Noun Phrase Anaphora and Referential Behaviour in Child-Directed Speech During the Child's First Year

David Pagmar

Noun Phrase Anaphora and Referential Behaviour in Child-Directed Speech During the Child's First Year

David Pagmar

Abstract

"Anaphora" is a label used for a referential expression that connects one entity (e.g. a pronoun) to another previously established entity (e.g. a proper name). The previously established entity is called an antecedent. The use of anaphora will, in this study, be referred to as referential behaviour.

The study was based around audio and video recordings of free play between a Swedish parent and his/her child. 10 parents and their children were recorded. The referential behaviour of the parents was analysed. The sessions took place when the children were 3, 6, 9 and 12 months old. Recent studies indicate that speech directed at children during a child's first six months contains a larger amount of pronouns than the speech directed at children between 6 and 12 months of age.

The purpose of the study was to examine if the decline of pronouns was visible in Swedish child-directed speech, and to see how different types of anaphora appeared in the same speech. Correlations between the visible changes of different types of referential expressions were also examined. A drop in the use of anaphoric pronoun with an explicit antecedent was found for the last two ages, which confirmed the study's hypothesis. The results were also compared to each child's vocabulary development.

Keywords

Sammanfattning

"Anaforer" är ett begrepp som används för referentiella uttryck som kopplar samman en entitet (t ex ett pronomen) med en annan tidigare introducerad entitet (t ex ett egennamn). En tidigare introducerad enhet kallas för antecedent. Användningen av anaforer kommer i denna uppsats stundtals att benämnas som referentiellt beteende.

Studiematerialet bestod utav ljud- och videoupptagningar av fri lek mellan en svensktalande förälder och dess spädbarn. Totalt medverkade 10 föräldrar och deras spädbarn. Det referentiella beteendet hos föräldrarna analyserades. Inspelningssessionerna tog plats då barnen var 3, 6, 9 och 12 månader gamla.

Den senaste tiden har forskningen visat indikationer på att tal riktat mot ett spädbarn under dess första 6 månader innehåller fler pronomen än tal riktat mot barn mellan 6 och 12 månader.

Syftet för denna studie var att undersöka ifall dessa indikationer var synbara i svenskt barnriktat tal och för att undersöka hur olika typer av anaforiska uttryck återfanns i talet. Korrelationer mellan de påträffade förändringarna av referentiella uttryck undersöktes också. En minskning av anaforiska pronomen med explicit antecedent påträffades vid de två senare åldrarna, vilket bekräftade studiens hypotes. Resultaten jämfördes även med varje barns ordförrådsutveckling.

Nyckelord

Anaforer, anaforiska pronomen, barnriktat tal, referentiellt beteende

Contents

1 Introduction

"Reference" is a challenging term. When scholars have tried to explain the relationship between a sign in a language and a referent, the arguments often lead to unwanted consequences, such as ambiguity (Mey 2008:52). Still, in the simplest understanding of the term, one could pose the following definition:

> A sign in a language refers to a given entity <u>if and only if</u> the sign is directing attention to the entity itself (and not only its position)

To completely rely on this definition, we would also need to define "directing attention". However, this simple understanding of the term might shed some light on a key element of this thesis: third person pronouns.

Third person pronouns (he, she, they, it) are empty. They are general terms that have the potential to be assigned to one or more referents. Standing alone without any context, they provide little, or no, semantic content. In natural discourse they are given semantic content, in the form of referents. Consider the following conversation:

> - He, said X.
>
> - Who?, asked Y.
>
> - David, X replied.

The first pronoun will at most provide the fact that there is a referent and that the referent is male. In the last phrase, the proper name gives the first pronoun a referent. Referents can be assigned to pronouns in different ways. One can first introduce a proper name and then use a pronoun to point back to the referent of the proper name. Gestures and situational events can also assign referents to pronouns. In theoretical attempts to clarify this relationship, problems like ambiguity can make the analysis difficult. In natural discourse, however, language users can effortlessly understand the referents of pronouns.

The language acquisition of children is another challenging topic in linguistics, as well as in related sciences. There have been numerous studies on the necessary structures that the acquisition possibly is dependent on (Baron 1990; Kitamura & Brunham 1998; Kuhl 2014).

The term child-directed speech (CDS) is a generic term for the features of the speech that adults direct at children. High pitch, long vowels and short sentences are some of the features that have been said are characteristic for CDS (Fernald & Kuhl 1987). Recently, studies have shown indications of a pattern in the use of pronouns in CDS (Rhode & Frank 2011; Smith & Frank 2012). These studies show that there are more pronouns in the speech directed at 6-month-old infants, than in the speech directed at 12-month-old children. One might find these indications peculiar. The understanding of pronouns in natural discourse is seemingly not only dependent on the understanding of the relationship between a sign and a referent, but also of the structure of assigning referents to general terms.

Can the drop of pronouns also be found in Swedish CDS? What other types of changes, that are connected to referential behaviour, can be observed? Are there any potential correlations between these changes?

2 Background

Referential expressions, such as pronouns, are fundamental elements of spoken discourse. Different languages can execute referential expressions in different ways. "Anaphora" is the label used in linguistics and related sciences for a referential expression that connects one entity (e.g. a pronoun) to another previously established entity (e.g. a proper name). In Swedish, as in many other languages, it is very common for a noun phrase (NP) to follow a lexical or pragmatic entity (e.g. a proper name, a definite description or a non-verbal situational phenomenon or action), referring back to the same lexical or pragmatic entity. The NP then takes on the role of an anaphor and the preceding entity is classified as an explicit, or possibly an implicit, antecedent based on its semantic or pragmatic status. In this study, the use of anaphora will be referred to as referential behaviour.

