem and can ensure high inter-rater reliability. Generally, it is easier and quicker to use data elicited with structured test material compared to using spontaneous speech data, at least if the child has a developed phonological system.

To use test materials with defined tasks enables quick mapping of the perceived phonological forms onto the targeted structures. On the other hand, using data elicited by a structured test material to map very early or deviant phonology to target forms may not be optimal as the child could be unwilling to try naming and therefore will produce too little data for a complete analysis. Thus, the method to collect speech affects assessment reliability, since amount and type of speech data may not be sufficient to reliably represent a child’s performance.
If the child has not yet developed a phonological system that can be related to mature phonology, it can be easier and also more fruitful to use an independent analysis, that is analysing occurring structures and describing features of the speech sounds present in the child’s output.

Speech samples are usually transcribed with the use of the International Phonetic Alphabet (IPA; Decker, 1999), and the quality of phonetic or phonological transcriptions is dependent on the test leader’s experience with a child’s speech. It is recommended that the test leader’s native language is the same as the assessed language, and multilingual children’s phonology needs to be tested separately for each language.

According to Shriberg & Lof (1991), phonological transcriptions are more reliable than phonetic transcriptions. As phonological assessment is always based on the test leader’s perception, the test will give no “true” answers, while they still may be useful. Shriberg & Lof (1991) describe a large number of potential variables that can interact with each other during the transcription process, and make it difficult to find agreement in transcriptions between test leaders. There is the risk of transcribers’ drift – that is a change from the original calibration to other perceptual criteria during transcription. Inter-rater agreement is important (Shriberg, Kwiatkowsky & Hoffman, 1984) especially when the speech sounds produced are deviant, distorted or intermediate, as this will give additional information about the child’s abilities. As long as speech sounds are perceived in relation to the target forms, this will not be problematic, regardless of production variability. And as long as speech sounds are assessed as either being produced or not, such a binary method of handling the child’s output will not take production accuracy into account. Assessing phonology is always qualitative and there is always a risk to either over- and/or underestimate a particular child’s performance.

1.5.2.4 Analysis

There are two ways of analysing expressive phonology: independent and relational. To describe expressive phonology of very young children or children with severe phonological deficiencies, independent analysis is typically used. Expressive phonology can be described even if no words are yet identifiable and the child’s speech is unintelligible. This applies in particular to unstable child phonology before the 50-word lexicon stage is reached.

Morris (2009) reported test-retest correlations regarding five phonological measures in 10 children from 18 to 22 months. These measures were 1) number of different consonants in word initial position; 2) number of different consonants in word-final position; 3) number of different word shapes; 4) syllable structure level, and 5) index of phonetic complexity (IPC; Jakielski, Maytasse & Doyle, 2006, see also section 1.5.3.1 below). Re-testing after one week revealed that the least stable measure was the number of different initial consonants (1), while the measures of syllable structure level (4) and phonetic complexity (5) were the most stable. Of these, only syllable structure reached
statistical significance in terms of stability. Moderately stable were the measures of number of different final consonants (2) and number of different word shapes (3).

An independent analysis can focus both on the segmental and supra-segmental levels. On the segmental level, it is typical to describe consonants and their word positions, but also descriptions of vowels can be included. The speech sounds present in the child’s individual phoneme inventory are mapped to their target phonemes and described by linguistic features such as place, manner and voice. On the supra-segmental level, occurrences of word patterns such as number of syllables or stress pattern, and syllable structures such as open or closed syllables, or consonant clusters are described. In independent analysis, correctness of production is not the central question. This, however, is the subject of relational analysis of phonology, which is suitable for typically developing older children who have an established phonological system.

In relational analysis, deviations from the target structure can be described by phonological errors or “processes”, on either the segmental or supra-segmental level. On the segmental level, context-independent errors, so called paradigmatic processes, describe reductions and substitutions of place, manner or voice, for example fronting of velars and stopping of fricatives. On the supra-segmental level, context-dependent errors, so called syntagmatic processes, describe phonotactic and prosodic structure, for example assimilation of features, metathesis, coalescence, reductions in consonant clusters, omission of initial or final consonant and omission of unstressed syllable. To determine if a process is context-dependent or not, the speech sample needs to be large enough to be able to find recurrent patterns of errors, and whether and if so how they are related to the context. For example, a fronting error, for example /k/ → [t], may occur because velars are still not acquired – a context-independent error – or because of assimilation to place of articulation – a context-dependent error. Ingram (1974) assigns speech errors to phonological processes and assumes that the child is gradually suppressing these phonological processes when developing speech towards phonological target structures that are presumed to be “underlying representations”. Regardless of the theoretical perspective on the basis of phonological development, context-independent and context-dependent processes are a practical way of describing a child’s phonology in a relational analysis.

1.5.3 Measures and tests to assess expressive phonology

There are not many materials available to assess the expressive phonology of children under the age of three. It is difficult to establish phonological norms for very young children, as expressive phonology up to age three shows a great variability (Holm, Crosbie & Dodd, 2007; Sosa & Stoel-Gammon, 2006). Measures discussed in this section are the Percentage Consonants Correct (PCC, Shriberg 1993; Shriberg, Austin, Lewis, McSweeny & Wilson, 1997),
the Phonological Mean Length of Utterance (PMLU; Ingram, 2002), The Word Complexity Measure (WCM; Stoel-Gammon, 2010) and The Word Complexity Measure for Swedish, WCM-SE (Marklund, Marklund, Schwarz & Lacerda, 2018).

One English test is described, the Profiles of Early Expressive Phonological Skills (PEEPS; Stoel-Gammon & Williams, 2013) as it forms the basis for the Swedish adaptation: the rest of the tests described in this section are Swedish: the Profiles of Early Expressive Phonological Skills for Swedish (PEEPS-SE; Marklund, Lacerda, Persson & Lohmander, 2018), the Swedish Test for Articulation and Nasalization (SVANTE; Lohmander-Agerskov & Forss, 2005; Lohmander, Lundeborg & Persson, 2017), the Assessment of Phonology (Bedömning av Fonologi, B.A.F; Frylmark, Edquist & Niemi, 2015), the Linköping Test (Linköpingsundersökningen, LINUS; Blumenthal & Lundeborg Hammarström, 2014), the Phoneme Test (Fonemtest; Hellquist, 1995) and Lunda-materialet (Holmberg & Stenvqvist, 2012).

