



UIT  
NORGES  
ARKTISKE  
UNIVERSITET

**The Null Subject Stage and Children's Referential Choice:**  
A Case Study of an English-Speaking Monolingual

by

Mitra Heravi Karimouy

*A thesis submitted in partial fulfillment of the requirements for the degree of  
Master of Philosophy in English Linguistics*

Course Code: ENG-3991

Faculty of Humanities, Social Sciences and Education

November 2014

Supervisor: Merete Anderssen





## Abstract

This thesis explores the null subject phenomenon in English-speaking monolingual children by taking a closer look at the effect of discourse-pragmatic features on the realization of subject arguments in the speech of an English-speaking child. For that purpose, first the three dominant approaches in the literature, namely grammatical, performance and discourse-pragmatic accounts, are reviewed. Then, the framework of Hughes and Allen's (2006, 2013, in press) studies is adopted to analyze the effect of four accessibility features, namely *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation* in the selected data.

The results show that the child's use of null subjects decreases as he moves towards the end of the null subject stage. They also indicate that the child is sensitive to the information flow and produces subjects based on the accessibility of different discourse-pragmatic features. The findings also reveal that such features have different weights as well as an incremental effect on the child's choice of subject arguments.

Finally, the other two approaches—the grammatical and performance perspectives—are brought back into discussion; by comparing all the three perspectives, it is concluded that an alternative unifying approach, as Allen (2006) maintains, would be able to provide a better more comprehensive explanation of the null subject phenomenon than any of the three can do alone.

## Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Merete Anderssen, for all her time, feedback, and insightful comments. I greatly appreciate her trust in my potential and her step-by-step advice throughout the project. She spent a lot of time patiently answering my questions, reviewing my drafts and improving them. Any remaining errors are mine alone.

I would also like to thank the Linguistics Department at the University of Tromsø for giving me the opportunity to study here and experience its warm academic atmosphere. My very special thanks go to Martin Krämer, Michal Starke, Thomas McFadden, and Marit Westergaard who kept me motivated with their great lectures. I am also grateful to Mary Hughes (Boston University) for kindly responding to my questions.

I am deeply indebted to Hamideh Marefat (Tehran University) who first introduced me to linguistics and I cannot thank Farhad Sasani (Alzahra University) enough for all the academic support as well as the encouragement he gave me to pursue my studies in this field.

I am also appreciative of the kindness and support Line Vråberg and my friends gave me in the ‘non-linguistic’ part of my life experience in Tromsø.

And last but not least, I want to thank my parents for always encouraging me and being there for me no matter what. I am wholeheartedly thankful for and proud of having their genuine and unconditional love. And a huge thank you to my one and only sister, my brother-in-law, and my little precious Ario whose presence means the world to me.

## Table of Contents

<b>CHAPTER 1: Introduction</b> .....	<b>1</b>
1.1. Background .....	1
1.2. Statement of the Problem .....	2
1.3. Research Questions .....	2
1.4. Research Design .....	3
1.5. Outline of the Thesis .....	3
<b>CHAPTER 2: Literature Review</b> .....	<b>4</b>
2.1. Introduction .....	4
2.2. Grammatical/Competence Account .....	4
2.2.1. Hyams' Pro-Drop Hypothesis .....	5
2.2.2. Rizzi's Truncation Hypothesis .....	7
2.2.3. Valian's Dual-Value Proposal .....	9
2.2.4. Orfitelli & Hyams' Competence Account .....	10
2.3. Processing/Performance Account .....	12
2.3.1. Bloom's Processing Account .....	13
2.3.2. Gerken's Metrical Account .....	14
2.4. Discourse-Pragmatic/Informativeness Account .....	16
2.4.1. (Non)Structural Pragmatic Account .....	16
2.4.2. Hughes & Allen's Discourse-Pragmatic Approach .....	17
2.5. Summary .....	20
<b>CHAPTER 3: Methodology</b> .....	<b>21</b>

3.1. Introduction .....	21
3.2. The Participant .....	21
3.3. Accessibility Features .....	22
3.3.1. Prior Mention .....	23
3.3.2. Animacy .....	23
3.3.3. Joint Attention .....	24
3.3.4. Disambiguation .....	25
3.3.5. Physical Presence .....	25
3.3.6. Person .....	25
3.3.7. Topicality .....	26
3.3.8. Explicit Contrast .....	26
3.3.9. Query .....	27
3.3.10. Summary .....	27
3.4. The Realization of Subjects .....	28
3.4.1. The Effect of Discourse-Pragmatic Features .....	29
3.4.2. Coding Strategies .....	30
3.5. Data Selection Criteria .....	32
3.6. Data Analysis .....	38
3.7. Summary .....	38
<b>CHAPTER 4: Results .....</b>	<b>40</b>
4.1. Introduction .....	40
4.2. The Effect of Individual Accessibility Features .....	41
4.2.1. Findings at Time 1 .....	41

4.2.1.1. Lexical NPs .....	42
4.2.1.2. Demonstratives .....	44
4.2.1.3. Pronouns .....	46
4.2.1.4. Null Forms .....	48
4.2.1.5. Time 1 Summary .....	50
4.2.2. Findings at Time 2 .....	51
4.2.2.1. Lexical NPs .....	52
4.2.2.2. Demonstratives .....	54
4.2.2.3. Pronouns .....	57
4.2.2.4. Null Forms .....	58
4.2.2.5. Time 2 Summary .....	60
4.3. Incremental Effect of Accessibility Features .....	61
<b>CHAPTER 5: Discussion and Conclusions .....</b>	<b>65</b>
<b>Bibliography .....</b>	<b>70</b>

## List of Figures and Tables

Figure 1: Distribution of accessible vs. inaccessible features for NPs at Time 1 .....	43
Figure 2: Distribution of accessible vs. inaccessible features for Dems at Time 1 .....	45
Figure 3: Distribution of accessible vs. inaccessible features for Pros at Time 1 .....	47
Figure 4: Distribution of accessible vs. inaccessible features for Nulls at Time 1 .....	48
Figure 5: Distribution of accessible vs. inaccessible features for NPs at Time 2 .....	52
Figure 6: Distribution of accessible vs. inaccessible features for Dems at Time 2 .....	55
Figure 7: Distribution of accessible vs. inaccessible features for Pros at Time 2 .....	57
Figure 8: Distribution of accessible vs. inaccessible features for Nulls at Time 2 .....	59
Figure 9: Cumulative effect of features at Time 1 .....	63
Figure 10: Cumulative effect of features at Time 2 .....	63
Table 1: Participant and data information summary .....	39
Table 2: Distribution of referential forms at Time 1 and Time 2 .....	40
Table 3: Accessibility of each referential form for the four features at Time 1 .....	42
Table 4: Accessibility of each referential form for the four features at Time 2 .....	51
Table 5: Cumulative effect of features at Time 1 .....	62
Table 6: Cumulative effect of features at Time 2 .....	62



# CHAPTER 1

## Introduction

### 1.1. Background

Young children at the early stages of language acquisition tend to (frequently) drop subject arguments from their utterances. This over-omission of subjects in child language is called the null subject phenomenon (e.g., Hyams, 2011). Children, at this stage, omit subjects not only in pro-drop languages such as Italian and Spanish, in which null subjects are a grammatical option, but also in non-pro-drop languages such as English and Danish, in which null subjects are not allowed in adult grammar. The following examples from Danish (Guasti, 2002), English (Allen, 2006), and French (Orfitelli & Hyams, 2012) show subject omissions in child language:

- (1) a. Se, blomster har. (Jens, 2;2)  
look flowers have/has
- b. No like celery. (Kathryn, 1;10, telling her mother that she does not like celery)
- c. Va sous la tabouret. (Philippe, 2;2)  
goes under the stool

Numerous studies have shown that children at this stage omit subjects more frequently than adult speakers of their target language would do (e.g., Hyams & Wexler, 1993; Valian, 1991). For instance, as Hughes and Allen (2013, p. 15) state, adult speakers of English omit subjects in only 5% to 10% of cases while this percentage is between 20% and 50% in English-speaking children. In pro-drop languages like Italian, children are reported to omit at least 70% of subjects in their utterances, while the equivalent proportion is 50% to 60% in their adult counterparts.

Another important characteristic of the null subject stage is the optionality of subject omission; while children omit subjects in some of their utterances, they use overt subjects in plenty of others (Orfitelli & Hyams, 2012, p. 564).

In order to provide an explanation for this non-adult-like behaviour, research has been carried out from three major theoretical perspectives—the grammatical, the performance and the discourse-pragmatic approaches. The grammatical approach argues that child grammar is different from adult grammar and children drop subjects because their grammar

allows them to do so. However, the performance-based accounts reject this assumption and state that children have the same grammar as their adult counterparts and the reason for the omission of subjects is related to performance factors such as processing/production difficulties. The discourse-pragmatic approach, on the other hand, tries to relate this phenomenon to the information flow in the context of the conversation, and the sensitivity of children to accessibility features (e.g., Bloom, 1990; Hughes & Allen, 2006; Hyams, 2011).

## **1.2. Statement of the Problem**

The majority of researchers working on the null subject phenomenon have taken one of the first two approaches—grammatical or processing accounts—in their studies. And those studies in which a discourse-pragmatic approach was taken have been mainly on pro-drop languages, with bilingual children, or limited to very few accessibility features (Hughes & Allen, 2013). To fill this gap, following previous studies by Hughes and Allen (2006, 2013, in press), the present study attempts to assess the utterances of another English-speaking monolingual child by examining the effect of four discourse-pragmatic features on his choice of subject arguments. However, the study does not merely focus on the discourse-pragmatic approach, but rather tries to bring back the other two approaches into discussion as well and questions whether any of these accounts alone can provide a comprehensive explanation of the null subject phenomenon in child language.

## **1.3. Research Questions**

This study aims to answer the following questions:

- i. Does the child's use of null subjects change over time?
- ii. Do the discourse-pragmatic features affect the child's choice of subject arguments?
- iii. Do the discourse-pragmatic features have equal weight? Do they work in isolation or interact with each other?
- iv. Is the discourse-pragmatic approach alone able to explain the null subject phenomenon?

The present thesis is inspired by Hughes and Allen's studies (2006, 2013, in press) and the first three questions are adopted from their works. However, since the present study is not limited to the discourse-pragmatic perspective, the fourth question is also posed in this thesis.

#### **1.4. Research Design**

In order to answer the above-mentioned questions, the utterances of an English-speaking monolingual from Western New York State, taken from one of the CHILDES archive files (MacWhinney, 2000; Weist & Zevenbergen, 2008), will be analyzed at two different periods, from ages 2;5.04 to 2;7.22 (Time 1) and 3;0.02 to 3;3.09 (Time 2). The data files are in audio format. The selected third person subject arguments are coded for the accessibility of four discourse-pragmatic features, namely *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*. The effect of these features will be assessed both individually and cumulatively on the realization of subject arguments. The analyses of the selected data will be done based on the framework of Hughes and Allen's (2006, 2013, in press) studies. Finally, the discourse-pragmatic approach will be compared to the other two approaches in order to see if any of them alone can explain the null subject phenomenon comprehensively.

#### **1.5. Outline of the Thesis**

The rest of the thesis is structured as follows. Chapter 2 reviews the literature on the three dominant approaches to the null subject phenomenon. For each approach, the most central works will be briefly discussed. Chapter 3 presents the methodology, the information on the participant, the data selection criteria and the data analysis procedure. In Chapter 4, the findings of the study will be reported. Finally, Chapter 5 discusses the implications of the reported results and summarizes the findings of the present study.

# **CHAPTER 2**

## **Literature Review**

### **2.1. Introduction**

The null subject (NS henceforth) phenomenon in young children (roughly 1;6-4;0 years) has been studied from three major theoretical viewpoints, namely a grammatical, a processing and a discourse-pragmatic perspective (Hughes & Allen, 2006, p. 293). The following provides an overall overview of the possible underlying explanations for this developmental stage according to each of these three perspectives. While this phenomenon has been found in both pro-drop (e.g., Spanish and Italian) and non-pro-drop (e.g., Danish and English) languages, the present thesis focuses on non-pro-drop languages—specifically English—in which overt subjects are obligatory and null subjects are not allowed, except in a few limited contexts.

The remainder of this chapter is organized as follows. Section 2.2 takes a closer look at some of the most influential works on children’s NS stage in grammar-based accounts; for that purpose, Hyams’ (1986) pro-drop hypothesis, Rizzi’s (1993/1994) truncation mechanism, Valian’s (1990) dual parameter setting, and Orfitelli and Hyams’ (2012) comprehension accounts are briefly reviewed. In Section 2.3, performance-based approaches, specifically Bloom’s (1990) processing account and Gerken’s (1991) metrical hypothesis, will be discussed. Section 2.4 explores the discourse-pragmatic take on the subject omission phenomenon, which argues that information flow affects young children’s production or elimination of external arguments. This section will mostly focus on Hughes and Allen’s (2006, 2013, in press) studies which provide the framework for the present study.

### **2.2. Grammatical/Competence Account**

One type of approach that attempts to explain missing subjects in child language is the grammatical account. This approach assumes that child grammar is different from adult grammar in some fundamental way; based on this assumption, children are believed to have a different set of linguistic rules and principles than adults, and as a result of this, they

produce subjectless sentences without considering them as syntactically ill-formed (e.g., Hyams & Wexler, 1993).

While there are numerous studies trying to provide a grammatical-based explanation for the NS stage, only four of them will be looked at in this section, namely Hyams' pro-drop hypothesis, Rizzi's truncation hypothesis, Valian's dual parameter setting, and Orfitelli and Hyams' comprehension account (but see also Guilfoyle's (1984) VP hypothesis, Hyams' (1996) underspecification account, Rohrbacher & Roeper's (1995) theory of economy of projection, Yang's (2002) variational model).

### 2.2.1. Hyams' Pro-Drop Hypothesis

On the basis of Chomsky's principles and parameters model of Universal Grammar, Hyams (1986) develops the pro-drop or parameter missetting hypothesis in order to explain the NS stage in young children. Her hypothesis, one of the most central works on this topic, assumes that all children start off with a pro-drop grammar like the one found in adult Italian.<sup>1</sup> Thus, the default/initial setting for the pro-drop parameter is assumed to be the unmarked [+pro drop] for all children. In other words, they start with a [+pronominal] AGR setting. AGR (agreement), part of the INFL (inflection) node, is an abstract element that carries phi-features. In languages like Italian, AGR is basically assumed to be an abstract subject pronoun which licenses empty subjects, because they can be identified or recovered by the phi-features the richly inflected verb carries in such languages (e.g., Hyams, 1986; Rizzi, 1982). But later on, when the children encounter evidence that the syntax of their target language (here, English) does not permit subjectless sentences, they reset the parameter's default value to [-pro drop], which means a grammar with the [-pronominal] Agr feature (Hyams, 2011). The presence of lexical expletives *it* and *there* may be one such piece of evidence in favor of resetting the default value since such structures are not allowed in pro-drop languages (Hyams, 2011, p. 17). In other words, encountering such contradictory evidence in English triggers this switch in parameter value in the target language.

There were some logical flaws in Hyams' original proposal and basic arguments were brought up against it which made her revise her initial thoughts several times. One of her revisions was in line with Jaeggli and Safir's (1989) 'morphological uniformity principle':

---

<sup>1</sup> Since finite verbs in such languages are richly inflected, i.e., they carry the phi-features of number, person (and sometimes gender), the presence of overt subjects is optional; therefore, both overt and empty subjects are allowed (Guasti, 2002).

- (2) *Morphological uniformity*: An inflectional paradigm P in a language L is morphologically uniform iff P has either only underived inflectional forms or only derived inflectional forms (Jaeggli & Safir, 1989, p. 30).

Based on this principle, the *null subject parameter* is defined as:

- (3) Null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms (Jaeggli & Safir, 1989, p. 29).

Accordingly, Jaeggli and Hyams (1988) proposed that children acquiring languages like Italian see their target language as a uniformly inflected one and presume null subject is a grammatical option for them. However, those acquiring “mixed” (non-uniform) languages like English will also assume that their target language is uniformly uninflected and hence subjectless sentences are syntactically acceptable. The difference between these two types of null subjects, as Jaeggli and Hyams (1988) put it, is related to the way they are identified. In languages like Italian, missing subjects are identified by agreement, i.e., they are realized in the form of the inflection on the verb, while in English they are identified by being topics, like in Chinese. One of the predictions that the *morphological uniformity* hypothesis—defined in (2)—makes is that when English-speaking children notice that verbs in their language have some inflections for tense and agreement and therefore is not morphologically uniform, they stop omitting subjects. But this prediction was not realized (Hyams, 2011, p. 20); several studies—in English as well as other languages such as Danish and French—showed that most children produce tensed verbs while they are still at their NS stage. Consequently, null subjects are not limited to uninflected forms in languages like English, which is what is expected given the *morphological uniformity* hypothesis.

The topic-drop hypothesis was another account that followed along the same vein. According to Hyams and Wexler (1993), children were assumed to start off with a topic-drop discourse-oriented grammar like the one found in Chinese and Korean. In topic-drop grammars, both external and internal arguments are allowed to be dropped when they are topics. But this approach was also problematic; according to some comparative studies of Chinese and English-speaking children by Wang et al. (1992), a noticeable difference was observed in the rates of subject and object drop in these two languages. The results showed that in early Chinese, subject and object drop were found at a rate of 46.5% and 22.5% respectively. However, the American children used null subjects at a rate of 33%, while object drop was reported in less than 4% of the cases. These findings suggest that object

drop is not an option in child English. This observation goes against the basic assumption of the topic-drop hypothesis.

There are some other properties related to early null subjects in non-pro-drop languages that are markedly different from those of adult pro-drop languages and have not been dealt with in approaches such as Hyams' parameter missetting hypothesis; Guasti (2002, p. 161) names three environments in which early NS phenomenon can rarely, if ever, be seen:

- (4) a. questions with a fronted wh-element,
- b. subordinate clauses,
- c. matrix clauses with some fronted XP other than the subject.

She continues to say that although missing subjects are not likely to be found in the contexts mentioned in (4) in languages like English, they are all possible in pro-drop languages such as Italian. Thus, she concludes that in non-pro-drop languages, early null subjects seem to occur only in clause-initial position and therefore, they are a 'root phenomenon' (Guasti, 2002). Another property of null subjects in child English is the close relationship between missing subjects and root infinitives (RIs) (Guasti, 2002, p. 166). These two features provide the basis for Rizzi's truncation account, discussed in the next section.