Anaphora as a linguistic topic has been examined by many scholars and from a variety of different perspectives. Anaphora has been connected to the Accessibility Theory (Ariel 1990) and anaphora resolution is discussed in *Lectures on Government and Binding* (Chomsky 1981). Anaphora can be analysed both as a syntactic and a pragmatic phenomenon. If an anaphoric pronoun appears in a clause that directly follows the antecedent, the relationship between them can be approached from a syntactic perspective. If the anaphor appears instead with a larger gap between it and its antecedent, then the analysis will tend to take a more pragmatic approach. Other properties of anaphora are also manifested in pragmatic structures, e.g. certain anaphoric expressions can change while the referent is intact, and vice versa[1]. With the pragmatic nature of anaphora in mind, it will at times be referred to as a complex referential behaviour.

There are still a lot of questions about the necessary properties CDS and its importance for a child's language acquisition. There have been some studies on how semantic-pragmatic features of language, such as anaphora and pronominalization, appears in English CDS (Rohde & Frank 2011) and in Japanese CDS (Smith & Frank 2012). Very few of the earlier studies are concentrated on the development of these features in CDS over the child's first year. The purpose of this study is to examine anaphora and referential behaviour in Swedish CDS, and compare the results to these previous studies.

2.1 The Organisation of the Thesis

First, some general statements about the features of CDS are presented. The features can be connected to recent studies and some of these studies are presented in this section. The definition of CDS is formulated and then follows definitions that categorize different kinds of anaphora, repetitions, and reintroductions. The categories provide the foundation for the analysis of referential behaviour. After an explanation of the term "coreference", the question formulations and a hypothesis are stated.

In the method section, the gathering, the annotation and the analysis of the data are accounted for. The analytical process and the application of the definitions are exemplified with data from the study. The parents in this study answered a survey throughout the project. This survey is also presented in this section.

---

[1] "It" and "that one" can both be applied to the same referent. At the same time, "it" and "that one" might as well point out two different referents.

In the result section, the individual outcome for each test subject connected to each established category (the different kinds of anaphora, repetitions, and reintroductions) is presented. The results of the survey are also presented.

In the discussion section, the method and the results are commented on. The strengths and weaknesses of the method are stated and some implications of the results are presented. In the last section of the thesis, the following is stated: answers to the research questions, suggestions for future studies, and a concluding remark.

2.2 Child-Directed Speech

The fact that CDS entails features that distinguish it from speech directed at adults should perhaps be held as an assumption. Whether these features are universal or a cultural phenomenon has not yet been determined. Even though there might be significant differences between CDS and speech directed at adults in, for example, Swedish, these differences might not be found in every language, or perhaps not to the same extent.

When listing features of CDS, Naomi S. Baron puts the use of nouns in lieu of pronouns at the top of the syntactic features (Baron 1990). This can be exemplified by the following sentences:

(1) Kripke was right when *he* talked about anaphora

(2) Kripke was right when *Kripke* talked about anaphora

Instead of having a pronoun point to the initial proper name, as in (1), the proper name is repeated, as in (2).

Baron states that the features (phonological, lexical, syntactic and conversational) are evident when adults address young children, especially between birth and about the age of 4. The same properties of CDS (such as a high F_0, diminutive marking, short sentences) can be found in other more recent studies, e.g. in Patricia K. Kuhl's paper on early language learning (2014), and in studies on the phonetics of infant-directed speech (Cristia 2013).

The assertion that nouns replace pronouns in CDS from birth can be questioned. In a study on zero-anaphora (a gap in a clause that has an anaphoric function) in Japanese CDS, the results indicate that there was a higher amount of pronominalization found in speech directed at younger infants, and that the use of pronouns dropped as the children grew older (Smith & Frank 2012). The study was made with corpus data that consisted of mother-infant interaction gathered when the children were 6, 12 and 18 months old. When comparing CDS from American and Japanese mothers, even though the referential behaviour seemed to differ between the two languages, the proportions of total pronoun references in the data showed a decline in the frequency of pronominalization, between the three ages and for both languages. The reason for this is still debated. The interaction behaviour of the parent has been connected to the parents' beliefs about the infant's current stage of development (Rohde & Frank 2011:4): "It is possible that caregivers use more adult-like rates of pronominalization with children who they believe are too young to be engaged in serious word learning". In other words, it may be the case that parents do not expect their infants to understand the semantic content of the speech directed towards them, and therefore the parents continue to use complex referential behaviour.

The amount of pronominalization and anaphora in speech directed at adults has not been considered as a variable to which the results of this study will be compared. The internal amount of anaphora and other referential behaviour categories are compared between the four different ages.

2.3 MINT and SECDI

This study relies on material from an on-going project on language acquisition, MINT[2], where 75 infants and their parents are recorded every three months, until the children's third birthday. The parents fill out an online survey, SECDI (Berglund & Eriksson 2000), concerning the development of their infant's vocabulary. The details of the recording sessions and the survey can be found under the method section in this report.

2.4 Definitions

Reference, as a linguistic concept, does not have a clear definition. In this study, "referring" means that there are connections that let us pair sound sequences and signs to concepts and entities in our environment.

The question of anaphora and what qualities they posses has been a recurrent topic in general linguistics. The phenomenon is discussed in major works on pragmatics (Levinson 1982) and there have been anthologies dedicated entirely to anaphora (Lasnik 1989). When an explicit antecedent (e.g. a proper name) precedes an anaphoric expression (e.g. a pronoun), the anaphoric expression is classified as *anaphora in the narrow sense* (anaphora$_{NS}$). In this study, some anaphoric expressions will be classified as *anaphora in the broad sense* (anaphora$_{BS}$). The idea is that the antecedent that is linked to the anaphoric expression is implicit, and is not an explicit sign in the discourse. Consider the following example: Y and Z are sitting in a room, when X enters the room. X dances for a while and then exits the room. Y turns to Z and says either (5) or (6).