1.5.3.1 Measuring tools
The measure Percentage Consonants Correct (PCC, Shriberg, 1993; Shriberg et al., 1997) is a measure to use on spontaneous speech that is transcribed phonetically. To classify deviations from target phonology according to consonants there are four severity classifications; mild (>90%), mild-moderate (65-85%), moderate-severe (50-65%) and severe (<50%). As the PCC is used on conversational samples, standardisation across children is not possible. The PCC cannot give the whole picture of a child’s phonology as PCC-scores may be high due to a very limited vocabulary. However, Shriberg and colleagues point out that a specific material to elicit speech could be useful when assessing children with very severe phonological disorders.

The PCC weighs all 24 English consonants equally no matter how complex they are to produce. This is addressed by the speech profile approach in the PCC (Shriberg, 1993), that uses subscales for three developmental consonant sound classes that could be especially useful for clinical research assessment: the early-eight sounds (m, b, j, n, w, d, p, h), the middle-eight sounds (t, η, k, g, f, v, tʃ, ʤ), and the late-eight sounds (ʃ, θ, s, z, ð, l, r, ʒ), defining individual child profiles by grouping different types of speech errors. The speech profile approach also uses subscales targeting percentage of individual sounds, clusters, class/manner features and absolute and relative percentage errors. The PCC weighs distortions (i.e., phonetic errors) in the same way as omissions and substitutions (i.e., phonological errors), and scoring is based on listeners’ subjective assessment of the severity of speech deviation. Severity classification is related to the percentage correct of articulated sounds – not the type of error. However, for clinical use, it is important to look closer into error types, to distinguish between substitutions and distortions and to differentiate between distortion types. Among 3- to 5-year-old children with or without speech delay, there are overlaps of speech error types. A way to calculate PCC
relative to distortion errors is with the complementary materials *Articulation Competence Index* (ACI) and *Relative Distortion Index* (RDI) (Shriberg, 1993). The ACI can differentiate between children with or without speech delay, and used together with the RDI, calculations of the percentage of errors that are distortions can be made. The PCC-A (adjusted) and PCC-R (revised) are versions that have added common and uncommon clinical distortions of consonants to the PCC measure. To measure vowels, the Percentage Vowels Correct (PVC) can be used, and for common and uncommon clinical distortions of vowels, the Percentage Vowels Correct revised (PVC-R). The PCC is traditionally used to weigh speech sound deviation as a function of frequency. Life-span reference data for PCC is available (Shriberg et al., 1997).

The measure *Phonological Mean Length of Utterance* (PMLU; Ingram, 2002) is a measure that, similarly to the Word Complexity Measure (see below, this section), takes whole-word productions into account. PMLU is applied to all segments produced in a word or an utterance as well as all correctly produced consonants. The PMLU can be used to calculate complexity of target words, to assess proximity of productions to target words, and to identify the child’s phonological developmental stage. One point is given for every segment in the word and also one point per each correctly produced consonant. Suggested sample size is at least 25, but preferably 50 words. Children’s word forms like “mommy” and “daddy” should not be used for calculations, as they have too low complexity and would inflate the PMLU if the child has many words that are comprised of reduplicated syllables. Compound words are only seen as one word if spelled as one. As children’s whole-word productions often vary in correctness, complexity and variability, Ingram proposes three additional measures: the Proportion of Whole-Word Proximity (PWP) that calculates proximity to target words, the Proportion of Whole-Word Correctness (PWC) that calculates correct produced words in the sample, and the Proportion of Whole-Word Variability (PWV) that measures how often produced words are in particular distinct phonological shapes (PWV). The PMLU can be used to measure early phonology in young children, to make cross-linguistic comparisons, to look closer into multilingual children’s phonological acquisition, and to assess children with phonological impairment.

*The Word Complexity Measure* (WCM; Stoel-Gammon, 2010) is used to measure phonological complexity in words. The WCM is particularly useful to assess children with early phonology and late talkers. It can give qualitative information regarding complexity of a child’s phonological production and can be used both in clinical practice or research to compare utterances over time and/or between children.

Each word receives a complexity score. There are eight complexity parameters in three domains; word patterns, syllable structures and sound classes. Higher scores indicate the presence of more complex parameters in the word, presumably acquired later.
The WCM does not assess accuracy, but focuses on the phonetic complexity of segments, syllables and words. It focuses on later-acquired structures and uses independent analyses. Even unintelligible words can be measured. For a relational analysis, comparisons with target words can be made, and a child’s avoidance patterns can be detected. Descriptions of a child’s phonology can be stated after using the instrument (for example, that the child has no fricatives or that most words have CVCV form) and children with phonological difficulties can be detected. Optimal sample size is not yet calculated for the WCM and lack of test-re-test-reliability due to differences in spontaneous speech samples needs to be considered.

In comparison to the PCC (Shriberg et al., 1997), the WCM provides information not only about segments, but also about word forms and syllable structures. Additionally, it can be used to assess unintelligible speech. While the PMLU (Ingram, 2002) offers the possibility to use both independent and relational analysis, it lacks the analysis of suprasegmental characteristics that is provided by the WCM. A similar measure to the WCM is the Index of Phonetic Complexity (IPC; Jakielski et al., 2006). Both measures give points to words with more than two syllables, final consonants, fricatives, liquids and velars. The differences are that the IPC also gives points to articulation place variegation of consonants within a word and within a cluster, whereas the WCM considers stress.

