University of Iowa
Iowa Research Online

Theses and Dissertations

2009

Acquisition of auxiliary and copula BE in young English-speaking children

Ling-Yu Guo
University of Iowa

Copyright 2009 Ling-Yu Guo

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/370

Recommended Citation
Guo, Ling-Yu. "Acquisition of auxiliary and copula BE in young English-speaking children." PhD (Doctor of Philosophy) thesis, University of Iowa, 2009.
http://ir.uiowa.edu/etd/370.

Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the Speech Pathology and Audiology Commons

ACQUISITION OF AUXILIARY AND COPULA BE IN YOUNG ENGLISH-SPEAKING CHILDREN

by

Ling-Yu Guo

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Speech and Hearing Science
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisors: Professor J. Bruce Tomblin
Assistant Professor Amanda J. Owen

ABSTRACT

This study tested the unique checking constraint hypothesis and the usage-based approach concerning why young children produced tense and agreement morphemes variably via three experiments. Experiment 1 investigated whether subject types influenced the production accuracy of auxiliary *'is'* in three-year-olds through an elicited production task. The rate of use of auxiliary *'is'* increased as children's tense productivity increased, but the pattern was different for each subject type. The rate of use increased more rapidly with tense productivity for lexical NP subjects than it did for pronominal subjects.

Experiment 2 further examined the role of subject types, predicate types, and predicate word frequency on the use of copula *'is'* in three-year-olds. The production accuracy of copula *'is'* was higher with nominal predicates than with permanent- or temporary-adjectival predicates, followed by locative predicates. Children also produced copula *'is'* more accurately with low-frequency predicate words than with high-frequency predicate words. Moreover, the effect of subject types on the use of copula *'is'* varied with children's tense productivity and predicate types. For sentences with nominal, permanent-adjectival, or temporary-adjectival predicates, children with lower tense productivity used copula *'is'* more accurately with lexical subjects than with pronominal subjects in. This pattern, however, reversed in children with higher tense productivity.

Experiment 3 extended Experiment 1 by exploring the degree of abstractness of representations of auxiliary BE via a structural priming task. The production accuracy of auxiliary *'is'* in three-year-olds increased above the baseline when the prime-target pair shared the same structure and <u>subject + auxiliary *'is'*</u> combinations, but not when the prime-target pair only shared the same structure. However, the production accuracy of auxiliary *'are'* did not change with prime types.

These experiments suggest that young children have only lexically-specific representations of auxiliary BE. Frequency, rather than structural properties, of sentence elements influenced the production accuracy of auxiliary and copula *'is'* in young children. These findings support the usage-based approach that young children use tense and agreement morphemes variably because they have not yet learned adult-like abstract representations and use highly frequent/ lexically-specific constructions for the production of these morphemes.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

_____
Thesis Supervisor

_____
Title and Department

_____
Date

ACQUISITION OF AUXILIARY AND COPULA BE IN YOUNG ENGLISH-SPEAKING CHILDREN

by

Ling-Yu Guo

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Speech and Hearing Science
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisors: Professor J. Bruce Tomblin
Assistant Professor Amanda J. Owen

Copyright by

LING-YU GUO

2009

All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ling-Yu Guo

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Speech and Hearing Science at the December 2009 graduation.

Thesis Committee: _____
J. Bruce Tomblin, Thesis Supervisor

_____
Amada J. Owen, Thesis Supervisor

_____
Prahlad Gupta

_____
Richard Hurtig

_____
Karla K. McGregor

To my parents.
Without your hard work on the farm, this thesis would have never been possible.

ACKNOWLEDGMENTS

The journey of my Ph.D. study began in a night of November 2002. I finally convinced myself to pursue a Ph.D. degree but I did not know whether I could make it, given that the deadline of application was only a month away and I had been out of school for four years. That night, I was sleepless.

Luckily, I came to Iowa. Looking back at my Ph.D. journey, I would like express my deepest gratitude to my advisors, Dr. J. Bruce Tomblin and Dr. Amanda J. Owen. I thank for their continued guidance, insights, inspiration, encouragement, and generosity throughout these years. Though Bruce and I have different working styles, I really appreciate his patience with, and flexibility for, me. His intellectual rigor has benefited me greatly. I always tell people that he has been my role model of research. I would also like to thank Amanda for her dedication to my dissertation work. She always responds to my questions quickly (far better than UPS!) and offers the help whenever I need it. Her sharp analytical skills and critical thinking always impress me. As a supervisor, she has also trained me how to efficiently carry out ideas into research studies and she taught me how to survive in the academic world. Because of her, I learned how a supervisor should work with students. I have to say it is my honor to work with both of them.

In addition, I thank Dr. Prahlad Gupta, Dr. Richard Hurtig, and Dr. Karla K. McGregor for serving in my dissertation committee. I have benefited from their insightful comments and discussions with them. I would like to express my special gratitude to Karla for letting me use her lab facilities in the past years and helping me realize that being a researcher can still be elegant! I am also grateful to numerous colleagues for their work in data collection, transcription, and/or analysis: April Amonson, Allison Bean, Jessica Colwell, Emily Diehm, Laural Everist-Lambert, Danielle Kazeos, Kenneth Marciniak, Emily Meier, Allison Otto, Sarah Raske, Sarah Stuck, Carrie van Zanten, and Beth Walker. I extend special thank to Dr. James Myers

for answering all the questions in experimental design and statistical analysis; and to Dr. Matthew Rispoli for encouraging me and sharing his ideas with me. Furthermore, I would like to acknowledge the funding from a Student Research Grant in Early Childhood Language from the American Speech-Language-Hearing Foundation and a Language Learning Dissertation Grant from *Language Learning: A Journal of Research in Language Studies*.