(5)   `den där X varje onsdag dansar hon`

"that X, every Wednesday she dances"

(6)   `varje onsdag dansar hon`

"every Wednesday she dances"

Antecedents are essential for anaphora. Without them an anaphoric expression has no semantic content. The question is if the semantic content of the anaphoric pronoun "she" is the same for (5) and (6). In (5), one can claim that the semantic content comes from the explicit antecedent "X". In (6), there is no explicit antecedent, but still the referent of the anaphoric pronoun "she" is clear, given the situation in which (6) is uttered. When the situation provides the connection to the pronoun and gives it its semantic content, we can call it an implicit antecedent. It is a very common phenomenon in

---

natural discourse (Poesio, Walde & Brew 1998). This is why anaphora$_{BS}$ is an important category to single out – to capture e.g. pronouns with implicit referents. This treatment is controversial to those who define anaphora on strict formal conditions and see the anaphoric expression as a placeholder for its head: the explicit antecedent (Chomsky 1981).

In their book *Anaphora and Definite Descriptions* (1985), the Finnish philosopher Jaakko Hintikka and the American philosopher Jack Kulas give examples on how the relationship between antecedents and anaphora is much more complex than the general idea that an anaphoric pronoun is nothing but a placeholder for its head (1985:79). Hintikka and Kulas exemplify how antecedents at times cannot be decided from strict formal conditions. They go on to state that, depending on the analysis, it can be argued that there are no differences between anaphoric and deictic expressions. Both anaphora and deictic expressions can single out referents and their position. In this study, however, deictic expressions are not analysed and the assumption that they are, at least in some situations, quite similar to anaphoric expressions will not affect the results of the study. Ken Safir formulated a general definition of anaphora as a "subsequent reference to an entity already introduced in discourse", but states that this definition also has problems (2004:2). That general definition will, however, be considered sufficient for this study.

Stipulative definitions for key concepts, such as child-directed speech and anaphora, will now follow[3].

I. Every utterance that the parent produces is considered to be child-directed speech if there is no one else in the room but the parent and the child.

II. An anaphoric pronoun is a pronoun or a pronominal entity that point to a previously introduced explicit antecedent (lexical NP, definite description or pronominal entity) or an implicit antecedent (e.g. a situational phenomenon).

III. Variations of a proper name, a lexical NP or a definite description, referring back to a previously introduced NP are considered to be lexical anaphora.

IV. An exact repetition of a lexical NP that appears within the next three utterances from the first one is considered to be a repetition. If the second lexical NP appears later than three utterances after a first one, it is considered to be a reintroduction.

V. If an anaphor has an explicit antecedent, it is labelled as anaphora in the narrow sense (anaphora$_{NS}$).

VI. If an anaphor has an implicit antecedent, it is labelled as anaphora in the broad sense (anaphora$_{BS}$).

Repetitions and reintroductions of lexical NPs are part of the categorization for this study because lexical NPs are the entities that anaphora$_{NS}$ hold as antecedents. If there is a change in the use of anaphoric pronouns, it will be interesting to see how lexical entities, not only appear, but also reoccur in contrast to that change. However, there is no assumption of a correlation between the amount of pronominalization and the amount of repetitions/reintroductions of lexical NPs. The categorization of these repetitions/reintroductions will to some degree overlap with lexical anaphora and anaphora$_{BS}$.

---

[3] The categorization of different kinds of anaphora and repetitions/reintroductions was determined under the council of a computational linguist (Kristina Nilsson-Björkenstam, personal communication).

An antecedent keeps its role over an entire recording session, although it can be reintroduced. The first time an expression that points to a referent is mentioned through the utterance of an NP, no matter what form it has (a proper name, definite description, pronoun) it takes on the role of an anchor (Fraurud 1990) and the analysis treats that NP as a starting point. Anaphora$_{BS}$ is closely related to presuppositions (utterance that provide implicit assumptions about the world, that are taken for granted). Anaphora$_{BS}$ will often take on the role of an antecedent for NPs that follow them.

Cataphora (i.e. an initial pronoun that corefers with a later expression in the discourse) have not been considered as a category for this study. In the data used in this study, no examples of cataphora are found.

A discussion concerning the strengths and weaknesses of these definitions can be found in section 6.1 of this essay.

2.5 Coreference

How do we know that a repeated anaphor is linked to a certain antecedent, and has not switched to an implicit antecedent in the middle of the discourse? A key concept for this study, which is connected to anaphora and referential behaviour, is coreference. A definition of coreference is used to set a logical structure for the connection between a first mention and a second mention (van Deemter & Kibble 1999):

$$\text{Coreference} =_{DEF} (\text{corefer}_{(\mu1, \mu2)} \Leftrightarrow (\text{reference}_{(\mu1)} = \text{reference}_{(\mu2)}))$$

The formal definition states that two expressions corefer if and only if the reference of one is the reference of the other. This formal definition provides a logical foundation for the analysis, but the analytic process still relies on interpretation of natural discourse. This formulation should be seen as a clarification that helps the interpretation part of the analysis.

3 Purpose, Research Questions and a Hypothesis

The purpose of the study is to examine referential behaviour in Swedish CDS, and to examine if the use of pronouns and lexical NPs change in any significant way in Swedish CDS during the child's first year.

In this study, free interaction in a controlled environment between parents and their children is recorded and the referential behaviour of the parent is analysed. The main concern of the study is referential behaviour connected to third person pronouns, proper names, and definite descriptions. The reason for this restriction is to narrow down the study to a suitable size for this thesis, while still focusing on the types of entities that anaphoric pronouns are connected to (Safir 2004).

The study revolves around the following questions:

- How do anaphoric pronouns occur in Swedish child-directed speech during the infant's first year?

- Are there significant changes in the use of anaphoric expressions during the infant's first year and, if so, do all the different kinds of anaphoric expressions change in the same way and to the same degree?

- Is there a correlation between changes in the amount of anaphoric pronouns and the amount of reoccurring lexical NPs in Swedish child-directed speech?

- Is there a correlation between the parent's referential behaviour during the child's first year and the child's vocabulary development?

Based on the indications of the study on zero anaphora in Japanese (Smith & Frank 2012), a hypothesis concerning anaphoric pronouns is formulated: a drop in the use of anaphoric pronouns is expected from the three- and six-month-old sessions to the nine- and twelve-month-old sessions.