The Word Complexity Measure for Swedish, WCM-SE. An advantage of the WCM is that it is easily modified depending on what structures are to be assessed and that it offers a picture of a child’s breadth of expressive phonology. The Word Complexity Measure for Swedish (WCM-SE; Marklund, Marklund, Schwarz & Lacerda, 2018) is an adaptation of the original Word Complexity Measure (WCM) developed for English by Stoel-Gammon (2010). The WCM-SE can be used to calculate the phonological complexity of Swedish words or vocalisations, based on phonological complexity parameters. It is suitable for the assessment of early expressive phonology or deviant phonology. Calculations can be used for both independent and relational analyses and comparisons between children and over time can be made. The child’s production is given a complexity score based on the complexity parameters present. The WCM-SE uses ten complexity parameters, selected on the basis of Swedish or universal patterns of phonology development and also considered from a language-general phonetic perspective (see Study 3).

1.5.3.2 Tests
There is a large variability between tests for expressive phonology. Macrae (2017) examined 12 American English tests and found a great deal of variability regarding number and type of stimuli: syllables, consonant singletons, consonant clusters, vowels, bound morphemes and phonological complexity. However, all have words as their target. Standardised tests of expressive pho-
ology are usually limited to target-only production of single phonemes. Another limitation is that the words used in these tests might not yet be acquired among young children. Only if the test words are acquired at the relevant age and the produced words have similar phonological properties as the target words, is a fair test result likely (Eisenberg & Hitchcock, 2010).

The *Profiles of Early Expressive Phonological Skills* (PEEPS; Stoel-Gammon & Williams, 2013) is a test of expressive phonology for English-learning 18 to 36-month-old children that uses a set of words acquired in these age groups. The words are organised in two word lists: a Basic Word List (BWL) with 40 words and an Expanded Word List (EWL), comprised of the BWL and an additional 20 words. The BWL is for assessment of children up to 24 months or children with up to 250 words in productive vocabulary. The EWL is for assessment for children aged 24 to 36 months or children with more than 250 words in productive vocabulary.

To facilitate a child’s production, the word selection is based on age of acquisition. Additionally, as a child’s phonology is tied to his or her vocabulary, the production is then more likely to be representative. Word selection is also based on phonological complexity, measured by the Word Complexity Measure. Words in the EWL are acquired later and are more phonologically complex compared to those in the BWL. PEEPS includes words from all sound classes, but not from all English phoneme classes. PEEPS is not designed to assess accuracy, rather it tests a range of domains such as sound classes, word and syllable shapes, stress patterns and error types, and can capture the breadth of a child’s phonological system. The target word selection for PEEPS also considers syntactic and semantic factors. Objects, body parts and actions are used to elicit the target words. PEEPS has been found to successfully assess consonant inventory, syllable structure, accuracy and speech errors among typically developing children as well as children with cleft palate (Scherer et al., 2012).

The *Profiles of Early Expressive Phonological Skills for Swedish* (PEEPS-SE; Marklund, Lacerda, Persson, Lohmander, 2018) is an adaptation to Swedish of the original PEEPS (Stoel-Gammon & Williams, 2013). It is a test of expressive phonology for Swedish-learning 18- to 36-month-old children, based on two word lists, a Basic Word List (BWL) of 40 words and an Expanded Word List (EWL) comprised of the BWL and an additional 20 words. Just like in the original English version of PEEPS, the BWL in the PEEPS-SE is for assessment of children up to 24 months or children with up to 250 words in productive vocabulary, and the EWL is for assessment for children aged 24 to 36 months or children with more than 250 words in productive vocabulary.

The target word selection for PEEPS-SE is based on two types of criteria; age of acquisition and word complexity. Swedish vocabulary data from the longitudinal language development study SPRINT was used to select words based on age of acquisition. Word complexity of suitable words was calculated with the Word Complexity Measure for Swedish (Marklund, Marklund,
Schwarz & Lacerda, 2018). The resulting two word lists for PEEPS-SE are slightly higher in mean complexity than the original PEEPS.

The *Swedish Test for Articulation and Nasalization (SVANTE; Lohmander-Agerskov & Forss, 2005; Lohmander et al., 2017)* is a Swedish test for articulation and nasalisation with norm data for speakers from three to 19 years. SVANTE provides information about the number of errors and also about the degree of deviation from the target pronunciation, which gives both quantitative and qualitative measures. A general phoneme analysis can be done, as well as the perceptual assessment of velo-pharyngeal function and intelligibility. SVANTE uses mainly picture-naming as a method to elicit specific phonological structures. For children younger than four years of age, a play situation can be set up; for children from four to ten years, a story-prompt picture is used. Different elicitation techniques such as spontaneous naming, sentence completion and repetition are used. Older participants are presented tasks in which they are asked to read.

The *B.A.F Test of Phonology (Bedömning Av Fonologi; Frylmark, Edquist & Niemi, 2015)* is a test of expressive phonology recommended for children from age four. B.A.F. has norm data for ages four and five (Olson & Ulander, 2017). The material can be used to collect information about expressive phonology also about basic information about simple concepts and grammar. The manual contains useful instructions for different elicitation techniques, assessment settings and basic information about phonological development and speech sound disorders. The instructions recommend an additional quantitative assessment by calculating PCC from recorded connected speech. The material is comprised of five thematic pictures and a booklet with smaller, more detailed pictures to elicit speech. One of the five theme pictures is designed for eliciting narrative connected speech.

The *Linköping Test (Linköpingsundersökningen, LINUS; Blumenthal & Lundeborg Hammarström, 2014)* is a Swedish test of expressive phonology with norm data for children aged three to seven years. LINUS can be used to analyse sound classes in terms of manner, place, voicing and aspiration of consonants, word patterns such as number of syllables and/or word stress and syllable structure (word final consonants and consonant clusters). LINUS not only gives information about deviations from target structures on the segmental level but also looks at the distribution of specific consonants or types of articulation (place/manner/voice) in order to find inconsistent substitutions and segment reductions. Some test words are selected to target metathesis, epenthesis, assimilation and coalescence in child production, and the analysis provides frequency information for correctly produced words and phonemes. Picture naming is used to elicit speech.

