Gesture-speech combinations in child language

Form, function, and how they relate to language acquisition

Freya Eriksson
Gesture-speech combinations in child language

Form, function, and how they relate to language acquisition

Freya Eriksson

Abstract

The present study investigates children’s production of gesture-speech combinations and how they relate to language acquisition. 16 children were observed at seven age points (0;9, 1;0, 1;3, 1;6, 1;9, 2;0 and 3;0), and their gesture-speech combinations were classified into the categories complementary, supplementary and discourse combinations. The results show that the production of gesture-speech combinations over time follows different patterns in children with high, average and low productive vocabulary. Furthermore, the amount of gesture-speech combinations produced at four age points predicted productive vocabulary at the age of 2;6, and the amount produced at the age of 1;6 predicted sentence complexity at 3;0. The use of complementary gesture-speech combinations at three age points was also linked to productive vocabulary. The age of onset of supplementary gesture-speech combinations predicted sentence complexity at 3;0, while the age of onset of discourse gesture-speech combinations predicted productive vocabulary at 2;6. The results support previous research suggesting that complementary and supplementary gesture-speech combinations play an important role in child language acquisition. Additionally, the results of the present study suggest that discourse gesture-speech combinations are also connected with language development.

Keywords

Gesture-speech combinations, language acquisition, complementary, supplementary, discourse, parent-child interaction
Gest-ordkombinationer i barns språk

Form, funktion, och hur de relaterar till språkutveckling

Freya Eriksson

Sammanfattning

Den här studien undersöker barns produktion av gest-ordkombinationer och hur dessa är relaterade till språkutveckling. 16 barns produktion undersöktes vid sju ålderspunkter (0;9, 1;0, 1;3, 1;6, 1;9, 2;0 och 3;0), och deras gest-ordkombinationer klassificerades i kategorierna komplementära, supplementära och diskursiva kombinationer. Resultaten visar att produktionen av gest-ordkombinationer följer olika mönster hos barn med högt, medel och lågt produktivt ordförråd. Mängden gest-ordkombinationer som producerades vid fyra ålderspunkter predicerade produktivet ordförråd vid 2;6, och mängden gest-ordkombinationer som producerades vid 1;6 predicerade meningskomplexitet vid 3;0. Användningen av komplementära gest-ordkombinationer vid tre ålderspunkter uppvisade ett samband med produktivt ordförråd vid 2;6. Tillägnandelandern för supplementära gest-ordkombinationer predicerade meningskomplexitet vid 3;0, medan tillägnandelandern för diskursiva gest-ordkombinationer predicerade produktivet ordförråd vid 2;6. Resultaten i den här studien ger stöd åt tidigare forskning, som har visat att komplementära och supplementära gest-ordkombinationer spelar en viktig roll i barns språkutveckling. Därutöver visar resultaten i den här studien att diskursiva gest-ordkombinationer också har ett samband med språkutveckling.

Nyckelord

Gest-ordkombinationer, språkutveckling, komplementär, supplementär, diskursiv, förälder-barninteraktion
Contents

1. Introduction ................................................................................................................. 1
2. Background .................................................................................................................. 2
   2.1 Language and multimodality .................................................................................. 2
   2.2 Cognitive and linguistic development ..................................................................... 2
   2.3 Gestures .................................................................................................................. 3
       2.3.1 Gestures in language evolution ...................................................................... 3
       2.3.2 Gestures in language acquisition .................................................................... 4
   2.4 Gesture-speech combinations in language acquisition ........................................... 5
3. Aim and research questions ....................................................................................... 7
   3.1 Aim ......................................................................................................................... 7
   3.2 Research questions ............................................................................................... 7
4. Method ......................................................................................................................... 8
   4.1 The MINT Project ................................................................................................... 8
   4.2 Data ....................................................................................................................... 8
       4.2.1 Participants .................................................................................................... 8
       4.2.2 Sample selection ......................................................................................... 8
       4.2.3 Recording sessions ..................................................................................... 8
       4.2.4 Ethical considerations .................................................................................. 9
   4.3 Annotation .............................................................................................................. 9
       4.3.1 Annotation in ELAN .................................................................................... 9
       4.3.2 Definition of utterances .............................................................................. 9
       4.3.3 Definition of gestures ................................................................................ 9
   4.4 Data analysis ........................................................................................................ 11
       4.4.1 Definition of gesture-speech combinations .................................................. 11
       4.4.2 Categorization of gesture-speech combinations ........................................... 11
       4.4.3 Productive vocabulary measure: SECDI-II .................................................. 11
       4.4.4 Sentence complexity measure: Unified predicate .......................................... 12
       4.4.5 Statistical analysis ..................................................................................... 12
5. Results ......................................................................................................................... 13
   5.1 Productive vocabulary and sentence complexity ................................................... 13
   5.2 Descriptive results ............................................................................................... 14
       5.2.1 Age of onset of gesture-speech combination categories .............................. 14
       5.2.2 Gesture-speech combinations over time ....................................................... 14
       5.2.3 Gesture types in gesture-speech combinations ............................................ 15
   5.3 Gesture-speech combinations in relation to later language skills .......................... 17
       5.3.1 Total amount of gesture-speech combinations ............................................. 17
5.3.2 Complementary and supplementary combinations ........................................19
5.3.3 Age of onset of combination categories ......................................................20

6. Discussion ......................................................................................................... 23
  6.1 Method discussion .......................................................................................... 23
    6.1.1 Validity ..................................................................................................... 23
    6.1.2 Reliability ................................................................................................ 23
    6.1.3 Generalizability ....................................................................................... 24
  6.2 Result discussion ............................................................................................ 25
    6.2.1 Order of acquisition of combination categories ...................................... 25
    6.2.2 Production of gesture-speech combinations over time .............................. 25
    6.2.3 Gesture types in gesture-speech combinations ......................................... 26
    6.2.4 Amount of combinations and later language skills .................................. 27
    6.2.5 Complementary and supplementary combinations and later language skills ... 27
    6.2.6 Age of onset and later language skills ..................................................... 28
    6.2.7 General discussion .................................................................................. 29
  6.3 Ideas for future research ................................................................................ 30

7. Conclusions ......................................................................................................... 31

References ............................................................................................................. 32

Appendices ............................................................................................................ 34
  Appendix 1. Information letter to parents ............................................................ 34
  Appendix 2. Annotation key for vocal and gesture tiers in ELAN ......................... 36
1. Introduction

All typically developed children learn language, but we are far from understanding exactly how they do it. Despite the fact that language in its nature is multimodal, the focus of language acquisition research has long been on the speech signal, its properties and content. In later years, however, the interest in other modalities than the spoken one, and these modalities’ relation to language acquisition, has grown.

The relationship between gesture and speech is not entirely understood, but according to Kendon (2004: 3) they are to be considered as two integrated forms of expression that work towards a common goal when produced together.

Gestures can provide insight into both child language acquisition and the evolution of language in the human species. There are several theories of the evolution of language; some argue that spoken language evolved from a manual language, others that speech and gesture have been used together from the start.

Gestures in child language acquisition have been a topic of interest during the last few decades and in later years there has been a focus on how gestures and speech are used together in children’s development. The main focus has been on two types of gesture-speech combinations, complementary and supplementary, and both of these types have been linked to language acquisition. Previous research into the topic of gesture-speech combinations has focused mainly on pointing and other representational gestures. This thesis aims to expand on this by including all gestures used by children during their first few years of life. Complementary and supplementary combinations are further investigated in the present study, and a category called discourse gesture-speech combinations is added to incorporate what McNeill terms discourse gestures (1992: 15-18).

The present study thus aims to give a comprehensive view of children’s use of gesture-speech combinations during development, and add to the growing body of research suggesting that gestures play an important role in child language acquisition.
2. Background

This chapter presents previous research. The first section presents language and multimodality, focusing on the different modalities that come into play in communication, followed by a short description of cognitive and linguistic development. The third section is on different aspects of gestures and gesture research, including how gestures are used, gestures’ role in language evolution and gestures’ role in language acquisition. The final section presents research on gesture-speech combinations and their role in language acquisition.

2.1 Language and multimodality

Language is multimodal. In interaction, humans use their bodies to make others aware of their intentions, mood, etc. An important part of this is the use of gesture – movements of the hands and the head that are part of what a person is saying (Kendon, 2004: 1).

During the twentieth century spoken language has been one main focus of linguistic research, but ever since the 1970s there has been a growing interest in research on other modalities, such as the manual, resulting in a surge in research about gestures. At the same time there was also a rise in the interest in signed languages (Kendon, 2004: 2).

An early description of nonverbal behavior comes from Ekman and Friesen (1969). They found that nonverbal behavior can be coded according to three major aspects: arbitrariness, i.e. the form of the expression is not linked to its meaning; iconicity, i.e. the form of the act or sign is linked to its meaning so that the meaning becomes apparent from form alone; and intrinsicness, i.e. the act is the same as what it signifies (e.g. hitting someone as an expression of anger). Ekman and Friesen (1969) further categorize five types of nonverbal behavior: emblems (gestures that have a verbal equivalent), illustrators (movements that illustrate the verbal), affect displays (body movements or facial expressions that show emotional states), regulators (movements that regulate the flow of conversation), and adaptors (grooming, flirtation etc.).

The present study will approach multimodality from the point of co-speech gestures.

2.2 Cognitive and linguistic development

Human development is complex, and many different aspects are intertwined. According to dynamic systems theory (Van Geert & Verspoor, 2015), language development involves a system that successively increases in complexity, and as it does so more components are integrated into it. The components, i.e. the motor-system, sensory system etc., then need to be coordinated for the system to become stable.

During the first few years of life children’s linguistic and cognitive skills develop simultaneously. At around 0;9 years they start directing others’ attention using pointing gestures (Kita, 2003: 2). Around their first birthday they can establish joint attention (Tomasello, 2003: 41), and at this time they also produce their first words (ibid., 2003: 36). At approximately 1;6, children start combining words into two-word sentences (Goldin-Meadow & Butcher, 2003). Around this time some children go through a period of rapid word learning, sometimes called the ‘vocabulary spurt’, coinciding with when the child’s vocabulary approaches 50 words (Benedict, 1979; Goldfield & Reznick, 1990). The increased speed of word learning might also be related to the ability to combine words and use rudimentary grammar (Tomasello, 2003: 50f). The vocabulary spurt has also been linked to handedness; children have a clearer hand preference in pointing after the vocabulary spurt period (Cochet, Jover & Vauclair, 2011).
At two years of age two-word sentences are quite common in children’s production (Özçalışkan & Goldin-Meadow, 2009).

2.3 Gestures

Gesture is an integrated part of language (McNeill, 2000: 9). Kendon (2004: 1-3) argues that gesture and speech are two modes of expression, and that there is a clear difference between them. While spoken language uses a syntactic system to organize words into utterances, gestures do not follow such a clear system. Gestures can be classified along a continuum, from the type of co-speech gestures that require speech in order to be understood on the left, to emblems and pantomimes, i.e. gesture that can be used and understood without speech, on the right (McNeill, 1992: 37; 2000: 5). When gestures are combined with speech, they can either be used as a complement, supplement or substitute to spoken language, or together with an utterance to modify it (Kendon, 2004: 1). Since gesture and speech are so closely interlinked, it has been suggested that gestures can provide an understanding of the processes that are the foundation of language.