4 Method

In this section the data, the selection of the studied sections, and the procedure of the analysis will be accounted for.

4.1 The Data

The data used in this study consisted of video and audio recordings of free play/interaction between 10 infants and their parents. The recordings were part of an on-going project on infant language acquisition – MINT – involving 75 infants and their parents. The recording sessions took place every three months during each infant's first living year, starting when the infants were three months old.

The test subjects were recorded in a 10x6 foot room by three stationary cameras and one in-action camera, worn by the parent (picture 1). Wireless clip-on microphones were attached to the clothing of the parent and to the infant's clothing for the first session, and then to a vest worn by the infant for the remaining three sessions. Both the parent and the infant could move freely around in the room. Each recording session lasted between 15 and 25 minutes.

Picture 1. Video stills from all four cameras during a recording session.

The parent was given no specific instructions for the interaction being recorded, other than that the main interest of the study was recording free play. The parent was, however, introduced to three objects (stuffed animals) that were in the room for each recording session and was told the specific names for each of the objects, mo, na and li. Photos of the objects and their names were framed and hung on the wall of the recording room.

In the recording room there was, besides the named objects, a basket with toys (a plastic mug, three plastic cows, a small blanket and a pillow), another stuffed animal and two books, one with pictures and one with the names of different gestures. The parent was told that the toys and books were at their disposal.

4.2 Transcription

The digital video and audio files were annotated in the program ELAN (Sloetjes & Wittenburg 2008). Every utterance by the parent was transcribed in accordance to a predetermined transcription-key. Utterances were picked out and determined based on a combination of natural pauses, clause length, and mutual agreement between annotators. Five minutes of transcribed data from each recording session was exported as a text file and then imported to a spread sheet for analysis. The videos were consulted throughout the analytical process, which has been key for similar studies (Smith & Frank 2012) and is considered a standard for this type of discourse analysis.

To avoid a researcher being in the room, and thereby not satisfying the stipulative definition of CDS, the five minutes of annotated data for analysis began with the first utterance stated after the three-minute mark of each recording, and the last utterance stated before the eight-minute mark. For eight of the annotated sessions, a different time gap was analysed, due to researchers entering the room. For these sessions, the first consistent five minutes of undisturbed parent-child interaction was chosen instead.

Each child was given an anonymous code, consisting of one letter and two digits.

The six-month-old session for A34 was excluded from the study, due to the parent speaking English during the recording session. As a consequence, only nine analyses make up the foundation for the results concerning the nine-month-old infants.

4.3 Analysis

The theoretical basis for picking out anaphora from the data is based on van Deemter and Kibble's definition of co-reference (1999) and the stipulative definitions mentioned above.

Each NP uttered by the parent was counted. Then all first and second person pronouns were excluded from the analysis. The remaining NPs were then analysed. Going back and forth between annotated lists of utterances and the original recordings, each NP was classified into one or more of the following categories: Anaphoric pronouns with an explicit antecedent were marked as A-PRO; lexical anaphora were marked A-LEX; anaphora with a implicit antecedent were marked as A-IMP; exact lexical repetitions were marked as RE-LEX; and reintroductions of exact lexical terms were marked as REINT-LEX. Each category was then compared to the total amount of NPs to generate a percentage value that represented that category's frequency. An NP that materializes in a discontinuous and repaired utterance was included in the analysis as long as the entire NP was intact before the reparation of the utterance started. The referential expressions that did not fit into one of the predetermined categories, or that stood out for any other reason, was categorized as OTHER.

4.4 Applying the Definitions

The following utterances were spoken by a parent who was touching a framed picture of the named objects.

(7)   aa jag hoppas dom har satt fast den där

"yeah, I hope they have attached that one"

9

(8)   aa där e dom på tavlan gubben

"yeah, there they are in the painting, honey"

The pronoun "they" in (7) is linked to a different antecedent, than the one that is linked to the same pronoun in (8), even though (8) immediately follows (7). The pronoun "they" in (8) is an anaphor for the explicit antecedent `kompisarna` ("the friends"), a collective term for the named objects that had been established previously during the recording session. The pronoun is therefore categorized as an anaphor$_{NS}$. In (7) the pronoun "they" is most likely used to single out the people that put up the framed picture on the wall. Since there is no candidate for an explicit antecedent for the pronoun "they" in (7), it is categorized as an anaphor$_{BS}$. The anaphoric expression "that one" in (7) is used for the framed picture. The framed picture has not been mentioned before this moment, so "that one" is considered to be an anaphor$_{BS}$. However, since "that one" in (7) has explicitly mentioned the framed picture, it acts as an anchor for "the painting" in (8), which makes it into a lexical anaphor.

The referent was always considered when picking out anaphora, but was not necessarily considered when picking out repetitions and reintroductions. Throughout the data, there were situations where the referent of a lexical term could not be decided. Names of animals, like "cow" and "elephant", could often be interpreted as referring to either an object in the room or to the whole extension of the lexical expression. To make up for this problem, the same proper name or lexical entity could occur at different times, for two different referents, or no specific referent at all[4], and still be considered a repetition/reintroduction. As a consequence, these categories differ from the others, in that they do not show a pattern of a specific part of the parent's referential behaviour. The repetition/reintroduction categories visualise to what extend specific lexical NPs reappear in the data.

Here is an example from A34:

(9)   en kossa

"a cow"

mu

"moo"

vad säger den

"what does it say?"

ja där e kossan ja

"yes there is the cow"

---

[4] E.g. the names of different animal types reoccur in the data, without specific referents.

```
här är en liten kossa
```

"here is a little cow"

The change between the indefinite and the definite article for "cow" is, in Swedish, represented by a suffix. The two forms are different enough to preclude marking them as an exact repetition. However, the same form of the first utterance appears in the last, although together with an adjective. The lexical entity "cow" in the last utterance is considered to be a reintroduction. In this example it also introduces a new object into the discourse. These consequences are discussed in section 6.2.3.