The *Phoneme Test (Fonemtest; Hellquist, 1995)* is a common test widely used in Swedish speech pathology clinics. Although it has no norms, it can be used from three years and up. It is used to map sound classes: consonant inventory, including occurrences in initial, medial and final word positions, and
vowel length and quality. Also, clusters with two or three consonants are targeted (CC and CCC). Additionally, word patterns (words with more than two syllables and words with late stress) and words sensitive to metathesis, epenthesis or assimilation are represented. The test uses pictures of objects to elicit speech. The New Phoneme Test exists in two versions. The long version has 99 pictures and each of them elicits one word, targeting one segment in one position or one structure. The short version has eight pictures, offering in total elicitation of 72 words, each of them targeting more than one segment or structure. This test maps a child’s phonology quickly and easily and identifies systematic or non-systematic deviations of Swedish phonology. The two versions offer the possibility of choosing a test according to both the child’s compliance and motivation, and the test leader’s experience.

The Lund Material (Lunda-materialet; Holmberg & Stenqvist, 2012) is a common screening test in Swedish speech pathology clinics. It is designed to screen for expressive language problems of phonology, morphology and syntax in Swedish-speaking children. Lunda-materialet is created for children from 2:6 to 4:5 years and parts of the material have age norms from 3:0 to 5:11 years (Gustafsson & Iranmanesh, 2013; Söderberg & Torstensson, 2013). The child is presented with pictures in a book and various strategies are used to elicit speech; questions, fill-in strategy, modelling and requests. The test maps all Swedish consonants in different word positions as well as initial consonant clusters (CC, CCC).
2. Studies

The studies of this thesis concern parental responsiveness and the assessment of early phonology in Swedish-learning children. The studies have resulted in three papers and one manuscript:


Reprinted with permission from Cambridge University Press.

**Marklund, U., Schwarz, I.-C & Marklund E.** Contingent turn-taking between parents and 6-month-olds: Primary caregivers respond faster than secondary caregivers.

Manuscript in progress.


Reprinted with permission from Taylor & Francis Group.


Reprinted with permission from Taylor & Francis Group.
2.1 Study 1

*Pause and utterance duration in child-directed speech in relation to vocabulary size*

This study investigated the relationship between the duration of parents’ utterances and pauses in interaction with their 18-month-old infants, and child vocabulary size at 18 months.

A sizeable number of studies has shown the relationship between parents’ responsiveness and child language outcomes (e.g., Hart & Risley, 1995; Baumwell, *et al.*, 1997; Rollins, 2003). Contingent vocal responses from parents correlate with infant turn-taking abilities (e.g., Kuchirko, Tafuro & Tamis LeMonda, 2018), vocabulary growth and use of complex language (e.g., Tamis-LeMonda, *et al.*, 2001; Rollins, 2003) and phonological development (e.g., Albert *et al.*, 2017). Even though responsive strategies are shown to be effective in parent-centred intervention programs for children with speech and language disorders (e.g., Girolametto *et al.*, 1996), the characteristics of responses and child language outcomes can be further studied, for example temporal contingency in responses in relation to child vocabulary.

In this study, we analysed vocal interaction between fifteen 18-month-old infants and their caregivers (eight girls) in relation to infant vocabulary size. Parent-infant dyads were selected on the basis of infant vocabulary size at age 18 months, as measured by SECDI, the Swedish version of McArthur Bates Communicative Developmental Inventory (Berglund & Eriksson 2000). Dyads were selected into three groups, with five in each group, according to SECDI results. The large vocabulary group contained infants with number of vocabulary items at the 90th-100th percentiles, the typically sized vocabulary group consisted of infants with upper medium results at the 50th-75th percentiles and in the small vocabulary size group, there were infants with results below the 25th percentile. Parents were instructed to make audio recordings of caregiver-infant interactions with Handy Zoom recorders H2 in the family homes in four different daily life situations: dressing, meal time, book reading and play. Each recording session lasted around 20 minutes and one or both caregivers were present. Instructions to parents were that they should behave as they usually would in these situations, without trying to perform in any particular way. Five minutes from each recording, starting with the first parental utterance judged as child-directed, were used for analysis. All infant and caregiver speech and vocalisations that were judged as having a communicative intent were marked up in Wavesurfer 1.8.2. Cries, fuss, laughter and vegetative sounds such as burping, sneezing, etc. were excluded, but singing was included. If two utterances or vocalisations from the same person were close in time, they were separated due to lexical content or prosodic features.
Silence between caregiver and infant vocalisations or between the two parents was defined as switching pauses. A parent-child pause was the silence between the parent’s utterance and the following child utterance, while a child-parent pause defined the silence after a child utterance, preceding a parent utterance. A parent-parent switching pause was the silence between two caregiver utterances. Silence between utterances within the same speaker was defined as intra-personal pauses. Overlaps were treated as negative pauses. Five minutes per situation per child were marked up and analysed, in total 20 minutes per infant, and duration of utterances and pauses were calculated. No variation between the two caregivers within the same infant was taken into account in this study. Instead, all parent data for each child was collated to a general parental input, as the main research question was how child vocabulary was related to temporal contingency in parent-child communication in general.

The results showed that parents in all vocabulary size groups used longer utterances when speaking to infants compared to speaking to adults. All parents also responded faster to infants compared to adults. No differences were found between vocabulary size groups according to caregivers’ utterance duration. The main finding from this study was that parents in the large vocabulary group responded faster to their infants compared to parents in the typically sized vocabulary group, who in turn responded faster than parents in the small vocabulary size group. The results from this study do not say anything about causality or directionality, but they clearly indicate a relationship between a child’s vocabulary size and his/her parents’ temporally contingent responses.