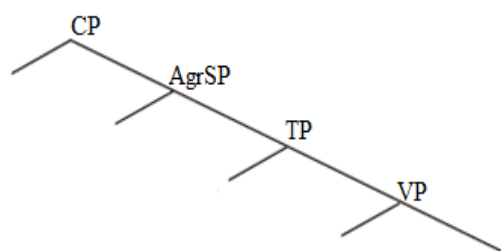
### 2.2.2. Rizzi's Truncation Hypothesis

Rizzi's truncation hypothesis is based on the assumption that in non-pro-drop languages, the NS phenomenon and RIs are two results of the same mechanism in child language (Guasti, 2002, p. 166). In order to explain the occurrence of subject drop in child language, Rizzi (1993/1994) refers to the following principle in adult grammar:

- (5) CP = root

Based on the principle in (5), every sentence in adult grammar always has a CP as its top node as the syntactic tree in (6) shows:

- (6)



According to Rizzi (1993/1994, p. 378), while the principle in (5) works in the adult grammar, it does not apply to child language; therefore, in child speech, other maximal projections lower than the CP layer can also be the root of a main clause. In other words, depending on the verb's morphosyntactic features, any of the nodes in the syntactic tree displayed in (6)—namely, CP, AgrSP<sup>2</sup>, TP/IP or VP—can be the starting point (root) of a child's clausal structure. For instance, if the top node is CP, it would be similar to adult grammar. If the starting point is AgrSP, it would result in root null subjects in tensed clauses. In this case, the top of the syntactic tree in (6) (i.e., CP) is 'chopped off' and a truncated structure would be created. If the top maximal projection is lower than TP, it would result in an RI structure. In other words, in RI structures, the TP and other nodes above it are not projected (Rizzi, 1993/1994, p. 379). Therefore, what is important in truncated structures is the fact that once the starting point is specified, everything above that specific node is truncated and everything below it must be fully projected (Deen, 2009, p. 277).

As noted earlier, although subject omission also occurs in finite declarative clauses, it is in infinitive clauses where most of subject drops happen; Guasti (2002, p. 178) refers to two types of null subjects that Rizzi has introduced in truncated structures, namely null constants (NCs) and PROs:

- i. An empty category called NC is assumed to be the null subject of finite clauses in children's early grammar, and
- ii. In RI structures, though the same NC can be considered as the null subject, PRO might be preferred more since PRO is licensed by nonfinite verbs in the adult grammar as well.<sup>3</sup>

Null subjects are licensed in the specifier (Spec) of the root (Rizzi, 1994). Therefore, NC is in Spec-IP, while in RIs, due to their truncated structures, PRO is in Spec-VP. The truncation account can explain why null subjects are unusual in the structures described in (4) (wh-questions, subordinate clauses, and topicalized structures), as they all represent structures in which the CP-layer is required and truncation cannot occur.

---

<sup>2</sup> AgrSP stands for 'subject agreement phrase'.

<sup>3</sup> This assumption is one of the problematic areas of Rizzi's truncation hypothesis (Guasti, 2002, p. 179). The theoretical arguments against it are not presented here as it is believed to be beyond the scope of this study.



Rizzi's truncation hypothesis has also come under criticism, due in part to its assumption about wh-questions in subjectless sentences; according to this hypothesis, null subjects in early child language are not expected to co-occur with wh-movement (Scott, p. 9). Rizzi (1993/1994, p. 390) refers to this property of truncated structures as "incompatibility with wh-preposing". The reason is that NS sentences have a truncated structure in which the CP node is missing as discussed above. But the CP position is required in wh-questions because it is the 'landing site' of wh-movement. However, the findings of several studies have shown that children produce structures such as 'Where going?' quite frequently (Rohrbacher & Roeper, 1995, p. 9).

### 2.2.3. Valian's Dual-Value Proposal

As noted earlier in Section 2.2.1, in Hyams' (1986) parameter missetting hypothesis it was assumed that there is a default value for the pro-drop parameter in child language acquisition. This parameter might be reset after exposure to the input from the target language, or might remain the same if parameter resetting is not triggered. But as opposed to this 'single-value' solution, Valian (1990) put forth a dual-value solution in which both [+pro drop] and [-pro drop] values are initially available to the child. Accordingly, children would have two mental grammars until they are exposed to sufficient amount of evidence that can help them choose the correct grammar over the other.<sup>4</sup>

Valian (1990) dismisses Hyams' single-value solution, regardless of which value (+ or -) is assumed to be the initial one—an Italian-like or an English-like value. One reason why this is not a convincing approach is because of restrictions on the child's parser. Valian (1990) believes that the child's parser can only analyze those data that are generated by the child's grammar, since the parser is fed by that grammar. As a consequence, a pro-drop grammar will not be able to parse an expletive subject (e.g., *there*), which is what Hyams (1986) has argued triggers the resetting of the parameter, simply because pro-drop languages do not have expletives in their grammars. Consequently, the value of the parameter cannot be reset. But as Valian (1990) goes on to argue, if it is assumed that there are initially two grammars available to the child, it means there are also two parsers which

---

<sup>4</sup> A more recent version of this proposal can be found in Yang's (2002) variational approach to language acquisition. According to this model, children have initially access to multiple competing grammars for analyzing the language input. For example, regarding the NS stage, Yang (2002, pp. 114-124) argues in favor of the coexistence of a pro-drop grammar (like Italian), a topic-drop grammar (like Chinese), and a non-pro-drop non-topic-drop grammar (like English) in young children.

can interpret all the relevant data and determine the final correct value of the parameter. A second reason why Valian rejects the parameter resetting approach is related to what she refers to as misleading input. Valian (1990, p. 110) mentions some contexts in which subject omission is considered to be acceptable in adult English; i.e., imperatives, expletive drop in conversational speech, and pronominal subject drop in informal questions, as shown in the following examples respectively:

- (7) a. Wash the dishes.  
b. Seems like she always has something twin-related perking.  
c. Want lunch now?

She considers such subjectless strings as misleading for a child with an initial [+pro drop] value, whose target language is English (a [-pro drop] language), because such data might lead the child to spuriously conclude that English is like Italian in the sense that null subjects are grammatically allowed. On the other hand, she maintains that the dual-value solution can resolve this problem as well since the two grammars are assumed to be available to the child at that stage.<sup>5</sup>

#### **2.2.4. Orfitelli & Hyams' Competence Account**

In their recent study on how English children interpret NS sentences, Orfitelli and Hyams<sup>6</sup> (2012) argue that young children, based on their non-adult grammar, have the possibility of interpreting NS clauses as grammatical declarative sentences—unlike English adult speakers. This supports the idea that the NS stage is a competence-based phenomenon. Orfitelli and Hyams (2012) go on to state that if it were merely performance-based, i.e., if null subjects were ungrammatical in the child language, children would—in an adult-like manner—assign an imperative reading to subjectless utterances. But as tested and illustrated in their experiments, which are discussed in this section, the children mostly did not identify

---

<sup>5</sup> While Valian's (1990) dual-value proposal has been referred to as an example of a grammatical account, it is worth mentioning that her other works such as Valian (1991) are cited in many studies in favor of the performance-based accounts. According to Valian (1991, p. 33), children are exposed to subjectless utterances in adults' casual speech, imperative structures, and cases like English diary drop. Based on the input, they think it is acceptable to omit subjects, but they don't know yet in which particular contexts and under what circumstances. Therefore, when a sentence is too complex for their production system, they tend to omit subjects. Valian (1991), in her study of 21 American children (1;10 to 2;8), concludes that the performance-based accounts can better explain the subject omission stage in English-speaking children. The performance-based approach will be discussed in more detail in Section 2.3.

<sup>6</sup> Orfitelli and Hyams do not really provide a competence account for NS in child language. However, it has been included in this subsection because it provides an argument in favor of competence approaches.

subjectless sentences as being commands but rather interpreted them as declarative. This can be seen as counter-evidence to the performance-based theories.

Orfitelli and Hyams (2012) used a version of a truth-value judgment task to test children's comprehension of subjectless utterances.<sup>7</sup> The participants were 30 children living in Los Angeles with ages between 2;6 and 4, divided into three age groups of 2;6–2;11, 3;0–3;5 and 3;6–3;11, with 10 kids in each group. The youngest group was expected to be in the NS phase, the oldest to have exited this stage and the middle group to have a combination of both types. Three kinds of sentences were used in this experiment; namely, finite declarative clauses, imperatives with initial *please*, and sentences with missing subjects.

The results indicated that in the youngest group, NS clauses were overwhelmingly interpreted as declarative (86% of the time) while the oldest group correctly assigned an imperative interpretation to NS sentences in more than 80% of the time. The middle group, however, had a more adult-like behavior compared to the youngest group. Children in this group assigned a declarative interpretation to NS clauses 60% of the time and an imperative interpretation 40% of the time. This shows that children in the middle group still prefer the non-adult declarative interpretation more than those in the oldest group.

Orfitelli and Hyams (2012) also found a direct correspondence between comprehension and production. According to this relationship, “children who were classified as being in the NS stage by the production measure were also classified as being in the NS stage by the comprehension measure” (p. 578). In other words, as children grow older, not only does their NS production decrease, the possibility of them assigning a declarative interpretation to subjectless sentences declines as well.

However, the fact that children, at this stage, in their everyday interactions produce and comprehend imperatives as well as subjectless declaratives raises a question about why children in these experiments preferred the declarative interpretation over the imperative one and not vice versa. Orfitelli and Hyams (2012, p. 579) refer to two possible explanations:

---

<sup>7</sup> Young children are generally assumed to be able to both understand and produce imperative structures. But according to Orfitelli and Hyams (2012, p. 569), there have not been enough quantitative studies in support of this assumption. Therefore, before starting their main experiments, Orfitelli and Hyams (2012) first conducted a separate study. They analyzed the speech of Alex (taken from the CHILDES database) when he was at the NS stage (from 2;6 to 3;6). The findings showed that Alex both produced and understood imperative utterances. This is in line with the above-mentioned general assumption.

- i. When children have to choose between imperative and declarative structures to solve the mood ambiguity they are facing, they probably might prefer the “less marked, or more economical, structure”, which is the declarative.
- ii. As another explanation, the participants’ behavior is attributed to the structure of imperative sentences in the experiments; since imperative sentences always had the word *please*, children at the NS stage might have taken it as a signal that only sentences containing this word can be imperatives.

To see if the second explanation could account for the high proportion of NS clauses interpreted as declarative, another experiment was conducted in which other imperative items without the word *please* were also included. The results, as stipulated, showed that the mood ambiguity (imperative vs. declarative) in production now emerges in comprehension as well, with a higher percentage of participants having both interpretations. Orfitelli and Hyams (2012) argued that processing factors might play a major role in resolving the mood ambiguity in this stage.

### **2.3. Processing/Performance Account**

Unlike the competence accounts discussed above, the performance-deficit accounts hold that there is no distinction between child syntax and adult syntax. According to this type of account, it is assumed that children possess and represent the same internal grammar and linguistic rules as adults; therefore, they are aware that subjects are syntactically obligatory (Guasti, 2002). However, due to performance factors and other extrasyntactic reasons—such as processing difficulties and memory limitations—young children often drop subjects and fail to produce them (e.g., Bloom, 1990; Gerken, 1991). The assumption in this approach is that children’s speech is not necessarily a direct reflection of their linguistic knowledge; for instance, as Aronoff (2003, p. 47) puts it, “processing constraints limit the actualization of the argument structure, causing elements to be dropped from sentences in order to reduce processing demands”.

This section will take a closer look at the different arguments put forth by Bloom (1990) in support of the processing approach, as well as the metrical account proposed by Gerken (1991), who argues that there is a direct relationship between prosody and subject drop in child language.

### 2.3.1. Bloom's Processing Account

As an alternative view to the grammatical approach, Bloom (1990) postulates that children have the same grammar as adults from the very beginning, and omit subjects because of performance limitations. He states that by performance limitations he means “an imperfect mapping from what they (young children) intend to say to what they actually say” (Bloom, 1990, p. 491); in other words, what young children say does not necessarily mirror what they know. Therefore, it is assumed that the longer the children's utterances are, the higher is the probability of constituent omission. Bloom (1990, p. 492) provides the following empirical support for this notion:<sup>8</sup>

- i. Children have difficulty even imitating long adult speech. Accordingly, because of the difficulty young children have in producing long utterances, they tend to shorten those sentences.
- ii. It is not just subject arguments that young children tend to omit; but other constituents such as objects, verbs, etc. can be eliminated from their speech too.
- iii. Some children reduce the subjects they might have difficulty producing to schwa, and do not omit them altogether. Aronoff (2003, p. 45) claims that the presence of schwa shows that the child is actually trying to fill the subject position, but at the same time minimize the related production effort. This subject reduction to schwa does not seem to be explicable by the pro-drop hypothesis according to which children either keep the subject or drop it.

Bloom (1990) refers to one of his studies specifically designed to test his hypothesis that sentences with subjects tend to have shorter verb phrases than sentences without subjects. In other words, if subjects are dropped due to processing limitations, it is to be expected that the longer the clause is, the more likely it is that the subject will be dropped. For that purpose, he studied 20-hour speech samples of three children—Adam, Eve and Sarah—from the ages of 2;3 to 2;7, 1;6 to 1;10 and 2;3 to 2;7 respectively.<sup>9</sup> The verbs analyzed in this study were either past tense verbs or non-imperatives (those that cannot have the imperative form). The results showed that the length of verb phrases in children's sentences

---

<sup>8</sup> He also refers to some studies in the field of sign language as further evidence in favor of processing limitations in children's language production.

<sup>9</sup> These children's language development was first studied by Brown (1973). The big age difference between Eve and the other two is because Eve was a very precocious talker. She had an MLU of 4.0 when she was 2;2 whereas Adam and Sarah reached this MLU when they were around 3;6 (Brown, 1973, p. 57).

had an inverse correlation with subjects' heaviness. This strongly confirms the processing prediction that "subjectless sentences tend to have longer VPs than sentences with subjects" (Bloom, 1990, p. 495).<sup>10</sup>

The aforementioned prediction about the correlation between VP length and the occurrence of null subjects in children's utterances can be put as follows (Bloom, 1990, p. 498):

- (8) Longest VP    No subject (null)  
                  ↓            Short subject (e.g., *you*)  
Shortest VP    Long subject (e.g., *the big mean lion*)

Accordingly, VPs are longest in sentences with empty subjects, shorter in sentences that have pronominal subjects, and shortest when the subject of a sentence is a full NP.

Bloom (1990) mentions that when children are at the NS stage, their use of long subjects (more than one word long) is very rare; therefore, he compares the use of pronominal subjects versus non-pronominal subjects and maintains that since the former group are phonetically shorter than the latter,<sup>11</sup> it must have some effect on VP length, in line with the previously-mentioned hypothesis. His data analysis supported the processing prediction once again.

### 2.3.2. Gerken's Metrical Account

From the perspective of the metrical hypothesis (Gerken, 1991), prosody plays an important role in the subject drop phenomenon; Gerken (1991) observes that young children tend to omit weak unstressed syllables when they occur in iambic stress patterns (weak-strong), but not when they occur in trochees (strong-weak). This is why, according to Gerken (1991, p. 436), young children are more likely to reduce "giRAFFE<sup>12</sup>" to "RAFFE" than "MONkey" to "MON", as the metrical structure of *giraffe* is iambic, whereas *monkey* has a trochaic foot.

---

<sup>10</sup> One of the arguments against this observation is that of Hyams and Wexler (1993); they found the same link in Italian adult speakers and concluded that since adults seem not to have the same processing difficulties as children do, this correlation is more probably related to pragmatic properties rather than production limitations.

<sup>11</sup> He does not provide any theoretical argument in support of this claim since all pronouns are not necessarily shorter than all NPs (for example, *you* vs. *Sue*).

<sup>12</sup> Capital letters are used to identify stressed syllables (-fe in 'raffe' is not pronounced).

The aforementioned examples show how this metrical template works at the multi-syllabic word-level, and the same can apply to sentence-level cases as well, leading to the omission of weak syllables in iambic feet. In other words, children at the NS stage prefer to keep the trochaic foot. Consider the following examples (Gerken, 1991, p. 437):

- (9) a. she KISSED + the DOG
- b. the DOG + KISSED her
- c. PETE + KISSED the + DOG

Assuming the division into metrical feet signaled by the plus marks (+) above,<sup>13</sup> in (9a) the subject pronoun belongs to an iambic foot and therefore may be the target of omission. However, the object pronoun in (9b), though a weak syllable too, is part of a trochaic foot and this significantly reduces the likelihood of the pronominal object being dropped from the sentence. The same prediction can be applied to articles; in (9c), the object article belongs to a trochaic foot, which makes it less likely to be omitted than in (9b), where the subject article is a part of an iambic foot. Gerken (1991) claims that this is an advantage for her approach as it uses the same mechanism to explain another case of omission, namely with articles.

As can be inferred from the examples in (9), Gerken's metrical account makes the following predictions with regard to the missing subject phenomenon in young children (Gerken, 1991):

- i. Pronominal subjects, being prosodically weak, are more likely to be the target of omission than lexical NP subjects,
- ii. Subjects, due to their iambic metrical structure, are expected to be dropped more often than objects, and
- iii. With respect to articles in lexical NPs, children are more likely to omit them from subject position than from object position, as articles in subject position typically end up as the unstressed syllable of an iamb.

In order to test the validity of the aforementioned predictions about the different behavior of iambic versus trochaic feet, Gerken (1991) conducted an experiment in the form of an imitation task. The study included 18 monolingual English-speaking children living in

---

<sup>13</sup> It should be noted that this is Gerken's division; today, it is generally assumed that all feet are trochaic and what used to be the unstressed syllable of an iamb in Gerken (1991) would now be considered an unfooted syllable and this is why it would be omitted.

New York, with ages between 23 and 30 months and a mean MLU of 2.54. Her data analysis showed that the participants omitted more pronominal subjects (32%) than lexical ones (24%).<sup>14</sup> It also illustrated that the number of omitted NP subjects (19%) was more than their object counterparts (0.3%). With regard to article omissions, the findings showed that the children omitted more articles in NPs from subject position (31%) than from object position (18%). The results were in line with the initial predictions.

## **2.4. Discourse-Pragmatic/Informativeness Account**

Another dominant approach to the NS stage in child language is the discourse-pragmatic account. According to this perspective, pragmatic constraints and the pattern of information flow within the context of interaction play a determining role in the child's overt realization of a subject argument vs. its omission (e.g., Hughes & Allen, 2013). Based on the first two accounts that have already been discussed in the previous sections, i.e., the grammatical and the performance accounts, it might be possible to divide the informativeness account into two major groups; these categories, which are referred to as the structural and the nonstructural pragmatic accounts (Hyams & Wexler, 1993), will be reviewed in Section 2.4.1. Then, in Section 2.4.2, Hughes and Allen's (2006, 2013, in press) studies will be discussed in more details.

### **2.4.1. (Non)Structural Pragmatic Account**

In terms of the 'nonstructural pragmatic account', some linguists such as Greenfield and Smith (1976, pp. 222–223) relate the subject omission in child language to children's sensitivity to the informational structure of an utterance and argue that children tend to keep those elements that, from their own point of view, are most informative and omit those that are less informative or carry presupposed information. Therefore, since subject arguments are "often taken for granted" and usually represent given/old information, they are more likely to be a frequent target of omission. Subjectless sentences, according to the nonstructural pragmatic perspective, are not a syntactically acceptable option in English child language and children drop subjects due to reasons such as their restricted production capabilities (Greenfield & Smith, 1976).