Gestures are context dependent and get their meaning from the context they appear in. McNeill (1992: 105) points out that while gestures have meaning, they differ from spoken language in that they are created in the moment, i.e. are not retrieved from a lexicon. Kendon (2004: 163) argues that a certain gesture on its own can have a general meaning, but in combination with speech that meaning is modified.

While there are several ways of classifying gestures, McNeill (1992: 12-18) identifies five main types of co-speech gestures: iconic (illustrate a concrete element of the action described), metaphoric (illustrate an abstract idea), deictic (pointing), beats and cohesives. McNeill describes beats and cohesives as discourse gestures that reveal the speaker’s notion of the narrative structure. Beats emphasize a word or phrase as important, while cohesives tie together temporally separate parts.

Kendon (2004: 176-198) describes six ways in which gestures can be used together with speech. Gestures with verbal equivalents (i.e. emblems) can be used both in combination with their verbal equivalent, making that word or phrase more prominent, or without their verbal equivalent, to replace a word or to express something that is not explicit (e.g. that the narrative happened in the past). Gestures can also be used to describe the manner in which a verb is performed, making the utterance more specific in meaning. Further, gestures can be used to illustrate a referent, or show it if it is present, and also to refer to the size, shape etc. of an object. In these ways gesture and speech combined make for a more complete whole. It is also possible to use gestures to create a referent that can then be indexed, e.g. drawing a map in the air and pointing to different places on it.

2.3.1 Gestures in language evolution

It has been argued that gestures have played a part in the evolution of language, and some have proposed that it is part of the origin of language as we know it today (Kendon, 2004: 4).

According to the frame/content theory of the evolution of speech production (MacNeilage, 1998), speech consists of two components: frames and content. The frame component is suggested to have evolved from movements such as chewing, sucking and licking, which became communicative facial gestures. These types of facial gestures, e.g. smacking the lips or tongue, or teeth chattering, are present in nonhuman primates. The content component is related to the system responsible for responding to external input. MacNeilage argues that these two components are different realizations of the motor control systems. He also states that the relationship between speech motor control and hand motor control is related to a specialization of the left hemisphere of the brain.

Gentilucci and Corballis (2006) suggest that spoken language evolved from manual gestures. They argue that the vocal tract and the hands are closely connected, and that spoken language is essentially vocal gestures. Thus, through a series of changes to the vocal tract, possibly facilitated by a mutation
of the FOXP2 gene,¹ which happened after the human lineage split from that of chimpanzees and bonobos, speech became the main modality of language.

Gesture usage in some of our closest relatives in the animal kingdom, chimpanzees and bonobos, can perhaps shed some light on the evolution of gestures in the human species. Graham, Hobaiter, Ounsley, Furuichi and Byrne (2018) compared bonobo and chimpanzee gestures to each other to examine whether the two species have a gestural repertoire that is mutually understandable. Around 90% of chimpanzees’ and bonobos’ gestures have the same form. By investigating the outcome of different gestures Graham and colleagues established their meaning, and when comparing the two species they found that there is also a large overlap when it comes to gesture meaning. Based on their results, Graham and colleagues suggest that the patterns of gesture that exist in bonobos and chimpanzees could be traced back to the last common ancestor of great apes and humans, and that gestures played an important part in the evolution of language.

At the very least, there is evidence suggesting that gestures have played some role in the origin of language. Fay, Lister, Ellison and Goldin-Meadow (2014) conducted a study where the participants had to create a common communication system in either the vocal, gestural or combined vocal-gestural modality. They found that gestures outperformed vocalization even when vocalizations and gestures were used in combination. Fay and colleagues thus argue that in light of gestures being such an effective means of communication it is unlikely that early humans did not make use of them.

While there are some who are critical towards the view that language evolved from manual gesture alone into spoken language, there is evidence that points towards language evolving through an interaction between the vocal and gestural modality (Perlman, 2017: 393).

2.3.2 Gestures in language acquisition

There is a growing body of research suggesting that children’s gesture production is related to language development. Children start communicating using gestures before they can speak, and the first gestures occur at around the age of 0:9 to 1:0 (Iverson & Goldin-Meadow, 2005: 367). Deictic gestures, i.e. pointing, are the first to appear (Kita, 2003: 2), and are the most common in children’s production (Özçalışkan, Adamson & Dimitrova, 2016). During the second and third year of life the gesture system continues to develop, and emblems and early iconic gestures appear. Discourse gestures, e.g. beats, cohesives and abstract pointing, are the last to develop, since they require the child to understand the metalinguistic and metanarrative features of language. These gestures appear around the age of five and are not developed fully until around the age of eleven (McNeill, 1992: 300-326).

Pointing gestures can be used both as declarative and imperative speech acts, and both types of pointing occur before children start using spoken language (Bates, Camaioni & Volterra, 1976). The imperative pointing can be used to get an adult to manipulate a toy is some way (1976: 56). The declarative pointing, however, is described as pointing to direct the interactor’s attention to an object (1976: 61).

Capirci, Contaldo, Caselli and Volterra (2005) investigated the relationship between gestures and language development in three children between the age of 0:10 and 1:11 years, and found that all children started to communicate during the first session, mostly using gestures. However, as the children became older speech became the preferred modality for communication.

Early gesture production can be indicative of language level later on in development. Lüke, Ritterfield, Grimminger, Liszkowski and Rohlfing (2017) investigated the use of deictic gestures in typically developed children and children with language delay. They found that at the age of 1:0 all children were pointing (with the whole hand or with index finger), but most of the typically developed children used index-finger pointing, while the children with language delay only pointed using the whole hand.

¹ The FOXP2 gene has been called the “grammar gene” (Pinker, 1994: 330), although it is most likely not the only gene responsible for grammar. Two functioning copies of the FOXP2 gene are required for a normal language acquisition, and although its function has not been fully established, it has been suggested to be related to articulation and the mirror system (Gentilucci & Corballis, 2006).
The degree of index pointing increased over time in all children, but this happened about two months later in children with language delay. This illustrates the important relationship between the conventional form of pointing (using the index finger) and language development (Lüke et al., 2017: 3193).

Typically developed children use deictic gestures to refer to referents before those words are produced in speech, and it has been shown that there is a link between when referents are expressed in gesture alone and the time at which those referents are expressed in speech alone (Iverson & Goldin-Meadow 2005). Özçalışkan, Adamson, Dimitrova & Baumann (2017) found that this also holds for children with autism spectrum disorder and Down syndrome, highlighting that gesture plays an important role in vocabulary development.

There are several developmental stages in the relationship between gesture and speech (Capirci & Volterra, 2008). In the early stages, before children can express what they want to say in speech only, gestures are used alone. This includes for instance pointing at objects instead of naming them and nodding the head for “yes”. Further on in development gestures are used in combination with speech, and this is “linked to language growth” (2008: 37).

### 2.4 Gesture-speech combinations in language acquisition

Before children start combining words, they start producing gesture-speech combinations, and the age of onset of gesture-speech combinations where the gesture and speech convey different information has been found to predict the onset of two-word utterances (Goldin-Meadow & Butcher 2003; Iverson & Goldin-Meadow, 2005).

Children produce vocalizations and movements together early in development. Iverson and Fagan (2004) found that 6-month-old children who have started to babble use vocalizations and rhythmic movements together at a higher rate than children of the same age who are not yet in the babbling stage. Thus, when children start to use gestures at around the age of 0;9-1;0 a strong link between the vocal and manual modalities is already present.

Furthermore, Esteve-Gilbert and Prieto (2014) investigated how four children between the ages of 0;11 and 1;7 used gestures in combination with speech during the babbling and one-word stage. They found that children mostly used deictic gestures in combination with speech, and that the deictic gestures could be either declarative or imperative. Additionally, children temporally synchronized gesture and speech in an adult-like manner already at the babbling stage.

When investigating children’s production of grunting, Roug-Hellichius (1998: 52-53) found that gestures are produced together with grunts from the age of 0;11. The amount of these combinations increase into a peak at the age of 1;2, when the amount of grunts accompanied by a gesture outnumbers the amount of word-tokens produced.

In a study investigating the gesture-speech production of 40 English children between the ages of 1;2 and 2;10, Özçalışkan and Goldin-Meadow (2009) found that gestures seem to indicate that children are in the process of learning a construction, i.e. multiword utterances. The study included conventional, deictic (pointing, showing and requesting an object) and iconic gestures. The gesture-speech combinations produced by the children were divided into three categories: reinforcing (gesture and speech providing the same information), disambiguating (gesture clarified the referent of a deictic word) and supplementary (gesture added semantic information to the speech). The supplementary gesture-word combinations were then divided into three constructions: argument-argument, predicate-argument and predicate-predicate. Deictic gestures were interpreted as arguments, and iconic and conventional gestures as predicates. The results showed that half of the children produced gesture-speech combinations at the age of 1;2, and at 1;6 months all children did. Each of the three types of constructions were first produced by the children using gesture-speech combinations, and later established in speech alone. The production of the constructions in gesture-speech combinations first increased, peaked at a level above the parents’ use, and then decreased towards the parents’ level.
Özçalıskan and Goldin-Meadow conclude that gestures are used to compensate for a lower linguistic level when the children are younger, and as they develop they move toward speech being the preferred modality. When this happens, the use of gesture-speech combinations moves towards an adult-like frequency. They also argue that gestures are important when learning constructions, but not to flesh out construction that they already know. When speech becomes the preferred modality gestures can instead be used to enrich a speaker’s communication (2009: 215).

Capobianco, Pizzuto and Devescovi (2017) followed ten children once a month from the age of 0;10-1;0 until the age of 1;11-2;1, to investigate the role of gesture-speech combinations in early language acquisition. The study included three types of deictic gestures (pointing, request and showing) and investigated how these co-occurred with speech. The gesture-speech combinations were classified as complementary (the gesture not adding new information to the speech) and supplementary combinations (the gesture adding a semantic element to the speech), and the different types of combinations were tested to see if they predicted age of onset of two-word utterances and general language development at the age of 2;0. They found that complementary gesture-speech combinations appeared around 12.5 months of age and the production of this type of combination peaked at the age of 1;6-1;10. Supplementary combinations appeared later, at around 17.5 months of age, and peaked around the age of 1;8-1;11 months of age. Furthermore, Capobianco and colleagues found a positive correlation between the amount of complementary gesture-speech combinations produced at 1;0 and 1;6 and verbal complexity and vocabulary size at the age of 2;0, while the amount of supplementary gesture-speech combinations at 1;3 and 1;6 was positively correlated with verbal complexity at the age of 2;0, but not with vocabulary size. They also found that the age of acquisition of supplementary combinations predicted age of onset of two-word utterances. Hence, Capobianco and colleagues conclude that children’s use of supplementary gesture-speech combinations indicate that they have reached a higher stage in development where they can express more complex thoughts.