4.5 SECDI

Starting from the third recording session, the parents were asked to fill out the online survey SECDI. In the survey, the parents were presented with a list of 414 words and were asked to mark each word that was included in the infant's active or passive vocabulary. Since the recording sessions were part of the on-going MINT-project, the answers that the parents gave after the sixth recording session, when the children were 18 months old, were available for analysis. The size of all 75 children's vocabulary at 18 months old was compared to the result of the parent's referential behaviour during the recording sessions. The entire group of 75 children was divided into quartiles, and each test subject's position within the quartiles was plotted.

5 Results

The results show a statistically significant drop in the use of anaphoric pronouns. At three months old, 23,1%, every fourth or fifth NP that appears in CDS, is an anaphoric pronoun with an explicit antecedent. At six months old, 19,4 % of all NPs that appears in CDS, is an anaphoric pronoun with an explicit antecedent. At nine and twelve months old, the amount of pronominal anaphora with an explicit antecedent has dropped to 12,6% and 12,5% (figure 1).
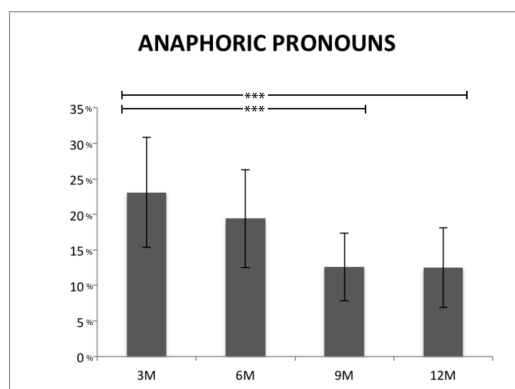

Figure 1.The mean of the total percentage of noun phrases that were anaphoric pronouns with an explicit antecedent for 3, 6, 9 and 12 months, presented with the standard deviation. The three stars marks a p-value lower than 0,001.

A t-test shows the statistical significance of the drop between 3- and 9-month sessions, t(9) = 8,7, p < 0,001, as well as between 3- and 12-month sessions, t(9) = 11,55, p < 0,001. The decline between 3 and 6 months show no statistical significance, t(9) = 1,57, p = 0,08. These results confirm the hypothesis: a drop in the use of anaphoric pronouns was expected from the three-month-old sessions and the six-month-old sessions to the last two recording sessions.

The remaining categories of anaphora did not appear to follow consistent patterns. The amount of lexical anaphora and anaphora$_{BS}$ seemed to materialize in flux (table 1).

Table 1. The total amount of noun phrases (NP), first and second person pronouns (1,2:SG,PL) and the total percentage of anaphoric pronouns with an explicit antecedent (A-PRO), lexical anaphora (A-LEX), anaphora$_{BS}$ (A-IMP), lexical repetition (RE-LEX) and lexical reintroductions (REINT-LEX) for each test subjects.

A06	3M	6M	9M	12M	A08	3M	6M	9M	12M
NP		115	48	35	44	NP		86	105	64	140
1,2:SG,PL		31	8	9	7	1,2:SG,PL		36	23	8	25
A-PRO (%)		15%	25%	7%	10%	A-PRO (%)		20%	12%	11%	13%
A-LEX (%)		32%	12%	27%	30%	A-LEX (%)		6%	22%	24%	9%
A-IMP (%)		2%	2%	11%	8%	A-IMP (%)		16%	5%	2%	9%
RE-LEX (%)		6%	2%	4%	5%	RE-LEX (%)		18%	19%	13%	11%
REINT-LEX (%)		23%	2%	0%	27%	REINT-LEX (%)		18%	7%	7%	9%

A11	3M	6M	9M	12M	A15	3M	6M	9M	12M
NP		121	101	109	125	NP		119	133	140	106
1,2:SG,PL		28	24	27	45	1,2:SG,PL		38	39	25	26
A-PRO (%)		37%	13%	13%	14%	A-PRO (%)		20%	26%	20%	21%
A-LEX (%)		15%	31%	0%	12%	A-LEX (%)		22%	26%	11%	14%
A-IMP (%)		4%	2%	1%	2%	A-IMP (%)		4%	0%	2%	4%
RE-LEX (%)		9%	28%	23%	17%	RE-LEX (%)		27%	24%	29%	7%
REINT-LEX (%)		2%	5%	4%	9%	REINT-LEX (%)		11%	13%	12%	9%

A25	3M	6M	9M	12M
NP	94	70	50	69
1,2:SG,PL	20	24	35	8
A-PRO (%)	20%	17%	17%	5%
A-LEX (%)	28%	4%	28%	10%
A-IMP (%)	5%	11%	3%	5%
RE-LEX (%)	20%	6%	11%	26%
REINT-LEX (%)	3%	13%	17%	20%

A26	3M	6M	9M	12M
NP	134	210	121	125
1,2:SG,PL	53	83	29	37
A-PRO (%)	12%	28%	17%	9%
A-LEX (%)	8%	9%	22%	16%
A-IMP (%)	6%	11%	8%	8%
RE-LEX (%)	18%	14%	12%	18%
REINT-LEX (%)	11%	5%	9%	9%

A27	3M	6M	9M	12M
NP	107	102	99	103
1,2:SG,PL	25	45	21	28
A-PRO (%)	30%	9%	9%	10%
A-LEX (%)	8%	9%	23%	4%
A-IMP (%)	8%	3%	1%	9%
RE-LEX (%)	5%	26%	22%	7%
REINT-LEX (%)	2%	5%	18%	8%

A34	3M	6M	9M	12M
NP	42	XXXXXXXXXX	31	85
1,2:SG,PL	13	XXXXXXXXXX	6	4
A-PRO (%)	24%	XXXXXXXXXX	16%	21%
A-LEX (%)	24%	XXXXXXXXXX	36%	9%
A-IMP (%)	0%	XXXXXXXXXX	12%	2%
RE-LEX (%)	14%	XXXXXXXXXX	12%	22%
REINT-LEX (%)	7%	XXXXXXXXXX	16%	21%