---

<sup>14</sup> The percentage of lexical subjects includes 11% proper NPs and 13% common NPs.



While in the aforementioned group it is assumed that the subject omission phenomenon in child language is “uniquely governed by pragmatic considerations”, there are other researchers such as Hyams (1986) who argue that it is the interaction of structural and pragmatic factors that influences subject realization and in fact determines when an overt subject will be used (Hyams & Wexler, 1993, p. 422). This group can be categorized under the label of ‘structural pragmatic accounts’. Unlike the ‘nonstructural pragmatic account’, and similar to the grammatical approach, in this category it is still assumed that child grammar differs from adult grammar. The ‘structural pragmatic account’ also generally assumes that it is the grammar of a language that sets the boundaries within which the pragmatic rules and principles work (e.g., Hyams, 2011, p. 42), and as such, it is both the grammar and the information structure of a language that influence the overt realization or omission of subject arguments.

#### **2.4.2. Hughes & Allen’s Discourse-Pragmatic Approach**

According to Hughes and Allen (2006, 2013, in press), children are sensitive to the information flow in utterances and tend to keep or omit arguments based on discourse-pragmatic features, such as *absence* and *newness*; the former feature shows whether or not a referent is present in the physical context of a conversation, while the latter indicates whether a referent has been previously mentioned in the discourse (given) or not (new) (Allen et al., 2008). Hughes and Allen (2006) point out that most of the studies taking this approach to subject drop in this field have been mainly on pro-drop languages. What makes the investigation on non-pro-drop languages challenging, they put forth, is the fact that overt subject realization rates are quite high in such languages and this makes it difficult to distinguish whether it is related to grammatical constraints or it is due to discourse-pragmatic constraints (Hughes & Allen, 2006, p. 294). Hughes and Allen (2013, p. 16) note that according to the discourse-pragmatic perspective, if the referent is accessible or identifiable from the context (which children sometimes overestimate), the subject is more likely not to be linguistically realized. On the other hand, if it is inaccessible or not recoverable from the discourse, the overt realization of the subject argument is more probable. But as Hughes and Allen (2013) maintain, it is not just the distribution of null vs. overt subjects that matters; there is variation in children’s subject realization and any of the four referential forms—null, pronominal, demonstrative and lexical NP forms—can be potential candidates for the subject position depending on the accessibility of different

discourse-pragmatic features and their interaction with one another. These four referential forms can be put in a hierarchy of accessibility markers with lexical NPs being the lowest and null forms being the highest in this hierarchy as presented in the following diagram (Hughes & Allen, 2013, p. 20):

(10) Low Accessibility Marker	Lexical NP	<u>Boy</u> go. (Brian 2;07.01)
	Demonstrative	<u>This</u> go round here. (Fraser 2;00.20)
	Pronoun	<u>It</u> goes there. (Annie 2;00.25)
High Accessibility Marker	Null	∅ want chips. (Eleanor 2;00.15)

In order to see how discourse-pragmatic features affect subject realization in early English and how such features interact, Hughes and Allen (2006) analyzed the utterances of a two-year-old monolingual English speaker called Annie, from the UK. Her conversations with her mother were videotaped for seven weeks, one hour once a week. In this study, the effect of six accessibility features on the realization of overt vs. null subject arguments was investigated. These features, namely *person*, *animacy*, *absence*, *newness*, *differentiation in context*, and *differentiation in discourse*,<sup>15</sup> can determine the degree of informativeness provided by a subject argument and show how easy it is to identify a target referent from the context. The more accessible a referent is, the more likely the child will be to use a null subject. And vice versa, the less accessible a referent is, the more likely the child will be to produce an overt subject (Hughes & Allen, 2006, p. 295).

Regarding the role of the aforementioned features, their data analysis showed that the feature *person* was problematic as it skewed the results due to the nature of the feature; it is argued that the first person subjects have no inaccessible value for any of the above-mentioned features, the second person subjects might have few features with inaccessible values such as *differentiation in discourse* for example, but for the third person, all the six features may be inaccessible. This supports the idea that third person utterances should be studied separately from the ones with the first and second person subjects. Therefore, the feature *person* was eliminated from their study and only those sentences with third person subject arguments were included for analysis, while sentences with first and second person subject arguments were removed from the data (Hughes & Allen, 2006, p. 299).

---

<sup>15</sup> The definition of all these features will be provided in Section 3.3.

The results of the analysis for the remaining five features confirmed that such discourse-pragmatic features affect subject realization, with the features *absence* and *newness* being the stronger predictors; when these features had an inaccessible value, overt subject arguments were mostly produced by the child (Hughes & Allen, 2006, p. 300).

In order to see if these findings can be generalized, Hughes and Allen (2013) conducted a more comprehensive study, this time with four participants, all of them monolingual English-speaking children. Besides Annie from the previous study, Brian, Eleanor, and Fraser were also added. Their data sets were collected at two different age ranges, 2;0–2;7 (called Time 1) and 3;0–3;1 (called Time 2). A random selection of their mothers' utterances was also analyzed in order to make a comparison of adult's vs. child's choice of referential forms. The aim of this study was to investigate the sensitivity of the children to six accessibility features (*animacy*, *physical presence*, *prior mention*, *joint attention*, *linguistic disambiguation* and *contextual disambiguation*) over time. The effect of these discourse-pragmatic features was examined both individually and cumulatively, with respect to the children's choices of referential forms (lexical NPs, demonstratives, pronouns, and null forms).

The results showed that children at both Time 1 and Time 2 were sensitive to discourse-pragmatic features, although at Time 2, being towards the end of the NS phase, their choice of referential forms was more adult-like. When referents were accessible, the children produced more pronominal and null subjects, and when inaccessible, they produced more demonstratives and lexical NPs. Moreover, as expected, the number of null subjects significantly decreased over time. The features *physical presence*, *prior mention*, and *joint attention* turned out to be the most significant and powerful individual features in this study.

In order to see if the discourse-pragmatic features work together, the cumulative effect of the features was also assessed for fully accessible vs. fully inaccessible referents. The results indicated that children, at both Time 1 and Time 2, were very sensitive to the effect of discourse-pragmatic features in combination. In their latest study, Hughes and Allen (in press) did a thorough analysis of the incremental effect of the aforementioned features on all the referents from the previous study with different accessibility patterns (not just fully accessible vs. fully inaccessible). The results showed that as referents' level of inaccessibility increased, children were more likely to produce highly informative forms such as lexical NPs, whereas low information forms such as pronouns were more likely to be used for highly accessible referents. These findings once again confirmed the prediction

that discourse-pragmatic features do not work in isolation, but interact with each other and work in combination.

## **2.5. Summary**

This chapter aimed to give an overview of the three dominant approaches in the null subject stage literature, i.e., the grammatical approach, the performance-based accounts, and the discourse-pragmatic perspective. The central works of each approach were surveyed. According to grammatical accounts, child grammar is different from adult grammar and children drop subjects at this stage due to some difference in their grammar (e.g., Hyams, 2011). Performance-based accounts, on the other hand, argue that children have the same grammar as adults, and that subject drop is due to processing and performance limitations children have at a young age (e.g., Bloom, 1990). The third approach (e.g., Hughes & Allen, 2006) relates the subject omission phenomenon to the information flow within the context of conversation and argues that children are sensitive to discourse-pragmatic features and omit those subject arguments that are highly accessible.

The present study chiefly adopts the framework of Hughes and Allen's (2006, 2013, in press) studies in order to test the validity of discourse-pragmatic predictions for another English-speaking monolingual child. The results will be also compared to some of the predictions that the other two approaches have made with regard to the null subject phenomenon.

## **CHAPTER 3**

### **Methodology**

#### **3.1. Introduction**

This chapter introduces the methodology used to sample and analyze the data for the present study. For that purpose, a brief introduction on the participant and the data set investigated in this study will be provided in Section 3.2. Section 3.3 gives some general information about all the discourse-pragmatic features studied in child language and explains why only some of them are assessed in the present study. In Section 3.4, the realization of subjects in their four possible referential forms will be explored. This section also takes a closer look at the potential relationship between the accessibility features selected for this study and children's choice of subject arguments. Next, the coding strategies of each individual feature investigated in this study will be explained with the help of some excerpts from the selected data. Finally, in Section 3.5, the selection criteria for the data will be discussed in detail, in order to make clear the motivation behind the inclusion and exclusion of examples.

#### **3.2. The Participant**

The present study analyzes the utterances of Matthew, a monolingual English-speaking child from Western New York State. The data were taken from one of the CHILDES (MacWhinney, 2000) archive files, and was part of the Fredonia Child Language Project directed by Richard M. Weist (Weist & Zevenbergen, 2008). Since one of the purposes of this study is to investigate how information flow affects subject realization in children over time, data from two different periods are analyzed separately; Time 1 consists of six files (files 7–12) from ages 2;5.04 to 2;7.22. Time 2 contains six files (files 17–22) from ages 3;0.02 to 3;3.09. At Time 1, the child is expected to be at the NS stage, and the proportion of subject omission is thus expected to be higher than at Time 2, at which he is towards the end of the NS stage and more adult-like behavior concerning subject realization is expected to be found. All the selected files were in audio format and each of them was approximately half an hour long. They were recorded in 2001–2002 with recording intervals of minimum

one week and maximum six weeks between sessions. The child's MLUw<sup>16</sup> range was from 2.5 to 3.6 at Time 1, and from 3.3 to 4.3 at Time 2, with the mean of 2.9 and 3.7 respectively. The selected data are the spontaneous interactions of the child with his mother (and an investigator) while playing or doing other daily activities, audio tape recorded in their own home. Matthew was a resident of Chautauqua County in Western New York State at the time of recordings. He came from a middle class family and his parents were professionals.

### 3.3. Accessibility Features

According to Chafe (1996), the notion of salience or accessibility of a referent can be defined in terms of "degrees of activation in consciousness" which can be either active, semi-active, or inactive at any given moment: active information is in the focus of a person's consciousness at a particular moment; semi-active ideas are in peripheral consciousness, which means that they are not fully active at the moment, while inactive information is not currently present in a person's consciousness though might be present in long-term memory (Chafe, 1996, p. 40). Based on Chafe's observation, and as discussed in other works such as Ariel (2001), it can be concluded that the more easily a referent's mental representation can be retrieved from memory or can be activated in the addressee's mind, the more accessible that referent is considered to be.

Multiple factors are involved in determining a referent's degree of accessibility/informativeness (Ariel, 2001); in the following subsections, nine accessibility features (in no specific order) studied in child language are discussed (Allen et al., 2008). It should be noted that these discourse-pragmatic features—though in reality gradient and complex in nature—are treated as binary (accessible versus inaccessible) in the literature in order to

---

<sup>16</sup> Mean Length of Utterance measured in words; in order to determine Matty's MLUw, his total number of words divided by his total number of utterances were calculated for each file at both age ranges (all the calculations were done manually). For that purpose, based on Wiczorek (2010), all contractions (such as it's, don't, etc.) were counted as one word. Fillers (e.g., um, oh) were not counted as words. Unintelligible words (transcribed as xxx in the data) were not counted. All repetitions were counted separately; the only exception was those cases marked by [/] in the transcripts which are considered as an indication of the child having trouble saying a word or phrase (MacWhinney, 2000). For example, consider the following utterances by Matty:

(11) CHI: yeah yeah yeah. (3;2.18)

(12) CHI: my [/] my room. (2;5.04)

While in (11), all the three 'yeah's are counted, in (12) only one 'my' is counted as a word.

make the statistical analysis of these values possible (Allen et al., 2008, p. 108; Hughes & Allen, 2013, p. 19).

### 3.3.1. Prior Mention

Recency of mention or *newness* refers to whether a referent has been mentioned in the prior discourse or not. However, as Allen et al. (2008) discuss, the cut-off point that makes a referent new or given is not fixed in the literature; for instance, some studies, such as Hughes and Allen (2006), use the preceding 20 utterances as the criterion to determine the accessibility of this feature; that is to say, a referent is given if it has been mentioned in the preceding 20 utterances, and it is new if no prior mention has been made in this range. There are some other studies in which this boundary varies from three to 10 preceding utterances, whereas researchers such as Serratrice et al. (2004) do not mention any specific range at all (Allen et al., 2008, pp. 109–110).

The present study follows Hughes and Allen's (2013) work in which the preceding five utterances were used as the cut-off point to determine the accessibility of *prior mention*; hence, if a referent has been mentioned within the preceding five utterances, it is not new and, therefore, will be coded as accessible for *prior mention*, and inaccessible otherwise.

### 3.3.2. Animacy

The feature *animacy* indicates whether a referent is animate or not. Allen et al. (2008, p. 115) describe this notion as “how alive or sentient an entity is”. They go on to state that this feature is an inherent, stable, semantic property of a referent and is not dependent on the discourse-pragmatic context of the interlocutors. This characteristic makes *animacy* distinct from other features such as *prior mention*, since the accessibility of the latter can vary due to different discourse-pragmatic contexts (Allen et al., 2008).

In the literature, *animacy* is seen more as a spectrum rather than a clear-cut binary distinction between animate and inanimate entities. However, in child studies—including the present study—this feature, similar to other features, is assigned a binary value. Human beings and animals are animate and therefore accessible for the feature *animacy*, while the rest taken as inanimate are considered inaccessible (Hughes & Allen, 2013).

Allen et al. (2008) mention that in some contexts, researchers code inanimate referents such as dolls or some particular toys as animate because of the human characteristics they are given in child discourse. In the present study, a similar strategy is applied; those

fictional movie/book characters, toys and stuffed animals that had the characteristics of humans/animate entities are coded as animate. Consider the following dialogues between Matty, his mother and the investigators:

- (13) CHI:<sup>17</sup> my pictures in my room.  
MOT: my pictures in my room, yeah check (th)em out.  
MOT: look at what Matthew's got in his room now.  
CHI: that one have one eye. (2;5.04)  
ERI:<sup>18</sup> ooh one eye (.) is that a monster?  
CHI: yeah.

- (14) RYA: oh my (.) what is that?  
%situation: child holds a small, plastic dinosaur up to the investigator in an attempt to scare him.  
CHI: a dinosaur.  
RYA: is it a nice dinosaur?  
CHI: it's a meat eater. (3;3.09)

In example (13), the referent of the underlined demonstrative ‘that one’ is the green one-eyed round monster called Mike, one of the main characters in the ‘Monsters, Inc.’ film. The underlined pronoun in (14) refers to a plastic dinosaur. Both of them are assumed to be animate in the present study and therefore have been coded as accessible for the feature *animacy*.

### 3.3.3. Joint Attention

The feature *joint attention* shows whether or not both a speaker and listener are attending to the same referent, while they are aware that the other is also visually focused on the same referent (Skarabela et al., 2013). It should be noted that this feature is different from *topicality* (discussed in Section 3.3.7); as Skarabela et al. (2013, p. 6) argue, while a referent is in the focus of attention, it does not necessarily mean that both the speaker and listener are visually focused on that same referent as well.

Head and body directions of the interlocutors, as well as their eye gaze and gesture (such as pointing) are some of the indicators of *joint attention* (e.g., Allen et al., 2008, p. 116; Skarabela et al., 2013, p. 9). As the selected files for the present study were not in video format, it was not possible to check out these indicators on the participant and determine the

---

<sup>17</sup> CHI stands for ‘child’ and MOT stands for ‘mother’.

<sup>18</sup> ERI and RYA (in example 14) refer to the investigators.



accessibility of this feature. Therefore, the feature *joint attention* is not investigated in the current study.

#### **3.3.4. Disambiguation**

The feature *disambiguation* evaluates “whether a particular referent has potential competitor referents in the linguistic or physical context” (Hughes & Allen, 2013, p. 17). In some studies, this feature is divided into two categories depending on whether the potential referents exist in the linguistic context or are found in the physical setting of an interaction. The former is called *linguistic disambiguation* and the latter is referred to as *contextual disambiguation*. There have been different terms and definitions assigned to this feature in the literature such as *contrast*, *interference* or *differentiation* (Allen et al., 2008, p. 112).

Regarding the feature *contextual disambiguation*, access to the physical context of the interlocutors’ conversation is required in order to determine the accessibility of this feature. Because the selected files for the present study were not in video format, it was not possible to examine the effect of this feature. Thus, *contextual disambiguation* is excluded from the analysis and only *linguistic disambiguation* is assessed.

Similar to *prior mention*, the threshold between ambiguous and unambiguous referents is the preceding five utterances; i.e., if there are no other possible referents for a particular subject argument within five prior utterances, it will be coded as accessible for *linguistic disambiguation*. But if the subject has other competitor referents within that range, it will be considered linguistically ambiguous and therefore inaccessible for this feature (Hughes & Allen, 2013, p. 20).

#### **3.3.5. Physical Presence**

As its name suggests, the feature *physical presence* indicates whether or not a particular referent is present in the physical context of an interaction. It is also referred to as *absence* in the literature. Regarding the accessibility of this feature, if a referent is physically present, it is coded as accessible, whereas its absence from the physical context of the interlocutors’ conversation makes it inaccessible for the feature *physical presence* (Allen et al., 2008, p. 111).

#### **3.3.6. Person**

The feature *person* indicates if a referent belongs to first person (e.g., I, we), second person (e.g., you) or third person (e.g., s/he, they) category. The first and second person referents

are coded as accessible for the feature *person* while the third person referents are coded as inaccessible (e.g., Hughes & Allen, 2006). The reason is that first and second person referents are given<sup>19</sup> as they are already known by the interlocutors. They are also always physically present, animate, etc.; therefore, the hearer does not have difficulty identifying them as opposed to third person entities, which can be fully inaccessible in some contexts, i.e., inanimate, absent from the physical context of the conversation, new to the interlocutors, etc. The fact that the discourse-pragmatic behavior of third person referents differs from the other two has been shown by several studies in child language (Allen et al., 2008, pp. 114–115). This is why some researchers have excluded the first and second person referents from their analyses (e.g., Hughes & Allen, 2013; Skarabela et al., 2013).

Because of the above-mentioned reasons, the present study is also restricted to third person only. Thus, the feature *person* is not included in the analysis either.

### **3.3.7. Topicality**

Allen et al. (2008) name a referent topical if it is the focus of the interlocutors' conversation. As such, it is coded as accessible for the feature *topicality*. They continue to argue that putting aside the extralinguistic contexts, this feature has so much in common with *prior mention* (discussed in Section 3.3.1) that makes it rather difficult to distinguish the two. One of the differences, however, is that identifying the feature *topicality* is less objective than the other features, including *prior mention*, because it is based more on "ones's impression of salience rather than on anything quantifiable" (Allen et al., 2008, p. 110).

