Joint attention and language acquisition

A longitudinal study of joint attention in parent-child interaction

Freya Eriksson
Joint attention and language acquisition

A longitudinal study of joint attention in parent-child interaction

Freya Eriksson

Abstract

Joint attention (JA) is the ability to coordinate attention between a conversation partner and an outside object, while being aware of the other’s attention. This study investigates JA in 14 parent-child dyads at the ages 0;9, 1;0, 1;3, 1;6, 1;9 and 2;0, and aims to examine how the initiation of JA develops with age, as well as the relationship between JA and later vocabulary size and syntactic level. The strategies for initiating JA were recorded for both parents and children, and the total amount of JA, as well as child-initiated JA, was calculated for each age point. The results show that children with a higher language level, calculated as a composite score of vocabulary at 4;0 and syntactic level at 3;0, spent on average more time in JA than children with a lower language level. In line with previous research, the present study found a positive relationship between JA and vocabulary. Furthermore, the results suggest a relationship between JA and syntactic development. Especially the amount of child-initiated JA was related to both vocabulary size at the age of 4;0 and syntactic level at the age of 3;0, which indicates the importance of this type of interaction for language acquisition.

Keywords

Joint attention, language acquisition, parent-child interaction, vocabulary development, syntactic development
Joint attention och språkutveckling

En longitudinell studie av joint attention i förälder-barninteraktion

Freya Eriksson

Sammanfattning

Joint attention (JA) är förmågan att koordinera sin uppmärksamhet mellan en konversationspartner och ett tredje objekt, och samtidigt vara medveten om den andres uppmärksamhet. Den här studien undersöker JA i 14 förälder-barn dyader vid åldrarna 0;9, 1;0, 1;3, 1;6, 1;9 och 2;0, med syftet att utröna hur initieringen av JA utvecklas med åldern, och hur relationen mellan JA och senare ordförrådssstorlek och syntaktisk nivå ser ut. Strategierna för att initiera JA undersöks för både föräldrar och barn, och den totala mängden JA och barninitierad JA räknades för varje ålderspunkt. Resultaten visar att barn med en högre språklig nivå, räknat som kompositpoängen av vokabulär vid 4;0 och syntaktisk nivå vid 3;0, i genomsnitt spenderade mer tid i JA än barn med en lägre språklig nivå. I linje med tidigare forskning påvisade denna studie ett positivt samband mellan JA och ordförråd. Därutöver tyder resultaten på ett samband mellan JA och syntaktisk utveckling. Särskilt mängden barninitierad JA var positivt relaterad till både ordförråd vid 4;0 och syntaktisk nivå vid 3;0, vilket indikerar att denna typ av interaktion är viktig för språkutveckling.

Nyckelord
Joint attention, språkutveckling, förälder-barninteraktion, ordförrådsutveckling, syntaxutveckling
## Contents

1. Introduction ............................................................................................................. 1

2. Background .............................................................................................................. 2
   2.1 Social interaction as a basis for language acquisition ........................................ 2
   2.2 The evolution of JA ............................................................................................ 2
   2.3 Components of JA ............................................................................................. 3
   2.4 Development of JA in children .......................................................................... 4
       2.4.1 JA and social-cognitive development .......................................................... 5
       2.4.2 JA and language acquisition ...................................................................... 5

3. Aim and research questions ................................................................................... 8
   3.1 Aim ..................................................................................................................... 8
   3.2 Research questions ............................................................................................ 8

4. Method ..................................................................................................................... 9
   4.1 The MINT Project ............................................................................................... 9
   4.2 Material ............................................................................................................... 9
       4.2.1 Sample ......................................................................................................... 9
       4.2.2 Recordings ................................................................................................... 9
       4.2.3 Annotation in ELAN .................................................................................. 10
       4.2.4 Vocabulary measure: SCDI-III ................................................................. 10
       4.2.5 Measure of syntactic level: Unified predicate score .................................... 10
       4.2.6 Ethical considerations ............................................................................... 10
   4.3 Definition and coding of JA .............................................................................. 10
   4.4 Data analysis ...................................................................................................... 11
       4.4.1 Statistical analysis ..................................................................................... 12

5. Results .................................................................................................................... 13
   5.1 Language measures ........................................................................................... 13
   5.2 JA over time ........................................................................................................ 14
       5.2.1 Initiation strategies for JA episodes ............................................................. 14
       5.2.2 Amount of JA episodes over time ............................................................... 15
       5.2.3 Length of JA episodes ............................................................................... 17
   5.3 Total JA and language measures ...................................................................... 18
       5.3.1 Total amount of JA and vocabulary ........................................................... 18
       5.3.2 Total amount of JA and syntactic level ....................................................... 19
   5.4 Child-initiated JA and language measures ....................................................... 19
       5.4.1 Amount of child-initiated JA and vocabulary ............................................. 20
       5.4.2 Amount of child-initiated JA and syntactic level ....................................... 21
   5.5 Age of onset of child-initiated JA and language measures ............................. 22

6. Discussion ............................................................................................................... 24
6.1 Method discussion

6.1.1 Validity

6.1.2 Reliability

6.1.3 Generalizability

6.2 Result discussion

6.2.1 Strategies for initiating JA

6.2.2 Number of JA episodes

6.2.3 Time spent in JA and length of JA episodes

6.2.4 Total amount of JA and language acquisition

6.2.5 Child-initiated JA and language acquisition

6.2.6 Age of onset of child-initiated JA and language acquisition

6.2.7 General discussion

6.3 Ideas for future research

7. Conclusions

References

Appendices

Appendix 1. Individual JA values

Appendix 2. Individual child-initiated JA values
1. Introduction

Humans are social beings, capable of complex cooperation, communication and culture (Tomasello 2014: 1-6). There are theories about the evolutionary steps that resulted in this, but at some point, humans developed the ability to have joint intentions and socially coordinate towards a common goal. Related to this is the ability to coordinate attention between a communication partner and an external object, i.e. joint attention (JA) (2014: 33-34).

JA is a social-cognitive ability which predates theory of mind in ontogeny, and research has suggested it to be a precursor to theory of mind (Sodian & Kristen-Antonow, 2015). It also seems intimately connected with complex language and has been suggested to have evolved either before or along with human language (Kwisthout, Vogt, Haselager & Dijkstra, 2008). JA abilities in children also emerge around the same time as their first linguistic symbols, e.g. gestures and first words, and continue to develop along with language in ontogeny (Tomasello, 1995).

The interest in JA in child development began in the 1970s (see e.g. Scaife & Bruner, 1975). Since then, research has connected JA with other developmental milestones, e.g. perspective taking, theory of mind and language acquisition.

Research into the role of JA in child language acquisition has mostly focused on its impact on vocabulary development. The present study aims to add to the growing body of research underlining the importance of JA for vocabulary development, and to further investigate a possible relationship with another aspect of language acquisition: the development of syntax. The present study also aims to extend the knowledge of how JA is established in parent-child interaction.
2. Background

2.1 Social interaction as a basis for language acquisition

There is a growing body of research underlining the importance of social interaction for children’s acquisition of language.

According to Tomasello (2003: 3-4), children are equipped with two sets of skills that make language acquisition possible: pattern-finding and intention-reading. Pattern-finding has to do with children’s ability to form categories of perceptually or conceptually similar items and is used to find (linguistic) patterns in adult utterances. Intention-reading abilities emerge around the age of 0;9 to 1;0 and include the ability to share, follow and direct attention, and to understand that others are intentional beings. Intention-reading skills are related to language because they allow children to use their understanding of others’ intentions to decipher what they are saying, i.e. the context that language is used in is meaningful and helps the child navigate their linguistic environment (ibid.).

Social interaction seems to be fundamental for all aspects of language acquisition. Kuhl, Tsao and Liu (2003) conducted a study to investigate the importance of social interaction when acquiring a phoneme inventory. American children at the age of 0;9 participated in either live play sessions with a Mandarin speaker, watched the same session on a screen or just listened to it, and were subsequently tested for language specific phoneme discrimination in Mandarin. The results revealed that only the children who participated in the live sessions showed phonetic learning, meaning that the social component was key. Furthermore, it has been found that children learn verbs in socially contingent interactions, both live interaction and video chats, but not from non-contingent pre-recorded interactive videos (Roseberry, Hirsh-Pasek & Golinkoff, 2014). Roseberry Lytle, Garcia-Sierra and Kuhl (2018) further found that children aged 0;9 who were exposed to a foreign language on an interactive screen had better phonemic learning when a peer was present than when they were exposed to the same setting alone. These studies suggest that pattern-finding (Tomasello, 2003: 3-4) is scaffolded by social interaction.

Clark (2018) argues that children learn language through conversation mainly with adults, and when interacting both parties rely on the Cooperative Principle (Grice, 1975) and the Principles of Conventionality (i.e language is conventional) and Contrast (i.e. new words equal new meaning). Using these principles, adults give feedback on children’s language by repeating or elaborating on their utterances. This behavior likely stimulates the development of intention-reading skills in children. Clark (2018) further suggests that this is successful because the interaction is based on cooperation, where both parties want to understand, and be understood by the other. Thus, through interaction and conversation children are provided with new words and constructions, while having opportunity to practice their own language.

Social interaction, both with adults and peers, is at the core of language acquisition.

2.2 The evolution of JA

JA is a social-cognitive ability, which is defined as two people simultaneously coordinating their attention between an external object and each other, while being aware of the other’s attention (Tomasello 1995: 105-7).

There are different theories about the evolution of JA. According to Racine, Wereha, Vasileva, Tafreshi and Thompson (2014), JA is a complex phenomenon that is connected to other social-cognitive capacities, such as intersubjectivity and theory of mind. They argue that JA is necessary for intentional social communication, and consequently for language, meaning it must have evolved
before language. Even though JA occurs together with a number of behaviors such as gaze following or pointing, explanations of the evolution of JA often focus on one behavior at a time.

In accordance with Racine and colleagues’ (2014) argument, research indicates that JA skills were crucial for the evolution of language. Kwisthout and colleagues (2008) argue that increasingly sophisticated JA skills may have evolved in parallel with language. They implicated the three stages in children’s JA development described by Carpenter, Nagell and Tomasello (1998), i.e. checking (sharing attention in Carpenter et al.), following and directing attention, in a language game model in which agents interact in a common environment and construct a lexicon. Kwisthout and colleagues (2008) tested how models with different configurations of the stages performed and found that a model which had all three stages performed the best, since JA narrows down the referential context. However, out of the models that only had one JA skill available, the checking attention was the best, followed by the following attention model. This is relevant for the order of evolution of these skills; they argue that and that the checking attention skill, which is the simplest and alone has the largest effect on language, likely evolved first, following attention likely came next and directing attention last.

2.3 Components of JA

There is a difference between passive JA, where children and parents focus on the same thing, but the child’s understanding of the adult’s attention is not clear, and coordinated JA, where it is clear that the child understands that the attention is joint. Passive JA emerges before coordinated in development (Bakeman & Adamson, 1984).

JA can be established in different ways; however, research suggests that adults mainly rely on speech, gestures and gaze to achieve JA, and that gestures play a big role in getting and holding attention (Clark, 2018). Children begin using pointing gestures towards the end of their first year of life (Camaioni, Perucchini, Bellagamba & Colonnese, 2004), around the same time as the first JA abilities.