Fasolo & D’Odorico (2012) found that the use of supplementary combinations at the age of 1;6 predicted sentence complexity at 2;0, and that the use of complementary combinations at the age of 1;6 predicted vocabulary size and MLU (mean length of utterance) at 2;0, but not sentence complexity. They also looked at the mothers’ response to the different types of combinations, and found that they answered supplementary combinations with a syntactically complete translation of the child’s utterance. In the case of complementary combinations the response was most often to fill in function words that the child left out. They argue that these different types of responses can be linked to why the different types of combinations predict different things.

Specific gesture-speech combinations can predict the development of the corresponding construction in speech alone. Cartmill, Hunsicker and Goldin-Meadow (2014) followed 18 children between the age of 1;2 and 4;10, and looked specifically at complementary combinations where deictic gestures were used to modify nouns (i.e. point + lamp = the lamp). They found that the age of acquisition of complementary gesture-speech combinations where the gesture acts as a determiner to the noun predicts the age of acquisition for the determiner + noun construction in speech alone. Based on this result the authors argue that gesture-speech combinations in children’s language development are tied to learning new constructions even at the phrase level.

The present study is expected to add to previous research suggesting that gesture-speech combinations play an important role in the acquisition of language.
3. Aim and research questions

3.1 Aim
The aim of the present study is to investigate children’s production of gesture-speech combinations during their first years of life, and how this relates to language development. To answer the research questions 16 children’s gesture-speech combinations at the age points 0;9, 1;0, 1;3, 1;6, 1;9, 2;0 and 3;0 have been analyzed.

3.2 Research questions
1a. When do children start producing complementary, supplementary and discourse gesture-speech combinations?
1b. How does the production of gesture-speech combinations develop over the observed age points?
1c. What gesture types are produced in complementary, supplementary and discourse gesture-speech combinations at different age points?
2. Does the total amount of gesture-speech combinations produced at the observed age points predict a) productive vocabulary size at 2;6 and b) sentence complexity at 3;0?
3a. Does the amount of complementary gesture-speech combinations produced at the observed age points predict productive vocabulary at 2;6?
3b. Does the amount of supplementary gesture-speech combinations produced at the observed age points predict sentence complexity at 3;0?
4a. Does the age of onset of complementary gesture-speech combinations predict productive vocabulary at 2;6?
4b. Does the age of onset of supplementary gesture-speech combinations predict sentence complexity at 3;0?
4c. Does the age of onset of discourse gesture-speech combinations predict later language levels, i.e. productive vocabulary at 2;6 and sentence complexity 3;0?
4. Method

The following chapter outlines the method of the thesis. First, the MINT Project and the data are presented, followed by a description of the annotation process and the definition of gestures and utterances used. Subsequently the data analysis is described, including a description of how gesture-speech combinations were categorized and analyzed and a description of the tools used to assess the children’s language level. Lastly, the statistical analysis is described.

4.1 The MINT Project

The present study used data from the MINT Project, an ongoing project at the Department of Linguistics at Stockholm University (Gerholm & Gustavsson, 2018, under review). An invitation letter (see Appendix 1) was sent to 2000 randomly selected families in the Stockholm area who had a child born in August or September of 2013. 85 families started out in the project, and when the children were three years old 72 families remained. During the first three years of life, beginning at three months of age and going on every three months, the children and their parents were recorded while interacting in the lab. After the age of three, the families came to the lab every six months.

4.2 Data

4.2.1 Participants

All families in the MINT Project are from the Stockholm area, or lived there when invited to participate in the project. All children but two have at least one parent with a university degree, and none have a parent with less than a high school education. All but one family have an annual income of 4-800 000 SEK. At the age of 2;3 all children included in the present study had started preschool.

4.2.2 Sample selection

The sample selection was partly randomized. In this thesis, 16 children were included (8 girls and 8 boys). Only children that had participated in all recording sessions, and had SECDI data from the 2;6 year session were included in this study. SECDI is a parental questionnaire used to measure the children’s language level (Berglund & Eriksson, 2000a). Furthermore, all children included in this study are monolingual, i.e. have only Swedish spoken in their home. Grandparents’ or other relatives’ languages have not been taken into consideration. None of the children have any known developmental disorders.

The age points included in the present study are 0;9, 1;0, 1;3, 1;6, 1;9, 2;0 and 3;0.

4.2.3 Recording sessions

The recording sessions include approximately ten minutes of free interaction, where the children play with one parent in the lab. The room is equipped with three stationary cameras, and the parent wore an action camera on their chest. The child and the parent each wore one lavalier microphone, and the

---

1 Modeling Infant Language Acquisition from Parent-Child Interaction, funded by the Marcus and Amalia Wallenberg Foundation (MAW 2011.007).

2 The stationary cameras are of the model Canon HDMI XA10 and the action camera is of the model GoPro Hero3.
The room was further equipped with a condenser microphone to synchronize the audio with the video files. The room was further equipped with some pillows, stuffed animals and age appropriate toys. The sessions from when the children were 0:9 consist of two parts. First the dyad played with the toys that were provided and after about 5-10 minutes a researcher came in with two books that they were instructed to use as they liked. For the present study only the time without the books was used. From the age of 1:0 the sessions consisted of approximately ten minutes of free interaction, without a researcher in the room. After that a researcher came in and, depending on the age of the child, conducted a semi-structured play session or an experiment with the child. In this study only the free interaction, i.e. without a researcher present, was included.

4.2.4 Ethical considerations

The MINT Project is conducted in accordance with the guidelines of the Swedish Data Protection Authority and the Research Ethics Committee at the Karolinska Institute (Dnr 2011/955-31/1), and follows the guidelines for research on humans (2003:460). The families taking part in the MINT Project have signed informed consent, and all participant information has been handled in accordance with PUL (Personuppgiftslagen, 1998:204). All children included in the present study have been anonymized.

4.3 Annotation

4.3.1 Annotation in ELAN

The material was annotated in ELAN, a free software for multimodal annotation (Brugman & Russel, 2004). The annotation was done by trained annotators following an annotation key (Gerholm, 2018, see Appendix 2), and the annotators also had weekly meetings to discuss difficult cases. Several different modalities were annotated for both the parent and the child, e.g. vocals, gestures, gaze, touch etc. For the present study only the children’s vocal and gesture tiers have been included.

4.3.2 Definition of utterances

When annotating utterances were defined as a string of sounds surrounded by silence, with a maximum duration limit of approximately five seconds. Utterances longer than that were divided into smaller units. For the present study only utterances that have a communicative meaning have been included, i.e. coughing, burping and other non-communicative noises have been excluded from the analysis.

When annotating the children’s speech the tag {word} was used for proto-words and wrong pronunciation (Gerholm, 2018, see Appendix 2). The target word was given within curly brackets. For example, if the child said pa for lampa (lamp), pa was written as {lampa} in the material. Target words are identified using the context, e.g. activity, gestures or parental response.

In the 0;9-year session the children’s vocal tiers were annotated using a controlled vocabulary, which is a predefined set of utterance types, such as cooing, babbling, grunting etc. Therefore, to find possible early proto-words, all files (at all age points) that had these types of tags were checked manually and proto-words were transcribed. Further, all files were manually checked for potential proto-words that had been transcribed for instance as /pa/ for lampa due to an earlier annotation key.

4.3.3 Definition of gestures

There are ten gesture categories listed in the annotation key for the MINT Project: deictic, iconic, emphatic, emotive, grooming, show, offer, action and other. The duration of the gestures is

---

4 The condenser microphone is of the model AKG SE 300 B and the lavaliere microphones are of the model Sennheiser eW 100 G2.
defined as from the start of the motion until the body part comes to a resting position, in accordance with McNeill (1992: 83). In addition to the gesture type every gesture in the material has a description. A gesture can be given more than one type, e.g. action/emblem. The definitions of the gesture types provided by the annotation key are given in table 1.

Table 1. Description of the ten gesture types included in the material of this study, as defined in the annotation key for the MINT Project.

<table>
<thead>
<tr>
<th>Gesture type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deictic</td>
<td>Pointing gesture (whole hand, index finger, middle finger etc.) This also includes gaze-pointing and pointing with an object – always mark which type of pointing it is.</td>
</tr>
<tr>
<td>Iconic</td>
<td>Descriptive gestures, e.g. symbolizing shape, distance, height etc. Can also include visual elements of actions, e.g. ”pulled back” with a motion of pulling something back with the hands.</td>
</tr>
<tr>
<td>Emblem</td>
<td>Conventional gestures, e.g. waving, clapping, put a finger to the mouth for ”husch” etc.</td>
</tr>
<tr>
<td>Emphatic</td>
<td>Gestures/movements/actions that mark rhythm, e.g. drumming fingers on the table; also other gestures used to indicate time aspects or emphasize part of an utterance.</td>
</tr>
<tr>
<td>Emotive</td>
<td>Gestures/movements that appear with emotional utterances of some sort (vocal, verbal, physical etc.) Positive and negative. Can sometimes apply to ”ordinary” gestures (like the ones above) but with particular force. In that case mark both, e.g. EMBLEM/EMOTIVE. Some facial expressions and actions count as gestures, e.g. smiling, hugging and kissing.</td>
</tr>
<tr>
<td>Grooming</td>
<td>Gestures/movements made to adjust clothing, hair, scratching, poking the nose, etc.</td>
</tr>
<tr>
<td>Show/Offer</td>
<td>Holding out an object with the intention of showing it or handing it over (often with gaze alternating between object and person).</td>
</tr>
<tr>
<td>Action</td>
<td>Movements where it is unclear if it is a conventional gesture or something else, e.g. movements/actions with toys.</td>
</tr>
<tr>
<td>Other</td>
<td>Gestures/movements that do not fit into the categories above.</td>
</tr>
</tbody>
</table>

From the age of 0;9 the children had two gesture tiers in ELAN, Gesture1-Child and Gesture2-Child. This is because young children have not yet developed a dominant hand, and therefore several gestures can be performed at the same time. If two or more gestures occurred at the same time, one was annotated in Gesture1-Child and the other in Gesture2-Child. Both gesture tiers were included in this study.

Since children are not expected to gesture in an adult-like way during their first months of life the younger ages were annotated using child gestures, such as “object in mouth”, “grabbing” etc. Annotators were instructed to use a combination of those tags and the adult gesture tags from the age of 0;9. Only adult gesture tags, i.e. the ones listed in table 1, have been included in this study. However, grooming gestures have been excluded from the data analysis since they are not considered linguistic.
4.4 Data analysis

4.4.1 Definition of gesture-speech combinations

Gesture-speech combinations were defined using the gesture as a reference point, since children coordinate gestures and speech from an early age (Iverson & Fagan, 2004; Esteve-Gilbert & Prieto, 2014). Any communicative utterance that occurred within the timeframe of a gesture was included, i.e. words, proto-words and other communicative sounds, as well as whole sentences when the children were older. Utterances that could not be understood either as a word or a proto-word, e.g. gaga, were excluded from the analysis. Further, gesture-speech combinations where the speech and the gesture were unrelated were excluded from the analysis, e.g. when the child was playing with a toy and answering a question that was not related to what the child was doing.