Thus, following Hughes and Allen's (2006, 2013, in press) studies, the feature *topicality* is another discourse-pragmatic feature that is not investigated in the present study because it is less objective, more general than the other features, and "subsumes several of them" (Hughes & Allen, 2013, p. 19).

### **3.3.8. Explicit Contrast**

When a referent is emphasized or explicitly contrasted by a speaker using tone of voice, stress, or gesture, it is coded as accessible for the feature *explicit contrast* (Hughes & Allen, 2013, p. 18). This feature is different from the feature *disambiguation* (discussed in Section

---

<sup>19</sup> It should be noted that in some studies, first and second person referents are coded as new (inaccessible for the feature *prior mention*) the first time they are introduced (Allen et al., 2008, p. 110).

3.3.4) in the sense that the emphasis or contrast used by the speaker, in fact, makes the referent unambiguous (Allen et al., 2008, p. 113).

In order to emphasize or contrast a referent in English, Hughes & Allen (2006, 2013) argue, it is necessary for the referent to be realized in an overt argument form. It means that null subjects cannot be an option for a referent with an accessible *contrast/emphasis* feature. Therefore, since the overtness of subject is fixed in such circumstances, this feature is not investigated in the present study.

### 3.3.9. Query

The feature query is coded as accessible if “a referent is the subject of or the response to a question” and inaccessible if “there is no interrogative context” (Serratrice et al., 2004, p. 191). However, there is no generally agreed upon way of coding the accessibility of this feature in the literature. Allen et al. (2008, p. 111) illustrate this considerable difference with an example: in the interaction ‘Who ate the cake? John did!’, some researchers code *who* and *John* as not accessible for the feature *query*, some code *John* and *cake* as not accessible while removing wh-words from the analysis. There are some other works in which only *John* is coded for *query* because questions are excluded from their analysis and this feature is only applied to responses.

As Hughes and Allen (2006, p. 302) put it, like contrasting, questioning a subject also always causes it to be realized. Therefore, this feature, for the same reasoning provided for *explicit contrast* in the previous section, is not investigated in the present study.

### 3.3.10. Summary

To sum up the accessibility features section, due to the linguistic nature of some of these features and because the data files were not in video format, only four out of the nine aforementioned features are investigated in the present study, namely *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*. The next section discusses what role accessibility features play in the realization of subject arguments. It also explores whether there is any established meaningful relationship between discourse-pragmatic features and children’s choice of subject arguments.

### 3.4. The Realization of Subjects

Different referring expressions can be chosen as the subject of an utterance by a speaker depending on the referent's degree of informativeness. It can be highly informative as in lexical noun phrases (e.g., *the Persian cat*), less informative as in deictics (e.g., *this*), or not very informative linguistic forms such as pronouns (e.g., *it*) (Graf & Davies, 2014, p. 161). To describe this with the notion of accessibility, the more accessible a referent is, the less linguistic information is needed for the referent to be identified from the discourse context. As such, according to Ariel's (2001) accessibility theory, in adult speech, lexical NPs carry the least accessible information. Demonstratives can be considered as medium accessibility markers, while pronouns and zero anaphora are high accessibility markers. In other words, when a certain piece of information can be retrieved by the addressee from the discourse context, the speaker does not have to put it in an elaborate linguistic form again. This is why a less informative reduced form is chosen by the speaker.

Ariel (2001, p. 31) suggests a detailed accessibility marker scale starting with the lowest in accessibility and ending with the highest, as illustrated in (15):

(15) Full name + modifier > full name > long definite description > short definite description > last name > first name > distal demonstrative + modifier > proximate demonstrative + modifier > distal demonstrative + NP > proximate demonstrative + NP > distal demonstrative (-NP) > proximate demonstrative (-NP) > stressed pronoun + gesture > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > zero

For the purpose of the present study, the elaborate scale in (15) has been reduced to four main categories, namely lexical NPs, demonstratives, pronouns, and zero forms (null subjects), the hierarchy of which was shown in diagram (10) in Chapter 2, repeated here as (16):

(16) Low Accessibility Marker	Lexical NP
	Demonstrative
↓	Pronoun
High Accessibility Marker	Null

Therefore, the effect of the four aforementioned accessibility features on these four referential forms will be analyzed in this paper.

### 3.4.1. The Effect of Discourse-Pragmatic Features

Before showing how subject arguments in the present study have been coded for the four discourse-pragmatic features, the relationship between these features and the realization of subjects in child language will be briefly discussed.

Regarding the feature *animacy*, controversial findings have been reported about the potential effect of this feature on children's subject choice; as Allen et al. (2008) put forth, one possible reason might be due to the rather different strategies researchers use for assigning accessibility values to animate vs. inanimate entities. As mentioned earlier in Section 3.3.2, in some studies inanimate objects such as dolls are coded as accessible for the feature *animacy*. Allen et al. (2008, p. 116) go on to refer to a probably weak relationship between this feature and subject realization in the sense that being animate or inanimate does not make a referent specifically salient or non-salient by itself. Therefore, it might be the case that the effect of *animacy* on subject choice is not 'strong' and 'significant'.

With respect to the feature *physical presence*, numerous studies have confirmed the prediction that highly informative argument forms are expected when a referent is not present in the physical context of the conversation. But the reverse is not always true; the physical presence of an entity, by itself, does not necessarily cause it to be realized in a low information form (Allen et al., 2008).

*Prior mention* is seen by most researchers to definitely have some influence on the realization of subject arguments in child speech. This is not only true of proponents of the discourse-pragmatic perspective, but also of grammatical as well as performance-based approaches (Allen et al., 2008, p. 109). A direct relationship is expected to exist between the accessibility of this feature and the subject arguments' degree of informativeness. A subject with a low information form such as a pronoun is expected when a referent has already been introduced in prior discourse (accessible value). But a referent with no prior mention (inaccessible value) is considered new and therefore is predicted to be realized by some highly informative form such as a lexical NP (e.g., Hughes & Allen, 2006, 2013).

With regard to *linguistic disambiguation*, as Allen et al. (2008) state, it is more likely that a speaker will use a subject argument with a high information form when a referent is linguistically ambiguous. As such, that particular referent will not be confused with some other potential referents. Having said that, the reverse is not generally correct; the mere unambiguity of a referent is not enough to make it salient to an interlocutor. Therefore, the

referent cannot be expected to be realized by a low information form only because it is accessible for the feature *linguistic disambiguation*.

### 3.4.2. Coding Strategies

As the last part of this section on the realization of subjects, the way the accessibility of the aforementioned features has been determined in the current data set will be explained in the following four dialogues, each of them representing one type of referential forms. The example in (17) illustrates the use of a subject realized as a null form:

(17) MOT: <is he> [//] does the teacher say sit down David?

CHI: sit down David.

MOT: oh my gosh now what?

CHI:  $\emptyset$  pulling them out. (2;5.04)

In (17), Matty has brought his David book<sup>20</sup> to his mother and is going to tell the investigator what happens in the book. His first utterance is not included as it is an imperative structure. However, the null subject in the final utterance refers to David. Thus, since he is considered to be animate, the subject is coded as accessible for *animacy*. The book and so the pictures of the characters are physically present. Therefore, the subject is coded as accessible for *physical presence*. Because David has been mentioned in the preceding utterances, the subject is also coded as accessible for *prior mention*. But there are two referents, 'David' and 'the teacher', in the preceding utterances that could potentially be the target referent of the null subject. Hence, the subject is coded as inaccessible for *linguistic disambiguation*. Thus, the null subject in this example has three accessible and one inaccessible features in total.

The example in (18) shows how one of the pronouns produced by Matty has been coded for accessibility:

(18) MOT: did you know that this was something that could make music?

MOT: I think they gave that to Jason when he was small because Jason used to play football.

MOT: did you know that (.) that your daddy used to play football?

CHI: he's a football. (3;1.29)

The subject of the child's utterance in (18) is the pronoun 'he'. It refers to his father, so it is accessible for the feature *animacy*. But his father is absent from the physical context of their

---

<sup>20</sup> A picture book for children with a mischievous boy called David as its main character.

conversation. Thus, it is coded as inaccessible for the feature *physical presence*. As he was previously mentioned, it is considered as accessible for *prior mention*. Regarding *linguistic disambiguation*, there are two potential referents for the subject, ‘Jason’ and ‘daddy’. Hence, it is coded as inaccessible for this feature. Therefore, the pronoun subject in this example has two accessible and two inaccessible features in total.

Example (19) shows the use of a demonstrative by the child (3;2.26):

(19) MOT: do you want to draw with it?

RYA: is that your chalkboard?

CHI: yeah.

RYA: I have a chalkboard too.

RYA: what are you drawing?

CHI: that's mountains.

As can be inferred from the interaction in (19), the deictic ‘that’ is inanimate (inaccessible) because it refers to mountains. Since Matty is drawing the mountains, it is considered to be physically present (accessible for the feature *physical presence*). No previous mention of the referent in the preceding five utterances can be found. Hence, it is coded as inaccessible for the feature *prior mention*. There is no linguistic ambiguity with regard to this referent. Therefore, the subject is coded as accessible for *linguistic disambiguation*. Thus, this demonstrative has two accessible and two inaccessible features in total.

In the last example, a case of NP subject is discussed for its accessibility features in (20):

(20) MOT: yeah what did we do at daddy's friend's house?

CHI: go to daddy's friends house.

MOT: we went to daddy's friends house.

CHI: and go outside.

MOT: and go outside +...

CHI: the ball went in the water. (2;5.18)

The first two of Matty’s utterances in (20) are in the first person and therefore have been removed from the data. But regarding the underlined lexical subject in the final utterance, it is inanimate. Therefore, it is coded as inaccessible for *animacy*. The ball is not physically present as they are talking about an event in the past. Thus, the subject is coded as inaccessible for the feature *physical presence*. Because this is the first time that the ball is being referred to in the preceding five utterances, the subject is new and therefore should be coded as inaccessible for *prior mention* as well. With regard to the coding for the feature *linguistic disambiguation*, what should be taken into consideration is all the other referents

in the preceding five utterances that might potentially be interfering if ‘the ball’ in (20) were not overtly realized; in other words, if the child’s last utterance in (20) were “∅ went in the water”, the null subject, based on the preceding utterances, would be interpreted as ‘we’. Therefore, since the underlined subject is ambiguous, it is coded as inaccessible for the feature *linguistic disambiguation*.<sup>21</sup> In total, this NP subject is fully inaccessible since all its four features have been assigned an inaccessible value.

The rest of the subject arguments in the current data were coded in the same way for all the four discourse-pragmatic features.

### 3.5. Data Selection Criteria

This section aims to give a detailed explanation regarding the selection criteria applied to the current data. Similar to what Hughes and Allen (2006, 2013, in press) did in their studies, only non-imperative, declarative, intelligible utterances containing a verb were chosen for this analysis. However, there were some cases, mainly at Time 1, where the child omitted the verb ‘to be’; in spite of lacking a verb, if relevant, those utterances have been included in the analysis as well. For example:

(21) CHI: that my bike. (2;5.04)

MOT: that's your bike (.) you wanna sit on it?

(22) CHI: the car on the snow. (2;6.11)

MOT: he gives him a hug and yes the car is on the snow.

(23) MOT: you can't see him?

MOT: is that the tractor part (.) can't see that part right?

CHI: it right there. (2;5.10)

It might be argued that there cannot really be any variation between null and overt subjects in examples such as the underlined in (21–23) because omitting the subject in these copula-less clauses would result in a noun phrase (‘my bike’), a prepositional phrase (‘on the snow’), and an adverbial phrase (‘right there’) respectively. Therefore, since the null subject version of such utterances would not be possible (as they contain no verbs), it is better not

---

<sup>21</sup> In other words, in order to determine the accessibility of the feature *linguistic disambiguation* for lexical NP subjects, all of them are treated as potentially zero.



to include them in the analysis.<sup>22</sup> But what has been taken into consideration regarding the inclusion or exclusion of such examples is the original versions prior to copula dropping, i.e., whether or not those utterances with potential copulas, which the child has happened to omit in examples (21–23), can typically allow for variation (null vs. overt). Since the overtness of subjects is not fixed in such examples, it means that they are in line with the precondition of this study. Hence, a total of 17 cases at Time 1 and two cases at Time 2 were included in the analysis.

As mentioned earlier at the beginning of this section, only non-imperative utterances were included for analysis in the present study. The reason why no structurally imperative utterances were included is the fact that subject omission is grammatically obligatory in this particular structure in English and therefore no variation between overt versus null subjects can be detected in imperatives (Hughes & Allen, 2013, p. 18). Accordingly, a total of 59 imperative utterances at Time 1 and 45 cases at Time 2 were excluded from the analysis.

In addition to declarative main clauses, which make up the major part of the data for this study, subordinate clauses were also included.<sup>23</sup> Thus, four cases of subordinate clauses at Time 1 and a total of 12 subordinate clauses at Time 2 were also included in the analysis.

As mentioned earlier about intelligibility, utterances with xxx (which indicates unintelligible speech) were mostly excluded from analysis. For example:

(25) MOT: what is this?

CHI: it's a xxx. (2;5.18)

(26) CHI: xxx is our little bed burrow. (3;1.29)

(27) CHI: and xxx goes there. (2;7.22)

Although the child's utterance in (25) contains a subject and a verb, since the referent of the pronoun is not known, its accessibility features cannot be determined. Therefore, such examples were excluded from analysis. The reason for exclusion of utterances such as (26–

---

<sup>22</sup> The only example in the selected files that contained no verb and still might be considered as a 'sentence-like' utterance even without a subject is the following:

(24) CHI: he still sick right there. (2;7.07)

MOT: he's still sick right there.

The potential null version of the child utterance in (24) would be "∅ still sick right there" which seems to be an acceptable case for the purpose of the present study, even without a verb.

<sup>23</sup> No null subjects were found in subordinate clauses of the current data. This is consistent with the results of other studies such as Valian (1991).

27) is the fact that it is not possible to know for sure if they had a subject or not in the first place, and if so, what kind of referential form they were. Thus, such examples were also excluded.

The only exception for inclusion of utterances with xxx was those cases which were full sentences despite containing xxx. Consider the following examples in (28–30):

(28) CHI: he xxx put that heat on. (2;7.07)  
MOT: he doesn't put the heat on you're right.

(29) CHI: this is a coin xxx. (2;7.22)

(30) CHI: xxx xxx he fell off the truck. (3;0.02)

Utterances such as (28–30) serve well for the purpose of the present study because they contain a subject and a verb, and the discourse-pragmatic features of the subject arguments can be coded for their accessibility. Therefore, a total of six cases at Time 1 and nine cases at Time 2 were included in this analysis.

Hughes and Allen (2013, p. 18) excluded their participants' "exact self-repetition, imitation, recitation of poems or songs, frozen forms, and routines" from their analysis. The following are some of the cases found in the current data that can be categorized under one of the above-mentioned types, and therefore, were also excluded:

(31) CHI: maybe that giraffe broke it!  
CHI: maybe that giraffe broke it. (3;1.29)

(32) MOT: yes he can.  
CHI: yes he can. (2;7.22)

(33) MOT: what kind of animal is this is guy?  
MOT: can you guess?  
MOT: it is a (..) fox.  
CHI: it's a fox. (2;5.18)

In (31), which shows an example of exact self-repetition, only the first utterance was counted and the repetition line was excluded from analysis. In examples (32–33), the child's lines are considered as imitation of his mother's and therefore were excluded as well.

With regard to imitation, there was an 'in-between' case which seemed not to be as straightforward as the previous ones. Consider the following dialogue between Matty and his mother in (34):

- (34) MOT: what happens here?  
 CHI: Ø throw (.) throw throw.  
 MOT: throw what?  
 CHI: Ø throw.  
 MOT: throw a banana?  
 CHI: Ø throw (b)anana.  
 MOT: throw an apple?  
 CHI: Ø throw an apple. (2;5.04)

In (34), the child's first utterance contains null subjects and therefore one of them was included in the analysis and the other two were removed as cases of self-repetition. His second utterance was also excluded because it is considered as a repetition of his previous line. Regarding the underlined utterances, although the child's intonation is different from his mother's, as indicated by the question marks and the periods, those utterances were considered to be a case of imitation and accordingly were removed from the data. The reason for this is that both of them involve the use of null subject forms for the third person referent.

Other examples of repetition are illustrated in (35–37):

- (35) CHI: he's hungry.  
 MOT: he's tiny or did you say he's hungry?  
 CHI: he's hungry. (2;6.11)
- (36) MOT: what did he bring for me?  
 CHI: he brought you xxx.  
 MOT: he brought some what?  
 CHI: he got some (.) he brought some pizza. (3;0.15)
- (37) CHI: they're gonna they're gonna.  
 MOT: they're gonna what?  
 CHI: you hide him then this time they're gonna see him. (3;2.18)

In cases such as (35–37), where the child is asked to repeat what he had just said, or complete his sentence, etc., only one of his utterances in each case was counted for this analysis and the rest were removed from the data.

Last but not least, consider the following examples in which immediate 'correction' or rewording seems to be done by the child:

(38) MOT: what does he do?

CHI: Ø gonna fly (.) this guy gonna fly. (2;6.11)

(39) CHI: hi cocoa.

CHI: are you doing good?

CHI: oh (.) he is (.) cocoa is not home. (3;1.29)

In such cases, only the child's ultimate choice of referent in an utterance is included which have been underlined in examples (38–39). As such, the first null subject in example (38) and the third person pronoun 'he' in example (39) were excluded from the study.

As discussed earlier in Section 3.3.6 about the feature *person*, discourse-pragmatically speaking, first and second person referents are different from third person referents in the sense that the former are always accessible; they are physically present, given, animate, etc. In other words, no distinction between present vs. absent, given vs. new, animate vs. inanimate can be found in first and second persons (Hughes & Allen, 2006, 2013). Thus, in order to prevent the first and second person referents from skewing the results, they have been excluded from the study. At Time 1, a total of 171 cases of first person and 28 cases of second person pronouns were omitted from the analysis and at Time 2, a total of 275 cases of first person and 73 cases of second person referents were excluded from the data.<sup>24</sup>

It is worth mentioning that a few cases were found in the selected data in which the child used his name or the third person pronoun to refer to himself as the following examples show:

(40) MOT: is it gonna be for you to open or for Erin to open?

CHI: Matthew open. (2;5.10)

(41) CHI: he's up on the stool. (3;2.18)

As Hughes and Allen (2013, p. 18) note, this strategy is frequently seen in child speech (and child-directed speech). In spite of the fact that the underlined subjects in the above examples (40–41) are grammatically third person, they are pragmatically first person as

---

<sup>24</sup> Taking a closer look at the first and second person results of the data shows that 37% of first person subjects and 33% of second person subjects were realized in null forms at Time 1. These percentages are (almost) twice as high as the third person cases (18%) at this time. At Time 2, for 6% of first person utterances and 4% of second person utterances, Matty produced null subject arguments. The same result was obtained for third person utterances (4%). The higher percentage of subject omission in first and second person utterances than in third person ones supports the idea of excluding the *person* feature from the analysis and assessing third person subject arguments separately. However, it also suggests that Matty drops subjects at a considerably higher rate than indicated by the current study at Time 1.

they are referring to the speaker himself, and therefore are considered as fully accessible. Hughes and Allen (2013) excluded such cases from their analysis as they might skew the results. In the present study, when it comes to such cases, the same strategy is applied. As a consequence, a total of five cases at Time 1 and two cases at Time 2 were also removed from this analysis.