The ability to follow gaze develops towards the end of the first year of life and is closely related with the development of JA (Corkum & Moore, 1995). Human eyes are suited for following gaze, because the sclera are white, which is not the case in other primates. It has been suggested that this feature has evolved because it makes behaviors such as gaze-following, and consequently JA abilities, easier (Tomasello, Hare, Lehman and Call, 2007). Brooks and Meltzoff (2005) found that children at the age of 0;10 and 0;11, look at an adult’s looking target when the adult’s eyes are open, but not when they are closed, while children at the age of 0;9 did not make this distinction. However, Corkum and Moore (1995) found that children between the age of 1;0 and 1;3 seem to rely more on head-orientation rather than eye-orientation, while children at the age of 1;6 rely more on the orientation of the eyes than the head. Tomasello and colleagues (2007) found the same tendency in great apes as in the younger children in Corkum and Moore (1995).

In complex situations including several objects that are close to each other, children at the age of 1;0 rely on the hand movements of their conversation partner to coordinate visual attention, rather than following gaze. This strategy is used more by children than by adults (Yu & Smith, 2013).

Touch likely also plays an important role in JA and can be used to attract an infant’s attention. It is also proposed to help the infant understand the difference between themselves and others in the earliest social interactions (Botero, 2016). Furthermore, Suarez-Rivera, Smith and Yu (2019) found that when parents and infants engage in joint visual attention, parents are very likely to add a behavior (speech or touching the object in focus) to help maintain the child’s attention.

While there is a clear focus on joint visual attention in the literature, gaze is not necessary for JA. JA has been observed in blind children; however, it seems to develop later in these children. Furthermore, they are more reliant on tactile and auditory information to respond to JA bids, and language to direct others’ attention. In blind children, as well as in sighted, JA helps the development of language, but language also helps the development of JA (Bigelow, 2003).
According to Clark (2018), initiating JA is a way for adults to start conversations with children. Adults then maintain the joint attentional focus using mainly speech and gestures to keep the child’s attention on the object or event in question. Within episodes of JA, the adult can “co-construct conversation” with the child, by scaffolding or framing the child’s utterances, which helps children move from one-word utterances to word sequences and early constructions. This argument is supported by a finding by Roth (1987) that mothers more often elaborate semantically on children’s utterances when they follow into the child’s focus of attention than when they have a different attentional focus.

2.4 Development of JA in children

Coordinated JA emerges towards the end of the first year of life, usually around the age of 0;9 to 0;10 (e.g. Tomasello, 1995; Adamson & McArthur, 1995). This is followed by a period of consolidation, which lasts approximately between the age of 1;1 and 1;6. Towards the end of their second year children use conventional symbols during JAA episodes, which enables more complex conversations and use of linguistic symbols (Adamson & McArthur, 1995). Carpenter and colleagues (1998: 69-70) suggest that JA abilities develop in the following order: sharing attention, following attention and directing attention. As JA skills become more sophisticated, the scope of JA expands through the first three years of life (Adamson, Bakeman & Deckner, 2004).

When JA skills first begin to emerge, it happens at the same time as other developmental milestones, e.g. intentional communication (both gestures and first words). Between the age of 0;9 and 1;6, children learn to follow and direct others’ attention, and between the age of 1;6 and 2;0, children’s JA abilities are apparent in several ways in learning and language use (Tomasello, 1995: 105). An example of this is the emergence of two-unit utterances around the age of 1;6, where children combine either a word and gesture or two words into a topic-comment structure. According to Tomasello, this requires an understanding that others’ attentional focus can be modified by language or gestures, which shows that children have reached a more “sophisticated level of social cognition” (1995: 117-8).

Bakeman and Adamson (1984) investigated JA in 28 children between the age of 0;6 and 1;6 during interaction with their mother or a peer. They found that coordinated JA, i.e. JA where the child’s understanding of the JA state is clear, made up about 2 percent of the interaction at the age of 0;6 and 0;9. Coordinated JA then increased with age and doubled between the age of 1;3 and 1;6. Children also spent more time in coordinated JA when interacting with their mother than when they interacted with a peer, although both increased with time. In light of these results, they argue that children between the age of 0;6 and 1;6 have “infant-based constraints on the flow of their attention between object and social domains”. However, adults can adapt their behavior to accommodate children’s attention capabilities, thus providing a social context for referential communication.

Mundy and colleagues (2007) studied the difference between responding to JA and initiating JA in a longitudinal study on 95 children between the age of 0;9 and 1;6. They found that responding to JA increased with age, except between the age of 1;3 and 1;6, where there was no significant increase. Initiating JA, however, did not increase linearly over time but instead showed a cubic main effect for age. They argue that the different kinds of JA abilities, responding and initiating, are part of different activation patterns in the brain and thus develop differently. Mundy and colleagues (2007) suggest that responding to JA is related to posterior activation, which is connected with the development of a more reflexive orienting to meaningful stimuli. This develops early in infancy and may be related to processing gaze shifts. Initiating JA, on the other hand, is suggested to be regulated by frontal activation, which is connected with intentional attention systems. This develops later and may be linked with the understanding of others as intentional agents. The present thesis will further investigate the development of JA initiation strategies and the development of the role as JA initiator.

Adamson and colleagues (2004) turn special focus to the development of what they call symbol-infused joint engagement, i.e. joint engagement (interaction while in a joint attention state) where the child is attending to and using symbols (language and gestures). Their results show that symbol-infused joint engagement increases with age and highlight that there is a large amount of variation in
this development, especially during the second half of the second year of life. When dividing the children (n=56) into an early and late language group, Adamson and colleagues (2004) found that while all children were able to engage in JA at 1;6, i.e. at the beginning of the study, there was a clear difference between the groups in terms of using symbols during joint engagement at this time.

The present study will contribute to previous research by investigating the development of JA in parent-child interaction every third month between the age of 0;9 and 2;0.

### 2.4.1 JA and social-cognitive development

Previous research has found that JA is related to other aspects of social-cognitive development, e.g. the development of perspective taking and theory of mind.

Nelson, Adamson and Bakeman (2008) investigated the relationship between JA and the development of theory of mind. They analyzed 42 children’s JA experiences with their mothers between the age of 1;6 and 2;6 and tested this against a false-belief understanding at the age of 3;6, 4;6 and 5;6. They found that more time spent in coordinated JA at the age of 1;6 and 1;9, as well as in symbol-infused JA and total JA at the age of 2;3 and 2;6 was related to higher false belief scores. Therefore, Nelson and colleagues (2008) conclude that JA experiences affect the timing of children’s development of theory of mind. As a possible explanation for this, they suggest that coordinated JA provides a setting for observing and investigating a communication partner’s reactions and perspectives on the object or event in focus.

Sodian and Kristen-Antonow (2015) conducted a study on 83 monolingual German children, investigating a possible link between declarative JA at the age of 1;0 and 1;3, mirror self-recognition at the age of 1;3, 1;6 and 2;0, Level 1 perspective taking at the age of 2;6 and theory of mind at the age of 4;2. They measured production of both declarative and imperative JA (i.e. declarative and imperative pointing, see e.g. Bates, Camaioni & Volterra 1975), and found that declarative JA skills, but not imperative or combined, were related to theory of mind at the age of 4;2 (measured by a false-belief test). Declarative JA at 1;0 was further related to mirror self-recognition at the age of 1;6 and 2;0, while declarative and imperative JA combined were related to Level 1 perspective taking. Sodian and Kristen-Antonow (2005) argue that this provides evidence for the claim that JA is essential for the development of theory of mind and highlights the importance of declarative JA specifically.

Camaioni and colleagues (2004) found similar patterns in a study on 40 typically developing children at the age of 1;0 and 1;3, i.e. that declarative pointing developed later than imperative, and that declarative pointing, but not imperative, was related to understanding others’ intentions. They argue that this is because declarative pointing shows an understanding that it is possible to effect others’ attentional states and thus see them as intentional agents.

Humans are social beings equipped with complex social-cognitive abilities. JA is one such ability which emerges quite early in ontogeny and it is intimately related to social-cognitive development in general.

### 2.4.2 JA and language acquisition

Though JA is not primarily a linguistic skill, several studies have investigated the relationship between JA and language acquisition, focusing on vocabulary development in particular (Tomasello & Todd, 1983; Tomasello & Farrar, 1986; Morales et al., 2000; Loy, Masur & Olson, 2018; Zampini, Salvi & D’Odorico, 2015; Adamson, Bakeman, Suma & Robins, 2019; Hirotani, Stets, Striano & Friederici, 2009). One study by Rollins and Snow (1998) has linked JA with grammar development. The present study aims to investigate JA in relation to both these aspects of language.

Tomasello and Todd (1983) investigated JA in six mother-child dyads, between the age of 1;0 and 1;5. They found that children who spent more time in JA with their mother had a larger vocabulary size at the age of 1;5. Furthermore, they found that joint interactions where mothers lead the interaction by directing the child’s attention to new objects were negatively correlated with the amount of object labels in the children’s vocabulary. However, JA where both parties lead the interaction was positively correlated to vocabulary.
Morales and colleagues (2000) investigated how responding to JA bids of others, i.e. following gaze, pointing, head turning or a combination, between the age of 0;6 and 2;0 was related to vocabulary development. They investigated responding to JA in 22 children at the age points 0;6, 0;8, 0;10, 1;0, 1;3, 1;6, 1;9 and 2;0, and tested against parental reports of expressive vocabulary at the ages 2;0 and 2;6, as well as test of receptive and productive language at the age of 2;6. Responding to JA at the ages 0;6, 0;8, 0;10, 1;0 and 1;6 was positively related to vocabulary size, and responding to JA at 0;10 and 1;6 was positively related to expressive vocabulary at the age of 2;6. However, responding to JA at the age of 1;9 and 2;0 was not related to vocabulary size.

More recently, Loy and colleagues (2018) investigated JA in 29 mother-child dyads at the age of 1;1 and 1;5, focusing on how initiatives and responses to JA were made by mothers and children and relating these skills to vocabulary size. They identified three types of initiatives, i.e. object oriented, partner oriented and combined JA initiatives, which included both an object and a partner and found that combined initiatives by mothers resulted in the largest number of JA episodes. Furthermore, they found a relationship between the amount of combined initiatives made by children at both the observed age points and vocabulary size at the age of 1;5, measured by a CDI. However, there was a positive correlation between both initiating JA and responding to JA initiatives and vocabulary size.

There is also evidence of JA having an effect on vocabulary development in children with Down syndrome. Zampini and colleagues (2015) found that the children with Down syndrome (n=18) who were more successful in achieving JA episodes with a communication partner at the age of 2;0 had a greater vocabulary size at the age of 2;6.

Adamson and colleagues (2019) further found that time spent in joint engagement, i.e. interaction while in a JA state, predicted later expressive vocabulary better than specific JA skills (ability to respond to and initiate JA) in children with developmental delay (n=46) and children with autism spectrum disorder (n=58). However, these children spent less time in joint engagement than typically developing children (n=40).