This study is one of the few that look purely into temporally contingency rather than conceptual contingency in parent-child interaction, and put this in relation to child language outcome. It gives a clear picture that there is a relationship between how fast parents respond to their children’s communicative initiatives, but it cannot answer the question of whether this relationship has its origins in parents’ sensitivity to children’s language skills, demonstrating how parents adapt to their child’s vocabulary size, or whether child vocabulary growth is a consequence of greater temporal contingency in parent responses, illustrating how fast parent responses foster greater language skills. In order to answer this question, a longitudinal study is necessary and is under way.
2.2 Study 2

Contingent turn-taking between parents and 6-month-olds: Primary caregivers respond faster than secondary caregivers

This study concerns vocal turn-taking between parents and their 6-month-old infants. The purpose was to find out if there are differences in vocal temporal contingency depending on caregiver status and on the sex of an infant.

Prior studies have shown differences in how parents interact with their infants depending on the sex of the infant, and that infant behaviour differs in interaction with the caregiver as a function of caregiver status. For example, it has been found that infants show preference for speech from their mothers before fathers during their first months of life. (Johnson et al., 2014). Differences in how parents act depending on the sex of an infant can be related to more or less conscious cultural beliefs on gender, parenting and socialisation goals (e.g., Harkness & Super, 1996; Rowe et al., 2016).

In this study, we examined whether and to what extent there were differences between primary and secondary caregivers’ vocal interaction with infant boys and girls. Interactions of 14 infants (seven girls) aged six months, and their caregivers were recorded and analysed. All infants lived in families with two caregivers. The primary caregiver was defined as the parent staying home during parental leave, thus spending most of the time with the infant. In this study, all primary caregivers were mothers and all secondary caregivers were fathers.

Ten minutes of audio-recordings of interaction between the infant and the primary caregiver and the infant and the secondary caregiver were collected in the family homes. Recordings were made with a Handy Zoom recorder H2, and were marked up by an experienced listener (first author) in Wavesurfer 1.8.2. All infant vocalisations were marked up, and so were the pauses preceding and following a vocalisation. Pause onset and offset was the end of the preceding caregiver utterance and the start of following caregiver utterance. Infant cries, fuss, laughter and vegetative sounds were excluded. Silence between caregiver and infant vocalisations was defined as a switching pause. Silences between infant vocalisations were marked up as infant intra-personal pauses. Parental intra-personal pauses were not marked up, as only caregiver vocalisations adjacent to infant vocalisations were used in the analysis. Duration of infant switching pauses (following a caregiver vocalisation) and caregiver switching pauses (following an infant vocalisation) were calculated. Overlaps of vocalisations were treated as negative pauses.

The results revealed shorter switching pauses in infant interaction with first caregiver – both for infants and caregivers. First caregivers responded faster to their infants than second caregivers, and there was a trend towards infants
responding faster to the first caregiver compared to the second caregiver. No differences were found regarding the sex of the infant—neither infant-parent nor parent-infant switching pauses.

Thus, it seems that the rhythm of vocal turn-taking between infants and their primary caregiver goes faster than between infants and their secondary caregiver.

2.3 Study 3

*Introducing WCM-SE: The Word Complexity Measure phonetically justified and adapted to Swedish*

There are not many test materials for assessing Swedish expressive phonology, and none that have taken universal phonetic principles into account. Current Swedish materials are all based on words, and also designed for assessment of children from age 3 and up.

The WCM-SE is an adaptation to Swedish of the Word Complexity Measure (Stoel-Gammon, 2010). The WCM-SE is based on phonological structures instead of words and suitable for assessment of very early and also deviant phonology. Individual as well as relational analyses of productions can be made, and it is suitable both in research and in clinical settings, enabling comparisons between children and over time. Like the WCM, the WCM-SE is based on acquisition of phonological structures, but the WCM-SE also takes universal aerodynamic and articulatory principles into account. For this reason, the WCM-SE is a valuable contribution to materials of expressive phonology.

This study describes the adaptation of the English Word Complexity Measure (Stoel-Gammon, 2010) to Swedish. The measures are used to calculate complexity of words or vocalisations by the use of complexity parameters. The parameters occurring in a word or production are given points, and the sum gives the complexity score. The measures are suitable for the assessment of early or deviant phonology, and can be used for independent and relational analysis. The original English measure has eight parameters selected on the basis of being acquired late among English-learning infants. The Word Complexity Measure for Swedish (WCM-SE), has ten complexity parameters, that are distributed over word patterns, syllable structures and sound classes just as in the original measure. The word pattern parameters *productions with more than two syllables* and *stress on any syllable but the first* and the syllable structure parameters *word final consonant* and *consonant cluster* are the same in both WCM and WCM-SE. The differences concern sound class parameters, as follows: *Liquids, fricatives* and *voiced fricatives* were kept as in the original
WCM. Syllabic liquids, rhotic vowels and affricates are not phonemic classes in standard Swedish, and were therefore excluded in WCM-SE. Trills and the typically Swedish long, front rounded vowels were added.

Besides the adaptation to Swedish phonemic classes, the WCM-SE measure also is phonetically justified. The study describes parameter selection on the basis of age of acquisition as well as language-general phonetic characteristics. The WCM-SE is a valuable contribution to Swedish speech assessment materials, as there are not many tools to assess early expressive phonology in Swedish.

2.4 Study 4

The development of a vocabulary for PEEPS-SE — profiles of early expressive phonological skills for Swedish

As mentioned above, there are not many materials for the assessment of early expressive phonology for Swedish-learning children, and most materials are based on words that also are selected primarily to target specific phonological structures.

The PEEPS-SE test is an adaptation of the English PEEPS test. PEEPS-SE is a test for assessment of expressive phonology in Swedish-learning children from 18 to 36 months. The test uses words that are selected according to two types of criteria. The criteria of age of acquisition were used in word selection to heighten the likelihood of the child being familiar with the selected target word, and the complexity criteria enables the use of test words with differently defined phonological complexity as measured with the WCM-SE. The PEEPS-SE consist of two word lists. The Basic Word List contains words earlier acquired and less complex words compared to the words in the Extended Word List. Which word lists are used during the assessment depends on the age and/or linguistic developmental level of the child.