A45	3M	6M	9M	12M
NP	119	135	118	129
1,2:SG,PL	41	44	52	30
A-PRO (%)	32%	23%	6%	6%
A-LEX (%)	5%	1%	12%	8%
A-IMP (%)	13%	8%	6%	3%
RE-LEX (%)	4%	10%	26%	15%
REINT-LEX (%)	10%	4%	7%	4%

A49	3M	6M	9M	12M
NP	166	143	114	130
1,2:SG,PL	61	43	21	15
A-PRO (%)	21%	22%	10%	16%
A-LEX (%)	18%	10%	10%	4%
A-IMP (%)	10%	2%	8%	3%
RE-LEX (%)	20%	21%	28%	27%
REINT-LEX (%)	10%	8%	12%	15%

Repetitions and reintroductions were grouped together (figure 2), to get a notion of how lexical entities appeared in the data. At three months old, 11,9% of all NPs is either a repetition or a reintroduction of specific lexical terms. At six months old, 11,7% of all NPs is either a repetition or a reintroduction of specific lexical terms. At nine and twelve months old, 14,1% and 14,3% of all NPs are either a repetition or a reintroduction of specific lexical terms. There was a tendency towards a rise in frequency found in the nine- and twelve-month-old sessions. However, the difference showed no statistical significance between the 3- and 12-month sessions, $t(19) = 1,24$, $p = 0,22$.
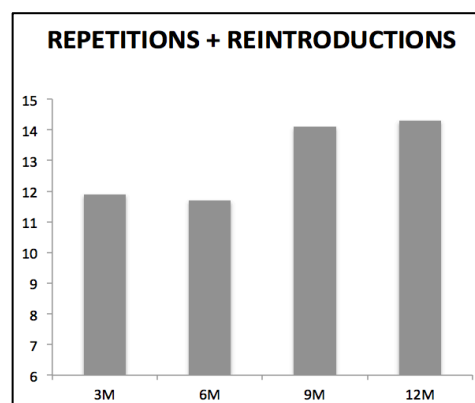

Figure 2. The mean of the total percentage of RE and REINT uttered by the parent during all the recording sessions.

In the total amount of NPs uttered by the parent, there was a trend of a decline in frequency at 9 months. At the three-month-old sessions, the mean of the total amount of NPs uttered by the parent during the five minutes of recorded interaction was 110,3 NPs. At the six-month-old sessions, the mean of the total amount of NPs uttered by the parent during the five minutes of recorded interaction was 116,3 NPs. At the nine- and twelve-month-old sessions, the mean of the total amount of NPs uttered by the parent during the five minutes of recorded interaction was 88,1 NPs and 105,6 NPs. The

decline found in the nine-month-old sessions was not statistically significant, but the difference between the six- and nine-month-old sessions showed a marginal significance, $t(9) = 2{,}16$, $p = 0{,}06$.
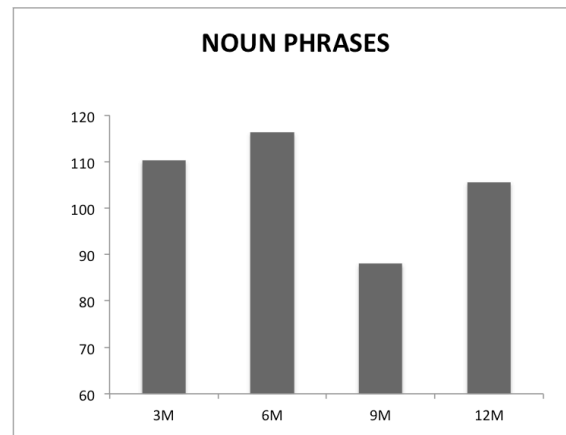
NOUN PHRASES

Figure 3. The mean of the total amount of noun phrases uttered by the parents during the five minutes from each of the 40 recording sessions.

The results of the online survey SECDI, concerning each child's vocabulary development up to and including, the sixth recording session, can be seen in Fig. 4. The result for each of the test subjects in this study show a wide spread over the entire group, with three children in the top three quartiles and one child in the lowest quartile. However, there is no significant correlation when comparing the referential behaviour of the parents. The drop of anaphora$_{NS}$ came earlier for the two test subjects that were ranked lowest in the survey. A significant drop in the parents' use of anaphoric pronouns with an explicit antecedent, started already at 6 months old for both A11 and A08. A similar drop was noticeable for A27 as well.
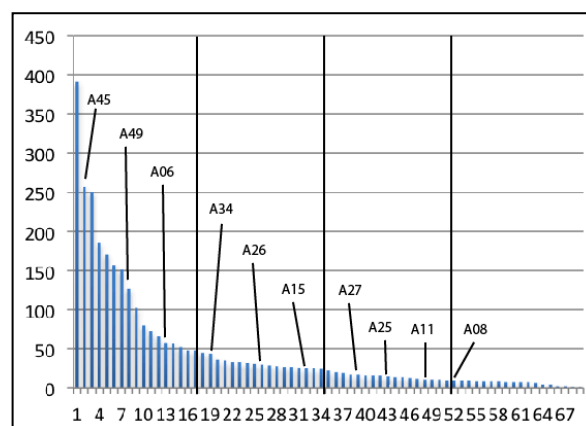
Figure 4. An overview of each test subjects position within the SECDI-quartiles, showing vocabulary development at 18 months old. The y-axis show total amount of words that the child produces.

14

6 Discussion

6.1 Method Discussion

In this section, some aspects of the method and the stipulative definitions will be discussed.