Tomasello and Farrar (1986) investigated how language was used within JA episodes in naturalistic recordings of 24 mother-child dyads at the age of 1;3 and 1;9, and subsequently tested whether JA affected word learning in a lexical training study. They calculated utterances per minute, MLU, words per minute, object labels, conversations, child turns, and mothers’ use of questions, comments and directives, during JA and outside JA, and found that the amounts of these measures differed between the times when the dyads engaged in JA and when they did not. The most significant difference between the inside and outside of JA episodes occurred at the age of 1;9. During JA the dyads talked more, had longer conversations and the mothers used shorter sentences and made more comments. The mother’s use of object references that followed the child’s attention correlated positively with vocabulary size, while none of the measures outside of JA correlated with measures of child language. The lexical training study supported these results; it was found that word learning was facilitated when children were presented with object labels inside JA episodes. Based on these results, Tomasello and Farrar (1986) suggest that JA scaffolds language acquisition. They argue that this might be because JA provides good conditions for learning, especially when the adult follows the child’s attention.

The claim that JA has a positive impact on vocabulary acquisition has also been supported by ERP (event-related potentials) evidence.1 Hirotani and colleagues (2009) conducted an EEG-study on 23 German children between the ages of 1;6 and 1;9, where the children were taught novel words in a JA and non-JA setting. The words were subsequently tested by presenting pictures on a screen and either congruent or incongruent words on a loudspeaker. Hirotani and colleagues (2009) found that congruent words taught in both settings elicited a negativity, but incongruent words that had been taught in the JA setting elicited an N400-like effect,2 which was interpreted as “failure in semantic integration”. Hence, they suggest that this indicates that the JA setting improves word learning, since it

---

1 Event-related potentials (ERP’s) are brain responses to stimuli measured by EEG (Kutas & Hillyard, 1980).

2 The N400 response is a negativity which typically follows semantically incongruent stimuli (Kutas & Hillyard, 1980).
facilitates a deeper semantic processing of words and thus “enhances the link between a lexical item and its meaning”.

While most studies on joint attention and language acquisition have focused on vocabulary development, a study by Rollins and Snow (1998) connected JA to the development of grammar. They investigated three types of pragmatic abilities (negotiating an ongoing activity, participating in social routines and regulating JA) at the age of 1;2 and the relationship between these abilities and grammatical level at the age of 2;7. In typically developing children (n=30), they found that regulating JA together with parental conversation style predicted grammar level, while the other pragmatic abilities did not. They then replicated the study on children with autism spectrum disorder (n=6) and found that only the ability to establish and maintain JA predicted grammar development in those children. Rollins and Snow (1998) suggest that only using language for instrumental purposes, e.g. negotiating an ongoing activity by requesting help, might not be related to grammar development. This is in line with the findings of Colonnese, Stams, Koster and Noom (2010) who found that declarative pointing, but not imperative, was positively related to both receptive and expressive language. More importantly, Rollins and Snow (1998) argue that their results indicate that pragmatic abilities have an impact on the development of grammar, and that JA seems to be essential for syntactic development.

The present thesis aims to investigate JA in parent-child interaction, both focusing on how it is initiated by parents and children and how JA is related to language acquisition. Previous research about JA and language acquisition has focused mainly on vocabulary size as a measure of language level and found a strong link between JA and vocabulary. The present study will investigate how JA is related to vocabulary size at 4;0, and is expected to support the findings of previous research. However, other aspects of language have not been given as much attention. Consequently, this thesis will build on the findings of Rollins and Snow (1998), which suggest that JA is related to syntactic development, and further investigate the relationship between JA and this aspect of language acquisition. In this study, syntactic level will be measured by an analysis of unified predicates at the age of 3;0 (Berman & Slobin, 1994), which will result in a percentage of well-formed clauses, following Tonér and Gerholm (under review). At this age the children are expected to have started building sentences, which is necessary for the syntactic analysis. Furthermore, the present study will investigate how child-initiated JA affects later vocabulary and syntactic level.
3. Aim and research questions

3.1 Aim
The aim of the present study is to investigate the development of JA in parent-child interaction during children’s first two years of life (at the age points 0;9, 1;0, 1;3, 1;6, 1;9 and 2;0). The study consists of a descriptive part which describes how JA develops over time, who initiates it and how this is done, and a part where it is tested whether JA is linked to syntactic level at the age of 3;0 and vocabulary size at the age of 4;0.

3.2 Research questions
1a. How is JA initiated by parents and children in free interaction, i.e. who initiates it and what strategies are used?
1b. Does the way JA is initiated change over time?
2a. Does the time spent in JA at the observed age points predict productive vocabulary at the age of 4;0?
2b. Does the time spent in JA at the observed age points predict syntactic level at the age of 3;0?
3a. Does the time spent in child-initiated JA (CJA) at the observed age points predict productive vocabulary at the age of 4;0?
3b. Does the time spent in child-initiated JA (CJA) at the observed age points predict syntactic level at the age of 3;0?
4a. Do children who initiate JA early have a larger productive vocabulary at the age of 4;0?
4b. Do children who initiate JA early have a higher syntactic level at the age of 3;0?
4. Method

In this section the method is presented, including a description of the MINT Project, material, language measures, coding of JA and data analysis.

4.1 The MINT Project

The material used in the present study is a subset of the material recorded in the MINT Project, a five-year research project conducted at the Department of Linguistics, Stockholm University (Gerholm, 2018). 1000 randomly selected families in the Stockholm area were invited to be part of the project. In the beginning of the project, 85 families participated, and at the age of 4;0 71 families remained. The children and parents were recorded every third month from the age of 0;3 until the age of 3;0, and twice a year until the age of 5;0.

4.2 Material

4.2.1 Sample

For this thesis, fourteen typically developed children (seven girls, seven boys) were chosen. All children included in this study have only Swedish spoken by the parents at home.

The majority of the children included in the present study started preschool before the age of 1;8, and by the age of 2;3 all were enrolled. All families except one had an annual income of at least 400 000-800 000 SEK. All children but two have at least one parent with a university degree, and all parents have finished high school. All families lived in the Stockholm area when the project started.

The sessions used in this study were recorded at the ages 0;9, 1;0, 1;3, 1;6, 1;9 and 2;0. All children included in this study were present at all six age points, as well as the session at the age of 3;0, which is the basis for the analysis of syntactic level (see section 4.2.5). Furthermore, they had SCDI-III, a parental questionnaire, filled out at the age of 4;0 (see section 4.2.4).

4.2.2 Recordings

The recordings were made at the Stockholm Babylab, located at the Department of Linguistics, Stockholm University. The recording studio was equipped with three stationary cameras mounted on the walls, and the parents wore an action camera on their chest. Both parents and children wore a clip microphone, and the room was further equipped with a condenser microphone, to synchronize the sound and the video.

The room was set up with pillows and a blanket, as well as three stuffed animals and some age appropriate toys and books.

At the age of 0;9, the recordings were approximately 15 to 20 minutes long, and divided into two parts of free interaction. In the present study, only the first five to ten minutes of free interaction were used.

From the age of 1;0 the sessions consisted of a free interaction part which lasted approximately ten minutes, where the parent and the child were instructed to play as they normally would. This was

---

1 Modelling Infant Language Acquisition from Parent-Child Interaction, funded by the Marcus and Amalia Wallenberg Foundation (MAW 2011.007).
2 The stationary camera model was Canon HDMI XA10 and the action camera model was GoPro Hero3.
3 The clip microphones were of the model Sennheiser eW 100 G2 and the condenser microphone model was AKG SE 300 B.
followed by a semi-structured game or test conducted by a researcher. For the purposes of this thesis, only the free interaction part was used.

4.2.3 Annotation in ELAN
The recordings are multimodally annotated in the software ELAN (Brugman & Russel, 2004) by trained annotators, following an annotation key (Gerholm, 2018). Each recording has several tiers in ELAN, where vocals, gestures, gaze, touch, facial expressions/mood, distance and activity type are tagged for the parent and child.

4.2.4 Vocabulary measure: SCDI-III
As a measure of vocabulary size at the age of 4:0, SCDI-III was used. SCDI-III is a Swedish adaptation of the MacArthur-Bates CDI-III (Fenson et al., 2007), a version of the MacArthur-Bates Development Inventories (Fenson et al. 1993) developed for children up to the age of 3:0. SCDI-III has been validated for the ages 2:6 to 4:0 by Eriksson (2017), and consists of a vocabulary list of 100 words in selected categories, e.g. “food words”, “body words”, “thought words”, etc., and sections on sentence structure, sentence complexity, pronunciation and metalanguage. For the purposes of the present study, only the score on the vocabulary list was used.

4.2.5 Measure of syntactic level: Unified predicate score
The syntactic level at the age of 3:0 was measured as the percentage of well-formed clauses in the recording session at that age. The children’s utterances in the recordings at the age of 3:0 were extracted and subsequently analyzed on clause level, i.e. a unit containing a unified predicate in accordance with Berman and Slobin (1994: 660). They define a unified predicate as “a predicate that expresses a single situation” (ibid.). Following Tonér and Gerholm (under review), each clause which contained a unified predicate was analyzed in relation to the target sentence, i.e. an assessment of what the child was trying to say. All phrases were subsequently scored between 0 and 4 for i) appropriate number of words, ii) appropriate inflections, iii) word order and iv) functionality (i.e. that the sentence works given that it is spoken language). This resulted in a percentage of well-formed clauses.

4.2.6 Ethical considerations
The MINT Project was conducted in accordance with the guidelines of the Research Ethics Committee at the Karolinska Institute and of the Swedish Data Protection Authority, and has been approved by the Regional Ethics Committee in Stockholm (Dnr 2011/955-31/1). Following the guidelines for research on humans (2003:460), all families participating in the MINT Project signed informed consent and the participant information was handled in accordance with the General data protection regulation (GDPR 2016/679). The children included in the present study have been anonymized.

4.3 Definition and coding of JA
In addition to the annotations that were previously done as part of the MINT Project, a tier was added to annotate JA for the purposes of this thesis. The definition of a JA episode used in the present study was based on that in Tomasello and Todd (1983), which was further developed by Tomasello and Farrar (1986) to contain three components:

a) A JA episode is initiated when one of the participants initiates interaction around an outside object or event with the other.

b) Both the child and the parent are focused on an object or activity for at least 3 seconds. However, they can look away briefly during a longer episode.

c) At some point during the JA episode, it must be overtly expressed that the child is aware of the JA, e.g. by looking at the parent, gesturing or commenting on the activity in a way that makes this clear.
The following is an example of a JA episode from the material:

The parent says “can you drive the car to me?” referring to a toy car that is on the floor while looking at the child. The child then pushes the car over the floor towards the parent, while looking at the car, then looks up at the parent who pushes it back. This was repeated several times.

For the purposes of this study, the annotation of JA episodes also included information about who initiated the episode (parent or child) and what strategy (i.e. vocalization, gesture, gaze, touch, or a combination) was used to do so.

It was possible for two JA episodes to follow directly after each other, if the focus was changed to some other object or activity. For example, if a parent initiated an episode in which the game was to look at what was in the basket and take out the toys, that game would constitute one episode. This could be directly followed by the child engaging with one specific toy and starting a different activity using it. Such situations were coded as two episodes, but directly adjoining.

Overt behaviors that signaled the child’s understanding of the JA were e.g. gaze alternating between the object and the parent, gestures that directed the parents attention or comments made about the joint activity or object in focus, which made it clear that the child understood that the parent was attending to the same thing, i.e. discussing the object or event in question. Furthermore, there needed to be some interaction between the participants. Following Tomasello and Farrar (1986), i.e. the child playing with an object and the parent watching without any kind of engagement with the child or object, or comments about it, was not counted as JA.