And finally, interrogative contexts were another group that is not examined in the present study. As discussed earlier in Section 3.3.9 on *query*, this feature is not investigated because when a referent is a response to a question or is the subject of a question, it is always linguistically realized (Hughes & Allen, 2006, 2013). Therefore, the variation between null vs. overt subjects which is a precondition for the present study does not basically exist in such cases. Thus, all the cases of yes/no questions and wh-questions found in the selected data were excluded from the analysis. Accordingly, a total of 13 yes/no questions at Time 1 and 10 yes/no questions at Time 2 were removed from the current analysis.<sup>25</sup> With regard to the wh-questions, a total of 32 cases at Time 1 and 30 cases at Time 2 were also removed from the data.<sup>26</sup> In addition, the following two utterances at Time 2 were the only cases in which the subject arguments were the response to a question:

(43) MOT: Roman came and &hmm (.) who else came?  
CHI: Sophie came too. (3;0.02)

(44) MOT: here here's the one that you (.) one of your favorite ones what's on this one?  
CHI: stars are. (3;0.02)

---

<sup>25</sup> No null subjects were found in these yes/no questions.

<sup>26</sup> Regarding the wh-questions, only one case of null subject form was found. Consider the following dialogue between Matty and his mother at Time 2 about a recording device:

(42) CHI: why it's not going right now? (3;2.18)  
MOT: it is going it's just recording.  
MOT: it is going right now.  
CHI: why's just recording?  
MOT: why is it just recording?  
MOT: (be)cause that's how it does it.

The appearance of a null subject in the underlined utterance might be explained by the fact that the child is not asking about the subject (the recording device), but the verb (the act of recording). It should be noted that from the grammatical perspective, such cases are expected to be rare if any; for instance, the occurrence of a missing subject in structures like (42) runs counter to what Rizzi's truncation hypothesis (Section 2.2.2) would predict.

Questioning about ‘Sophie’ and ‘stars’ in the above examples, as Hughes and Allen (2013, p. 19) maintain, “necessitates the use of an overt subject in English”. Thus, they have not been included in this study.

### **3.6. Data Analysis**

For the analysis of the current data set, Hughes and Allen’s (2006, 2013, in press) framework was adopted; all the selected third person subject arguments at Time 1 and Time 2 were first assigned a binary value (accessible or inaccessible) for each of the four discourse-pragmatic features under investigation, namely *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*. Then, the distribution of accessible vs. inaccessible features for each referential type—lexical NP, demonstrative, pronoun, and null form—was compared in order to determine whether there is any link between accessibility of features and the child’s choice of subject arguments. This analysis shows the effect of the discourse-pragmatic features individually and another analysis is needed to test whether or not they really work in combination. In order to investigate the incremental effect of these features, first a score of 0 through 4 was assigned to each subject argument at both Time 1 and Time 2. These scores show the overall level of accessibility for each subject argument; the score 0 indicates a fully accessible argument while the score 4 represents a fully inaccessible subject. Therefore, the numerals 0–4 demonstrate the number of features that are inaccessible for a particular subject argument. In other words, when an argument has been coded accessible for all four discourse-pragmatic features, it gets a 0 score. Subjects with only one accessible feature are categorized with a score of 3 and so on.

In the current analyses, the accessibility factor (accessible vs. inaccessible or number of features inaccessible) is the independent variable and the argument form (with four possibilities) is the dependent variable.

### **3.7. Summary**

To give a summary of the final data, before moving on to the results section, it consists of the utterances of Matthew, a monolingual English-speaking child from Western New York State. A total of 260 utterances from six files at Time 1 and 240 utterances from six files at Time 2, with third person referents as their subject arguments were analyzed for the present

study. The subject arguments, which were categorized under four possible referential forms, were coded as accessible or inaccessible for the four previously-defined discourse-pragmatic features. The effect of these features on the realization of subject arguments was investigated both individually and cumulatively. The information about the participant and the data set is summarized in Table 1:

	Time 1	Time 2
Age	2;5.04–2;7.22	3;0.02–3;3.09
Sessions	6×30 min.	6×30 min.
Total # of subject contexts (3 <sup>rd</sup> person)	260	240
MLUw	Range: 2.5–3.6 Mean: 2.9	Range: 3.3–4.3 Mean: 3.7
Referential forms	NPs, demonstratives, pronouns, null forms	
Accessibility features	Animacy, physical presence, prior mention, linguistic disambiguation	

Table 1: Participant and data information summary

## CHAPTER 4

### Results

#### 4.1. Introduction

This chapter reports on the results of the analysis carried out for the selected data. First, the distribution of each referential form at Time 1 and Time 2 is illustrated in Table 2.

Table 2: Distribution of referential forms at Time 1 and Time 2

Period (total # of subj. arguments)	NP	Dem	Pro	Null
Time 1 (260)	48 (18%)	51 (20%)	115 (44%)	46 (18%)
Time 2 (240)	52 (22%)	41 (17%)	137 (57%)	10 (4%)

As shown in Table 2, the proportion of null subjects at Time 1 is more than four times higher than at Time 2. At Time 1, Matty is at an age (2;5.04–2;7.22) when children, according to numerous studies (e.g., Hughes & Allen, 2013; Hyams, 2011) tend to systematically drop subjects. Therefore, the higher percentage of null subjects at this time is in line with the prediction that children omit subjects at a particular stage of development. At Time 2, Matty has just turned three (3;0.02–3;3.09) which is considered to be towards the end of the NS stage. Thus, the significant drop in the number of null forms is expected. While the number of null subject arguments has decreased over time, the use of pronouns has increased, as expected, which suggests a more adult-like grammar. This increase in the number of pronouns over time is consistent with what is expected under some versions of the grammar-based approach.<sup>27</sup> For example, Hyams and Wexler (1993, p. 444) refer to it as a “trade-off between null subjects and pronouns”. According to such grammatical accounts, empty subject arguments produced by children at the NS stage in non-pro-drop languages are in fact those arguments that older children as well as adults would realize as pronouns (Allen et al., 2008, p. 105). As shown by the results in Table 2, the total proportion of the pronominal and null forms, together, is roughly the same over time (61% of the data at both stages). This is also what the grammatical approach would expect (Hyams & Wexler, 1993).

---

<sup>27</sup> As mentioned earlier in Section 2.3.2, from the performance-based perspective, Gerken (1991) also claims that dropped elements are more likely to be pronouns as they are prosodically weak.



With respect to the proportion of lexical and deictic subjects, as shown in Table 2, the change in the child's choice of demonstratives and NPs was not substantial from Time 1 to Time 2. The percentages, however, suggest that the child uses lexical items slightly more at Time 2 while his use of deictic forms has dropped a little. The sum of the proportions of the lexical and demonstrative forms has also remained the same over time (38% and 39% of the data at the two age ranges). The steady proportion of the lexical items over time is another prediction that the grammatical model has made (Hyams & Wexler, 1993). It is good to note that studies such as Guerriero et al. (2006) also obtained similar results for the English-speaking children from ages 1;9 to 3;0, in terms of an increase in the use of pronouns as well as a decrease in subject omission, while the number of lexical NPs stayed about the same over time.

The remainder of this chapter is structured as follows. Section 4.2 reports the findings on the effect of each individual discourse-pragmatic feature on the child's use of referring expressions at Time 1 and Time 2. This section also examines whether the participant is sensitive to information flow and whether this changes over time. Comparing the results can reveal which one of these features is more influential with respect to the subject choice. In Section 4.3, the results for the incremental effect of accessibility features are given. The findings indicate how these features interact and work in combination. The results of this section also show how the total level of accessibility affects subject realization over time.

## **4.2. The Effect of Individual Accessibility Features**

### **4.2.1. Findings at Time 1**

This section reports the findings on the effect of individual discourse-pragmatic features at Time 1. Table 3 represents to what extent each of the four referential types, namely lexical NPs, demonstratives, pronouns, and null forms, are accessible at Time 1 for the features *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*. The accessible results for each feature are marked by a plus (+) and the inaccessible results are indicated by a minus (-) in this table. The triple hyphen/minus (---) is used in those cases where that particular form is not attested. The 'total' row shows how the proportion of accessible and inaccessible values is distributed for all the four referential forms altogether, for each discourse-pragmatic feature.

Table 3: Accessibility of each referential form for the four features at Time 1

Feature Referential form	Animacy	Physical presence	Prior mention	Linguistic disambiguation
NP	+31 (65%) -17 (35%)	+30 (62.5%) -18 (37.5%)	+19 (40%) -29 (60%)	+19 (40%) -29 (60%)
Dem	+12 (24%) -39 (76%)	+51 (100%) ---	+15 (29%) -36 (71%)	+34 (67%) -17 (33%)
Pro	+74 (64%) -41 (36%)	+97 (84%) -18 (16%)	+100 (87%) -15 (13%)	+106 (92%) -9 (8%)
Null	+38 (83%) -8 (17%)	+42 (91%) -4 (9%)	+41 (89%) -5 (11%)	+33 (72%) -13 (28%)
Total	+155 (60%) -105 (40%)	+220 (85%) -40 (15%)	+175 (67%) -85 (33%)	+192 (74%) -68 (26%)

Taking a closer look at the total number of accessible and inaccessible referents for each feature shows that the proportion of accessible items is higher than the inaccessible ones for all the four features. However, this difference is less clear for the features *animacy* and *prior mention* which means that there are more inaccessible referents for these two features in the current data at Time 1 than the other two, i.e., *physical presence* and *linguistic disambiguation*. The fact that Matty uses a higher proportion of accessible referents in his speech at this time is not surprising because young children generally tend to talk mostly about objects/persons that are known, physically present, concrete, etc.

In order to get a better picture of the results, the distribution of accessible vs. inaccessible features for each of the four referential forms is reported in a separate chart, under separate subsections.

#### 4.2.1.1. Lexical NPs

The results for the lexical NP subjects in relation to each discourse-pragmatic feature at Time 1 are displayed in Figure 1; each column represents the accessibility of all 48 lexical NP subjects for a particular feature.

With respect to the use of lexical forms, Matty produces more NPs when referents are inaccessible for the features *prior mention* and *linguistic disambiguation*. The proportion of lexical items, however, is higher when referents are animate and physically present, which means they are accessible for the features *animacy* and *physical presence*.

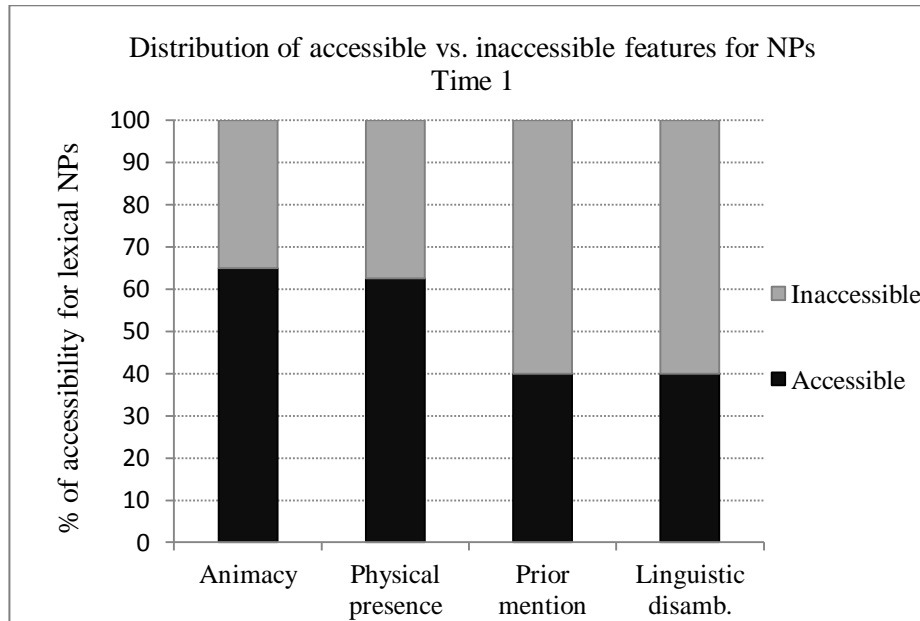


Figure 1

As discussed earlier in Section 3.4 on the realization of subjects in adult speech, the more accessible a referent is, the less linguistic information is needed for the referent to be identified with (Ariel, 2001); in spite of the fact that lexical NPs are highly informative forms and carry the least accessible information, Matty at Time 1 produces them even for some highly accessible referents. This, as Hughes and Allen (2013, p. 21) state, can in part be related to the fact that children usually repeat third person lexical forms in their speech, as the underlined lexical subject in the following conversation between Matty (2;7.22) and his mother shows:

- (45) MOT: let's see (.) yep there he goes.  
 MOT: who is this?  
 CHI: Donald Duck.  
 MOT: oh it's Donald Duck (.) Donald Duck what are you going to buy at the store today?  
 MOT: what do ducks eat do you know?  
 CHI: ducks eat (.) some (.) rocks.  
 MOT: they eat some rocks.

Although the underlined subject can be easily identified with a low information form such as a pronoun, which is what his mother uses afterwards, Matty repeats the lexical NP. This, as a common feature of child speech, can be one of the reasons for the higher than expected rate of accessibility in the lexical category at Time 1.

Considering the fact that *prior mention* and *linguistic disambiguation* have the higher inaccessible percentages in this group, it can be inferred that these features play a determining role here and have a stronger effect on the realization of lexical subject arguments. The following dialogue between Matty (2;5.18) and his mother can help make this point clear:

- (46) CHI: yeah (.) it's mine.  
MOT: it's your present?  
MOT: oh is it Matty's birthday?  
CHI: yeah [=! whispering].  
MOT: what do you do on your birthday?  
CHI: Buddy going poop on the floor.  
MOT: oh (.) Buddy's going poop on the floor.

In (46), the child and his mother are talking about birthday and presents that Matty suddenly changes the topic and starts talking about Buddy—their dog who is very ill and is going to die and see his late grandfather. Buddy is alive and physically present. Thus, the referent is accessible for the features *animacy* and *physical presence*. There is no mention of Buddy in the preceding five utterances. Therefore, it is new and inaccessible for the feature *prior mention*. It is also inaccessible for the feature *linguistic disambiguation*, because if the underlined subject were not linguistically realized, it could refer to Matty himself. Therefore, the child's use of an NP in such a context can be explained by the fact that referents which have not been previously mentioned are considered new and therefore are expected to be realized in a highly informative form. Moreover, when there are some potential interfering referents in the linguistic context of a conversation, using a highly informative form can guarantee that the intended referent would not be confused with any other potential referents.

Accordingly, it can be concluded that *prior mention* and *linguistic disambiguation* seem to be the most influential features with respect to Matty's lexical subject choice, in the sense that these forms are preferred when a referent is inaccessible for the features *prior mention* and *linguistic disambiguation*.

#### 4.2.1.2. Demonstratives

The findings on the distribution of accessibility features for demonstrative subject arguments are reported in this section. Each column in Figure 2 represents the accessibility of all 51 demonstrative subjects for a particular feature.

With regard to the distribution of deictic forms, the results show that no demonstrative with inaccessible value for the feature *physical presence* is found at Time 1. It means that referents of all demonstratives produced by Matty at Time 1 were present in the physical context of the conversation. This is not surprising as deictic forms, by nature, rely on the presence of their referents to be realized. The number of demonstrative forms is also quite high for referents with accessible value for the feature *linguistic disambiguation*. However, Matty produces more demonstratives when subject arguments are inaccessible for the features *animacy* and *prior mention*.

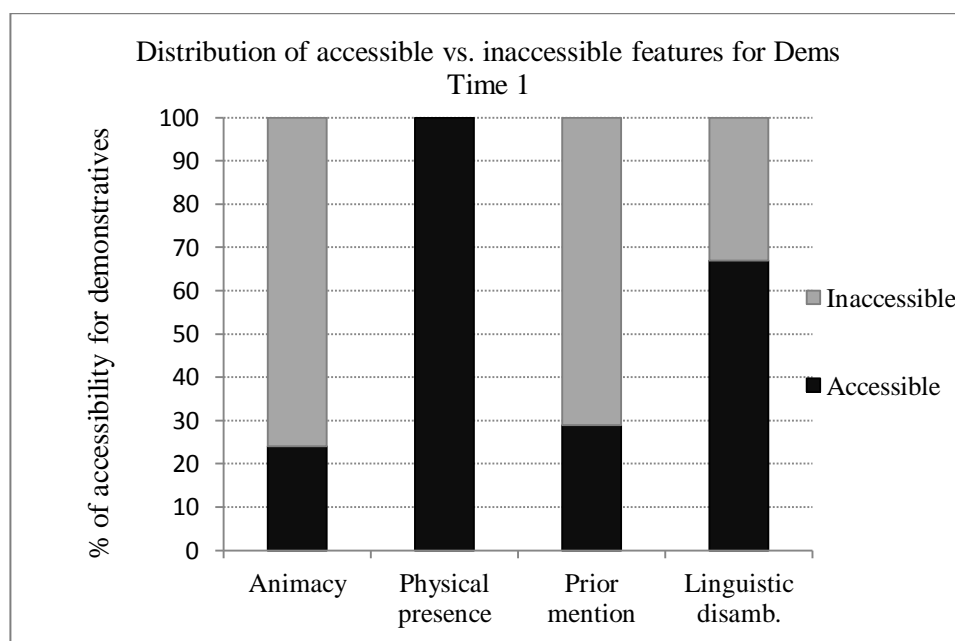


Figure 2

As mentioned earlier in Chapter 3, demonstratives in adult speech are considered medium accessibility markers because they are not as highly accessible as pronouns and zero anaphora, nor are they as inaccessible as lexical NPs (Ariel, 2001); this property of deictic forms is also reflected in Figure 2. For two (*physical presence* and *linguistic disambiguation*) out of the four discourse-pragmatic features studied in this paper, the majority of subject arguments in this referential form have accessible value and for the other two (*animacy* and *prior mention*), the majority have inaccessible value. For example, the following dialogue between Matty (2;7.07) and his mother can show this pattern:

(47) MOT: dump them out.  
CHI: maybe Erin can play with these.  
MOT: maybe Erin can play with those (.) maybe.  
CHI: wanna play with these (.) Erin?  
MOT: oh I think (.) you wanna play with these with Erin then you have\_to show her  
what's she gotta do what's she supposed to play with them?  
CHI: this is a sock.  
MOT: it's a sock?