This study forms part of the creation of the Swedish version of the American English Profiles of Early Expressive Phonological Skills (PEEPS; Stoel-Gammon & Williams, 2013). This work describes the selection of words for the Swedish version of the test, PEEPS-SE, that is in accordance with the selection procedure for the original PEEPS. The words are selected on the basis of two types of criteria: age of acquisition and word complexity. The words selected also need to be easy to elicit in a test situation. The age of acquisition criteria were set by Stoel-Gammon and Williams (2013) to make sure that selected test words should be familiar to children in relevant ages and therefore more likely to be produced. Stoel-Gammon and Williams also set complexity criteria, so that the BWL would contain less complex words, compared
to the EWL. The criteria were followed as closely as possible in the Swedish
word selection procedure.

To find words that meet the age of acquisition criteria, Swedish vocabulary
data from the prospective longitudinal language development study SPRINT
was used. Complexity in words was calculated by the Word Complexity
Measure for Swedish (WCM-SE; Marklund, Marklund, Schwarz & Lacerda,
2018) (see Study 3). For details about the selection criteria, see paper 4.

With the PEEPS-SE test, a child’s production can be assessed by the use of
words with various complexity levels, word forms and syllable structures.
With respect to sound classes, only consonants are targeted. Toys are used to
elicit words. The PEEPS-SE test is useful to assess early or deviant phonology
in Swedish-learning children, both in the clinic setting and in research.
Through the adaptation to child vocabulary, by offering target words with vari-
ous complexity, and being able to be used from as early as 18 months, the
PEEPS-SE is a valuable contribution to the word-based materials for assess-
ment of expressive phonology in Swedish learning children. This study has
contributed to the development of the PEEPS-SE test by creating the word
lists. Further development work for the test and norm data collection is cur-
rently ongoing.
3. Discussion

In this work, infant communication and speech development is considered from an ecological, functional theoretical perspective. With parent-child interaction in focus, it discusses specifically infant development of non-vocal and vocal turn-taking abilities, as well as the development and measurement of expressive phonology. The findings shed light on patterns of parent responsiveness and its potential impact on child language and communicative outcomes. They also contribute substantially to the available means of assessing early expressive phonology in Swedish-learning children.

Child vocabulary growth is known to be related to many factors in the child’s environment (e.g., Goodman et al., 2008; Brooks & Meltzoff, 2008), and in this thesis parents’ prompt responses to their 18-month-olds were found to be positively related to child vocabulary size (Study 1; Marklund et al., 2015). Young infants turn-taking with their caregivers have in previous studies showed reciprocal patterns of several properties, for example gaze (Rutter & Durkin, 1987), smiles (Messinger & Fogel, 2007) and durational properties in vocal exchange (Dominguez et al., 2016). This thesis showed that caregiver status influenced pause duration in both and caregivers and their 6-month-old infants, with the shortest pauses in interaction between primary caregiver and infant (Study 2).

To find out more about parents’ utterances and responses, how they are related to infant vocal turn-taking and child vocabulary parental responsiveness, and how they contribute to infant and child language learning, there are several approaches and perspectives to discuss.

In Study 1, the purpose was to investigate how infant vocabulary size was related to the infant’s general language input from parents, not specifically how mothers or fathers talk. Therefore, data from the mother and the father in the same family were collated and treated as parental data. In some of the recordings, siblings were present, but they were not included in the analyses. Other interesting aspects could have been revealed if we had treated mother and father data separately, and/or if we had included turn-taking patterns with siblings. Since these factors were not controlled for in the study, this was not possible to do.

In Study 2, we aimed to find out if there were differences in vocal temporal contingency in caregiver responses to their 6-month-old infant as a function of caregiver status. In this study, all primary caregivers were mothers and all
secondary caregivers were fathers. Similar to Johnson and colleagues (Johnson et al., 2014), we found that responses from primary caregivers (mothers) were faster compared to secondary caregivers (fathers). Additionally, in both Study 2 and in Johnson’s study (Johnson et al., 2014) it was found that infant responses were faster to the primary caregiver (mother) in comparison to the secondary caregiver (father). Although there have been some previous findings of differences between mothers and fathers in their speech to infants, differences in interactional patterns may primarily be related to how much time the caregiver has spent with the child, not to the sex/gender of the caregiver. Further studies on responsiveness in relation to caregiver status could shed more light on this topic, for example longitudinal studies which would cover for example the change to secondary caregiver going on parental leave; or studies of infants at ages when he or she has more experience of conversations with other adults than the parents. Since far from all children grow up with a mother and a father, further research on vocal contingency in infant-caregiver interaction in other types of family constellations (e.g., families with single caregivers, same-sexed caregivers, live-in grandparents or other related adults, families with non-related nannies, commune families) could be valuable contributions to the research field.

Different settings in the recording situation might also affect turn-taking patterns. In Study 1, the recordings were made in four different daily situations, and in Study 2 caregivers were instructed to play with hand dolls. More naturalistic situations like in Study 1 may result in data that is more ecologically valid compared to the situation in Study 2, but on the other hand less controlled and therefore less comparable. Responsiveness can vary according to the type of activity infant and parent are engaged in (Gros-Louis et al., 2016). Conversational settings can vary according to the parent’s and the infant’s personal preferences of activity, and thus affect interactional patterns. Other dynamic factors such as the parents’ stress level and mental state are also likely to have impact on parent-infant interaction. If the parents in Study 2 had had the opportunity to choose the activity themselves, this could have led to other results. Finding out more about which situations parents prefer and which situations they find most challenging when communicating with infants could tell us more about the child’s linguistic environment and how it affects child language learning.

Even though parent responses mostly seem to be multimodal, it has been suggested that there are some differences according to which modalities (e.g., touch or speech) are most frequent in parent responses across different cultures (Kärtner et al., 2008). Video-recording also presents more information about interactional properties such as gaze, gestures and distance between infant and parent.