### 4.4 Data analysis

The JA tier was extracted from ELAN for all sessions and subsequently analyzed.

The strategies used to initiate the JA episodes by parents and those used by children were counted separately and are presented on group level for the whole group.

The age of onset of child-initiated JA was recorded for each child as the age point where they first successfully initiated a JA episode.

For the descriptive results showing the number of JA episodes and time spent in JA, the children were divided into a high and a low language level group. Language level was calculated as a composite score of the SCDI-III score and the unified predicate score, meaning that it includes both vocabulary and syntax. This was calculated for each child by adding the percentage of the score on the SCDI-III and the percentage of well-formed clauses and subsequently calculating the mean of the two scores. Based on the results of this analysis, the children were divided into a high language level group (n=7, mean=0.84, SD=0.06) and a low language level group (n=7, mean=0.59, SD=0.1).

The mean number of JA episodes was counted for both groups at each age point, and subsequently calculated as frequency per minute, to account for the fact that all files were not the same length. The mean number of parent- and child-initiated episodes for the groups were also calculated as frequency per minute.

The mean amount of time spent in JA was calculated for the high and low language level groups at each age point as a percentage of the total time, by adding the time spent in JA and dividing it by the total recorded time.

The mean length of JA episodes was subsequently calculated in seconds for the high and low language level group at each age point. This was calculated by adding the time spent in JA and dividing it by the number of JA episodes, for each group respectively.

The total time spent in JA was also calculated for each parent-child dyad and age point. To account for the fact that the files are not all the same length, the total time in JA was then calculated as a percentage of the total file length.
Furthermore, the time spent in child-initiated JA was calculated for each parent-child dyad and age point. This was also adjusted to account for the different file lengths, resulting in a percentage of the total file length.

**4.4.1 Statistical analysis**

To test the relationship between the total amount of JA and vocabulary at the age of 4;0, linear regression analyses were conducted for each of the observed age points using JA as the predictor.

To test the relationship between the total amount of JA and syntactic level at the age of 3;0, linear regression analyses were conducted for each of the observed age points using JA as the predictor.

To test the relationship between the amount of child-initiated JA and vocabulary at the age of 4;0, linear regression analyses were conducted for each of the observed age points using child-initiated JA as the predictor.

To test the relationship between the amount of child-initiated JA and syntactic level at the age of 3;0, linear regression analyses were conducted for each of the observed age points using child-initiated JA as the predictor.

To test the relationship between the age of onset of child-initiated JA and vocabulary at the age of 4;0 and syntactic level at the age of 3;0, two linear regression analyses were conducted using the age of onset as the predictor.
5. Results

In this chapter, the results are presented in the same order as the research questions. First, the language measures used are presented, followed by a presentation of the initiations strategies and amount of JA over time, corresponding to the first research question. Then the results of the total amount of JA in relation to the language measures are presented, corresponding to the second research question, as well as the results of the child-initiated JA in relation to the language measures, corresponding to the third research question. The results of the fourth research question, i.e. age of onset of child-initiated JA in relation to the language measures, is presented last.

5.1 Language measures

The children’s language levels were measured as score on SCDI-III (vocabulary) at the age of 4;0 and percentage of well-formed unified predicate clauses at the age of 3;0 (syntactic level), as shown in figure 1.

The vocabulary scores ranged between 36 at the lowest and 96 at the highest (out of a maximum score of 100). The scores were spread out in between. The children’s syntactic level, i.e. the percentage of well-formed clauses, ranged between 36 and 91 percent.

![Figure 1. The graph shows the children’s vocabulary score (SCDI-III) at the age of 4;0 (x-axis) and the syntactic level at the age of 3;0 as measured by unified predicate score (UP, y-axis). There is a correlation of 0.725 (p < .01) between the two language measures.](image)

A Pearson correlation test was conducted to investigate the correlation between the vocabulary score at 4;0 and the unified predicate score at 3;0. There was a significant correlation of 0.725 (p < .01) between the language measures.
5.2 JA over time

The first research question addressed how JA is initiated in parent-child interaction and whether this changes over time. The results of this are presented on group level below.

5.2.1 Initiation strategies for JA episodes

The strategies used by parents and children to initiate JA were counted and are presented on group level as the percentage of total initiations for each of the observed age points (table 1 and 2), to account for the different file lengths.

The by far most frequent strategy to initiate JA episodes used by the parents was a vocalization in combination with a gesture. Vocalizations alone were also quite common, except at the age of 0;9, where they accounted for approximately 9 percent of parent initiations compared to 80 percent vocalization plus gesture.

Table 1. Initiation strategies used by the parents at the observed age points, presented as percentage of the number of parent initiations. The abbreviations used are: vocalization (V), gesture (Ge), gaze (Ga), and touch (T).

<table>
<thead>
<tr>
<th></th>
<th>V+Ge</th>
<th>V</th>
<th>Ge</th>
<th>V+T</th>
<th>V+Ge+T</th>
<th>V+Ga</th>
<th>V+Ge+Ga</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0;9</td>
<td>80</td>
<td>9.1</td>
<td>3.6</td>
<td>0</td>
<td>5.5</td>
<td>0</td>
<td>0</td>
<td>1.8</td>
</tr>
<tr>
<td>1;0</td>
<td>76.7</td>
<td>19.4</td>
<td>1.9</td>
<td>1.0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1;3</td>
<td>72.6</td>
<td>18.9</td>
<td>6.7</td>
<td>1.2</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>1;6</td>
<td>65.0</td>
<td>32.5</td>
<td>2.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1;9</td>
<td>70.3</td>
<td>27.1</td>
<td>0.6</td>
<td>0.6</td>
<td>0</td>
<td>1.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2;0</td>
<td>70.6</td>
<td>23.5</td>
<td>2.9</td>
<td>0.6</td>
<td>1.2</td>
<td>0.6</td>
<td>0.6</td>
<td>0</td>
</tr>
</tbody>
</table>

Children mostly used the vocalization-plus-gesture strategy to initiate JA episodes (vocalizations were most often speech, but sometimes younger children produced vocalizations that were not speech per se to catch the parent’s attention).

At the age of 0;9 gesture alone or vocalization-plus-gesture were used equally, and at the age of 1;0 the gesture alone strategy was used to the same extent as the vocal plus gesture strategy. At the age of 1;3 the vocal-plus-gesture strategy was used more than twice as often as the others.
Table 2. Initiation strategies used by the children at the observed age points, presented as percentage of the total number of child initiations. The abbreviations used are: vocalization (V), gesture (Ge), gaze (Ga), and touch (T).

<table>
<thead>
<tr>
<th></th>
<th>V+Ge</th>
<th>Ge</th>
<th>V</th>
<th>V+T</th>
<th>V+Ga</th>
</tr>
</thead>
<tbody>
<tr>
<td>0;9</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1;0</td>
<td>51.2</td>
<td>41.5</td>
<td>7.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1;3</td>
<td>65.7</td>
<td>29.3</td>
<td>5.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1;6</td>
<td>67.4</td>
<td>16.3</td>
<td>15.5</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>1;9</td>
<td>74.1</td>
<td>12.0</td>
<td>13.9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2;0</td>
<td>73.0</td>
<td>7.8</td>
<td>18.6</td>
<td>0</td>
<td>0.6</td>
</tr>
</tbody>
</table>

When only one modality was used, children and parents used different strategies: parents preferred the vocal alone strategy to the gesture alone strategy. Children on the other hand relied more on the gesture alone strategy at the three younger ages. At the age of 1;6, vocal alone and gesture alone were used to a similar extent, and later vocal alone was more frequent. At the age of 2;0, the distribution of the children’s initiation strategies looked very similar to the parents’, with vocalization-plus-gesture dominating, followed by vocal alone and gesture alone.

Both children and parents used gaze and touch sparsely to initiate JA episodes.

5.2.2 Amount of JA episodes over time

The dyads were divided into two groups based on the children’s language level, measured by a composite score of unified predicate score at 3;0 and vocabulary score at 4;0, resulting in a high (n=7) and low (n=7) language level group.

The total number of JA episodes, as well as the number of parent-initiated and child-initiated episodes was calculated for each age point and normalized as number of episodes per minute, to account for the different file lengths. The results of this in the high and low language level groups are presented in Figure 2.

There was an average of 0.5 JA episodes per minute at the age of 0;9 in both groups (SD=0.2 and SD=0.4 in the high and low groups respectively). The high language level group tripled this amount by the age of 1;0 (SD=0.7), while the low language group level was still under one episode per minute at this age (SD=0.3).

At the age of 1;3, the high level group had an average of more than two episodes per minute (SD=0.7), while the low level group reached two episodes per minute six months later, at the age of 1;9 (SD=0.9). The high level group’s curve leveled out at an average of just over 2.5 JA episodes per minute at the age of 1;9 (SD=0.8), and the low level group stayed just over two episodes per minute on average (SD=0.9).
Figure 2. The figure shows the mean number of JA episodes per minute at the observed age points in the high (left) and low (right) language level group, as well as the mean number of parent- and child-initiated episodes in both groups. The error bars show the standard error.

At the age of 0;9 almost all JA episodes in the material were parent-initiated, and the amount of parent-initiated episodes increased with age in both groups until the age of 1;3. In the high language level group, the average amount of parent-initiated episodes leveled out at the age of 1;3 at approximately 1.3 episodes per minute (SD=0.4). In the low language level group, the average amount of parent-initiated episodes also leveled out at the age of 1;3, at close to 1.1 episode per minute (SD=0.4).

The average amount of child-initiated JA episodes increased over time in both groups, and in the high language level group the children successfully initiated more JA episodes than the parents at the age of 1;9 and 2;0. However, in the low language level group the mean number of child-initiated episodes never surpassed the number of parent-initiated ones, though they approached the same mean amount of episodes per minute at the later age points.

The mean amount of time spent in JA, both total, child-initiated and parent-initiated, was calculated for each age point. This is presented for the high and low language level groups in Figure 3 as a percentage of total time. The mean amount of time spent in JA increased over all age points in both groups. While the high group had consistently more JA, the groups had a similar total time at the age points 1;9 and 2;0.

In the high language level group, the average amount of time spent in parent-initiated JA increased until the age of 1;6, where the child-initiated JA caught up. At the age of 1;9, there was on average more time spent in child-initiated JA than in parent-initiated JA in the high language level group, and at the age of 2;0, half the JA time was spent in child-initiated and parent-initiated JA respectively.
Figure 3. The average amount time spent in total, parent-initiated and child-initiated JA at the six age points for the high (left) and low (right) language level groups. The error bars show the standard error.

The low language level group followed a slightly different pattern, where both the average amount of time spent in parent-initiated and child-initiated JA increased with age. However, there was consistently less time spent in child-initiated JA across all age points and only at the age of 2;0 the average amount of time spent in child-initiated JA exceeded a third of the total JA time.

5.2.3 Length of JA episodes

The mean length of the JA episodes (in seconds) was calculated for the high and low language level groups for each age point as the total amount of time spent in JA divided by the number of JA episodes in the two language level groups respectively, and the results of this are presented in Table 3. Parent-initiated JA episodes were on average longer than child-initiated episodes, both in the high and low language level group.