The underlined subject argument in (47) refers to a ‘sock’ which is inanimate. It is present in the physical setting of the conversation. No prior mention of the referent has been made in the preceding five utterances which makes the subject inaccessible for the feature *prior mention*. The referent is linguistically unambiguous as there is no potential competitor referent for this subject argument in the linguistic context. Therefore, this subject is accessible for *physical presence* and *linguistic disambiguation*, and inaccessible for *animacy* and *prior mention*. It is interesting to note that once the ‘sock’ has been introduced in the context, Matty’s mother, as the final utterance in (47) shows, uses a pronoun (a low information form) to refer to the same referent.

With respect to the effect of these four features on the child’s choice of demonstratives, the proportion of accessible vs. inaccessible features for this referential type suggests that all four features do have some meaningful influence on the child’s subject realization with *physical presence* being the strongest one due to its 100% accessible rate.

#### **4.2.1.3. Pronouns**

The distribution of accessible vs. inaccessible features for all 115 pronouns at Time 1 is displayed in Figure 3.

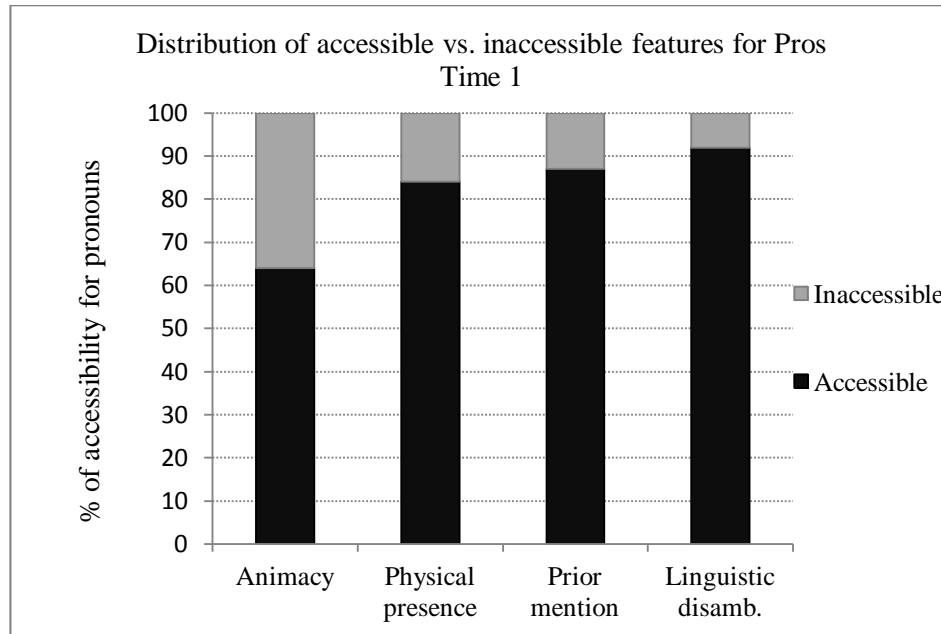


Figure 3

As shown in Figure 3, Matty produces more pronominal forms when referents are accessible; this is in accordance with what is expected from such low information forms. Since they are highly accessible, they can be easily identified by interlocutors, and therefore there is no need to use redundant linguistic information to refer to them (Ariel, 2001).

The findings also show that *animacy* has the lowest proportion of accessible referents. This suggests that being animate or inanimate does not have a determining role in the realization of subjects in this group as there are lots of inanimate entities in the selected data that Matty uses pronominal forms to refer to. For example, consider the following conversation between Matty (2;5.18), his mother, and an investigator (Erin):

- (48) ERI: is this the mean train?  
 CHI: No.  
 MOT: what is that?  
 CHI: a phone.  
 MOT: a phone (.) hello (.) hello.  
 CHI: it's mine.

The underlined subject 'it' in (48) refers to 'a phone'. It is inanimate, but physically present, given, and linguistically unambiguous. In 61% of the cases with this accessibility pattern (28 out of a total of 46 cases) in the current data at Time 1, the child uses pronouns as the subject argument of his utterances.

With respect to the effect of discourse-pragmatic features on the realization of pronominal subject arguments, taking the high percentage of accessible referents for *physical presence*, *prior mention*, and *linguistic disambiguation* into consideration, it seems that all of the three have an influence on the child's choice of referential forms, with the last two features being the strongest ones. In other words, pronouns are used when referents are highly accessible.

#### 4.2.1.4. Null Forms

The distribution of all 46 null subjects at Time 1 is illustrated in Figure 4. Each column represents the accessibility of the null subject forms for a particular feature. Similar to pronouns, all of the four features in this category have a considerably higher percentage of accessible values as well.

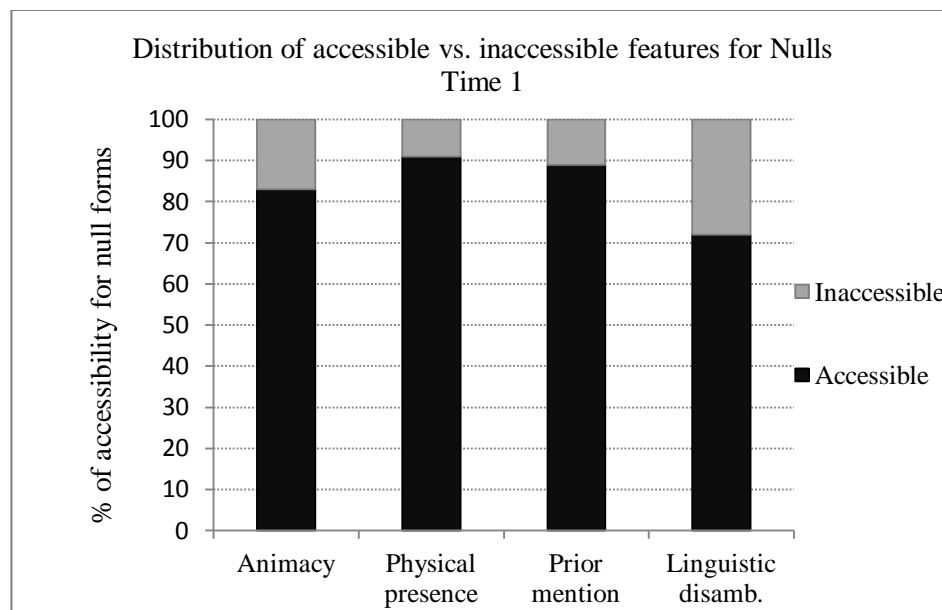


Figure 4

Based on the results shown in Figure 4, all null subjects at Time 1 are accessible for the features under investigation at a rate of 72% or more. This suggests that the child does have some sort of sensitivity to discourse-pragmatic features, even when he is dropping a syntactically-obligatory subject argument. For instance, consider the continuation of example (46) in the following dialogue:



- (49) MOT: oh is it Matty's birthday?  
 CHI: yeah [=! whispering].  
 MOT: what do you do on your birthday?  
 CHI: Buddy going poop on the floor.  
 MOT: oh (.) Buddy's going poop on the floor.  
 CHI: ∅ going to grandpa angel. (2;5.18)

The underlined null subject refers to Buddy, their dog. As discussed earlier in example (46), the first time Matty wants to refer to the dog, he uses a lexical NP because it was new information and linguistically ambiguous. But once it is previously mentioned in the linguistic context, the child omits the subject which is fully accessible, i.e., animate, physically present, given and linguistically unambiguous.<sup>28</sup> Cases such as (49) once again support the idea that children, as young as Matty, are aware of the information flow even when they are at the NS stage.

Let us consider one last point before moving on to the findings at Time 2. As the *linguistic disambiguation* column in Figure 4 shows, this feature has the highest percentage of inaccessible referents among the four. This can be partly explained by the fact that more than 60% of the null subject arguments at Time 1 with inaccessible value for the feature *linguistic disambiguation* belong to a context where Matty (2;5.04) has brought one of his picture books and is telling its story. Earlier in Chapter 3, part of the following dialogue was analyzed in example (17), repeated here as (50):

- (50) MOT: <is he> [//] does the teacher say sit down David?  
 CHI: sit down David.  
 MOT: oh my gosh now what?  
 CHI: ∅ pulling them out.  
 MOT: pulling them out and then what happens here?  
 CHI: ∅ reading the book.  
 MOT: reading the book okay and then what?  
 ...

The underlined empty subjects in (50) are accessible for all the features except *linguistic disambiguation*; because they can refer to either David or his teacher. But using a linguistically ambiguous referent in such a context by Matty does not necessarily refute what was stated earlier in terms of his sensitivity to discourse-pragmatic features. The

---

<sup>28</sup> The null subject in (49) can grammatically refer to Matty himself too, and as such, it should be considered as inaccessible for the feature *linguistic disambiguation*. But semantically, it cannot be the case since by “going to grandpa angel”, they mean “dying”. Therefore, the null subject has been coded as accessible for *linguistic disambiguation*.

reason is that there are some other factors at play which could not be examined for the current data set since the selected files were not in video format. For instance, perhaps, both the child and his mother are jointly attended to the same referent, e.g., by pointing to the picture of David or his teacher, which again can be in favor of the conclusion earlier made about Matty's sensitivity to the information flow, as *joint attention* is another discourse-pragmatic feature that affects children's use of referential forms (e.g., Skarabela et al., 2013).<sup>29</sup>

Regarding the effect of discourse-pragmatic features on the realization of null subjects, the distribution of accessibility in Figure 4 suggests that all four features are influential with *linguistic disambiguation* being the least salient one.

#### 4.2.1.5 Time 1 Summary

To sum up the findings at Time 1, it can be concluded that the child is aware of the information flow when it comes to the use of referential forms. As the results show, he uses low information forms, i.e., pronouns and null subjects, when referents are highly accessible, and medium accessibility markers, i.e., deictic forms, when referents are neither highly accessible nor highly inaccessible. The pattern for lexical NPs is not as clear nevertheless since a higher than expected rate of accessible values is found for this referential type in spite of the fact that lexical NPs are highly informative referring expressions and are expected to be realized in contexts with a low level of accessibility. As explained in Section 4.2.1.1, this is partly related to a child speech feature according to which children tend to repeat third person lexical NPs regardless of their level of accessibility (e.g., Hughes & Allen, 2013).

With regard to the weight of the features investigated in this study at Time 1, the distribution of accessible vs. inaccessible values indicates that these features are not equally significant in determining the child's choice of referential forms and some of them might be stronger or weaker. The findings suggest that *prior mention* and *linguistic disambiguation* with inaccessible value are highly influential on the realization of lexical NPs. In other words, referents that are linguistically ambiguous and have not been previously mentioned are more likely to be realized as lexical NPs. For demonstrative forms, all the features seem to play some role with *physical presence* with accessible value being the most determining

---

<sup>29</sup> It can also be argued that in (50), David is mentioned last before the null subjects, and it seems more likely that a null element should refer back to last mentioned.

one. In the realization of pronominal forms, all the features except for *animacy* are quite effective. The features *prior mention* and *linguistic disambiguation* had the highest percentages of accessibility and therefore seem to be more influential. In other words, when referents have been previously mentioned and there are no competitor referents for them, they are more likely to be realized as pronouns. Regarding the use of null forms, *linguistic disambiguation* seems to be the least significant feature in determining Matty’s choice of subject arguments at Time 1.

In the next section, the findings at Time 2 are reported. The results will also be compared to those of Time 1 to see whether there is any change or consistency in the individual discourse-pragmatic features’ patterns and their possible effects on the realization of subject arguments over time.

#### 4.2.2. Findings at Time 2

The results for individual features at Time 2 are given in Table 4. These findings show to what extent each of the four referential types, namely lexical NPs, demonstratives, pronouns, and null forms, are accessible for the features *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*. The accessible results for each feature are marked by a plus (+) and the inaccessible results are indicated by a minus (-) in Table 4. The triple hyphen/minus (---) is used in those cases where that particular form is not attested. The ‘total’ row shows how the proportion of accessible and inaccessible values is distributed for all the four referential forms altogether, for each discourse-pragmatic feature.

Table 4: Accessibility of each referential form for the four features at Time 2

Feature Referential form	Animacy	Physical presence	Prior mention	Linguistic disambiguation
NP	+25 (48%) -27 (52%)	+23 (44%) -29 (56%)	+15 (29%) -37 (71%)	+21 (40%) -31 (60%)
Dem	+2 (5%) -39 (95%)	+36 (88%) -5 (12%)	+21 (51%) -20 (49%)	+26 (63%) -15 (37%)
Pro	+87 (63.5%) -50 (36.5%)	+90 (66%) -47 (34%)	+114 (83%) -23 (17%)	+127 (93%) -10 (7%)
Null	+6 (60%) -4 (40%)	+5 (50%) -5 (50%)	+10 (100%) ---	+9 (90%) -1 (10%)
Total	+120 (50%) -120 (50%)	+154 (64%) -86 (36%)	+160 (67%) -80 (33%)	+183 (76%) -57 (24%)

As the total numbers of accessible and inaccessible referents for each feature show, the proportion of accessible items at Time 2 is still higher than the inaccessible ones for the features *physical presence*, *prior mention*, and *linguistic disambiguation*. However, this proportion has become equal for the feature *animacy*. Compared to Time 1, it is interesting to note that the percentages of accessible and inaccessible referents have remained the same for the features *prior mention* and *linguistic disambiguation*, while for the other two features, i.e., *animacy* and *physical presence*, the amount of inaccessible referents has increased. With regard to *physical presence*, the proportion of inaccessible referents has more than doubled. Such increases support the fact that when children get older, they talk more about entities that are not necessarily present in the physical context of their conversations as well as inanimate objects including abstract concepts.

Similar to the previous section, the results for each of the four subject types are reported in a separate chart, under separate subsections.

#### 4.2.2.1. Lexical NPs

Figure 5 demonstrates the results for all the 52 lexical subjects at Time 2 in terms of each discourse-pragmatic feature.

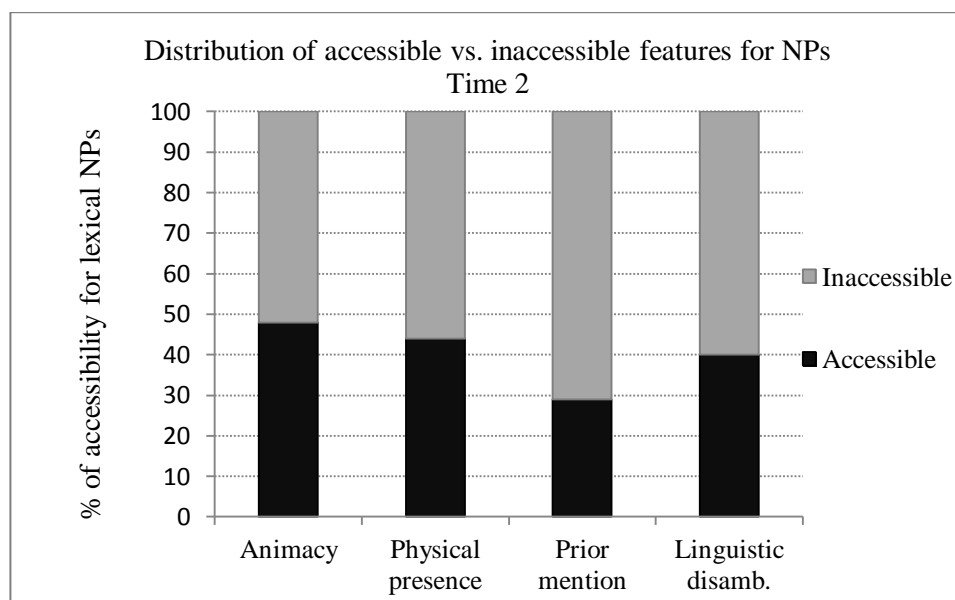


Figure 5

The findings at Time 2 show that Matty produces more lexical forms when referents are inaccessible. This is in line with the prediction that when a referent cannot be easily identified by a listener because it does not carry enough accessible information, the child

uses a highly informative form (here a lexical NP) to make the identification easier for the listener. Comparing the results with those at Time 1 shows that the proportion of inaccessible referents has increased for almost all the features, the only exception is *linguistic disambiguation* for which percentage of inaccessible values has remained the same. Since such a pattern with a higher proportion of inaccessibility is what is expected in adult speech, it shows that Matty's choice of lexical forms has become more adult-like over time.

Taking a closer look at the distribution of accessibility in Figure 5 also reveals that the subject arguments in this group are nearly equally accessible and inaccessible for the features *animacy* and *physical presence*, while *prior mention* and *linguistic disambiguation* still are inaccessible at a higher rate. Therefore, it can be concluded that similar to Time 1, *prior mention* and *linguistic disambiguation* continue to be significantly influential in the realization of lexical subjects. However, the role of the other two features cannot be ignored; it is interesting to note that more than one third of the lexical subjects at Time 2 are inaccessible for the features *prior mention* and *linguistic disambiguation* plus one other feature, *animacy* or *physical presence*. For example:

- (51) CHI: I think I got them all for me.  
MOT: you got them all for you.  
MOT: where did you get (th)em?  
CHI: actually found (.) were on the play+ground.  
MOT: oh actually.  
CHI: and the silver one came from the ground. (3;2.26)

In (51), Matty is talking about a collection of rocks which is present in the physical context. The underlined subject refers to a specific rock. Therefore, it is inanimate. It is physically present. The referent has not been mentioned in the preceding five utterances, therefore, it is new.<sup>30</sup> And it is also inaccessible for the feature *linguistic disambiguation* because if it were not linguistically realized, it would be assumed to be referring to all the rocks rather than a specific one ('the silver one'). Thus, in this case, the NP subject is inaccessible for the features *prior mention* and *linguistic disambiguation* plus *animacy*.

---

<sup>30</sup> It is good to note that although the underlined NP in (51) has not been previously mentioned, it refers to a member of a collection that has already been mentioned in the preceding utterances.

As an example of referents with inaccessible values for the features *prior mention* and *linguistic disambiguation* plus *physical presence*, consider the following conversation between Matty (3;1.29) and his mother:

(52) MOT: &um &hmm the little cats.

MOT: and what happened with the little cats when we were there?

CHI: one of the cats got (.) hurt.

MOT: one of the cats got hurt.

MOT: how'd it get hurt?

CHI: &ah when (.) &ah (.) when the (.) the hound dog was mean.

The underlined subject argument in (52) refers to a dog which is not physically present. Therefore, it is animate but not physically present. It has not been previously mentioned and it is linguistically ambiguous. Hence, the lexical subject argument in this case is inaccessible for the features *prior mention* and *linguistic disambiguation* plus *physical presence*. This makes sense from the point of view that lexical NPs carry a lot of information themselves. However, it is also important to point out that in this case Matty and his mother clearly are discussing referents that are accessible (familiar) to the two of them even though they are inaccessible with the relevant features. When Matty uses the definite NPs in the target utterance, this is because they are sharing context and Matty assumes that his mother can identify those referents. But this is not reflected in the division made by Hughes and Allen (2006, 2013).<sup>31</sup>

Examples such as (51–52) are in favor of the assumption that features overlap and work in combination. More on this topic will be provided in Section 4.3 on the cumulative effect of discourse-pragmatic features.