The gesture-speech combinations were extracted from ELAN into Excel. Since there were two gesture tiers, only combinations with one utterance and one gesture at a given time point was included, i.e. no overlapping gestures. If two gestures occurred at the same time during an utterance, the one most closely related to the speech was kept. However, if two gestures occurred after each other, i.e. not overlapping, during one utterance both were included.

4.4.2 Categorization of gesture-speech combinations

Gesture-speech combinations were categorized into the categories complementary, e.g. pointing to a book and saying “book”, or supplementary, e.g. pointing to a book and saying “give” following previous research (Capobianco et al., 2017).

Furthermore, another category was added, called discourse gesture-speech combinations, which included combinations that could not be classified as either complementary or supplementary, but where the gesture still added something to or modified the utterance. This category included the kind of gestures that McNeill (1992: 15ff) termed discourse gestures, e.g. nodding for emphasis.

All gesture-speech combinations included in the material were checked manually, which allowed for a more context dependent interpretation of the gestures. For example, deictic gestures can be both declarative and imperative (Esteve-Gilbert & Prieto, 2014; Bates et al., 1976), and following this, deictic gestures were interpreted as arguments or predicates based on the context. A point could for instance be translated as either an object name, e.g. if a child says kossa (cow) and points to a cow, or a verb, if the child says pappa (daddy) and points to a pillow with the meaning “lay down”. In the first case the result is a complementary gesture-speech combination and in the second it is a supplementary combination.

The total amount of gesture-speech combinations was counted for each child and age. To compensate for the fact that all recordings were not the same length, a mean value for number of gesture-speech combinations per minute was calculated. Additionally, the children were divided into three groups based on vocabulary size (high, average and low), and a mean value for each group and age point was calculated.

For each child the gesture-speech combinations were categorized as either complementary, supplementary or discourse. These were also adjusted to frequency per minute. Furthermore, the age of onset of each gesture-speech category was established for each child. The frequency of gesture types used in the gesture-speech categories was calculated on group level, and these were adjusted to gesture type per minute to compensate for the different file lengths.

4.4.3 Productive vocabulary measure: SECDI-II

SECDI, Swedish Early Communicative Development Inventories, (Berglund & Eriksson, 2000a) is a parental questionnaire used for measuring children’s language acquisition, adapted for Swedish from the MacArthur-Bates Development Inventories (Fenson et al., 1993). SECDI-I (Words & Gestures) is developed for the ages 8-16 months and focuses on perception and production. SECDI-II (Words & Sentences) is adapted for the ages 16-28 months and focuses mainly on productive vocabulary. CDIs
are widely used and are generally considered reliable for assessing children’s language level in research, though parents sometimes over- or underestimate their child’s abilities (Law & Roy, 2008). SECDI-II has been validated and is considered a reliable tool for assessing productive vocabulary (Berglund & Eriksson, 2000b). In the present study SECDI-II was used as a measure of productive vocabulary at 2;6 years, which was the latest age point where productive vocabulary data had been collected.

**4.4.4 Sentence complexity measure: Unified predicate**

As a measure of syntactic complexity, a percentage of well-formed sentences in the 3;0-session was calculated for all the children. This age point was chosen because it was the latest that had been annotated. The children’s speech was extracted from ELAN and imported into Excel. Only utterances that included a unified predicate, defined as “a predicate that expresses a single situation” (Berman & Slobin, 1994: 660), were included in the analysis. All clauses containing a unified predicate were given a score between 0 and 4 based on how close they were to the target sentence, following Tonér and Gerholm (2018, submitted). The target sentence is an interpretation of what the child is trying to say. All clauses were scored, in relation to the target sentence, on i) appropriate number of words, ii) appropriate number of inflections, iii) word order, and iv) functionality (i.e. if the sentences ‘works’ given that it is spoken language). Subsequently, a percentage of clauses that got the maximum score was calculated for each child, resulting in a percentage of well-formed sentences.

**4.4.5 Statistical analysis**

Research question 2, whether the total amount of gesture-speech combinations predicts later language levels, was tested as follows:

The productive vocabulary and sentence complexity was tested against the total amount of gesture-speech combinations produced at the observed age points, using linear regression analyses.

Research questions 3a and 3b, whether the amount of complementary and supplementary combinations predicted later language levels, were tested as follows:

The productive vocabulary at 2;6 was tested against the amount of complementary gesture-speech combinations produced at the observed age points, using linear regression analyses.

The sentence complexity at 3;0 was tested against the amount of supplementary gesture-speech combinations produced at the observed age points, using linear regression analyses.

Research questions 4a-c, regarding whether the age of onset of the three combination categories predicted later language skills were tested as follows:

The productive vocabulary at 2;6 was tested against the age of onset of complementary gesture-speech combinations, using a linear regression analysis.

The sentence complexity at 3;0 was tested against the age of onset of supplementary gesture-speech combinations, using a linear regression analysis.

The productive vocabulary at 2;6 and sentence complexity at 3;0 were tested against the age of onset of discourse gesture-speech combinations, using linear regression analyses.
5. Results

In the following chapter the results of the present study are presented. First, the language measures used are presented, followed by the descriptive results corresponding to research questions 1a-c. Then the results of the statistical testing are presented, corresponding to research questions 2-4c.

5.1 Productive vocabulary and sentence complexity

The two language measures used to assess the children’s language levels are presented below. The children’s productive vocabulary, as measured by SECDI-II at the age of 2;6 and the sentence complexity scores, as measured by the percentage of well-formed sentences at 3;0, is displayed in figure 1. The highest possible score on productive vocabulary in SECDI-II is 719 words. The sample is quite diverse, with vocabulary scores ranging from 164 words at the lowest to 696 words at the highest.

![Productive vocabulary and sentence complexity graph]

*Figure 1. The figure shows the productive vocabulary scores at 2;6, as measured by SECDI-II, and the sentence complexity scores at 3;0, as measured by the percentage of well-formed sentences in the 3;0-session, for all children. The vocabulary scores range from 164 to 696 (the maximum possible score was 719). The sentence complexity scores range from 36 % at the lowest to 91 % at the highest.*

As a measure of sentence complexity, a percentage of well-formed sentences in the 3;0-session was calculated for all the children. The child with the lowest score had 36 % well-formed sentences, while the child with the highest score had 91 %. In between, the children are spread out. This measure also shows a quite diverse sample.
5.2 Descriptive results

5.2.1 Age of onset of gesture-speech combination categories

The age of onset of the different types of gesture-speech combinations, complementary, supplementary and discourse, are presented in figure 3. Complementary gesture-speech combinations were found earliest at 0;9. By 1;0 most children (all but four) were producing complementary combinations, and by the age of 1;3 all children had started producing this combination type. The mean age of acquisition of complementary gesture-speech combinations was 12.4 months of age.

![Age of onset of gesture-speech categories](image)

*Figure 2. The figure shows the age of onset of the three types of gesture-speech combinations (complementary, supplementary and discourse) for each child. The children's age of onset is presented in months on the y-axis. The children are presented in the order of their productive vocabulary score (low to high).*

The earliest age of onset of supplementary gesture-speech combinations was 1;0, at which point seven children were producing supplementary combinations. By the age of 1;9 all children had started producing supplementary combinations. The mean age of acquisition of supplementary gesture-speech combinations was 15.9 months of age.

Discourse gesture-speech combinations were the last to appear. The earliest age of onset of discourse combinations was 1;3 (two children), and the latest was 3;0. At that age all children were producing gesture-speech combinations of this type. The mean age of acquisition of discourse combinations was 24.6 months of age.

As is shown in figure 3, the order of acquisition of the gesture-speech combinations follows a general pattern. Complementary combinations appear first, followed by supplementary combinations. Combinations of the discourse category appear last in all children. In some cases, complementary and supplementary combinations appear at the same age point. However, the order is never reversed, i.e. supplementary combinations can appear at the same time as complementary ones, but not before, and discourse combinations always appeared after the children had started producing the other types of combinations.

5.2.2 Gesture-speech combinations over time

The children were divided into three groups based on their productive vocabulary score at 2;6, resulting in a low (n=6), average (n=5) and high scoring group (n=5). A mean total amount of combinations per minute was calculated for each group and age point (figure 4). The three groups displayed different patterns. The high scoring group’s usage of gesture-speech combinations increased
from the age of 0;9 to 1;3, at which point the increase became steeper. Their production of combinations peaked at the age of 1;9 and then decreased.

**Figure 3.** The figure shows the mean total amount of gesture-speech combinations per minute (y-axis) over the seven age points in the children with low, average and high vocabulary score respectively. The high scoring group’s production peaked at 1;9 and the average group’s production peaked at 2;0. The low scoring group’s production did not exhibit a peak during the observed age points. The groups’ production frequency converged at the age of 3;0.

The average scoring group’s production of gesture-speech combinations also increased, but not as steeply, and peaked at the age of 2;0, i.e. at a later age than the high scoring group, before decreasing. The low scoring group displayed a different pattern; the production increased over all the age points, and there was no peak in their production of gesture-speech combinations during the observed age points. At the age of 3;0 all three groups converged around the same mean amount of gesture-speech combinations per minute.

### 5.2.3 Gesture types in gesture-speech combinations

An analysis of what gesture types occur in the different gesture-speech categories was conducted on group level. The results were adjusted to compensate for the sessions’ different lengths, and a ratio of gesture type per minute was calculated. Complementary combinations were more common in the material than supplementary. Discourse combinations were the least used, except at 3;0, when discourse combinations were more common than supplementary combinations. For complementary combinations (figure 4), deictic gestures were the most common at all age points. Showing, offering, action and emblematic gestures also occurred in complementary gesture-speech combinations.
The figure illustrates the distribution of gesture types in complementary gesture-speech combinations over the seven age points (normalized to mean amount per minute). Deictic gestures were the most commonly used, followed by emblems and showing gestures.

In supplementary combinations (figure 5) deictic gestures were also the most predominant, except at 0;9 and 3;0, where showing gestures were more common. Offering and emblematic gestures also occurred in supplementary combinations. Furthermore, both complementary and supplementary combinations displayed a peak in production at the age of 2;0 (complementary) and 1;9 (supplementary).

The discourse combinations differed significantly from the other two categories in what gesture types were used (figure 6). Emphatic gestures were the most common in the discourse category, followed by emotive and iconic gestures. Additionally, the use of discourse gesture-speech combinations increased as the children got older and did not exhibit a peak, unlike the complementary and supplementary combinations.
Figure 6. The figure shows the distribution of gesture types in discourse gesture-speech combinations over the seven age points (normalized to mean amount per minute). Emphatic gestures were the most commonly used in this type of combination, followed by emotive and iconic gestures.

5.3 Gesture-speech combinations in relation to later language skills

5.3.1 Total amount of gesture-speech combinations

Research question 2 (a and b) addressed whether the total amount of gesture-speech combinations produced at the different age points predicts productive vocabulary at 2;6 and sentence complexity at 3;0. The total amount of gesture-speech combinations produced by the children at all age points was calculated and adjusted to compensate for the different length of the recordings. This resulted in a mean total amount of gesture-speech combinations per minute for each child and age.