The mean length of child-initiated episodes increased over time in both groups. However, in the high language level group the mean length of child-initiated episodes approached the mean length of parent-initiated episodes (approximately 13 seconds) at the age of 1;9. At the age of 2;0 the mean length of parent-initiated JA episodes was approximately 16 seconds and the child-initiated episodes were on average just over 15 seconds. In the low language level group, the child-initiated episodes were on average shorter than the parent-initiated ones, even at the later age points.

Table 3. The table shows the mean length (in seconds) of JA episodes for the high (H) and low (L) language level groups at the observed age points. Both parent-initiated (PJA) and child-initiated (CJA), as well as total mean length of JA episodes for each group are presented. Standard deviations are given in brackets.

<table>
<thead>
<tr>
<th></th>
<th>0;9</th>
<th>1;0</th>
<th>1;3</th>
<th>1;6</th>
<th>1;9</th>
<th>2;0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PJA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>15.5</td>
<td>8.2</td>
<td>11.4</td>
<td>8.3</td>
<td>14.9</td>
<td>11.9</td>
</tr>
<tr>
<td>(21.5)</td>
<td>(3.7)</td>
<td>(6.4)</td>
<td>(4.4)</td>
<td>(13.3)</td>
<td>(10.7)</td>
<td>(12.5)</td>
</tr>
<tr>
<td>L</td>
<td>10.1</td>
<td>0</td>
<td>7.8</td>
<td>6.9</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>(2.3)</td>
<td>(0)</td>
<td>(5.0)</td>
<td>(3.3)</td>
<td>(6.7)</td>
<td>(6.6)</td>
<td>(8.5)</td>
</tr>
<tr>
<td><strong>CJA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>10.1</td>
<td>0</td>
<td>7.8</td>
<td>6.9</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>(2.3)</td>
<td>(0)</td>
<td>(5.0)</td>
<td>(3.3)</td>
<td>(6.7)</td>
<td>(6.6)</td>
<td>(8.5)</td>
</tr>
<tr>
<td>L</td>
<td>15.1</td>
<td>8.2</td>
<td>10.2</td>
<td>8.1</td>
<td>12.4</td>
<td>10.8</td>
</tr>
<tr>
<td>(19.5)</td>
<td>(3.7)</td>
<td>(6.1)</td>
<td>(4.2)</td>
<td>(11.6)</td>
<td>(9.5)</td>
<td>(11.1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>15.1</td>
<td>8.2</td>
<td>10.2</td>
<td>8.1</td>
<td>12.4</td>
<td>10.8</td>
</tr>
<tr>
<td>(19.5)</td>
<td>(3.7)</td>
<td>(6.1)</td>
<td>(4.2)</td>
<td>(11.6)</td>
<td>(9.5)</td>
<td>(11.1)</td>
</tr>
<tr>
<td>L</td>
<td>16.1</td>
<td>10.2</td>
<td>12.4</td>
<td>10.8</td>
<td>14.5</td>
<td>12.5</td>
</tr>
<tr>
<td>(22.8)</td>
<td>(16.2)</td>
<td>(11.4)</td>
<td>(14.7)</td>
<td>(14.1)</td>
<td>(19.4)</td>
<td>(15.6)</td>
</tr>
</tbody>
</table>
5.3 Total JA and language measures

The second research question addressed whether the total amount ofJA at the observed age points predicted vocabulary size at the age of 4;0 and syntactic level at the age of 3;0. The total amount of JA was calculated for each parent-child dyad and age point (Appendix 1).

The development of the mean amount of time spent in JA for the high and low language level groups is presented in figure 4.

![Figure 4](image)

*Figure 4. The figure shows the development of the mean amount of total time spent in JA at the six age points for the children with high (n=7) and low (n=7) language level. The error bars show the standard error.*

The high language level group had on average more JA than the low language level group at all six age points. At the age of 0;9 both groups had approximately 10 percent of the total file time in JA, 13 and 8 percent in the high and low language level groups respectively. The high language level group then increased steeply to a mean of 58 percent at the age of 1;6, where the curve leveled out and continued to increase more slowly until the age of 2;0.

The low language level group’s curve stayed below a mean of 10 percent at the age of 0;9 and 1;0, and then increased with age. At the age of 1;9 the high and low language level groups converged around a mean of 55 percent, 58 and 54 percent in the high and low language level group respectively.

5.3.1 Total amount of JA and vocabulary

To test the relationship between the total amount of JA at the observed age points and vocabulary size at the age of 4;0, linear regression analyses were conducted for each age point.

The results showed a significant positive relationship between the total amount of JA and vocabulary size at the age of 4;0 at the age points 1;0 ($F(1,12) = 7.955; p < .05; R^2 = .3987$) and 1;6 ($F(1,12) = 9.52; p < .01; R^2 = .4424$), as shown in figure 5.
5.3.2 Total amount of JA and syntactic level

To test the relationship between the total amount of JA at the observed age points and syntactic level at the age of 3;0, linear regression analyses were conducted for each age point. However, no statistically significant relationships were found.

5.4 Child-initiated JA and language measures

The third research question addressed whether the amount of child-initiated JA at the observed age points predicted vocabulary size at the age of 4;0 and syntactic level at the age of 3;0. The amount of child-initiated JA was calculated for each dyad and age point (Appendix 2).

The development of child-initiated JA in the high and low language level groups (figure 6) showed a somewhat different pattern than the total amount of JA.

The high language level group had more time in child-initiated JA at all age points, though the difference was small at the age of 0;9, where there were almost no child-initiations. However, at 1;0 months the high language level group spent on average approximately 7 percent of the time in child-initiated JA, while the low language level group had an average of about 1 percent.

At 1;3 both groups spent an average of approximately 10 percent of the time in child-initiated JA (12 percent in the high group and 9 percent in the low group). The low language level group then increased steadily, almost doubling the time in child-initiated JA between the age of 1;3 and 1;9, and increased more rapidly between the age of 1;9 and 2;0.
Figure 6. The figure shows the development of the mean amount of time spent in child-initiated JA at the six age points for the children with high (n=7) and low (n=7) language level. The error bars show the standard error.

The high level group had a steady increase between the age of 0;9 and 1;3, and then the mean amount of child-initiated JA increased with over 20 percent until the age of 1;9. The curve leveled out between the age of 1;9 and 2;0 months.

The difference between the groups was largest at the age of 1;9, and then the groups converge slightly at the age of 2;0.

5.4.1 Amount of child-initiated JA and vocabulary

To test the relationship between the amount of child-initiated JA (CJA) at the observed age points and vocabulary size at the age of 4;0, linear regression analyses were conducted for each age point.

The results showed significant relationships between the amount of child-initiated JA and vocabulary score at 4;0 at the age points 1;0 ($F(1,12) = 4.915; p < .05; R^2 = .2906$), 1;6 ($F(1,12) = 8.481; p < .05; R^2 = .4141$), 1;9 ($F(1,12) = 28.09; p < .001; R^2 = .7007$) and 2;0 ($F(1,12) = 4.916; p < .05; R^2 = .2906$).

The relationship was strongest at the age of 1;9, where the variance in the amount of child-initiated JA explained 70 percent of the variance in vocabulary score at the age of 4;0. The results of the regression analyses are presented in figure 7.
Figure 7. The figure shows the relationship between amount of child-initiated JA and vocabulary (SCDI-III) at the age of 4;0 at the points 1;0 ($F(1,12) = 4.915; p < .05; R^2 = .2906$), 1;6 ($F(1,12) = 8.481; p < .05; R^2 = .4141$), 1;9 ($F(1,12) = 28.09; p < .001; R^2 = .7007$) and 2;0 ($F(1,12) = 4.916; p < .05; R^2 = .2906$).

5.4.2 Amount of child-initiated JA and syntactic level

To test the relationship between the amount of child-initiated JA at the observed age points and syntactic level at the age of 3;0, linear regression analyses were conducted for each age point.

The results showed significant positive relationships between the amount of child-initiated JA and syntactic level at 3;0 at the age points 1;6 ($F(1,12) = 6.301; p < .05, R^2 = .3443$) and 1;9 ($F(1,12) = 11.29; p < .01; R^2 = .48$), as is shown in figure 8.
5.5 Age of onset of child-initiated JA and language measures

The fourth research question addressed whether children who begin initiating JA episodes early have a larger vocabulary at the age of 4;0 and a higher syntactic level at the age of 3;0.

The age of onset of child-initiated JA was recorded as the age at which the child first successfully initiated a JA episode. Two children began initiating JA at the age of 0;9, six began at the age of 1;0 and five at the age of 1;3.

To test the relationship between age of onset of child-initiated JA and later language ability, two linear regression analyses were conducted (one for each language measure).

Figure 8. The figure shows the relationship between the amount of child-initiated JA and syntactic level at the age of 3;0, measured as unified predicate score (UP), at the age points 1;6 (F(1,12) = 6.301; p <.05, $R^2 = .3443$) and 1;9 (F(1,12) = 11.29; p <.01; $R^2 = .48$).

Figure 9. The figure shows the relationships between the age of onset of child-initiated JA and vocabulary
The results showed a significant negative relationship between the age of onset of child-initiated JA and vocabulary at the age of 4;0 \((F(1,12) = 11.89; p < .01; R^2 = .4976)\) and syntactic level at the age of 3;0 \((F(1,12) = 9.277; p < .05; R^2 = .436)\), as showed in figure 9. This means that children who learn to successfully initiate JA early have a larger vocabulary at the age of 4;0 and higher syntactic level at the age of 3;0.
6. Discussion

In this chapter, the method and results are discussed in relation to previous research.

6.1 Method discussion

In this section, the method of the present study is discussed in relation to the concepts of validity, reliability and generalizability.

6.1.1 Validity

The material used in this study comes from the MINT Project, where parents and children were recorded playing in a lab environment. While the recording situation is different from a recording set in the participants’ home, the instructions given to the parents were to play as they normally would. Since JA is a basic human ability, a cornerstone of human interaction, there is no reason to believe that the recording situation affected the validity of the JA measures (see section 6.1.3 for further discussion of the recording situation).

The recordings were made every third month, which affects the measurement of the age of onset of child-initiated JA. Three months is a long time from a developmental perspective, and this spread means that the age of onset measurement is not very precise. A tighter interval of recordings would have resulted in a more accurate measurement of when the children began to initiate JA with a parent.

The definition of JA used in this study was based on the definition given in Tomasello and Farrar (1986), i.e. that one person must initiate interaction around an object, there must be at least 3 seconds of JA and the child has to overtly show that they are aware of the JA. However, defining JA this way implies that it is possible to understand from watching an interaction that someone (especially the child) understands that the attention is in fact joint. As Tomasello (1995) states, gaze is not a completely reliable way to determine if there is JA, since a child who is alternating their gaze between an object and a parent can simple be alternating the gaze, without understanding the adult’s attentional state. In accordance with Tomasello and Farrar (1986), any overt behavior directed at the other which made it clear that the child understood the JA state was accepted in this study, including gaze, gestures and comments about the object in focus. Including more behaviors than gaze raises the validity of the coding.

Furthermore, the opposite scenario was of course also possible, i.e. that the child was aware of the JA without showing it overtly, which would then fall outside of the definition used in this study and not be coded as JA.