6.1.1 Lexical Anaphora

Lexical anaphora is dependent on the referent. If two lexical entities are used to address the same referent, then the first one will be labelled as an antecedent for the second one, and the second one will be labelled as a lexical anaphor$_{NS}$. An anticipated problem lies in formulaic speech, e.g. a common lexical form that is repeated as a nickname or a more general attention holder. It is possible that certain referential expressions are more direct, towards the referent and others are more general and, therefore, produced and perceived in a different way.

(11)   `titta gumman titta här gumman`

"look sweetie, look here sweetie"

(12)   `titta vilken konstig alice`

"look, what a strange*, Alice"

The referent of $_{gumman}$ in (11) and the referent of the proper name in (12) are one and the same. The proper name would, in this case, be marked as a lexical anaphor with $_{gumman}$ as its antecedent. An intrinsic property of a proper name is that it can single out a referent by itself. General diminutive expressions like $_{gumman}$ (≈sweetie/darling) can be applied to a number of referents. In this sense, $_{gumman}$ acts more like a pronoun in its need to specify the referent (especially when talking about a third person). Perhaps there is room for a more fine-grained categorization, concerning lexical anaphora.

6.1.2 Is There an Implicit Antecedent?

To call a pronoun without an explicit antecedent an anaphor$_{BS}$ is controversial to those who advocate a formalistic point of view or, rather, the claim that there is an implicit antecedent is controversial. One reason for calling it anaphora is that, arguably, pronominal reference will, in the case of an implicit antecedent, not differ from the case of an explicit antecedent.

Consider the example mentioned earlier: Y and Z are sitting in a room, when X enters the room. X dances for a while and then exits the room. Y turns to Z and says either (13) or (14).

(13)   `den där X varje onsdag dansar hon`

"that X, every Wednesday she dances"

<div align="center">

(14)   `varje onsdag dansar hon`

"every Wednesday she dances"

</div>

If one has no problem with calling the pronoun in (13) that is referring back to the explicit antecedent, an anaphor, it would be reasonable to conclude that the same pronoun in an identical clause, (14), would be an anaphor, even if the referent of the pronoun has not been explicitly mentioned in the discourse. X can be considered an element in the discourse for both (13) and (14); this would be an argument for not determining anaphora based on formal criteria, but rather based on the function of pronouns and certain lexical entities.

6.1.3 Reintroductions

Reintroduction and the stipulative definition of the category can at times seem somewhat forced. If a lexical entity reappears every fourth utterance, perhaps it would be more sufficient to classify the lexical entities as repetitions instead of reintroductions. The category is however supposed to give an indication of how frequent lexical terms reappear in child-directed speech, not as direct repetitions (e.g. saying the same proper name over and over again), but rather as a reoccurrence of specific terms and themes. Together with the lexical repetitions, the categorisation should be sufficient for supplying a notion on the frequency of lexical terms in the data.

6.1.4 SECDI reliability

The result of the survey relies on each parent's answers, and there is no control of the parents' understanding or interpretation of the survey questions. Therefore, it is important to be careful when drawing conclusions about the result that the survey generates.

6.2 Result Discussion

In this section the result of the analysis of referential behaviour and the other defined categories, will be discussed, as well as the result of the survey on vocabulary development.

6.2.1 Changes in Referential Behaviour

The referential behaviour in CDS undergoes changes during the child's first year. During the first 6 months every forth or fifth NP in the data was an anaphoric pronoun with an explicit antecedent. During the last 6 months of the child's first year, the amount of these anaphoric pronouns drops to every eight NP. There were, overall, a lot fewer anaphoric pronouns with an implicit antecedent in the data, and the ones that were found, did not show the same pattern as the anaphoric pronoun with an explicit antecedent.

The results correlate with the result of previous studies, and could be interpreted as an indication of the speculation made by Rohde and Frank (2014): that parents in general change their referential behaviour to meet their infants language-acquisition needs. It might be the case that CDS up to a certain point in the language accusation process allows more complex structures, than it later will. At the same time, the results showed a trend of a decline in the total amount of NPs at the nine-month-old sessions. Are there possibly other changes in CDS when the child is nine months old?

Other aspects of CDS will arguably be noticeable from the start. The speech that is directed at infants from birth will most likely differ from the speech directed at adults, if one looks at the features of CDS, e.g. stated by Baron (1990) and Kuhl (2014). However, the results show that the amount of anaphoric pronouns with an explicit antecedent in speech directed at infants changes, dropping to half its initial size, during the child's first year.

6.2.2 SECDI Results

There was no correlation found when comparing the referential behaviour of the parents to the infant's vocabulary development. The drop of anaphoric pronouns with an explicit antecedent came earlier for the two test subjects that were ranked lowest in the survey. The decline in frequency of anaphoric pronouns can be seen already at 6 months for both A11 and A08. It is hard to say what this might imply. It would be interesting to examine a larger group of test subjects to see whether referential behaviour and vocabulary development go hand-in-hand.

6.2.3 Lexical Anaphora, Anaphora$_{BS}$ and Repetitions in Flux

It is noteworthy that the anaphoric pronouns with an explicit antecedent show such a consistency in its decline over the child's first year, while lexical anaphora and anaphora$_{BS}$ show no consistent patterns at all. When examining the total percentage of these categories, the use of lexical entities with or without an explicit antecedent seem to appear in flux. Even though there are differences between parents when it comes to pronominalization and the use of anaphoric pronouns with an explicit antecedent, the general pattern can be seen in the data from every parent.

Due to the ambiguous nature of certain lexical terms in the data, the referent was not always considered when picking out repetitions and reintroductions. The consequence of this, being that the trend of a rise of these categories during the child's first year, is not sufficient to suggest that lexical NPs take the place of the decreasing anaphoric pronouns. Still, the fact that there is not a statistically significant increase of repetitions and reintroductions for the last sessions, might indicate that the drop of anaphora is not only an effect of using nouns instead of pronouns in CDS. There was a tendency towards a decline of the total amount of NPs at the nine-month-old sessions. These two notions might suggest that the referential behaviour undergoes a more complex change, syntactically. An analysis of the phrase structures in CDS during the child's first year could possibly show syntactic changes.