Subsequently, linear regression analyses were conducted to determine whether the amount of combinations produced at the observed age points predicted productive vocabulary at the age of 2;6. The results showed that the amount of gesture-speech combinations produced at the age of 1;0 (F(1,14) = 7.568; p < .05, R² = .351), 1;6 (F(1,14) = 11.884; p < .01, R² = .459), 1;9 (F(1,14) = 5.793; p < .05; R² = .293) and 2;0 (F(1,14) = 5.306; p < .05; R² = .275) respectively predicted productive vocabulary at 2;6, as presented in figure 7. The relationship between total amount of gesture-speech combinations per minute and productive vocabulary was strongest at the age of 1;6.
Figure 7. The figure shows the relationship between productive vocabulary at 2;6 (y-axis) and the total amount of gesture-speech combinations per minute (x-axis) at the age points (in order from top left to bottom right) 1;0 ($F(1,14) = 7.568; p < .05, R^2 = .351$), 1;6 ($F(1,14) = 11.884; p < .01, R^2 = .459$), 1;9 ($F(1,14) = 5.793; p < .05; R^2 = .293$) and 2;0 ($F(1,14) = 5.306; p < .05; R^2 = .275$).

The amount of gesture-speech combinations produced at the other age points did not predict productive vocabulary at 2;6.

Furthermore, linear regression analyses were conducted to test if the total amount of gesture-speech combinations per minute at the observed age points predicted sentence complexity at the age of 3;0. The results showed a significant relationship between the amount of gesture-speech combinations produced at the age of 1;6 and sentence complexity at 3;0 ($F(1,14) = 11.595; p < .01; R^2 = .453$), as presented in figure 8.
The amount of gesture-speech combinations produced at the other age points analyzed in this study did not show any relationship with sentence complexity at the age of 3:0.

5.3.2 Complementary and supplementary combinations

Research question 3a addressed whether the amount of complementary gesture-speech combinations produced at the observed age point predicted productive vocabulary at 2:6. The amount of complementary combinations per minute was calculated for each child and age, and linear regression analyses were conducted. The results, presented in figure 9, showed that the amount of complementary combinations produced at the age of 1:0 (F(1,14) = 5.331; p < .05; R² = .276), 1:6 (F(1,14) = 13.001, p < .01; R² = .482) and 1:9 (F(1,14) = 5.457; p < .05; R² = .280) predicted productive vocabulary at 2:6. Again, the relationship was strongest at the age of 1:6.
The amount of complementary combinations produced at the rest of the observed age points did not predict productive vocabulary at 2;6.

Furthermore, research question 3b addressed whether the amount of supplementary combinations produced at the different age points predicted sentence complexity at 3;0. The results of the linear regression analyses did not show any such relationship.

5.3.3 Age of onset of combination categories

Research question 4 (a, b and c) addressed whether the age of onset of complementary, supplementary and discourse gesture-speech combinations predicted later language skills. The age of onset of the different types of gesture-speech combinations was established for all the children, and subsequently tested against later language skills using linear regression analyses.

The age of onset of complementary gesture-speech combinations was tested against productive vocabulary at the age of 2;6, but the results did not show any relationship.

The age of onset of supplementary gesture-speech combinations was tested against sentence complexity at 3;0. The results of the test showed a negative relationship, i.e. a low age of onset predicted a high level of sentence complexity at 3;0, (F(1,14) = 6.350; p < .05; R^2 = .312). The result of the linear regression analysis is presented in figure 10.
Figure 10. The figure shows the relationship between sentence complexity at 3;0 (y-axis) and the age of onset (in months) of supplementary gesture-speech combinations (x-axis). A low age of onset of supplementary combinations predicted a high sentence complexity score ($F(1, 14) = 6.350; p < .05; R^2 = .312$).

Lastly, linear regression analyses were conducted to establish whether there was a relationship between the age of onset of discourse gesture-speech combinations and productive vocabulary and sentence complexity. The results showed a negative relationship with productive vocabulary at 2;6, meaning a low age of onset of discourse gesture-speech combinations predicted a high vocabulary score, ($F(1, 14) = 10.747; p < .01; R^2 = .434$). The results are presented in figure 11.
onset (in months) of discourse gesture-speech combinations (x-axis). A low age of onset of discourse combinations predicted a high productive vocabulary score at the age of 2;6 (F(1,14) = 10.747; p < .01; \( R^2 = .434 \)).

However, there was no relationship between the age of onset of discourse gesture-speech combinations and sentence complexity at the age of 3;0.
6. Discussion

In this chapter the method of the present study is discussed in terms of validity, reliability and generalizability. Furthermore, the results are discussed in relation to previous research, and ideas for future research are presented.

6.1 Method discussion

6.1.1 Validity

The material used in the present study came from the MINT Project, and consisted of recordings made every third month during the children’s first three years of life. The recording studio was equipped with toys and made to be as comfortable and natural as possible, and the instructions given to the parents were to play as they normally would. While the recording situation might not be as natural as recordings made in the participants’ own homes there is no reason to assume that the lab environment affected the children’s production of gesture-speech combinations. However, certain gesture types might occur more often in other situations, such as meal time and more ritualized situations, which could affect the results of the investigation into which gesture types occur in the different types of combinations. Since all communication is context dependent and it would be impossible to cover all possible situations, we have to assume that the recording environment is representative enough for the purposes of this study.

Research question 4a-c concerned the age of onset of the three types of gesture-speech combinations investigated in the present study. Since the recording sessions consist of about ten minutes of interaction recorded every three months (and there is a gap of one year between the two last sessions, 2;0 and 3;0), the age of onset as measured in the material might not correspond entirely to the actual age of onset of different combination types. However, the results of the present study do still show an effect of age of onset of supplementary combinations on sentence complexity and of discourse combinations on productive vocabulary.

6.1.2 Reliability

All files were annotated by trained annotators. When annotating the children’s speech, a special tag was used to indicate proto-words, or words that were pronounced wrong. Files that had been annotated using an early transcription key, however, did not have this tag. Furthermore, in the 0:9-sessions a controlled vocabulary was used, where utterances were not written as they sound but rather given tags such as “babbling”, “grunting” etc. Since the present study included proto-words, all gesture-speech combinations in all 112 files were checked manually, to make sure that all proto-words were included.

Checking the gesture-speech combinations manually also enabled a more detailed analysis of them. Gestures in the material were given a type (or several if a gesture was interpreted as belonging to more than one category), and a description, e.g. “DEICTIC_points to the toy on the floor in front of her”. When the combinations were viewed in context, the descriptions of the gestures were sometimes complemented with additional information, so that the child’s intention was captured. Thus, the context in which the gesture-speech combinations occurred in the material was taken into consideration. Since gesture meaning is context dependent, this method made the analysis more robust.

Özçalışkan and Goldin-Meadow (2009: 195) interpreted deictic gestures as arguments only, but in the present study deictic gestures have sometimes been interpreted as predicates based on the context they occur in. This is because previous research (Bates et al., 1976; Esteve-Gilbert & Prieto, 2014) has established that deictic gestures can be both declarative and imperative. All gestures included in the present study were looked at in the context where they occur, thus allowing deictic gesture to be
interpreted either as a declarative, or argument using Özçalışkan and Goldin-Meadow’s term, or as an imperative, or predicate in Özçalışkan and Goldin-Meadow’s sense of the word. This in turn affected the categorization of gesture-speech combinations. A declarative deictic gesture (e.g. pointing to the lamp) combined with the word lampa (lamp) would be categorized as a complementary combination, but combinations such as pappa (daddy) and a point to a pillow (meaning “lay down”) was categorized as a supplementary combination. As a measure of vocabulary size, SECDI-II from when the children were 2;6 was used. CDIs are a common tool for measuring vocabulary size. They are cost efficient and easy to administer, but have the downside that parents sometimes have a tendency to either over- or underestimate their child’s abilities. Despite this, CDIs are generally considered to be reliable estimates of children’s vocabulary level (Law & Roy, 2008). More specifically, SECDI-II has previously been shown to have a high validity and reliability (Berglund & Eriksson, 2000b).

As a measure of sentence complexity at the age of 3;0 a percentage of well-formed sentences was calculated for each child, following Tonér and Gerholm (2018, submitted) and using unified predicate as a basis of the analysis (Berman & Slobin, 1994). The method was chosen to get a genuine estimate of the children’s sentence complexity level, since it includes several aspects of sentence construction; that the children do not skip function words, that they use correct inflection and word order, and it also takes into consideration if the sentence is acceptable given that it is spoken language. Furthermore, this method does not rely on parental reports, but rather on what the children actually produce. One difficulty with the method was that for every sentence containing a unified predicate a target sentence had to be determined. The target sentence was an interpretation of what the child was trying to say, and thus an approximation of what the same sentence would look like if it had been uttered by an adult. While it was sometimes difficult to establish what the target sentence should be, it was fairly clear if the actual sentence produced was grammatically well-formed or not. Since the measure of sentence complexity used in this study was the percentage of well-formed sentences, the resulting score should be considered a fair estimate of the children’s sentence complexity levels.

6.1.3 Generalizability

The sample selection was made based on a few criteria; the children had to be monolingual, defined in this study as having only one language spoken at home by the parents, had no known developmental disorders and had SECDI filled out at the 2;6 year session. This resulted in a sample of 16 children (8 girls and 8 boys), whose gesture-speech production was analyzed at seven age points.

The sessions were about ten minutes long, resulting in approximately 18.5 hours of material. While sample size poses a problem for the generalizability of the study’s results, the children included in the study are quite diverse when it comes to language level at the two age points used as reference, i.e. productive vocabulary at 2;6 and sentence complexity at 3;0. When it comes to vocabulary size the children range from 164 words at the lowest to 696 at the highest, a quite large interval with vocabulary scores spread out relatively evenly in between. The sentence complexity measure also shows a diverse sample, with scores ranging from 36 % well-formed sentences at the lowest to 91 % at the highest. The fact that the children are spread out when it comes to language level should increase the study’s generalizability.

All families are from the Stockholm area, or lived there when the project started, and are therefore only a cross-section of a small part of Sweden. Another aspect that should be mentioned is that the sample is not representative for the population as a whole when it comes to socio-economic status (SES). Most of the families are middle class, earning 4-800 000 SEK per year. Furthermore, most of the parents have a higher education, which is often the case when it comes to research participants but is not representative of the whole population. This could also affect the generalizability of the study. However, by the age of 2;3 all children included in the present study had started preschool, which could even out the effect of language use in the home.
6.2 Result discussion

6.2.1 Order of acquisition of combination categories

The results of the present study show that the order of acquisition of the three categories of gesture-speech combinations follows a general pattern. First, the complementary combinations are acquired, followed by supplementary combinations. The discourse combinations appear last. The order of acquisition of complementary and supplementary combinations is in line with previous research (Özçalışkan & Goldin-Meadow, 2009; Capobianco et al., 2017). Capobianco and colleagues (2017) found that complementary combinations appeared at 12.5 months of age on average, while supplementary combinations appeared on average at 17.5 months of age. In the present study the mean age of acquisition of complementary combinations was 12.4 months, comparable to Capobianco and colleagues’ findings. For supplementary combinations the mean age of acquisition in the present study was 15.9 months of age, which can be explained by the fact that supplementary combinations were found as early as at the age of 1:0. This is earlier than has been reported in previous research. However, the reason for this is not entirely clear.