6.1.2 Reliability

CDIs (and parental questionnaires in general) are common tools used to assess vocabulary development and are both cost efficient and easy to administer. Although CDIs rely on parents’ reports of their own children’s linguistic abilities, which can be over- or underestimated, CDIs are generally considered reliable assessments of child vocabulary (Law & Roy, 2008). The vocabulary measure used in this study was SCDI-III, a Swedish version of the MacArthur-Bates Development Inventories (Fenson et al., 1993), adapted for Swedish children aged 2;6-4;0 (Eriksson, 2017). SCDI-III consists of several parts, and the vocabulary list of 100 items is the only part used in this study. The vocabulary checklist has been validated by Eriksson (2017), who found it to be a reliable measure of vocabulary that was not impacted by socio-economic status and did not exhibit any ceiling effects.

As a measure of syntactic level at the age of 3;0, an analysis of clauses containing unified predicates (Berman & Slobin, 1994) was conducted on the children’s utterances from the recording session at that age point following Tonér and Gerholm (under review). This measure is based on what the
children actually produced in the session at the age of 3;0 and does therefore not rely on the parents’ judgement of the children’s syntactic abilities. The children’s utterances were analyzed in relation to a target sentence, which was an assessment of what the child was trying to say, and each unified predicate clause was given points based on number of words, inflections, word order and functionality. This resulted in a percentage of unified predicate clauses which had the full score, i.e. were syntactically well-formed.

The assessment of the target sentences sometimes posed problems, since it was not always clear what the child was trying to say. However, this should not have affected the percentage of well-formed clauses, since the analysis was based on whether the clauses were grammatical or not. Target sentences that could not be understood were not considered to be functional, and therefore not well-formed.

The coding of JA was made by the author of this thesis, and followed the definition given in Tomasello and Farrar (1986). A test of interrater agreement would have improved the reliability of the coding, but this was not done due to time limitations. However, this definition of JA has been used in several previous studies (Tomasello & Todd, 1986; Tomasello & Farrar, 1986; Loy et al., 2018; Zampini et al., 2015).

6.1.3 Generalizability

There are certain aspects of the present study which could have an impact on the generalizability of the results. The sample size is small, including only 14 children recorded for approximately ten minutes at six age points, resulting in approximately 14 hours of material. Ten minutes every third month is a very small portion of the children’s lives, which has implications for the generalizability of the study.

As is common in research, the participants in this study are not entirely representative when it comes to socio-economic status (SES). Most of the families are middle class and earn 400 000-800 000 SEK per year. Furthermore, all children except two have at least one parent with a university education, and all parents have completed high school. However, the sample still shows a large variance in both vocabulary score (ranging between 36 and 96 items) and syntactic level (ranging between 36 and 91 percent well-formed clauses). It should also be mentioned that most of the children had started preschool by the age of 1;8, and all had started by the age of 2;3, which can have a mediating effect on the language used at home.

It can also be discussed whether the recording situation constitutes a representative sample of the children’s normal days, JA-wise. Being recorded by visible cameras is not part of normal interaction, which could affect the generalizability of the study. However, the MINT Project is a longitudinal study, meaning that the participants can get used to the environment. Furthermore, the recordings were made in a small room, which was equipped with age-appropriate toys, where the child and the parent played undisturbed. This might not be a completely fair representation of the everyday play sessions these families take part in, since there were no distractions and the context for the interaction was so small. This could have affected the amount of JA that occurred during the sessions, since these are good conditions for establishing JA. This point is also made by Carpenter and colleagues (1998) when discussing the generalizability of recordings made in lab environments. However, the children’s basic abilities to establish and maintain JA with an interaction partner is not affected by the recording situation, and the setup was the same for all dyads and age points. This means that even though these sessions might have a larger percentage of JA than there is during a regular day, the numbers should be relative to each dyad’s abilities, i.e. dyads who normally have more JA are likely to have more in the recorded sessions, while dyads who have less JA are likely to have less in the recorded sessions. Despite the fact that the recording situation in itself presents good conditions for JA, there was still a lot of variation between the dyads and age points.
6.2 Result discussion

In this section, the results of the present study are discussed in relation to previous research, and ideas for future research are presented.

6.2.1 Strategies for initiating JA

The first research question addressed how JA is initiated in parent-child interaction, and how this changes over time. Concerning the strategies to initiate JA, the results showed a clear preference by both parents and children to use a combination of a vocalization (most often speech) and a gesture to start JA episodes (see table 2 and 3, section 5.2.3). This is in line with Clark (2018), who states that adults frequently use gestures, speech and gaze to achieve JA.

For the parents, vocalizations alone were the second most common, followed by gesture alone. For the children, gesture alone was more common than vocalization alone except at 1;9 and 2;0, where speech alone was more common than gesture alone. At the age of 2;0, the distribution of initiation strategies was similar for the parents and the children, indicating that the children are developing toward a more adult-like behavior.

Touch and gaze were sparsely used by both parties in the initiation of JA in the present study, though touch can be used to attract an infant’s attention (Botero, 2016). However, Suarez and colleagues (2019) found that parents often add behaviors such as touching the object in focus during JA. It is possible that while touch was sparsely used to initiate JA, it might be used to maintain it.

The fact that gaze was not used much in the initiation of JA episodes was surprising, considering the amount of previous research emphasizing the importance of gaze following abilities for JA (e.g. Corkum & Moore, 1995; Brooks & Meltzoff, 2005; Tomasello et al., 2007). However, Yu and Smith (2013) found that children, and to a slightly lesser extent adults, prefer to follow hand movements over eye movements in contexts with several objects, since gaze becomes unreliable in those types of situations. This could explain why the strategy of using vocalization plus gesture was so common in the material, since the recordings are made in a quite small room with many different toys. Using speech and gestures would then be a more reliable way to direct attention to an object, since gaze would be harder to follow.

Furthermore, gaze following is likely to be more cognitively complex than e.g. following pointing, since it is only around the age of 1;6 that children rely on eye-orientation over head-orientation to determine gaze direction (Corkum & Moore, 1995). This could also explain the initiation strategies used. Assuming, in line with the argument made by Clark (2018) that both the parent and the child want to be understood as easily as possible and make the communication as efficient as possible (relying on the Cooperative Principle formulated by Grice, 1975), it seems unlikely that they would use such a subtle strategy as e.g. gaze pointing alone in a situation where it is not necessary. The vocalization-plus-gesture strategy makes the object in focus less ambiguous than both vocalization alone and gesture alone.

6.2.2 Number of JA episodes

The results of the present study showed that the average number of JA episodes increased with age, both in the high and low language level groups. However, apart from at the age of 0;9, the high language level group had on average more JA episodes than the low language level group. The high level group reached a mean of over two JA episodes per minute at the age of 1;3, which then increased further. The low level group, however, reached two JA episodes per minute at the age of 1;9, six months later than the high group, and stayed just over a mean of two episodes per minute. The high language level group thus managed to achieve JA more often than the low-level group. However, the present study did not take failed initiations into consideration and it is therefore not clear why this is. It might be the case that the high language level group simply initiated more episodes, or both the parents and the children in that group were better at or more willing to responding to JA bids made by their conversation partner, making them more successful in establishing JA whenever there was an initiation attempt.
Morales and colleagues (2000) found that responding to JA bids made by a conversation partner at the age point 0;6, 0;8, 0;10, 1;0 and 1;6 was positively correlated with vocabulary. With regard to the present study, this could mean that the high language level group was indeed better at establishing JA when an initiation was made. However, this could also have been combined with a larger number of initiations, both successful and unsuccessful. Furthermore, Bigelow (2003) suggests that JA and language help each other along during development. It could be the case that the dyads which included children with a higher language level were able to take advantage of their language skills to establish JA more often, which then advanced their language development further. Thus, in accordance with Bigelow’s (2003) argument, JA and language scaffold each other.

The average number of parent- and child-initiated JA episodes differed between the high and low language groups, and the groups showed slightly different developmental patterns. In the high language level group, the number of parent-initiated episodes increased with age and leveled out at the age of 1;3, and then decreased a bit. The mean number of child-initiated episodes in the high-level group also increased with age, and at the age of 1;9 the mean number of child-initiated episodes surpassed the number of parent-initiated ones. At the age of 2;0, the number of parent and child-initiated JA episodes were about the same in the high-level group, meaning that the parents and the children initiated around the same mean number of episodes (approximately 1.3 per minute, respectively).

In the low language level group, the amount of parent-initiated episodes followed a similar pattern as in the high-level group. The average number of parent-initiated JA episodes increased with age until the age of 1;3, and then leveled out. However, in contrast with the high language level group the children initiated consistently fewer initiated episodes than the parents, and the mean number of child-initiated episodes in the low language level group never reached the level of the parent-initiated episodes.

This indicates that the children in the high language level group take more responsibility for the interaction than the children in the low language level group, which is in accordance with the findings by Tomasello and Todd (1983), that JA in which both parties lead the interaction was positively correlated with vocabulary size.

6.2.3 Time spent in JA and length of JA episodes

The mean amount of time spent in (total) JA, parent-initiated JA and child-initiated JA was also calculated for both the high and low language level groups. The high language level group spent on average more time in JA at all age points, but the low language level group approached the same mean total amount as high level group at the age of 1;9 and 2;0. This can be discussed in relation to the number of JA episodes in the groups. The high language level group had on average more and longer JA episodes between the age of 0;9 and 1;6 than the low language level group. At the age of 1;9 and 2;0 the high language level group had more, and on average shorter, JA episodes than the low language level group, yet the mean total time spent in JA was similar between the groups at the two latest age points (1;9 and 2;0). When investigating the mean length of JA episodes, JA episodes were indeed found to be longer in the low language level group at these ages (Table 1, section 5.2.2).

Furthermore, the developmental patterns for the amount of time spent in parent and child-initiated JA respectively differed between the two groups. In the high language level group, the majority of time was spent in parent-initiated JA between the age of 0;9 and 1;6. At 1;9, however, the high language level group spent more time in child-initiated JA, and at 2;0 there was an equal amount of parent- and child-initiated JA. In contrast, the low language level group spent more time in parent-initiated JA than child-initiated JA across all age points. Adamson and McArthur (1995) state that JA skills consolidate between the age of 1;3 and 1;6, which could explain the fact that the high language level group has a similar mean amount of child- and parent-initiated JA after that age. The children in the low language level group will likely reach this stage at a later age, since JA abilities develop at least through the first three years of life (Adamson et al., 2004).

Furthermore, the results showed that at the age of 1;9 and 2;0 the low-level group had approximately six seconds longer parent-initiated episodes than the high language level group did at these ages.
which might be too early to have en- 
mutual attention. This is in contrast with Rollins and Snow (1998) who found a positive relation- 
s. However, the results did not show any significant relationships between these two variables.

The total amount of
considering development. The results at the later age points in the present study could p-

Loy and colleagues (2018) found that responding- 

The developmental pattern for the total amount of JA was presented for the high and low language 
level groups respectively (figure 4, section 5.3). This showed that the amount of JA increased over 
time and that the high language level group had more time in JA at all age points. However, the groups 
converged at the age of 1:9. According to Adamson and McArthur (1995), JA skills consolidate 
between the age of 1:1 and 1:6, which might explain why the difference between the average amount 
of JA in the high and low language level groups seems to disappear after the age of 1:6.

A number of previous studies have linked the amount of JA during infancy with vocabulary 
development, both in typically developed children (Tomasello & Todd, 1983; Tomasello & Farrar, 1986; Morales et al., 2000; Loy et al., 2018; Hirotani et al., 2009), children with Down syndrome 
(Zampini et al., 2015), and children with autism spectrum disorder and developmental delay 
(Adamson et al., 2019).