#### 4.2.2.2. Demonstratives

With respect to the child's choice of demonstratives, Figure 6 shows the distribution of accessibility for all 41 cases of this referential type at Time 2.

---

<sup>31</sup> Such examples indicate that lexical NPs do not all behave the same and that the question of which referents are accessible is really more complex than assumed here. This is also true of lexical NPs when it comes to definiteness in the sense that definites are more reliant on prior mention than indefinites. The majority of the lexical NPs used by Matty in the current data set were definite (90% of the cases at Time 1 and 79% at Time 2). The small proportion of indefinites made it impossible to compare the behavior of definite and indefinite NPs with regard to the accessibility features, especially *prior mention*.

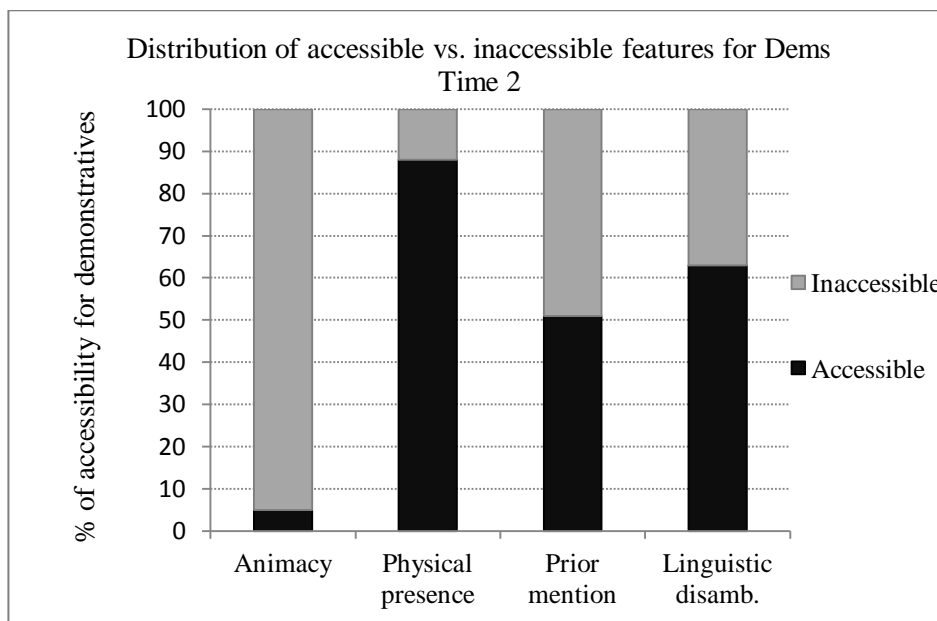


Figure 6

As displayed in Figure 6, the results show that almost all the demonstratives at Time 2 are inanimate and close to 90% of them are physically present. The proportion of demonstratives is almost equally accessible and inaccessible for the feature *prior mention*. In addition, Matty produces more demonstratives for linguistically unambiguous referents. Comparing these findings with the results at Time 1 shows that the proportion of inaccessible referents has increased for the features *animacy*, *physical presence*, and *linguistic disambiguation* while the percentage of demonstrative forms with inaccessible value for the feature *prior mention* has decreased more than 20%.

The high percentage of inaccessible referents for the feature *animacy* can be explained by the assumption that demonstratives typically refer to inanimate entities. The pattern of accessibility for the rest of the features in Figure 6 is in fact consistent with the idea that demonstratives, assumed to be inherently definite (Lyons, 1999), are used in a way more similar to definites in adult speech, although demonstratives are deictic forms and generally rely on the presence of their referents, which the high proportion of accessible values for the feature *physical presence* here also shows. The decrease in the proportion of demonstratives with accessible value for this feature at Time 2 suggests that the child does not merely rely on the physical presence of referents in order to realize them in the form of demonstratives. In fact, they might be known to the interlocutors even without being previously mentioned. This is another similarity between demonstratives and definites. Alternatively, demonstratives might be found in contexts where, as Lyons (1999, p. 53) notes, they have

been previously mentioned, but this was far back in the discourse, and this makes the identification of an intended referent difficult for a listener due to the presence of other potential interfering referents.

Taking a closer look at the distribution of accessibility features for demonstratives at Time 2 also reveals that this referential type, similar to at Time 1, is still treated as medium accessibility markers when it comes to the child's choice of referential forms.

Regarding the salience of the discourse-pragmatic features in the category of demonstratives, it seems that *animacy* with inaccessible value and *physical presence* with accessible value have the most influential role when it comes to the child's choice of subject argument in this group. However, the role of the other two features cannot be ignored. For example, consider the following dialogues between Matty (3;2.18) and his mother:

(53) MOT: they're gonna what?

CHI: you hide him then this time they're gonna see him.

MOT: okay so they're gonna hide him and this time they're gonna see him okay.

CHI: where you hide him?

MOT: oh you want me to hide him?

CHI: (be)cause that's just a truck.

(54) MOT: oh I see.

CHI: there's (.) the motorcycle.

MOT: there's the motorcycle.

CHI: and the wheel.

MOT: okay try.

CHI: this can't stay on there.

In (53), the underlined subject argument is inaccessible for the feature *animacy*, but accessible for the feature *physical presence*. It has not been previously mentioned, therefore, it is inaccessible for *prior mention*, but accessible for *linguistic disambiguation* as there is no competitor referents for it in the preceding five utterances. In (54), the underlined demonstrative is again inaccessible for the features *animacy* and accessible for the feature *physical presence*. But since it has been previously mentioned, it is accessible for *prior mention* while it is inaccessible for *linguistic disambiguation* as it can be referred to either 'the motorcycle' or 'the wheel'. In the current data set, close to half of the demonstrative subjects at Time 2 have one of these two accessibility patterns, mentioned in (53–54).



#### 4.2.2.3. Pronouns

Close to 60% of all subject arguments at Time 2 are realized in the form of pronouns. The distribution of informativeness features for all 137 pronouns at Time 2 is displayed in Figure 7.

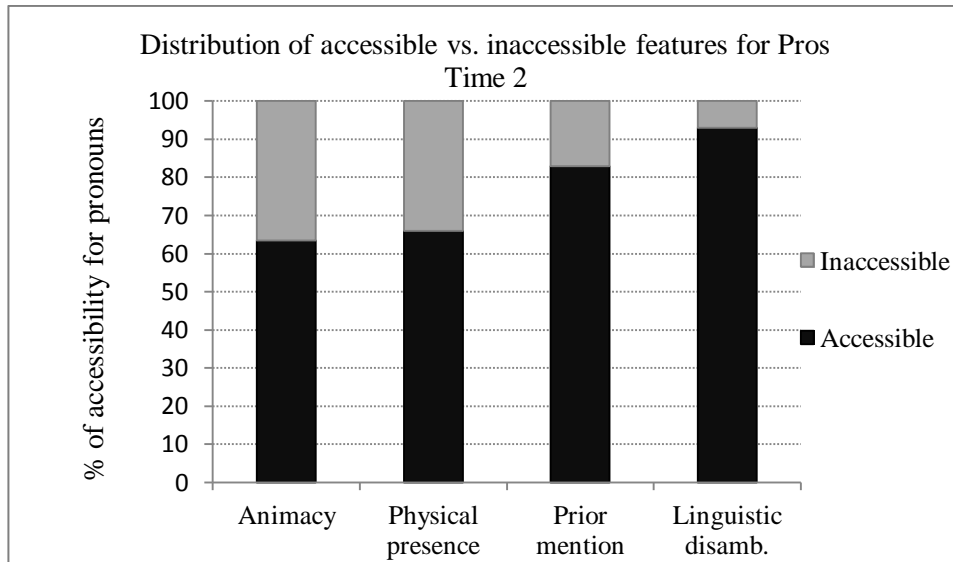


Figure 7

As can be seen in Figure 7, Matty produces more pronouns when referents are accessible. The high proportion of accessible features in this category demonstrates that the child pays attention to the degree of informativeness when he uses pronouns. As expected, the more accessible a referent is, the more easily it can be identified and therefore, there is no need to use highly informative overt subjects. Comparing the results of this category over time shows that the proportion of accessible vs. inaccessible referents has remained almost the same at both times, the only exception is the feature *physical presence* for which the proportion of accessible referents has decreased over time.

As Figure 7 shows, the features *animacy* and *physical presence* have approximately the same distribution of accessible vs. inaccessible referents for pronouns at this time. However, more than 50% of all the pronouns at Time 2 are either fully accessible or have only one inaccessible feature, *animacy*. For example, consider the following two dialogues between Matty and his mother:

(55) MOT: is that his mommy in there?

CHI: no.

CHI: he's looking for his momma.

MOT: oh he's looking for his momma.

CHI: no not in here.

MOT: so where is his momma?

CHI: she's right here. (3;0.15)

(56) CHI: umm (.) I don't know (.) what?

MOT: what does it look like?

MOT: it's what?

CHI: what?

MOT: it looks like a shard [?] of sea glass.

CHI: it's not called sea glass. (3;2.26)

In (55), Matty and his mother are talking about a tortoise and his mom which are physically present. The underlined pronoun is fully accessible as it is animate, present, given, and linguistically unambiguous.

In (56), the underlined subject refers to a specific rock. Therefore, it is coded as inanimate. But for the rest of the three features, this referent is accessible, i.e., it is physically present, given, and linguistically unambiguous. Cases such as (56) show that the accessibility of the feature *animacy*, similar to Time 1, does not have a determining role in the realization of subject arguments in the form of pronouns.

Therefore, it can be concluded that excluding *animacy*, the other features seem to play a more significant role in the realization of pronominal forms with *prior mention* and *linguistic disambiguation* being the strongest ones as they have the highest percentage of accessible referents in this group.<sup>32</sup>

#### 4.2.2.4. Null Forms

As the last category of referential forms at Time 2, this section reports the results on Matty's use of null forms. Each column in Figure 8 displays the accessibility of one of the four discourse-pragmatic features under investigation for all and only 10 null subjects found at this age range. As mentioned earlier at the beginning of this chapter, the number of subject omissions has dropped considerably at Time 2 which shows a more adult-like

---

<sup>32</sup> In the category of pronouns, no referent with inaccessible value for *linguistic disambiguation* and accessible value for the other three features was found in the data at Time 2. Regarding *prior mention*, only 6% of subjects were inaccessible for this feature while accessible for the rest. This can support the idea that these two features play a more determining role than the other ones in this group.

choice of subjects since the child is towards the end of the null subject stage. As a result of this substantial decrease, null subjects make up only 4% of the whole data set at this time. This should be taken into consideration for the interpretation of the percentages in Figure 8.

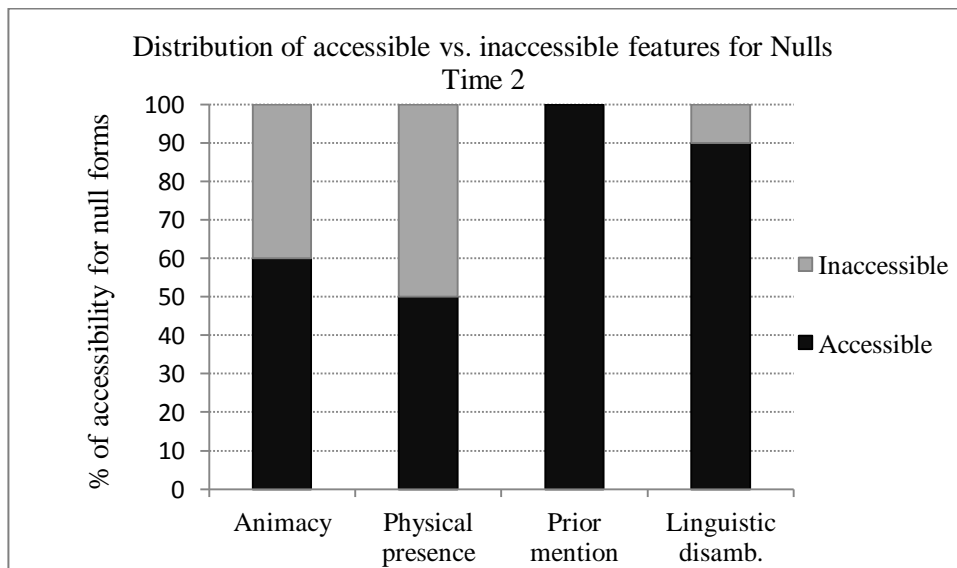


Figure 8

As illustrated in Figure 8, the findings show that all the null subject forms are accessible for the feature *prior mention* and almost all of them are also accessible for the feature *linguistic disambiguation*. For the features *animacy* and *physical presence*, the proportion of accessible and inaccessible referents is (almost) equal. It should be noted that in 80% of the cases in this group, null subjects are inaccessible for either *animacy* or *physical presence*, as can be inferred from the following dialogues:

(57) CHI: I got those all (.) for you.  
 MOT: you got these all for me?  
 CHI: I think I got them all for me.  
 MOT: you got them all for you.  
 MOT: where did you get (th)em?  
 CHI: actually found (.)  $\emptyset$  were on the play+ground. (3;2.26)

(58) RYA: a mean guy?  
 MOT: yes (.) what does Randall do in the movie?  
 CHI: &ah (.) (.) I don't know.  
 MOT: well we were just watching it yesterday right?  
 MOT: remember what did Randall do in the movie (.) that was not very nice?  
 CHI:  $\emptyset$  take Boo. (3;3.09)

In (57), the underlined empty subject refers to a collection of rocks. It is inanimate, physically present, previously mentioned, and linguistically unambiguous. In (58), the referent of the null subject is Randall, a monster character in the ‘Monsters, Inc.’ film. It is considered animate. Since Matty and his mother are talking about a movie that they had watched before, the referent of the null subject is not physically present. But it is accessible for the features *prior mention* and *linguistic disambiguation*.

Based on the much higher percentages of accessible referents for the features *prior mention* and *linguistic disambiguation* in the data and as examples (57–58) show, it can be concluded that these two features with accessible value have a more influential role in the realization of subject arguments in this group.

#### 4.2.2.5 Time 2 Summary

To summarize the findings at Time 2, it can be concluded that the child’s choice of referential forms has become more adult-like as Matty omits significantly fewer subject arguments than at Time 1. He mainly produces pronominal forms when referents are highly accessible and more lexical NPs and demonstratives when referents are inaccessible. In addition, the total proportion of accessible vs. inaccessible values for all the features assessed in this study shows that Matty uses more referents with inaccessible value at Time 2. This indicates that his linguistic ability has improved over time so that he can talk about entities that are new, absent from the physical setting of the conversation, as well as inanimate entities and abstract concepts.

Regarding the salience of the discourse-pragmatic features investigated in the present study, the results support the view that these features are not of equal weight and that some of them may be more influential than the others when it comes to the child’s choice of subject arguments. The findings suggest that in the realization of lexical NPs, all the features seem to be influential with *prior mention* and *linguistic disambiguation* with inaccessible value having the strongest effect on them. With regard to the use of demonstratives, *animacy* (with inaccessible value) and *physical presence* (with accessible value) seem to play the most significant role in the realization of this referential type although the effect of the other two features cannot be ignored. When it comes to the pronominal forms, the feature *animacy* seems to still have the weakest effect and *prior mention* and *linguistic disambiguation* with accessible values play the most influential role. And finally, the distribution of accessibility in the category of null forms suggests that *prior*

*mention* and *linguistic disambiguation* play a more determining role with this argument type as with pronouns.

The findings at both Time 1 and Time 2, all in all, confirm the prediction that children are sensitive to the information flow. Based on different patterns of accessibility, they realize subjects in different referential forms. The results are also in favor of the hypothesis that some discourse-pragmatic features have a stronger and more salient effect on the child's choice of subject arguments than others.

### **4.3. Incremental Effect of Accessibility Features**

In the previous sections of this chapter, the dominant assumption for the analysis of the results was that discourse-pragmatic features can work in isolation and independently of one another. But this is not what really happens when it comes to children's choice of subject forms; when children want to use a referential form, they take some or even all the discourse-pragmatic features into consideration (Hughes & Allen, 2013, p. 25). Accordingly, the hypothesis in this section is that the child will be more likely to produce a highly informative subject form such as a lexical NP when a referent is more inaccessible. Therefore, the fully inaccessible subjects are more likely to be realized in the form of lexical NPs and the fully accessible subject arguments are more likely to be realized in pronominal or null forms (Hughes & Allen, 2006, 2013, in press). In order to test the validity of this hypothesis, the cumulative effects of the four discourse-pragmatic features, namely *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*, have been assessed at both Time 1 and Time 2, the results of which are presented in Tables 5 and 6 respectively. The triple hyphen/minuses in the tables denote that no such cases were attested in the data.

As mentioned earlier at the end of Chapter 3, the number of features inaccessible for a particular referential form indicates to what extent a referent is inaccessible. The scores 0 to 4 in the first column of Tables 5 and 6 show the number of inaccessible values, followed by the total number of subject arguments found for that specific category in parentheses. For instance, the 0-row in Tables 5 and 6 displays the fully accessible referents, while the 4-row refers to fully inaccessible referents in the current data set.