One such early supplementary combination was the word titta (look) combined with showing a toy, resulting in a gesture-speech phrase that can be translated as “look at this” or “look at the book”. This is a quite common phrase in the input to children, which could explain why it occurs early as a gesture-speech combination. Another example of an early supplementary combination was pappa (daddy) combined with a pointing or showing gesture, resulting in a phrase like “daddy, a crocodile”. At the age of 1;0 children can establish joint attention with adults (Tomasetto, 2003: 41). In this example perhaps joint attention is established using a word to make sure that dad is listening, and then expressing the rest of the sentence using a gesture.

The discourse category was the last to be acquired for all children. This could be explained by the fact that these combinations do not help children in their language development in the same way as the complementary and supplementary ones. Özçalışkan and Goldin-Meadow (2009) found that combining gestures with speech helps children learn new constructions, and both complementary and supplementary constructions are used by children to express concepts that they cannot express in speech alone. However, discourse combinations are essentially different from the others, since the gesture does not add a word but rather emphasis, emotion or some other dimension that cannot be translated into a word. Consequently, these combinations would not be used to learn new constructions. Rather, they can be said to add “flavor” to the children’s speech. Discourse gesture-speech combinations can perhaps be said to be more adult-like in their nature. McNeill (1992: 21) states that discourse gestures, i.e. beats and cohesives, are the last gestures to appear. He further posits that they generally appear around the age of five and that they are close to an adult-like level around the age of eleven. However, the present study shows that children as young as 1;3 can use these gestures, although not on an adult-like level, and that their use of these gestures increase over time.

6.2.2 Production of gesture-speech combinations over time

Turning to the amount of gesture-speech combinations produced at the observed age points (figure 3, section 5.2.2), a pattern emerges. The children were divided into three groups based on their productive vocabulary level (low, average and high), and the mean total production of gesture-speech combinations per minute was calculated for each age and group. Previous research has found that gesture production increases during the first two years of life and then peaks at 2;2 years of age before decreasing (Özçalışkan & Goldin-Meadow, 2009). The results of the present study seem to support this claim, but since the children were divided into groups based on vocabulary score an additional perspective can be added. The high-scoring group’s production of gesture-speech combinations peaked at 1;9 years, while the average-scoring group peaked at 2;0 years. The low-scoring group’s production did not peak during the observed sessions, but rather showed a steady increase. At 3;0 years of age all three groups had converged around the same frequency of gesture-speech combinations per minute. Given that previous research has found that children’s gesture production
peaks it is possible that the low-scoring group does have a peak in production somewhere between the age of 2;0 and 3;0, but that cannot be asserted from the results of the present study.

It is noteworthy that the peak in production is earlier, and higher, in the high-scoring group than the average- and low-scoring groups. This might indicate that they are further along in development, but they also seem to make more use of gesture-speech combinations all together, since their curve is steeper and on a higher level. Furthermore, the average group’s production is generally higher than the low-scoring group’s, and based on the trend lines it is likely that their peak is also more pronounced than a possible peak of the low-scoring group. These results are in line with previous research, which shows that gestures act as a stepping-stone in language development (Özçalışkan & Goldin-Meadow, 2009). The present study shows that children who make more use of gestures in combinations with speech, and who peak earlier in their production, have higher language levels later on in development.

Previous research has also stated that the production of gestures declines towards a more adult-like level as speech becomes the preferred modality of expression (Capirici et al., 2005), and that the same goes for gesture-speech combinations (Özçalışkan & Goldin-Meadow, 2009). This could be an explanation to why the high-scoring group peaks first, followed by the average-scoring group, since speech probably becomes their preferred modality earlier. At 3;0 years of age all children are most likely using speech as their preferred modality, which could explain why the production of gesture-speech combinations of the three groups seems to converge around the same level at this age.

The results further show that both complementary and supplementary gesture-speech combinations seem to increase into a peak at 2;0 and 1;9 respectively, and then decrease slightly (figure 4 and 5, section 5.2.3). This result goes against what was found by Capobianco and colleagues (2017), i.e. that complementary combinations peaked earlier than supplementary ones (1;6-1;10 and 1;8-1;11 respectively). The age at which supplementary combinations peak is however consistent with their findings. However, the difference could be explained by the fact that Capobianco and colleagues only included deictic gestures, whereas the present study included all gestures produced by the children.

The use of discourse combinations instead seems to increase over time. This could be a consequence of complementary and supplementary gesture-speech combinations being more “true” combinations, i.e. the gesture can be translated into a word that is combined with the speech. This has been suggested to act as an aid for children when learning new constructions (Özçalışkan & Goldin-Meadow, 2009)

The discourse combinations, on the other hand, do not work in the same way. As opposed to the other types of combinations they are not multi-unit constructions comparable with multi-word constructions. Rather, they add something on a discourse-pragmatic level, which in turn can be considered more adult-like. McNeill (1992: 319ff) states that these types of gestures are not used by young children, since they require a sense of the metanarrative and metalinguistic level of an utterance, which small children do not understand. That is a likely explanation to why these gestures increase over the observed age point in this study; they have not reached their full capacity yet. At the age of 3;0 all children in this study were using discourse gesture-speech combinations to some extent, mainly using emphatic gestures. While they probably do not fully understand the potential of these gestures yet, they have most likely picked up that adults around them use gestures in this way, and thus they are practicing the behavior.

**6.2.3 Gesture types in gesture-speech combinations**

The results of the analysis of which gesture types occurred in the different gesture-speech combinations showed that deictic gestures were the most common in both complementary and supplementary combinations. Showing, offering, action and emblematic gestures were also used in both complementary and supplementary combinations. In discourse combinations emphatic gestures were by far the most commonly used.

The distribution of gesture types in complementary and supplementary combinations is quite similar, which can be explained by the fact that the gestures that occur in these combinations can be translated into words, which gestures occurring in the discourse category cannot. It is not strange that deictic gestures are the most commonly used in complementary and supplementary combinations. Deictic gestures are the first communicative gestures to appear (Kita, 2003: 2), and they are the most common
gesture type in children’s production all together (Özçalışkan et al., 2016). The use of deictic gestures has also has been linked with language development both in typically developed children (Iverson & Goldin-Meadow, 2005), children with language delay (Lüke et al., 2017) and children with autism spectrum disorder and Down syndrome (Özçalışkan et al., 2017).

6.2.4 Amount of combinations and later language skills

The results of linear regression analyses showed that the total amount of gesture-speech combinations produced at 1;0, 1;6, 1;9 and 2;0 predicted productive vocabulary at 2;6. Previous studies have shown that gesture production plays an important role in vocabulary development (Iverson & Goldin-Meadow, 2005; Özçalışkan et al., 2017). Additionally, Capirci and Volterra (2008: 37) found that when gestures are combined with speech it is “linked to language growth”. The results of the present study seem to be in line with this, since a large production of gestures in combination with speech at four age points was linked to a large productive vocabulary later in development. This could be because the use of gesture-speech combinations is a manifestation of a maturity in the developmental process, as posited by Özçalışkan and Goldin-Meadow (2009), since the children have understood that language consists of smaller units that can be combined into larger constructions.

There was also a relationship between the total amount of gesture-speech combinations produced at 1;6 and sentence complexity at 3;0. The relationship between the total amount of combinations and productive vocabulary was also the strongest at the age of 1;6. Together, this indicates that something happens around this time in the children’s development, since this is the only time point that predicts both language measures. Previous research has also found that both amount of complementary and supplementary combinations used at the age of 1;6 correlate with later language measure (Capobianco et al., 2017; Fasolo & D’Odorico, 2012). The results of the present study suggest that the children who are good at combining elements at the age of 1;6 have a higher level of sentence complexity at 3;0. Perhaps this can be linked to the fact that it is around the age of 1;6 that children start combining words into two-word utterances (Goldin-Meadow & Butcher, 2003). Gesture-speech combinations have been suggested to be a stepping-stone in development towards multi-word utterances (Özçalışkan & Goldin-Meadow, 2009; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005), and the results of the present study seem to support this claim. Children who have understood the concept of combining elements and thus are doing it to a larger extent at 1;6 could be using the gesture-speech combinations to express concepts that they cannot yet formulate in speech alone, indicating that they are further along in development and are soon ready to take the step into the two-word stage fully.

Children develop at different rates, and though the effect of gesture-speech production is strongest at 1;6 it is also present at other ages. There is an effect already at 1;0, which is related to productive vocabulary. Capobianco and colleagues (2017) also found a correlation between complementary combinations at the age of 1;0 and vocabulary size at the age of 2;0. Since most of the combinations found at 1;0 are complementary it is possible that this result could be explained in light of that.

At 1;9 and 2;0 the relationship between the total amount of gesture-speech combinations and productive vocabulary was weaker than at 1;6. Özçalışkan and Goldin-Meadow (2009) argue that children use gestures more when learning new constructions, but not to flesh out constructions that they already know. The fact that the relationship with productive vocabulary is not as strong at these age points, could be explained by that the children at this time were moving towards speech becoming their preferred modality.

6.2.5 Complementary and supplementary combinations and later language skills

The results of the present study showed a significant relationship between the amount of complementary combinations used at the ages 1;0, 1;6 and 1;9 and the productive vocabulary at 2;6. This is in line with previous research that has found a connection between the use of complementary combinations at 1;0 and 1;6 and vocabulary size (Capobianco et al., 2017) and at the age of 1;6 with vocabulary size and MLU (Fasolo & D’Odorico, 2012).
Deictic gestures were the most commonly used in both complementary and supplementary gesture-speech combinations in the present study. Research has shown that children use deictic gestures to express referents before those referents are produced in speech alone. Furthermore, the time at which a referent starts being produced in gesture alone predicts when that word appears in speech alone (Iverson & Goldin-Meadow, 2005). In light of this, the use of complementary combinations could indicate that a child has a concept that they cannot yet produce fully in speech alone, resulting in the use of a gesture-speech combination. One example of this is the deictic word dār (there) and a point to an object, which is a common complementary combination in the material, especially at the early ages. In this case it could be that the children cannot produce the referent, so instead they produce a deictic word that is close to a determiner and combine it with a pointing gesture that expresses the referent. Another example from the material is when a child says e.g. lampa (lamp) and points to a lamp. In this case the referent can be expressed, but the word is not inflected in any way. Cartmill and colleagues (2014) found that deictic gestures can be used in combination with speech to modify nouns, and that the age of acquisition of complementary combinations where the gesture acts as a determiner predicts the age of onset of determiner + noun constructions in speech alone. In the present study it seems that gestures both modify nouns, as was found by Cartmill and colleagues, but they can also express the referent in combinations where the speech modifies the gesture (i.e. dār (there) and a point to an object).