A possible reason for this is discussed by Tomasello and Farrar (1986), who compared a number of 
language measures (e.g. amount of utterances, mean length of utterance, utterance types, amount of 
words, etc.) inside and outside of JA episodes, and found that during JA there was more talk and 
longer conversations. Furthermore, mothers used shorter sentences and made more comments about 
the objects in focus during JA. Based on this, they argue that JA provides a good learning context for 
presenting new words, which could explain why JA has been shown to have a positive impact on 
vocabulary development. Bakeman and Adamson (1984) argue along the same line, suggesting that JA 
provides a social context for referential communication.

The positive effect of JA on vocabulary has also been found in an EEG-study, where words that had 
been learned in a JAal setting elicited an N400-like response when presented along with an 
incongruent picture. Thus, JA was suggested to facilitate a deeper semantic processing of words, 
which strengthens word learning (Hirotani et al., 2009).

Thus, the results of the present study are in line with previous research showing a positive relationship 
between the amount of joint attention in interaction and vocabulary development; linear regression 
analyses showed that the total amount of JA at the age of 1:0 and 1:6 was positively related to 
vocabulary size at the age of 4:0, with the strongest relationship present at the age of 1:6. Similarly, 
Loy and colleagues (2018) found that JA at 1:1 and 1:5 predicted vocabulary size at the age of 1:5.

The average amount of JA was very similar in the high and low language level groups at 1:9 and 2:0, 
which explains why there were no relationships with vocabulary at those age points. Morales and 
colleagues (2000) found that responding to JA at the age of 1:9 and 2:0 was not related to vocabulary 
development. The results at the later age points in the present study could possibly be explained 
considering this.

The total amount of JA was also tested against syntactic level at the age of 3:0, using linear regression 
analyses. However, the results did not show any significant relationships between these two variables. 
This is in contrast with Rollins and Snow (1998) who found a positive relationship between regulating 
mutual attention and grammar development, both in typically developed children and in children with 
autism spectrum disorder. The syntactic level in the present study was measured at the age of 3:0, 
which might be too early to have enough variance to result in significant effects.

6.2.4 Total amount of JA and language acquisition

The second research question addressed whether the total amount of JA at the observed age points is 
related to vocabulary size at the age of 4:0 and syntactic level at the age of 3:0 respectively.

A number of previous studies have linked the amount of JA during infancy with vocabulary 
development, both in typically developed children (Tomasello & Todd, 1983; Tomasello & Farrar, 
1986; Morales et al., 2000; Loy et al., 2018; Hirotani et al., 2009), children with Down syndrome 
(Zampini et al., 2015), and children with autism spectrum disorder and developmental delay 
(Adamson et al., 2019).

A possible reason for this is discussed by Tomasello and Farrar (1986), who compared a number of 
language measures (e.g. amount of utterances, mean length of utterance, utterance types, amount of 
words, etc.) inside and outside of JA episodes, and found that during JA there was more talk and 
longer conversations. Furthermore, mothers used shorter sentences and made more comments about 
the objects in focus during JA. Based on this, they argue that JA provides a good learning context for 
presenting new words, which could explain why JA has been shown to have a positive impact on 
vocabulary development. Bakeman and Adamson (1984) argue along the same line, suggesting that JA 
provides a social context for referential communication.

The positive effect of JA on vocabulary has also been found in an EEG-study, where words that had 
been learned in a JAal setting elicited an N400-like response when presented along with an 
incongruent picture. Thus, JA was suggested to facilitate a deeper semantic processing of words, 
which strengthens word learning (Hirotani et al., 2009).

Thus, the results of the present study are in line with previous research showing a positive relationship 
between the amount of joint attention in interaction and vocabulary development; linear regression 
analyses showed that the total amount of JA at the age of 1:0 and 1:6 was positively related to 
vocabulary size at the age of 4:0, with the strongest relationship present at the age of 1:6. Similarly, 
Loy and colleagues (2018) found that JA at 1:1 and 1:5 predicted vocabulary size at the age of 1:5.

The average amount of JA was very similar in the high and low language level groups at 1:9 and 2:0, 
which explains why there were no relationships with vocabulary at those age points. Morales and 
colleagues (2000) found that responding to JA at the age of 1:9 and 2:0 was not related to vocabulary 
development. The results at the later age points in the present study could possibly be explained 
considering this.

The total amount of JA was also tested against syntactic level at the age of 3:0, using linear regression 
analyses. However, the results did not show any significant relationships between these two variables. 
This is in contrast with Rollins and Snow (1998) who found a positive relationship between regulating 
mutual attention and grammar development, both in typically developed children and in children with 
autism spectrum disorder. The syntactic level in the present study was measured at the age of 3:0, 
which might be too early to have enough variance to result in significant effects.
6.2.5 Child-initiated JA and language acquisition

The third research question addressed whether there is a relationship between the amount of child-initiated JA at the observed age points and vocabulary size at the age of 4;0 and syntactic level at the age of 3;0 respectively.

The development of child-initiated JA in the high and low language level groups showed that the high-level group had on average more child-initiated JA than the low language level group (see figure 6, section 5.4). However, in contrast with the development of the total amount of JA, the difference between the groups grew after the age of 1;3 and was largest at the age of 1;9, i.e. where the groups converged in total amount of JA. This indicates that child-initiated JA follows a different developmental course.

As previously mentioned, JA skills consolidate between the age of 1;1 and 1;6 (Adamson & McArthur, 1995), yet the results of this study showed a big difference in child-initiated JA between the high and low language level groups after the age of 1;6. The amount of child-initiated JA represents the child’s ability to direct the parent’s attention to an outside object or event, and generate interaction around it, which Carpenter and colleagues (1998) claim is the last and most complex stage in JA development. Thus, since JA skills have been linked with language development, it appears plausible that the children who master this developmental step to a higher degree also are further along in their language development, in regard to both vocabulary and syntax.

It should be pointed out, however, that child-initiated JA, measured in this study as percentage of file time, shows both children’s ability to successfully direct another’s attention and lead the interaction, and the parents’ willingness to follow and lead the children. It has been found that interaction where parents frequently redirect children’s attention is negatively correlated to vocabulary size, while following the child’s attention is positively correlated with vocabulary (Tomasello & Todd, 1983). Since JA by definition is a joint endeavor, parents have an important role in the establishing of JA, even when they are not the ones initiating it.

The results of the present study showed that there were significant positive relationships between vocabulary and child-initiated JA at the age points 1;0, 1;6, 1;9 and 2;0. This is in line with previous research, which has found a positive relationship between vocabulary and JA where mothers referred to objects that were in the children’s focus (Tomasello & Farrar, 1986). Loy and colleagues (2018) also found a positive correlation between the amount of JA initiatives including both an object and the conversation partner made by children at the age of 1;1 and 1;5 and vocabulary size at 1;5.

The relationship between amount of child-initiated JA and vocabulary size was strongest at the age of 1;9, where the $R^2$ was approximately 0.7, meaning that the amount of child-initiated JA explained 70 percent of the variance in vocabulary size. In contrast to this, Morales and colleagues (2000) found that responding to JA was not related to vocabulary development at the age of 1;9 and 2;0. However, the present study shows that the amount of child-initiated JA, but not total amount of JA, is related to vocabulary size even at these ages, and especially at the age of 1;9. Tomasello and Farrar (1986) found the biggest difference in language use between the inside and outside of JA episodes at the age of 1;9, which could explain why the relationship between the amount of child-initiated JA and vocabulary in the present study is strongest at that age.

The results of the present study further showed that the amount of child-initiated JA at the age of 1;6 and 1;9 was positively related to syntactic level at the age of 3;0, measured by unified predicate score, with the strongest relationship present at the age of 1;9.

Few studies have looked into the relationship between JA and syntactic development, since the focus of the relationship between JA and language acquisition has been on vocabulary development. However, Roth (1987) notes that mothers more often elaborate semantically on their children’s utterances when they follow the child’s focus, which could possibly offer a perspective on the role of child-initiated JA in syntactic development. Clark (2018) also describes how adults co-construct conversations with children by elaborating on and framing their utterances, and that this helps children move from one-word utterance to simple constructions. Assuming that this indeed happens more often
in JA, which is supported by Tomasello and Farrar (1986), it is likely that this provides a good setting for learning syntactic aspects of language.

Rollins and Snow (1998) also connected JA with grammar development. Investigating pragmatic development in 30 typically developed children, they found that regulating mutual attention (i.e. establishing and maintaining JA) together with parental conversation style at the age of 1:2 predicted grammar development at the age of 2:7 in typically developing children. When investigating the same thing in children with autism spectrum disorder, they found that only the ability to establish and maintain JA predicted grammar development. Based on these results, Rollins and Snow argue that JA is essential for grammar development. The results of the present study support this argument and indicate that it is the child’s ability to lead the interaction that is the most important for this aspect of language acquisition.

Tomasello (1995) states that between the age of 1:6 and 2:0, children’s JA abilities are shown in their use of language. In light of this, it is noteworthy that the statistically significant relationships between child-initiated JA and syntactic level are present at the age of 1:6 and 1:9, since it is around this time that children start combining words into simple constructions (1995: 117-118). It is possible, since JA and language help each other along (Bigelow, 2003), that children who are further along in their social-cognitive development, and thus are better at directing attention, use this to direct attention to an object or event that they are interested in. Thus, they do not have to waste cognitive resources trying to figure out what the adult is referring to, since they themselves can decide on the object of conversation and thus be certain that the adult’s utterances refer to that. This might then free up cognitive resources to focus on building sentences, furthering the syntactic development. This argument is supported by Adamson and colleagues (2004), who found that at the age of 1:6, children who are early and late in language development are significantly different when it comes to the use of symbol-infused JA, i.e. using language and gestures within JA states.

### 6.2.6 Age of onset of child-initiated JA and language acquisition

The fourth research question addressed whether children who start to initiate JA early have a larger vocabulary at the age of 4:0 and a higher syntactic level at the age of 3:0.

The results showed positive relationships between the age of onset of child-initiated JA and vocabulary and syntactic level respectively, meaning that children who learn to direct others’ attention early on are further along in their language development at the later age points.

According to Carpenter and colleagues (1998), the development of JA skills follows a specific pattern. First, sharing attention without an initiator occurs, then the ability to follow attention develops, and, lastly the ability to direct others’ attention. These developmental steps are seen as increasingly cognitively complex. All dyads in the present study except two engaged in parent-initiated JA before child-initiated (see appendix 1 and 2). The remaining two started initiating at the age of 0:9, which was the earliest age included in this study, and thus had both parent- and child-initiated JA at that age.

Since JA skills have been linked with language development, it appears plausible that the children who master the most complex developmental step early also are further along in their language development, in regard to both vocabulary and syntax. According to Mundy and colleagues (2007), initiating JA is related to a part of the brain which controls intentional attention, which they suggest is linked with the ability to understand others as intentional agents. When children begin using pointing gestures towards the end of the first year of life (Camaioni et al., 2004), declarative pointing, i.e. directing attention simply to share information, develops after imperative pointing. Declarative pointing, but not imperative, has further been found to predict later language level (Colomnesi et al., 2010) and theory of mind (Sodian & Kristen-Antonow, 2015; Nelson et al., 2008). Rollins and Snow (1998) also suggest that an instrumental use of language, which imperative pointing is an example of, is not related to grammar development. Since the earliest child initiations of JA in the present study was by gestures, the children in this study who initiate JA episodes at an earlier age might also master declarative pointing earlier. To point declaratively, children must understand that their communication partner’s attention will be affected by their gesturing, i.e. they must understand something of their intentional state. This understanding is essential for initiating JA.