Socio-cultural contexts are known to influence parental responsiveness (e.g., Kärtner et al., 2008; Ochs & Schieffelin, 2011). Differences in parenting
practices and parental knowledge about a child’s development can affect parental responsiveness, and in turn, the language development of the child. In addition, the attitudes and beliefs of parents to children with speech difficulties are interesting to study and relate to responsive behaviour of parents in interaction with their child. The Parent-Child Communication Survey (PCCS: Marklund Schwarz & Marklund, 2016) is a questionnaire surveying parent attitudes and beliefs about child language development. It explores to what extent parents use specific strategies in their communication with young children to support their language learning.

In Study 1, the measure utterance duration (in seconds), not utterance length (e.g. MLU), was used to investigate parental utterances. As we wanted to know not only the actual time the infants heard parental language, but also how much linguistic content they were exposed to, we calculated the amount of linguistic content per time unit in parents’ utterances, giving a measure of speech rate. No differences between vocabulary groups were found. However, qualitative estimations of parental speech in the recordings revealed differences in clarity of parental speech. To find out more about how infant development of expressive phonology is affected by parents’ vocal responses, it would be interesting to investigate the phonetic aspects of parents’ speech and how this is related to early speech learning (e.g., Liu et al., 2003). Previous studies have shown that contingent vocal responses cause infants’ vocalisations to become more mature (Goldstein & Schwade, 2008) and that more mature infant vocalisations are more likely to receive a parent response (e.g., Goldstein & West, 1999; Gros-Louis, West & King, 2014). The relationship of specific phonetic aspects of child-directed speech on the development of early expressive phonology is also a possible avenue to follow up from the studies in this thesis. Clarity of parental speech in terms of hypo-hyperarticulation could be studied in relation to children’s early word productions, and also to what extent the parent speech mirrors infant utterances phonetically and vice versa.

In this thesis, early phonological development is described from a language-general as well as language-specific perspective. The specific developmental course of Swedish expressive phonology, and the importance of age-appropriate assessment tools are highlighted. In Study 3, the development of the Word Complexity measure for Swedish (WCM-SE) is described. The WCM-SE, an adaptation of the English Word Complexity Measure, is a measure based on universal and language-specific acquisition of phonological structures. The ten complexity parameters in the WCM-SE are selected on the basis that they are complex to produce, and that they are acquired late. The selection of structures is also phonetically justified by the descriptions of articulatory and aerodynamic principles. Future work should focus on instrument validation.
Even though a phonetic perspective is taken when selecting the parameters it is not obvious which structures to include. There are likely to be other candidates of sound classes, word patterns or syllable structures considered sufficiently phonetically complex to be included in the measure. This can depend on contextual factors that vary across different dialects. However, the measure is easily adapted to other languages or dialects. Furthermore, nine out of ten parameters are assigned the same complexity level, and are given one point for each occurrence in the calculations. This is somewhat problematic as all parameters presumably vary in difficulty depending on adjacent speech sounds and on supra-segmental structures. The trill /r/, on the other hand, is given three points, as common substitutions of /r/ in early speech in Swedish-learning children are found as other parameters (e.g., a voiced fricative) given points in the WCM-SE. Assigning the /r/ three points, and the voiced fricatives two points is a step towards a more nuanced gradation.

Development of a complexity measure based on aerodynamic and articulatory variables where degrees of freedom in the vocal tract are considered is currently in progress. This measure will describe the complexity of different speech sounds, but in contrast to WCM-SE, it is not language-specific. It is an attempt to create a language-general measure based on general phonetic principles (Lacerda & Marklund, in preparation). The WCM-SE is a relatively crude measure, but easy to use and easily adapted to other languages and dialects. It is useful for both individual and relational analyses for comparisons over time or between children and easy to use in both the clinical setting and in research.

The selection of words for the PEEPS-SE test, described in Study 4, was based on the selection criteria for the original PEEPS (Stoel-Gammon & Williams, 2013). By using age-of-acquisition criteria in the selection of words, the likelihood of having words in the test that are familiar to children in targeted ages is heightened. The use of word complexity criteria – as measured with WCM-SE – in the selection to the two word lists, enables the test leader to easily expose even the youngest children or children with severe speech problems to test words that are relatively easy to pronounce.

Assessing deviations in speech is important as it enables early intervention. Early language skills are known to be associated with later language outcomes (e.g., Marchman & Fernald, 2008) school performance (e.g., McCormack et al., 2011), life success and psycho-social wellbeing (e.g., Whitehouse, et al., 2009; Law et al., 2009). There are not many tools or tests to assess early expressive phonology in Swedish-learning children, and especially not young children, therefore the WCM-SE and PEEPS-SE are valuable contributions. The WCM-SE has the advantage of not being word-based, but based on phonological structures instead, whereas PEEPS-SE’s word selection is tailored to suit children of ages 18-36 months and norm data will shortly be collected for this age group.
Early detection of speech and language disorders enables early intervention, and for young children, parent-based intervention programs have been found to be particularly effective (Roberts & Kaiser, 2011; Ha, 2015). However, it may be difficult to predict later outcomes very early, and according to a recent study, 15 months is found to be too early to reliably predict outcomes (Bornstein, Hahn & Putnick, 2016). Around 50% of young children who are identified with speech and language difficulties self-recover within a couple of years (Norbury, C., personal communication 2018-10-12). However, in other cases, language disorder may be persistent, and no full recovery will occur since language skills are fairly stable over time and the developmental rate of language skills will not exceed that of a typically developed child (Bornstein et al., 2016).

There are strong genetic influences on language skills (Bishop, Laws, Adams & Norbury, 2006), and parent-based intervention programs can be difficult for parents who are likely to have language problems themselves (Norbury, C. personal communication 2018-10-12).

Still, parental responsiveness can positively influence infant and child language and communication development. Parent-based intervention programs can help increase parental awareness of simple strategies to use in daily life that are highly valuable for scaffolding interaction, and can also boost parental beliefs, that they are able to support their child’s language development even if they themselves experience language difficulties (Girolametto, et al., 1996).