However, the amount of supplementary combinations produced at the observed age points did not show a significant relationship with sentence complexity at 3;0. This goes against the results of Capobianco and colleagues (2017), who found that the amount of supplementary combinations at the age of 1;3 and 1;6 correlated with verbal complexity at 2;0, and Fasolo and D’Odorico (2012), who found that the use of supplementary combinations at 1;6 predicted sentence complexity at 2;0. Capobianco and colleagues (2017) calculated verbal complexity as token of multi-word utterances (2017: 61), which is a measure that is different from the sentence complexity measure used in the present study, i.e. percentage of well-formed sentences. This, together with the fact that the sentence complexity measure in this study is at a later age point could account for the difference in results.

Another explanation is that there were too few supplementary combinations in the material, complementary combinations far outnumbered supplementary combinations. It could be the case that the setting in which the recordings for the present study was made, i.e. a lab environment equipped with toys and books, affected the frequency of the supplementary gestures. Furthermore, the supplementary combinations in the present study were counted as tokens, and perhaps the results would be different if types had been counted instead.

6.2.6 Age of onset and later language skills

Though the amount of supplementary combinations used by the children did not show any relationship with sentence complexity, the age of acquisition did. A low age of acquisition of supplementary gesture-speech combinations predicted a high level of sentence complexity at 3;0. Previous research has found a link between the age of onset of supplementary combinations and age of onset of two-word utterances (Capobianco et al., 2017; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005).

Supplementary combinations are multi-unit utterances, and an early production of this type of gesture-speech combination is indicative of an early understanding of how language works, i.e. that smaller units can be combined into larger ones to convey a more complex meaning. Özçalışkan and Goldin-Meadow (2009) argue that children use gestures in combination with speech to learn new constructions, and this idea is relevant when considering the results of this study. Children who have a low age of onset of supplementary combination are learning or practicing multi-unit utterances early on, showing a more advanced stage in development. It is therefore not unlikely that they would also be further along in development at the age of 3;0.

It should be pointed out that supplementary combinations were found in the present study in seven children at the age of 1;0, earlier than in previous research, which has found supplementary combinations earliest at the age of 1;3 (Özçalışkan & Goldin-Meadow, 2009; Capobianco et al., 2017).
Despite this, age of acquisition of supplementary combinations still showed a relationship with sentence complexity. The age of onset of complementary combinations did, however, not predict productive vocabulary.

The results of the present study also showed a relationship between the age of onset of discourse gesture-speech combinations and productive vocabulary at 2;6, but not with sentence complexity at 3;0. In discourse combinations the gesture cannot be interpreted as a word or a semantic element, so in that they differ from both complementary and supplementary combinations. Rather, it adds something that is perhaps more related to discourse or pragmatics, such as emphasis or emotion. The reason why there is a relationship between these combinations and productive vocabulary size could be related to the conclusion that Özçalışkan and Goldin-Meadow (2009: 215) come to, i.e. after speech becomes the preferred modality gestures can be used to enrich the speaker’s communication. The discourse combinations can well be said to be used in that way. They are not necessary when it comes to making oneself understood in the way that complementary and supplementary combinations are. Rather, they add something on a more abstract level, which is most likely why they occur after the children have already mastered the other two types.

As previously stated, McNeill (1992: 319ff) argues that these types of gestures are related to the speaker’s notion of the metanarrative and metalinguistic structure of language, and that this is the reason why they occur late in development. However, it might be that children will start using these gestures in combination with speech when they have enough “raw material”, i.e. words, to be able to make use of them. This could explain why they seem to be related to vocabulary size rather than sentence complexity.

6.2.7 General discussion

The relationship between gesture-speech combinations and later language levels, measured both as productive vocabulary at 2;6 and as sentence complexity at 3;0, was strongest at the age of 1;6. Previous research has found that there is often a rapid increase in children’s vocabulary at this time, sometimes called the ‘vocabulary spurt’ (Benedict, 1979; Goldfield & Reznick, 1990). Furthermore, children often start making early two-word utterances around this time (Goldin-Meadow & Butcher, 2003), and Tomasello (2003: 50f) suggests that the vocabulary spurt could be linked with this ability to combine words and use simple grammar.

The results of the present study seem to support that something of importance happens in children’s development around the age of 1;6, since the link between use of gesture-speech combinations and vocabulary size was strongest at this age. 1;6 was also the only age point at which use of gesture-speech combinations was linked with sentence complexity. Since the vocabulary spurt has also been linked to the development of handedness, i.e. that children have a clearer hand preference in pointing after the vocabulary spurt than before (Cochet, Jover & Vauclair, 2011), the motor (gesture) system seems to be involved in this developmental step. The results of the present study seem to support that the gesture system is involved. This could in turn be viewed in light of a dynamic systems theory approach to language acquisition (Van Geert & Verspoor, 2015), which states that as components are integrated into a system, i.e. the language system in this case, it increases in complexity. For the system to become stable the components need to be properly coordinated. It is possible that around the age of 1;6, the gesture system and the language system are becoming so coordinated that a developmental step occurs, and that this is what manifests as strong relationships between the use of gesture-speech combinations and later vocabulary size and sentence complexity.

 Gestures and speech are indeed part of an integrated system, as posited by McNeill (2000: 9). Since gestures are so widely used for different purposes, and also help children in their language development, it seems unlikely that gestures should not have played a role in the evolution of language. The gesture repertoires of our closest relatives, chimpanzees and bonobos, are largely overlapping and could be traced back to our last common ancestor, suggesting that our gesture system is very old (Graham et al., 2018). Furthermore, it has been shown that gestures are effective in building a communication system from scratch (Fay et al., 2014). The results of the present study, along with previous research, suggest that gestures play an important role in the acquisition of
language. The vocal and manual modalities are closely linked even from a very young age, thus a likely scenario could be that speech and gestures have evolved together, perhaps helping each other along.

6.3 Ideas for future research

The present study has found a link between the age of acquisition of discourse gesture-speech combinations and productive vocabulary, suggesting that the types of gestures used in these combinations are linked to language development. These combinations were also the only category to increase in usage over all age points. Future research focusing on these types of gestures in children’s production, even at young ages, could give more insight into the role of gestures in language development, and perhaps into the development of discourse-pragmatic features of children’s language. It could also be of interest to have an adult base-line to investigate how children’s use of discourse gestures differs from that of adults, and if they are moving towards an adult-like pattern of use.

The results of the present study, along with previous research, suggest that something of importance happens in children’s development at the age of 1;6. Future research into this could perhaps tie together several important mile stones that occur around this age, to get an idea of why the use of gestures at the age of 1;6 seems to be so closely linked to later language development.
7. Conclusions

The present study investigated children’s use of complementary, supplementary and discourse gesture-speech combinations during their first two years of life and at three years of age, and their role in language acquisition.

The results showed that complementary combinations appeared first, at a mean age of 12.4 months. Supplementary combinations occurred second, at a mean age of 15.9 months, and lastly discourse combinations appeared at a mean age of 24.6 months.

The children with the largest productive vocabulary scores peaked earlier and more distinctly in their use of gesture-speech combinations than the children with average and low vocabulary scores. The children in the average group in turn peaked earlier and higher than the children in the low vocabulary group, whose combination production did not peak during the observed age points. The total production of gesture-speech combinations then converged at the same level at the age of 3;0.

The use of both complementary and supplementary combinations increased over time and peaked at the age of 2;0 and 1;9 respectively, and then declined. The use of discourse combinations increased over the observed sessions.

Deictic gestures were the most commonly used gestures in both complementary and supplementary combinations, while emphatic gestures were the most prominent in the discourse category.

The total amount of gesture-speech combinations produced at the age points 1;0, 1;6, 1;9 and 2;0 predicted productive vocabulary at 2;6, while the amount of gesture-speech combinations produced at 1;6 predicted sentence complexity at 3;0.

The amount of complementary gesture-speech combinations at the age points 1;0, 1;6 and 1;9 predicted productive vocabulary at 2;6. However, the amount of supplementary combinations showed no significant relationship with sentence complexity at 3;0.

While there was no relationship between age of onset of complementary combinations and productive vocabulary, a low age of onset of supplementary combinations was related to a high sentence complexity at 3;0. Furthermore, a low age of onset of discourse combinations was related to a high productive vocabulary at 2;6.

The results of the present study show that gesture-speech combinations are related to the development of both productive vocabulary and sentence complexity. The present study thus supports previous research suggesting that gestures play an important role in child language acquisition.
References


32


Perlman, M. (2017). Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies, 18*(3), 376-401. doi: 10.1075/ts.18.3.05per


Appendices

Appendix 1. Information letter to parents

Bästa föräldrar!

Vill ni delta med ert barn i ett forskningsprojekt om barns språkutveckling? Med detta brev vill vi presentera vårt forskningsprojekt ”MINT - Modellering av interaktion” vid Institutionen för lingvistik, Stockholms universitet (Frescati).


Trots att resultaten är av stort vetenskapligt intresse kan vi inte ersätta er ekonomiskt för deltagandet. Däremot kan vi erbjuda kopior av alla ordlistor som ni fyller i som en dokumentation av ert barns ordförråd. Vi hoppas att dessa, och även våra kontinuerliga kontakter, ska kunna ge värdefull insikt i ert barns språkutveckling.

Nedan följer mer omfattande information om projektet, metoden och informationsbehandlingen, samt detaljerad information om hur ni hittar till oss på Stockholms universitet. Om ni är intresserade av att delta i vårt projekt anmäler ni ert intresse genom att fylla i och skicka den medföljande amnälningsblanketten. När vi fått blanketten, kontaktar
vi er för tidsbokning och eventuellt ytterligare information per telefon. Vill ni komma i kontakt med oss går det bra att lämna ett meddelande på vår telefonsvarare, tel. nr 08-16 19 32, så ringer vi upp, alternativt via e-post: eevak@ling.su.se.

**Information om forskningsprojektet MINT**

I forskningsprojektet* undersöks olika aspekter av barns tidiga utveckling av språk. Undersökningarna utförs av personal och studenter under handledning vid institutionen för lingvistik, Stockholms universitet. Ansvariga för projektet är Tove Gerholm och Eeva Klintfors.

**Studiens syfte**

Under de första levnadsåren börjar barnet förstå att ord refererar till olika föremål eller situationer i omgivningen. Föräldrar och andra vuxna i barnets omgivning är delaktiga i denna process genom att interagera och samtala med barnet. Projektets syfte är att studera barnets inlärning av ord, joller- och språkutveckling, samt samspelet mellan förälder och barn.

**Etiska aspekter och sekretess**

Följande personuppgifter ingår i undersökningen: barnets namn och kön, beräknat och faktiskt födelsedatum, födelsevikt, föräldrars namn, adress och telefonnummer, språk som talas hemma hos barnet, antal äldre och yngre syskon samt eventuell förekomst av öroninflammation. Ni har rätt att när som helst och utan vidare förklaringar ta del av informationen eller begära att era personuppgifter tas bort. Vid bearbetning och presentation av resultaten i vetenskapliga sammanhang kommer materialet att vara aidentifierat och sammanställt per åldersgrupp.

**Hur visar jag mitt intresse?**

Är ni intresserade av att delta med ert barn i denna undersökning anmäler ni er genom att fylla i medföljande blankett och skicka den till oss i det bifogade förfrankerade kuvertet. Vi kommer att kontakta er så fort som möjligt för att boka in tid och kan då svara på eventuella frågor. Om ni kommer med bil får ni ett särskilt tillstånd för fri parkering av oss.