30
6.2.7 General discussion

Children learn language through interaction (Clark, 2018; Kuhl et al., 2003; Roseberry et al., 2014; Roseberry Lytle et al., 2018). To help children navigate their linguistic environment, Tomasello (2003) theorizes that children are equipped with two sets of skills: intention-reading and pattern finding. JA skills are included in the intention-reading skill set, and the ability to establish JA can be viewed as a basis for understanding others as intentional agents (Tomasello, 2003; 1995). The assumption that others have intentions that motivate their actions and language use means that the context in which language is used becomes meaningful, which in turn helps children understand others’ utterances during their acquisition of language (Tomasello, 2003).

Language development and social-cognitive development go hand in hand, and have been suggested help each other along (Bigelow, 2003) and JA, specifically, has been suggested to be necessary for language (Racine et al., 2014). This is supported by the results of previous research, as well as the present study, which has connected JA with language acquisition, both vocabulary and syntax.

The results of this study indicate that the amount of JA, and child-initiated JA in particular, during the second half of the second year of life seems especially important for subsequent language development. Tomasello (1995) points out that it is during this time, i.e. between the age of 1;6 and 2;0, that children start using constructions with a topic-comment structure, either by using a combination of a word and gesture or two words. He further argues that this is the result of the development of the understanding that they can affect their communication partner’s attention by gesturing or speaking, and then comment on the object in focus.

This has implications for the interpretation of the results concerning child-initiated JA in the present study. Child-initiated JA was related to later vocabulary size at more age points than the total amount of JA, and the amount of child-initiated JA, but not the total amount of JA, was related to syntactic level. This indicates that children’s ability to initiate JA, i.e. direct the communication partner’s attention and initiate interaction around it, might be a better predictor of later language abilities than total amount of JA.

A similar argument is made by Kwisthout and colleagues (2008), who used Carpenter and colleagues’ (1998) JA steps to test which of several models was better at developing a vocabulary during a language game. They found that a model which has all three steps (checking, following and directing attention) performed the best, and argue that this is because the context in which a referential expression is used is narrowed down when all these are in place, which makes it easier for the listener to understand what the speaker is referring to. This could also explain the findings of the present study, i.e. that there seems to be a strong relationship between child-initiated JA and vocabulary and syntax respectively.

6.3 Ideas for future research

The present study has only taken successfully initiated JA episodes into consideration. However, it would be interesting to also take the failed initiation attempts into account, to see whether there is a difference regarding this in the high and low language level groups.

Furthermore, it would be interesting to investigate how JA is maintained by parents and children, and if this also changes over time.

Previous research about JA and language acquisition has been largely focused on vocabulary development, and the present study along with the results of Rollins and Snow (1998) found that JA also affects syntactic development. Hence, future research about JA and child language acquisition should take more aspects of language into account.
7. Conclusions

The present study investigated JA in parent-child interaction at the age points 0;9, 1;0, 1;3, 1;6, 1;9 and 2;0, focusing on how it is initiated and whether it has an effect on later vocabulary size and syntactic level.

The results showed that the strategies used by the parents for initiating JA were primarily, in order of frequency, vocalization plus gesture, vocalization alone and gesture alone.

The initiation strategies used by the children were, in order of frequency, vocalization plus gesture, gesture alone and vocalization alone. However, at the two latest age points the order of frequency was the same for the parents (V+Ge, V and Ge).

There was a difference in the mean amount of JA episodes, mean amount of time spent in JA and mean length of JA episodes between the high and low language level groups. The high language level group had on average more JA over all, both time and number of episodes, and more time in parent-initiated JA between the age of 0;9 and 1;6, after which there was a more equal distribution of parent and child-initiated JA, both in time and number of episodes. The low language level group had on average less time in JA over all, more time in parent-initiated than child-initiated JA across all age points, and longer parent-initiated JA episodes at the age of 1;9 and 2;0, possibly to compensate for the lack of child-initiated JA.

The total amount of JA at the age of 1;0 and 1;6 was positively related to vocabulary size at the age of 4;0. However, there was no significant relationship between the total amount of JA and syntactic level at the age of 3;0.

The amount of child-initiated JA at the age of 1;0, 1;6, 1;9 and 2;0 was positively related to vocabulary size at the age of 4;0. The amount of child-initiated JA at the age of 1;6 and 1;9 was positively related to syntactic level at the age of 3;0.

The age of onset of child-initiated JA was positively related to both vocabulary size at the age of 4;0 and syntactic level at the age of 3;0.

The results of the present study support previous research claiming that JA is important for vocabulary development and suggest that it also plays a role in syntactic development. Furthermore, time spent in child-initiated JA seems to be more important than the total amount of JA. The results further add to the research concerning how JA is established in parent-child interaction.
References


## Appendix 1. Individual JA values

Table 4. Total amount of JA for each dyad and age point, in percent.

<table>
<thead>
<tr>
<th></th>
<th>0;9</th>
<th>1;0</th>
<th>1;3</th>
<th>1;6</th>
<th>1;9</th>
<th>2;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>7M0Z</td>
<td>35</td>
<td>35.3</td>
<td>18.3</td>
<td>53.7</td>
<td>56.5</td>
<td>66.1</td>
</tr>
<tr>
<td>8M0Z</td>
<td>0</td>
<td>4.6</td>
<td>19.3</td>
<td>13.2</td>
<td>74.5</td>
<td>72.2</td>
</tr>
<tr>
<td>0M1Z</td>
<td>14.7</td>
<td>10.2</td>
<td>7.4</td>
<td>28.0</td>
<td>19.3</td>
<td>18.0</td>
</tr>
<tr>
<td>1M1Z</td>
<td>9.8</td>
<td>9.9</td>
<td>30.7</td>
<td>32.6</td>
<td>50.7</td>
<td>64.0</td>
</tr>
<tr>
<td>5M1Z</td>
<td>3.7</td>
<td>19.2</td>
<td>39.0</td>
<td>56.5</td>
<td>55.3</td>
<td>64.1</td>
</tr>
<tr>
<td>1M2Z</td>
<td>6.6</td>
<td>18.6</td>
<td>55.2</td>
<td>58.9</td>
<td>78.3</td>
<td>62.3</td>
</tr>
<tr>
<td>2M2Z</td>
<td>10.4</td>
<td>30.7</td>
<td>45.1</td>
<td>44.8</td>
<td>41.4</td>
<td>62.2</td>
</tr>
<tr>
<td>4M2Z</td>
<td>10.5</td>
<td>20.0</td>
<td>35.2</td>
<td>66.6</td>
<td>54.5</td>
<td>67.6</td>
</tr>
<tr>
<td>5M2Z</td>
<td>7.5</td>
<td>10.3</td>
<td>41.2</td>
<td>64.5</td>
<td>49.9</td>
<td>83.7</td>
</tr>
<tr>
<td>6M2Z</td>
<td>2.0</td>
<td>11.7</td>
<td>23.8</td>
<td>38.4</td>
<td>47.0</td>
<td>68.4</td>
</tr>
<tr>
<td>7M2Z</td>
<td>13.0</td>
<td>12.3</td>
<td>51.8</td>
<td>51.2</td>
<td>73.7</td>
<td>71.0</td>
</tr>
<tr>
<td>1M3Z</td>
<td>16.6</td>
<td>24.4</td>
<td>61.8</td>
<td>51.0</td>
<td>53.9</td>
<td>69.0</td>
</tr>
<tr>
<td>1M4Z</td>
<td>8.4</td>
<td>3.8</td>
<td>43.9</td>
<td>32.7</td>
<td>66.0</td>
<td>83.0</td>
</tr>
<tr>
<td>5M4Z</td>
<td>6.9</td>
<td>36.2</td>
<td>49.7</td>
<td>74.5</td>
<td>66.3</td>
<td>79.9</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
<td><strong>10.3</strong></td>
<td><strong>17.7</strong></td>
<td><strong>37.3</strong></td>
<td><strong>47.6</strong></td>
<td><strong>56.2</strong></td>
<td><strong>66.5</strong></td>
</tr>
</tbody>
</table>
### Appendix 2. Individual child-initiated JA values

Table 5. Amount of child-initiated JA for each dyad and age point, in percent.

<table>
<thead>
<tr>
<th></th>
<th>0;9</th>
<th>1;0</th>
<th>1;3</th>
<th>1;6</th>
<th>1;9</th>
<th>2;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>7M0Z</td>
<td>0</td>
<td>8.9</td>
<td>2.3</td>
<td>21.5</td>
<td>27.2</td>
<td>43.1</td>
</tr>
<tr>
<td>8M0Z</td>
<td>0</td>
<td>0</td>
<td>9.3</td>
<td>2.6</td>
<td>16.7</td>
<td>2.6</td>
</tr>
<tr>
<td>0M1Z</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>15.2</td>
<td>0</td>
<td>2.8</td>
</tr>
<tr>
<td>1M1Z</td>
<td>0</td>
<td>0</td>
<td>7.1</td>
<td>9.6</td>
<td>12.9</td>
<td>18.4</td>
</tr>
<tr>
<td>5M1Z</td>
<td>0</td>
<td>1.4</td>
<td>14.9</td>
<td>16.9</td>
<td>30.2</td>
<td>20.9</td>
</tr>
<tr>
<td>1M2Z</td>
<td>0</td>
<td>1.2</td>
<td>5.8</td>
<td>19.0</td>
<td>41.9</td>
<td>23.6</td>
</tr>
<tr>
<td>2M2Z</td>
<td>0</td>
<td>6.5</td>
<td>18.9</td>
<td>26.3</td>
<td>22.4</td>
<td>28.5</td>
</tr>
<tr>
<td>4M2Z</td>
<td>2.8</td>
<td>15.7</td>
<td>12.9</td>
<td>34.1</td>
<td>52.6</td>
<td>43.6</td>
</tr>
<tr>
<td>5M2Z</td>
<td>0</td>
<td>0</td>
<td>10.2</td>
<td>11.5</td>
<td>23.1</td>
<td>30.3</td>
</tr>
<tr>
<td>6M2Z</td>
<td>0</td>
<td>1.1</td>
<td>8.0</td>
<td>7.9</td>
<td>14.1</td>
<td>31.7</td>
</tr>
<tr>
<td>7M2Z</td>
<td>0</td>
<td>6.2</td>
<td>11.3</td>
<td>23.0</td>
<td>20.6</td>
<td>28.1</td>
</tr>
<tr>
<td>1M3Z</td>
<td>0</td>
<td>1.7</td>
<td>14.4</td>
<td>19.7</td>
<td>19.6</td>
<td>46.5</td>
</tr>
<tr>
<td>1M4Z</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>16.5</td>
<td>25.2</td>
<td>58.2</td>
</tr>
<tr>
<td>5M4Z</td>
<td>2.9</td>
<td>14.2</td>
<td>13.0</td>
<td>19.4</td>
<td>37.1</td>
<td>29.2</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>0.004</td>
<td>4.1</td>
<td>10.3</td>
<td>17.4</td>
<td>24.5</td>
<td>29.1</td>
</tr>
</tbody>
</table>