6.2.4 OTHER

No referential expression was categorized only as OTHER. The stipulative definitions were sufficient to categorize every referential expression in the data. Two types of phenomena were, however, classified as OTHER: pronouns that changed referents, and referents that changed pronouns.

(15)   `aa jag hoppas dom har satt fast den där`

"yeah, I hope they have attached that one"

(16)   `aa där e dom på tavlan gubben`

"yeah, there they are in the painting, honey"

In the connected utterances in (15) and (16), the same pronoun is used for two different referents. The switch between them is made without a gestural or verbal attempt to clarify the antecedent. Since there is a location mentioned with the second pronoun, it provides a direction towards the referents.

(17)   kommer du ihåg han + henne + honom + hen

"do you remember he*, her, him, this one"

In (17), a parent is directing the attention of the infant towards one of the named objects (a stuffed animal). When the parent cannot decide the gender of the object, several different pronouns are used, one after the other. The final pronoun hen is a neutral, genderless pronoun. One can either draw the conclusion that the object could be the referent for all of these pronouns, or that each new pronoun should be interpreted as reparation of the last one.

7 Conclusions and Further Research

7.1 Answering the Research Questions

- How do anaphoric pronouns occur in Swedish child-directed speech during the infant's first year?

The result showed that the amount of anaphoric pronouns with an explicit antecedent in Swedish CDS changed during the child's first year. During the first 6 months every forth or fifth NP in the data was an anaphoric pronoun with an explicit antecedent. During the last 6 months of the child's first year, the amount of these anaphoric pronouns drops to every eight NP. There were fewer occurrences of anaphoric pronouns with an implicit antecedent, and the ones that were found, did not show the same pattern as the anaphoric pronoun with an explicit antecedent.

- Are there significant changes in the use of anaphoric expressions during the infant's first year and, if so, do all the different kinds of anaphoric expressions change in the same way and to the same degree?

The change of anaphoric pronouns with an explicit antecedent was statistically significant. Lexical anaphora and anaphora$_{BS}$ show no consistent patterns and no significant changes.

- Is there a correlation between changes in the amount of anaphoric pronouns and the amount of reoccurring lexical NPs in Swedish child-directed speech?

There were no correlation between the amount of anaphoric pronouns and the amount of repetitions and reintroductions of lexical NPs in Swedish child-directed speech. The total amount of repetitions and reintroductions were higher for the nine- and twelve-month-old sessions, but the difference did not show any statistical significance.

- Is there a correlation between the parent's referential behaviour during the child's first year and the child's vocabulary development?

The drop of anaphoric pronouns with an explicit antecedent appeared already at the six-month-old sessions for the two children that scored the lowest in the vocabulary development-survey. No direct claims or indications where drawn from this correlation.

7.2 Further Studies

In further studies, one could analyse the syntactic structures of child-directed speech, to see if there is a significant change in this structure, which correlates with the drop of anaphora. One could possibly examine the relationship between referential behaviour and vocabulary development, to see if there is a correlation between the two. It would also be possible to include first and second person pronouns and examine how they appear in CDS, during the child's first year.

7.3 Concluding Remark

If the use of nouns instead of pronouns, in the long run, is important for the learning and establishing of words, ideas and concepts, one can wonder why the significant drop in the use of anaphora$_{NS}$ is still to come, when the infant is 6 months old? Perhaps it is beneficial for the infant that he or she first meets a natural language, with complex structures unaltered.

References

Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.

Baron, Naomi. 1990. *Pigeon-Birds and Rhyming Words*. Centre for Applied Linguistics and by Prentice-Hal Inc. Englewood Cliffs, New Jersey.

Berglund, E and Eriksson, M. 2000. Communicative development in Swedish children 16-28 months old: The Swedish early communicative development inventory—words and sentences. *Scandinavian Journal of Psychology*, vol. 41, pp. 133-144.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Cristia , Alejandrina. Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Language and Linguistics Compass*, Volume 7, Issue 3, pp. 157–170, March 2013.

van Deemter, Kees and Kibble, Roger. 1999. What is coreference and what should coreference annotation be? In Amit Bagga, Breck Baldwin, and S Shelton (eds.), Procs. of ACL workshop on Coreference and Its Applications. Maryland.

Rohde, Hannah and Michael Frank. 2011. Markers of discourse structure in child-directed speech. Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.

Fernald, Anne and Patricia K. Kuhl. (1987). Acoustic determinants of infant preference for motherese speech. *Infant behavior and Development*, *10*(3), pp. 279-293.

Fraurud, Kari. 1990. Definiteness and the Processing of Noun Phrases in Natural Discourse. *Journal Semantics* 7: pp. 395-433.

Hintikka, Jaakko and Jack Kulas. 1985. *Anaphora and Definite Descriptions*. Dordrecht. D. Reidel Publishing Company.

Kitamura, Christine and Dennis Burnham. 1998. The infant's response to maternal vocal affect. *Advances in infancy research*, 12, pp. 221-236.

Kuhl, Patricia K. 2014 Early Language Learning and the Social Brain Published in Advance 2015 2014 Cold Spring Harbor Laboratory Press.

Lasnik, Howard. 1989. *Essays on Anaphora*. Dordrecht. Kluwer Academic Publishers.

Levinson, Stephen 1983. *Pragmatics*. Cambridge. Cambridge University Press.

Poesio, Massimo, Sabine Schulte im Walde, and Chris Brew. 1998. Lexical clustering and definite description interpretation. In Proceedings of the AAAI Spring Symposium on Learning for Discourse, pp. 82–89, Stanford, CA, March. AAAI.

Safir, Ken 2004. *The Syntax of Anaphora*. New York. Oxford Unversity Press.

Sloetjes, Han and Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).

Smith, Cybelle and Michael C. Frank. 2012. Zero Anaphora and Object Reference in Japanese Child-Directed Speech. Proceedings of the Cognitive Science Society Annual Meeting, pp. 989- 994.