Välkomna till vårt fonetiklaboratorium och ett stort tack för att ni hjälper oss inom forskningen om tidig språkutveckling!

Tove Gerholm

Eeva Klintfors

FD

FD

Appendix 2. Annotation key for vocal and gesture tiers in ELAN

Vocal: orthographic transcription but close to the actual sounds for *ja* (*jag*), *de* (*det*), *va* (*var/vad*), *å* (*att/och*) and *dom* (*de/dem*). Otherwise, keep to orthography in order to make it searchable. If an adult has deviant speech in any way, such as creaky voice, do not mark this in the tier but add information as a comment to the whole file in “Information from annotator on ready file”.

**IMPORTANT:** caps are only used for loud speech (see below). Names of persons, places, Mo/Na/Li are written with lower-case letters.

### Vocal/Verbal tier

<table>
<thead>
<tr>
<th>Label</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[??]</td>
<td>A message from the annotator to have the tag checked, e.g. word [??].</td>
</tr>
<tr>
<td>?</td>
<td>Question intonation, placed after the intended word/phrase.</td>
</tr>
<tr>
<td>(?)</td>
<td>Unsure transcription.</td>
</tr>
<tr>
<td>( )</td>
<td>Swallowed/omitted sounds, e.g. &quot;den h(är) (het)te Mo&quot;; &quot;vi(lk)en fin&quot; [Eng: &quot;loo(k) a(t) (th)at&quot;; &quot;i do(n)t wan(t) it&quot;]').</td>
</tr>
<tr>
<td>&amp;</td>
<td>Interrupted (word), e.g.: &quot;och &amp;ko kolla här då&quot; [Eng: &quot;and &amp;lo look here&quot;]').</td>
</tr>
<tr>
<td>&amp;(phrase)</td>
<td>Interruption (phrases), e.g.: &quot;det har vi sagt &amp; (för att han hela tiden eftersom han alltid droglar på fjärrkontrollen&quot; [Eng: &quot;we said that &amp; (because the entire time) because he was always drooling on the remote control&quot;]').</td>
</tr>
<tr>
<td>xxx</td>
<td>One or more unknown/inaudible words, e.g.: &quot;ta den då ta xxx bollen&quot; [Eng: &quot;take that take xxx the ball&quot;]').</td>
</tr>
<tr>
<td><em>a-z</em></td>
<td>Nonwords with communicative function, e.g. “huh?”, imitation or vocal illustration (“nam nam nam” to eat), or sound effects (“hå!”), e.g.: &quot;nam nam nam&quot;.</td>
</tr>
<tr>
<td>:</td>
<td>Extended sounds are marked with colon, e.g.: “hå:”.</td>
</tr>
<tr>
<td>Ex ee</td>
<td>Filled pauses are transcribed as the sound, e.g.: &quot;Ee mm mm aa öh&quot;.</td>
</tr>
<tr>
<td>-</td>
<td>Continuation intonation, e.g.: “Ja ska gå å-&quot;.</td>
</tr>
<tr>
<td>-</td>
<td>Disfluency due to hesitation, e.g.: “jjjjja det tror jag” [Eng: “yyyyes, i think so”]. “neej” are marked as “j_a det tror jag”, “ne_j” [Eng: &quot;nooo&quot; are marked as “y_es”, &quot;n_&quot;].</td>
</tr>
<tr>
<td>’</td>
<td>Marks typical (or atypical) reductions, e.g: “har’u”, “är’e”, “var’e” [Eng: &quot;y’ave&quot;, &quot;there’re&quot;, &quot;would’n’t’ve&quot;]').</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>Speech clearly louder than the surrounding speech.</td>
</tr>
<tr>
<td>.word.</td>
<td>Speech clearly softer than the surrounding speech. Not whispering.</td>
</tr>
<tr>
<td>#VI</td>
<td>Whispering, e.g.: Den e [#VI jättefin] [Eng: it’s [#VI really nice].</td>
</tr>
</tbody>
</table>
| {word} | A child’s deficit pronunciation for a word or when the child imitates the intonation of a word, e.g.: ’appå’ but it is evident in the context that the
<table>
<thead>
<tr>
<th><strong>&quot;word/sound&quot;</strong></th>
<th>Irritation/anger.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>%word/sound%</strong></td>
<td>Distorted speech (changed intonation pattern, sudden use of (other) dialect, cartoon figure speech, etc.).</td>
</tr>
<tr>
<td>~</td>
<td>Creaking voice, marked on both sides of the word/phrase, e.g.: hon kom <del>så</del> långt [Eng: she came <del>so</del> far].</td>
</tr>
<tr>
<td>+</td>
<td>Paus &lt;0,5 sec.</td>
</tr>
<tr>
<td>(h:)</td>
<td>Audible in or out breathing.</td>
</tr>
<tr>
<td>®word/sound®</td>
<td>Crying/whining voice. For children 3 months CV is applied instead. Crying/whining adults (imitation?), use approximate spelling, e.g.: ®uhhhhhha®</td>
</tr>
<tr>
<td>!word/sound!</td>
<td>Excited speech, screaming etc.</td>
</tr>
<tr>
<td>/approximation of sound/</td>
<td>Used for coughs, hummings, panting, whissling, kissing sounds, etc. /grunting x 2/ = repeated sound. Also used for other sources of sounds than vocal, e.g. /clapping sound/. Also used for /sound/ for 12-month olds’ (and older) sounds in between babbling and words proper.</td>
</tr>
<tr>
<td>#word/sound#</td>
<td>Laughter. Put # around the utterance produced while laughing, e.g.: #ja det va de värsta# ja varit med om. When only laughter: ###.</td>
</tr>
<tr>
<td>&quot;word/sound&quot;</td>
<td>Singing, humming.</td>
</tr>
<tr>
<td>#VR</td>
<td>Adult directed speech (usually in interaction with experimental leader); child directed speech is unmarked/standard. Place [ ] around the utterance, e.g.: [#VR ska jag läsa i boken?] ja ska vi läsa? [Eng: [#VR shall I read the book?] yes shall we read?].</td>
</tr>
<tr>
<td>#IN</td>
<td>Ingressive speech. Place [ ] around the utterance if it is part of a longer sequence: [#IN ja de vore ju] en nåd att stilla bedja om [Eng: #IN yeah that would sure be] a consummation devoutly to be wished].</td>
</tr>
<tr>
<td>#LA</td>
<td>Word or phrase in other language than Swedish. [#LA the thing] du vet [Eng/Spa: [#LA la cosa] you know]. If you can’t identify the language [#LA xxx].</td>
</tr>
<tr>
<td>#FS</td>
<td>Formulaic Speech (frozen phrases). A FS could be an idiom like &quot;better late than never&quot;, but also (in CDS) expressions that reoccur among many parents and that you recognize, e.g. &quot;titta lampå&quot;, &quot;kossan säger… [#FS_å hur låter kossan/vad säger kossan?]. [Eng: &quot;look lamp&quot;, &quot;the cow says&quot;… [#FS_oh what sound does the cow make/what does the cow say?].</td>
</tr>
</tbody>
</table>
Ungrammatical or unsemantic utterances:

[#UG_utterance]. Example: [#UG_den va in honh hand][Eng: [#UG_is was in she's hand]]; [#UG_ni har två många][Eng: [#UG_you have three many]]; [#UG_nelly hon ramla bakom på stolen][Eng: [#UG_nelly she fall behind on the chair]]; [#UG_fast ja har långt här för ja e ju blond][Eng: [#UG_: well I have long hair 'cause I’m blond]]; [#UG_ja kunde skriva de självt utan å stav][Eng:[#UG_I could write it myself without spelling]]; etc.

---

### Controlled Vocabularies

<table>
<thead>
<tr>
<th>CV VOCAL TIER (children 3, 6 and 9 months)</th>
<th>Grunting; Panting; Cooing; Babbling; Laughing; Whining/Crying; Screaming/Shrieking; Other.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV GESTURE-BABY TIER (children 3 and 6 months; from 9 months use labels but without CV)</td>
<td>Wiggie whole body; Wiggie with arms; Wiggie with legs; Grab r-hand; Grab l-hand; Grab both hands; Object in mouth; Other.</td>
</tr>
<tr>
<td>OBS! The hierarchy for Gesture Baby is: mouth → hands/upper body → feet/lower body.</td>
<td></td>
</tr>
<tr>
<td>CV GESTURE-CHILD 1 and GESTURE CHILD 2 (children from 9 months)</td>
<td>Tier 1 is used for the most prominent movement (hands → head → feet).</td>
</tr>
<tr>
<td>If both hands are used for different purposes, use Tier 1 for the right hand and Tier 2 for the left hand.</td>
<td></td>
</tr>
<tr>
<td>Use a combination of terms from CV GESTURE BABY and GESTURE ADULT (see below).</td>
<td></td>
</tr>
</tbody>
</table>

---

### Gesture tier

One per child (from 9 months, use two per child, and from 12 months, use the annotation tags from adult AND child). Mark beginning and end time for gesture. Right and Left hand respectively. The dominant hand is used for GESTURE 1. Describe the gesture in words (e.g. both hands out from body, palms facing up, “it’s all gone”). Annotate FUNCTION whenever possible (DEICTIC, ICONIC, EMBLEM, EMPHATIC, GROOMING, EMOTIVE, OTHER). Example: DEICTIC_points with right hand index toars the bunny on the floor. One gesture can have many functions like ACTION/EMBLEM_makes peek-a-boo 4x.

DEICTIC = pointing gesture (whole hand, index finger, middle finger etc. Inlude pointing with gaze and with object – mark which kind of pointing it is).
ICONIC = decriptive gestures symbolising e.g. shape, distance, height etc. Visual element of actions could also appear, e.g. “pull back” together with a movement of pulling something.

EMBLEM = conventional gestures like waving, clap hands, put index-finger in front of lips for “hush”, etc.

EMPHATIC = gestures/movements/actions marking rhythm, e.g. drumming with fingers on table; other gestures used for marking time or to emphasize something said.

EMOTIVE = gestures/movements appearing together with emotional utterances of some sort (vocal, verbal, physical, etc). Positive and negative; could sometimes be “regular” gestures (like the ones above) but used with a different force. In these cases, mark as both EMOTIVE and, for example, EMBLEM.

- Some facial expressions are marked as gestures, e.g. /smiles/: EMOTIVE_smile.
- Some actions are marked as gestures, e.g. /hugs/: EMOTIVE_hugs child; EMOTIVE_/kissing sound 3x).

GROOMING = gestures/movements like adjusting hair, clothes; scratching, pick the nose, etc.

SHOW/OFFER = a child/adult holds out an object in order to show it (often gaze alternates between object and person). Sometimes show and offer are similar. Use SHOW/OFFER if uncertain, otherwise pick one of them.

ACTION = movements where you are uncertain if it is a conventional gesture or something else, e.g. movements/actions with toys. Describe the activity in words. Example: rhymes and clapping games (“itsy bitsy…” etc).

OTHER = gestures/movements not fitting in any description above.