Table 5: Cumulative effect of features at Time 1

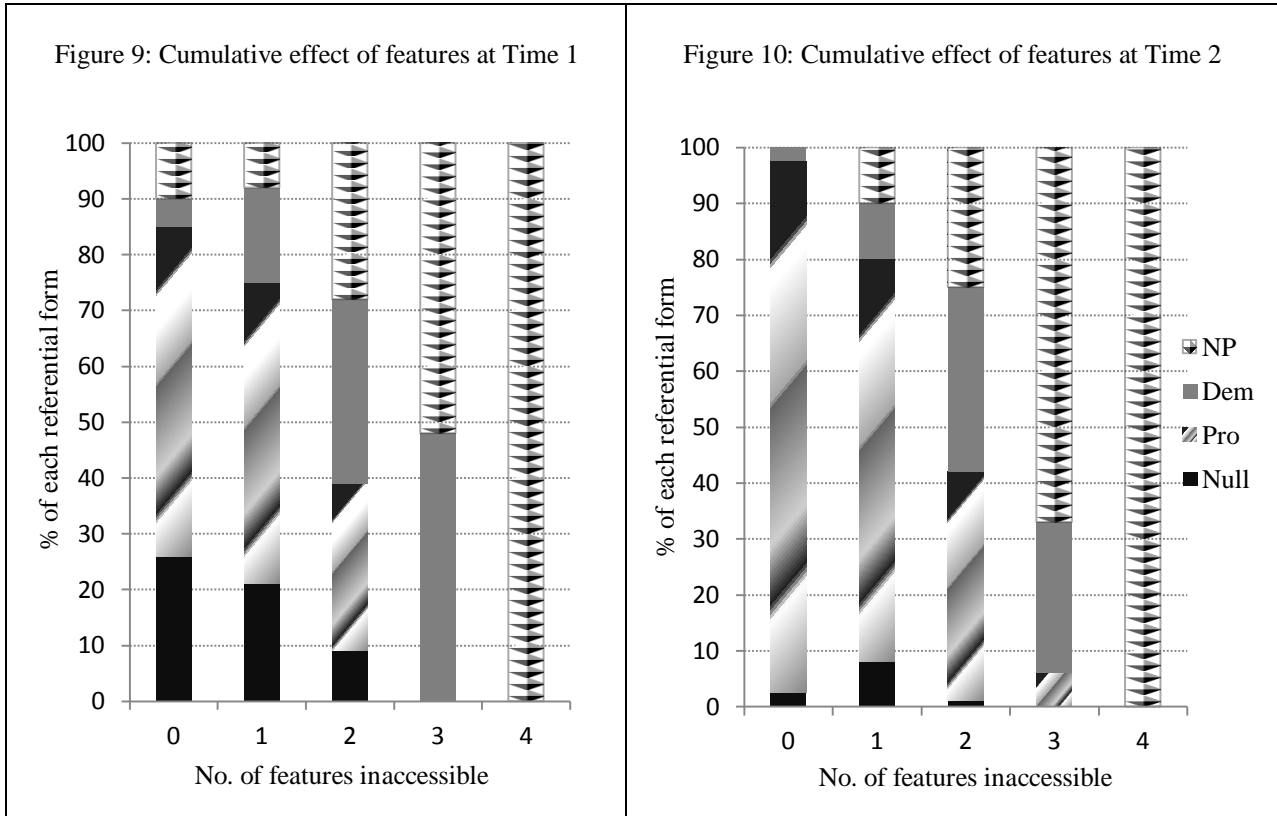
# of features inaccessible (total # of subj. arguments)	NP	Dem	Pro	Null
0 (83)	8 (10%)	4 (5%)	49 (59%)	22 (26%)
1 (90)	7 (8%)	15 (17%)	49 (54%)	19 (21%)
2 (57)	16 (28%)	19 (33%)	17 (30%)	5 (9%)
3 (27)	14 (52%)	13 (48%)	---	---
4 (3)	3 (100%)	---	---	---

Table 6: Cumulative effect of features at Time 2

# of features inaccessible (total # of subj. arguments)	NP	Dem	Pro	Null
0 (39)	---	1 (2.5%)	37 (95%)	1 (2.5%)
1 (100)	10 (10%)	10 (10%)	72 (72%)	8 (8%)
2 (64)	16 (25%)	21 (33%)	26 (41%)	1 (1%)
3 (33)	22 (67%)	9 (27%)	2 (6%)	---
4 (4)	4 (100%)	---	---	---

Taking a closer look at the results at both times shows that while 83 cases of fully accessible (the 0-row) referents are found at Time 1, which equals more than 30% of the data at this time, this proportion is halved at Time 2. The proportion of referents in the other categories (with 1–4 score), however, has increased over time. This suggests that Matty talks more about inaccessible entities at Time 2 which shows a more adult-like linguistic behavior. Referents with only one inaccessible feature comprise the majority of the referents at both stages. Most of the referents with fully accessible features are realized in the form of pronouns both at Time 1 and Time 2, but at a considerably higher proportion at Time 2. Moreover, as expected, only lexical NPs are found with referents that are fully inaccessible, although there are few examples in both periods. These findings support the hypothesis of this section according to which the accessibility features are expected to have an incremental effect on the child's choice of subject arguments.

In order to make the comparison between Time 1 and Time 2 more convenient, the results are displayed in the following charts; the referential form (NP, Dem, Pro, Null) is the dependent variable and the summed value of accessibility (number of features inaccessible) is the independent variable in this analysis as the cumulative effect of discourse-pragmatic features on the child's choice of subject arguments is assessed.



As Figure 9 illustrates, the higher the number of inaccessible features at Time 1, the lower the proportion of pronouns and null forms, and the higher the percentages of demonstratives and lexical NPs are. The only exception is the proportion of fully accessible lexical NPs, which is slightly higher than the proportion of those with one inaccessible feature. That said, fully accessible referents are not expected to be realized by a highly informative form such as a lexical NP. As discussed earlier in Section 4.2.1.1, this result can be explained by a typical feature of child language, that is, the fact that children tend to repeat third person referents regardless of their degree of informativeness (Hughes & Allen, 2013, in press).

A similar, but more mature, pattern is observed for the subject arguments at Time 2 as displayed in Figure 10. The higher the number of inaccessible features, the lower the proportion of pronouns and null forms is. The only exception is the proportion of empty subjects with one inaccessible feature which comprise the highest percentage in this group. As mentioned earlier in Section 4.2.2.4, the proportion of accessible vs. inaccessible referents in this category is (almost) equal for the features *animacy* and *physical presence* which led to 80% of the null forms at this time to be inaccessible for one of these two features. But considering the fact that null subjects comprise only 4% of the data at Time 2, and that referents with only one inaccessible feature are still considered to be highly

accessible, this difference can be considered negligible. Regarding the distribution of demonstratives, it is interesting to note that the highest proportion in this category of referential forms belongs to referents with the inaccessibility score of 2 which exactly supports the idea that demonstratives are medium accessibility markers.

Overall, it can be concluded that discourse-pragmatic features do work together and Matty's sensitivity to the incremental effect of the four accessibility features investigated in this study is considerable, at both age ranges.



## CHAPTER 5

### Discussion and Conclusions

The present study set out to explore the null subject phenomenon in English-speaking monolingual children and their sensitivity to discourse-pragmatic features by assessing their choice of referential forms. Along the lines of Hughes and Allen's recent studies (2006, 2013, in press), the current study examined another English-speaking child's sensitivity to four discourse-pragmatic features and the possible effect of these features on the child's choice of subject arguments. The significance of individual accessibility features was also assessed as it is believed these features do not have equal weights when it comes to the realization of subjects. In another analysis, the incremental effect of discourse-pragmatic features was examined in order to see if such features interact and work together. The analysis of the child's speech was carried out at two different age ranges so that the developmental pattern of the child's choice of referential forms could be observed over time, both when he is believed to be at the null subject stage and when he is towards the end of it. For that purpose, the utterances of Matty, an English-speaking monolingual from Western New York State, were analyzed from ages 2;5.04 to 2;7.22 (Time 1) and 3;0.02 to 3;3.09 (Time 2). The data files were in audio format and were taken from one of the CHILDES archive files (MacWhinney, 2000; Weist & Zevenbergen, 2008). The following predictions were made:

- i. The child was expected to omit more subject arguments at Time 1 than at Time 2. At Time 2, he was towards the end of the null subject stage and therefore was expected to produce fewer null subjects. Conversely, the proportion of pronominal forms was expected to rise significantly at Time 2.
- ii. The child was expected to show sensitivity to the discourse-pragmatic features under investigation by using highly informative referential forms for inaccessible referents and low information forms for accessible ones.
- iii. The accessibility features were expected to have different weight and significance in determining the child's choice of referential forms.
- iv. The discourse-pragmatic features were expected to have an incremental effect on the child's choice of subject arguments.

In order to test the validity of the above-mentioned hypotheses, 260 third person subject arguments at Time 1 and 240 third person subject arguments at Time 2—including lexical NPs, demonstratives, pronouns and null forms—were coded for the features *animacy*, *physical presence*, *prior mention*, and *linguistic disambiguation*.

The results showed that Matty's use of null subjects at Time 1 was more than four times higher than at Time 2. While the number of null subjects has dropped substantially, the proportion of pronominal forms has increased from 44% at Time 1 to 57% at Time 2. Therefore, the null forms were "traded off" with pronouns over time, as the grammatical accounts such as Hyams and Wexler (1993), and Gerken's (1991) metrical account would expect. Therefore, the significant drop in the number of null subjects as well as the increase in the use of pronominal forms over time support the first hypothesis.

In order to assess the child's sensitivity to the information flow, the effect of discourse-pragmatic features on the child's referential choice was determined by coding all the selected subject arguments for their accessibility values. The findings at Time 1 showed that the child was aware of the information flow when he produced high and medium accessibility markers. But with regard to lexical NPs, the pattern was not quite in line with what is expected from a highly informative referential form; more than 30% of the NP subjects were used for referents with a high accessibility level. As Hughes and Allen (in press) state, this more than expected proportion of NP subjects for accessible referents is due to a feature of child speech according to which children repeat lexical NPs in their utterances for a number of reasons. For instance, the production and comprehension of NPs are "less cognitively demanding" for children than demonstrative and pronominal forms (Hughes & Allen, in press, p. 14). At Time 2, the child showed a more adult-like pattern in his subject realization as he produced more lexical NPs and demonstratives for inaccessible referents and more pronouns for accessible referents.

Regarding the different weight of discourse-pragmatic features, results from both age ranges supported the prediction that these features do not have equal significance. At both times, *prior mention* and *linguistic disambiguation* with inaccessible value turned out to be the most influential features in determining the realization of lexical NPs. For demonstratives, the features *animacy* with inaccessible value and *physical presence* with accessible value were the most salient features at both age ranges. With respect to the realization of pronominal forms, it seemed that *animacy* had the weakest effect, and *prior mention* and *linguistic disambiguation* with accessible values had the most influential role at

both Time 1 and Time 2. Regarding the null forms, different behaviors were detected at Time 1 and Time 2; while *linguistic disambiguation* with accessible value was the least salient feature at Time 1, it became one of the strongest along with *prior mention* at Time 2.

With respect to the incremental effect of discourse-pragmatic features, the results supported the hypothesis that as referents become more and more inaccessible, the child is more likely to use high information forms such as lexical NPs to refer to them.

The current findings also showed that as the child got older, his use of more appropriate referential forms increased. This more adult-like behavior over time supports the view that children from early on are able to assess the discourse participants' attentional states (their centers of attention) and take their knowledge of the context into consideration for referential choices (Graf & Davies, 2014, p. 169; Hughes & Allen, in press, p. 16). In addition to showing the developmental pattern of argument realization in particular, the current findings also have some implications on the child's linguistic knowledge and cognitive skills in general; as Hughes and Allen (in press, p. 16) state, such findings strongly indicate that children have a "well-grounded though still developing pragmatic system [...] that involves not only linguistic knowledge, but also the coordination of logical, psychological, and social knowledge."

With regard to how the discourse-pragmatic approach explains the null subject phenomenon as a whole, the findings of the current study show which arguments are more likely to be the target of omission at this stage. According to the discourse-pragmatic approach (Hughes & Allen, 2006, 2013) taken in this analysis, highly accessible referents are more likely to be realized as null forms. But what this approach does not really discuss is the reason why children at this particular stage tend to omit subjects in the first place, and why later on, as they get older, they do not drop those highly accessible subjects any longer but use pronouns instead. As reviewed earlier in Chapter 2, the other two approaches, the grammatical and performance accounts, provide an answer to this question.

According to the grammatical perspective (e.g., Hyams, 1986), children omit subjects at this particular stage because their child grammar allows them to do so. For instance, the results in Orfitelli and Hyams (2012) suggest that subject drop at this stage is a competence-based phenomenon. As the children get older, their grammar becomes more target-like due to a number of reasons, such as exposure to triggering linguistic input (Allen, 2006). On the other hand, the performance-based studies reject the assumption that child grammar is different from adult grammar. On the contrary, they argue that child grammar is the same as

adult grammar and that children tend to omit subjects because of performance limitations, processing difficulties (Bloom, 1990), or prosodic constraints (Gerken, 1991). As children get older and their production capabilities get more mature, their language becomes more and more adult-like.

However, even though Orfitelli and Hyams (2012) convincingly argue that subject omissions are permitted by child grammars, it is a weakness of the grammatical accounts that they do not explain when children omit subjects, which arguments are more likely to be the target of omission “out of the many ones for which omission would be allowed under the general conditions” (Allen et al., 2008, p. 100). As mentioned before, studies from the discourse-pragmatic perspective deal quite well with such questions. With regard to the performance approach, although it tries to specify targets of omission in its works, as Allen (2006) states, it fails to offer explanations for all the relevant contexts such as observed cases of subject omission in contexts with short VP. This approach also ignores the grammatical facts related to this stage such as the occurrence of null subjects in *wh*-questions or the likelihood of its occurrence with non-finite verbs (Allen, 2006, p. 243). Besides, the analyses based on the performance approach mainly consider the speakers and their processing limitations. The addressees’ knowledge and their assessment of the context of conversation are not taken into consideration. Even in their pragmatic versions (as discussed in (non)structural pragmatic account in Section 2.4.1), the grammatical and performance approaches do not define and provide the appropriate tools to investigate all the discourse-pragmatic features which have an effect on the realization of overt vs. null subject arguments. It is not possible to measure an argument’s degree of informativeness in such accounts either.

As can be inferred from the previous paragraphs and as Allen (2006) maintains, none of the three approaches alone can provide a full explanation of the null subject phenomenon. Each of them covers parts of the issue, but there are certainly some gaps in their arguments and perspectives that the other two can fill. Allen (2006, p. 236) describes these approaches as “largely complementary rather than competing”. Thus, they need to be considered together, rather than in isolation and independent of one another. As such, Allen (2006) suggests that the study of the null subject phenomenon needs an alternative unifying approach which does not focus on only one perspective, but brings all the three together. For example, Allen (2006) mentions a prediction that such an alternative view can make

based on the findings of the three approaches so far in order to explain the optionality of subject omission at this stage:

Omission would occur for all subjects of non-finite verbs that have long VPs or are in weak prosodic contexts, and that are accessible from the discourse or situational context. Conversely, omission would never occur for subjects of finite verbs that have short VPs or are in strong prosodic contexts, and that are not accessible from the discourse or situational context. (Allen, 2006, p. 247)

Therefore, based on Allen's (2006) proposal further research on the interaction between different grammatical, processing and discourse-pragmatic approaches can help us improve our understanding of the null subject phenomenon in child language as a whole.

## Bibliography

- Allen, S. E. M. (2006). Formalism and functionalism working together? Exploring roles for complementary contributions in the domain of child null arguments. In R. Slabakova, S. A. Montrul, & P. Prévost (Eds.), *Inquiries in linguistic development: In honor of Lydia White* (pp. 233–255). Amsterdam: Benjamins.
- Allen, S. E. M., Skarabela, B., & Hughes, M. (2008). Using corpora to examine discourse effects in syntax. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 99–137). Amsterdam: John Benjamins.
- Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schliperoord & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: John Benjamins.
- Aronoff, J. M. (2003). Null subjects in child language: Evidence for a performance-based account. In G. Garding & M. Tsujimura (Eds.), *WCCFL 22 proceedings* (pp. 43–55). Somerville, MA: Cascadilla Press.
- Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, 21(4), 491–504.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Chafe, W. (1996). Inferring identifiability and accessibility. In T. Fretheim & J. K. Gundel (Eds.), *Reference and referent accessibility* (pp. 37–46). Amsterdam: John Benjamins.
- Deen, K. (2009). The morphosyntax interface. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 259–280). Cambridge: Cambridge University Press.
- Gerken, L. (1991). The Metrical basis for children's subjectless sentences. *Journal of Memory and Language*, 30, 431–451.
- Graf, E., & Davies, C. (2014). The production and comprehension of referring expressions. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 161–181). Amsterdam: John Benjamins.
- Greenfield, P. M. & Smith, J. H. (1976). *The structure of communication in early language development*. New York: Academic Press.
- Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. Cambridge: MIT Press.
- Guerriero, S., Oshima-Takane, Y., & Kuriyama, Y. (2006). The development of referential choice in English and Japanese: A discourse-pragmatic perspective. *Journal of Child Language*, 33(4), 823–857.

- Guilfoyle, E. (1984). The acquisition of tense and the emergence of lexical subjects in child English. *The McGill Working Papers in Linguistics*, 2(1), 20–30.
- Hughes, M., & Allen, S. E. M. (2006). A discourse-pragmatic analysis of subject omission in child English. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 293–304). Somerville, MA: Cascadilla Press.
- Hughes, M. E., & Allen, S. E. M. (2013). The effect of individual discourse-pragmatic features on referential choice in child English. *Journal of Pragmatics*, 56, 15–30.
- Hughes, M. E., & Allen, S. E. M. (In press). The incremental effect of discourse-pragmatic sensitivity on referential choice in the acquisition of a first language. *Lingua* (2014), <http://dx.doi.org/10.1016/j.lingua.2014.03.001>
- Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- Hyams, N. (1996). The underspecification of functional categories in early grammar. In H. Clahsen (Ed.), *Generative Perspectives on Language Acquisition* (pp. 91–127). Amsterdam: John Benjamins.
- Hyams, N. (2011). Missing subjects in early child language. In J. De Villiers & T. Roeper (Eds.), *Handbook of generative approaches to language acquisition* (pp. 13–52). Dordrecht: Springer Netherlands.
- Hyams, N., & Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, 24(3), 421–459.
- Jaeggli, O., & Hyams, N. (1988). Morphological uniformity and the setting of the null subject parameter. *Proceedings of the NELS*, 18, 239–255.
- Jaeggli, O., & Safir, K. J. (1989). The null subject parameter and parametric theory. In O. Jaeggli & K. J. Safir (Eds.), *The Null subject parameter* (pp. 1–44). Dordrecht: Kluwer.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Orfitelli, R., & Hyams, N. (2012). Children's grammar of null subjects: Evidence from comprehension. *Linguistic Inquiry*, 43(4), 563–590.
- Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, L. (1993/1994). Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition*, 3(4), 371–393.

- Rizzi, L. (1994). Early null subjects and root null subjects. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (pp. 151–176). Amsterdam: John Benjamins.
- Rohrbacher, B., & Roeper, T. (1995). Null subjects in early child English and the theory of economy of projection. *University of Pennsylvania Working Papers in Linguistics*, 2(1), Article 6. Retrieved from <http://repository.upenn.edu/pwpl/vol2/iss1/6/>
- Scott, K. (2005). Child null subjects. *UCL Working Papers in Linguistics*, 17, 1–25.
- Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in Italian-English bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, 7(3), 183–205.
- Skarabela, B., Allen, S. E. M., & Scott-Phillips, T. C. (2013). Joint attention helps explain why children omit new referents. *Journal of Pragmatics*, 56, 5–14.
- Valian, V. (1990). Null subjects: A problem for parameter-setting models of language acquisition. *Cognition*, 35, 105–122.
- Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40, 21–81.
- Wang, Q., Lillo-Martin, D., Best, C., & Levitt, A. (1992). Null subject versus null object: Some evidence from the acquisition of Chinese and English. *Language Acquisition*, 2(3), 221–254.
- Weist, R. M., & Zevenbergen, A. (2008). Autobiographical memory and past time reference. *Language Learning and Development*, 4(4), 291–308.
- Wieczorek, R. (2010). Using MLU to study early language development in English. *Psychology of Language and Communication*, 14(2), 59–69.
- Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.