My interest in parental responsiveness and how it can support child language development was born in clinical work at the speech clinic at Danderyd’s Hospital. As a Hanen-certified speech language pathologist working with parental programs my aim was to increase parents’ awareness of responsive and supportive talking styles in interaction with their toddlers with speech and language difficulties. This work was very different from other clinical work I had previously done. I felt it was important to support the parents in discovering that their toddlers were actually communicating even if they weren’t talking. I was happy to see the parents putting effort in learning how to support their children and to follow the process of how the parents learned how to encourage their toddler to communicate.

Regardless of how children learn, the importance of learning language is hard to question, and child language development is therefore always something to support unconditionally – albeit in different ways. As the ultimate goal for a child is not to learn to communicate only with the parents, but over time with more, and also unknown, people, parents simply form the natural link between the restricted family environment and future social interaction out in the world. Parental behaviour has strong implications for further child language development, but it is important to see the child as developing towards independence in social communication, even though the role of being a parent never ends.
4. References


solve?urn=urn:nbn:se:liu:diva-109308


of age. Monographs of the society for research in child development, 63(4) (pp. 1-174).


Field, T., Diego, M., & Hernandez-Reif, M. (2009). Depressed mothers’ infants are less responsive to faces and voices. *Infant behavior and Development, 32*(3), 239-244.


MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and Performance 13: Motor*


Appendix: General definitions of terms

*Speech* can be described in terms of phonetic and linguistic properties. Speech is produced in the *vocal tract*, which is comprised of the oral and nasal cavities and the *articulators*: lips, tongue, jaw and soft palate. Voice is produced in the *larynx*, by vibrating vocal folds.

A *phonetic* description of speech concerns articulation and acoustics. Articulation concerns physiology, how speech is produced in the vocal tract and how the aerodynamic energy of this action transforms into acoustic energy. Acoustics concern the physical properties of the speech signal, described in physical terms: frequencies, intensity and duration. Frequencies determine pitch and resonance, intensity the loudness, and duration the extent in time.

A *phone* is any speech sound regardless of whether it bears meaning or not for a given language. A phone can be described in terms of the distribution of acoustic energy over the frequencies – spectral information. Different speech sounds vary according to spectral information.

*Linguistic* properties of speech are those linked to language and can be described in terms of features, segments, syllables, words or utterances. A speech segment, for example a phoneme, morpheme or utterance can bear linguistic meaning.

A *phoneme* is an abstract representation of a speech sound that is perceptually distinct from other speech sounds in a given language. The phoneme is the minimal unit that distinguishes one word from another, for example the initial speech sounds in each of the words “cat” and “hat”. Some speech sounds can be interchanged without changing the meaning of a word in the specific language, for example the alveolar /r/ and the uvular /R/ in Swedish. These different variants of the same phoneme are called *allophones*.

Phonemes are classified as *vowels* and *consonants*. Vowels are produced with voice and an open vocal tract and with minimal airflow constriction. Vowels differ according to the position of the tongue in two dimensions: the front-back dimension, and the open-closed dimension. In addition, vowels can differ according to lip rounding. Consonants comprise airflow constriction applied
by articulators and can differ according to articulation place, articulation manner and voicing. Place of articulation refers to the place in the vocal tract where the articulators are involved or active. Manner of articulation refers to how the articulators work or interact to produce a specific speech sound type. Some consonants are produced voiced (with vocal folds vibrating) and some unvoiced.

*Phonology* concerns speech sounds in relation to language and can be described on a segmental level (isolated phonemes), or on supra-segmental level, where phonemes are related to other properties such as prosody and syllable structure. *Prosody* refers to stress, intonation and rhythm. *Syllable structure* concerns the position of a segment within a word and also in relation to other segments in the word. Each language has its specific *phonological system* depending on the speech sound inventory and the *phonotactics* – the rules of how speech sounds can be combined.

Besides writing with letters (orthographically), speech can be *transcribed* phonetically or phonologically. Most often the *International Phonetic Alphabet* (IPA) system is used for transcriptions (see section 1.5.2.3). A phonetic transcription is based on phones, written within square brackets: [], and can be on different levels, from broad to narrow, depending on how much details are captured. A phonological transcription is based on phonemes (the contrastive sounds of the specific language), is closer to a broad transcription, written within slashes: //, and does not typically include allophones. The IPA symbols used for transcribing phonetically and phonemically are the same (see Figure 1).

Speech sounds are not isolated phenomena, but affected by context, for example *coarticulation* - the transitions between adjacent speech sounds - and suprasegmental characteristics. Therefore, phonetic descriptions of produced speech presumably vary according to the context. Additionally, how the speech is perceived is always coloured by the native language of the listener and essentially captured by phonological transcription. Still, phonology is a practical way of systematically describing a phonetic continuum in terms of linguistic categories. It is based on conventions that are necessary for linguistic descriptions of speech sounds, and the IPA system offers a useful system for this purpose.

*Phonological processes* refer to transformation rules that describe how speech sounds are affected by context and how speech production is changed in relation to target (see also section 1.2.1.1). Regardless of the theoretical perspective on phonological development, and the possible causes of a “deviant” production, phonological processes are useful as descriptions of speech errors in relation to target productions. A phonological process means that there is a deviation from the target phonology that can involve for example a change
from one structure to another, a reduplication or a deletion. The processes are mainly at segmental level, but they can also include word forms and syllable structure. Changes on the segmental level may concern articulation place, manner and/or voice.

Phonological processes can be independent of context – *paradigmatic* – or dependent on context – *syntagmatic* – or both. Examples of paradigmatic processes are fronting: /k/ → [t], backing: /t/ → [k], stopping: /s/ → [t], and unvoicing: /d/ → [t]. Examples of syntagmatic processes are cluster reduction: /kr/ → [k], and contact assimilation: /saks/ → [sask].

Certain deviations from the target phoneme may still be interpreted as representative of the underlying phoneme and are viewed as *distortions*, for example lisp.
Fig. 1 International Phonetic Alphabet. Provided by International Phonetic Association under creative commons license (no changes made).  
https://www.internationalphoneticassociation.org/content/full-ipa-chart#ipachartdoulos