I usually joked with friends that my life in Iowa was full of "interesting" experiences, such as tornado, bicycle accident, and stolen pants. Yet, I enjoyed the life in Iowa because of the warmth and happiness that my friends brought to me. I thank Fuh-cherng Jeng for telling me all the information about Iowa and reducing my anxiety way back when I was still in Taiwan; Li-kuei Chiou and Ming-ta Chia for giving me free food, solving all the problems in life and offering me a place to go when I have nothing to do; Mary Jo Yotty, for being my "caregiver" and dealing with all my troubles; Marlea O'Brien, for organizing things for me and giving me excellent advice; Linda Spencer for talking with me whenever I feel happy or sad; and Julie Hsin-Jen Hsu for being generous and teaching me how to look at the bright side of life. I would like to extend my sincere thank to the following people for their help and friendship, which supported me during this journey: Dr. Sheng-Hwa Chen, Dr. Yu-chao Hsiao, Janice Fon, Shaun Ho, Hui-tsu Huang, Lillian Yeh, Li-chu Hung (Yeh mama), Alfred Li, Jackie Lin, Daphne Sieh, Shuman He, Yu-Hsiang Wu and Chin-yi Chen, Dr. Ruth Bentler, Dr. Carolyn Brown, Dr. Jean K. Gordon, Dr. Robert Owens, Allison Bean, Amanda Berns, Alice Berube, Dawna Duff, Nichole Eden, Lucien and Lucienne Loiselle, Michelle Quinn, Gwyneth Rost, and Vicki Samuelson. Without these people, I am not sure whether I would have survived in Iowa.

Special thanks go to two important people in my research career: Dr. Eric Zee and Dr. Ming-jang Chiu. Though Dr. Zee has never officially taught me, he was the first person to let me know that I had the potential in research when I was not very sure about

my future. He also constantly encourages me to pursue a career in research. I am also deeply indebted to Dr. Chiu for his guidance, patience, and generosity. He has been a great mentor ever since I was his research assistant eleven years ago. I almost gave up my research career once. It was his understanding and encouragement that convinced me to resume my research. He also continued offering assistance that I needed to conduct studies even after I left his lab.

Finally, I would like to thank my family. I am indebted to my sisters—Hsiu-mei Guo, Hsiu-wen Guo, and Hsiu-li Guo—for taking care of my parents in the past years. I especially thank my eldest sister, Hsiu-mei Guo, for sharing my responsibility to deal with everything in the family so that I can study abroad. Though I do not know whether the word "thank" can adequately express how I feel, I would still like to thank my parents—Kuai-hsiung Guo and Ji Jian—for their continuing love and support, which make this dissertation and everything possible. I still remember those days when I studied at school and they had to work on the farm in the sun. I hope that this dissertation can at least honor their hard work on the farm and comfort all the pain they endured in the past years.

This dissertation is dedicated to my parents.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER I INTRODUCTION 1

CHAPTER II EXPERIMENT 1: ELICITED PRODUCTION AUXILIARY *'IS'* IN YOUNG CHILDREN 4

Overview 4
The Unique Checking Constraint (UCC) Hypothesis 5
The Usage-based Approach 11
Limitations of the Existing Evidence 16
The Scope of the Current Study 18
Methods 18
Participants 18
Stimuli 20
Procedures 23
Transcription and Coding 24
Reliability 27
Statistical Analysis 27
Results 29
Sentence Length of Target Responses 29
Effect of Subject Types on the Production of Auxiliary *'is'* 29
Comparison with Other Studies 31
Discussion 38
Limitations of the Current Study 39
Effect of Subject Types on the Production of Auxiliary *'is'* Varies in Children with Different Developmental Levels 40
Tense Productivity Scores are More Sensitive to Children's Use of Auxiliary 'is' than Finite Verb Morphology Composites 45
Clinical and Experimental Implications 46
Conclusion 48

CHAPTER III EXPERIMENT 2: ELICITED PRODUCTION OF COPULA *'IS'* IN YOUNG CHILDREN 49

Overview 49
The Unique Checking Constraint Hypothesis 50
The Usage-based Approach 57
Summary of Research Questions 65
Methods 66
Participants 66
Stimuli 68
Procedure 73
Transcription and Coding 75
Reliability 78
Statistical Analysis 78
Results 79
Sentence Length of Target Responses 79

Effects Subject Types, Predicate Types, and Predicate Word Frequency 81
Discussion 87
Limitations of the Current Study 87
Predicate Types and Predicate Word Frequency Affect the Production Accuracy of Copula *'is'* 88
Effect of Subject Types on the Production of Copula *'is'* Varied with Predicate Types and Tense Productivity 91
Experimental Implications 95
Conclusion 97

CHAPTER IV EXPERIMENT 3: REPRESENTATION OF AUXILIARY BE IN YOUNG CHILDREN 98

Overview 98
Representations of Tense and Agreement Morphemes in the Usage-based Approach 98
The Structural Priming Paradigm 100
Research Questions and Predictions 106
Method 108
Participants 108
Stimuli 109
Procedure 111
Transcription and Coding 113
Reliability 115
Statistical Analysis 115
Results 117
Effect of Test Order 117
Effect of Priming Condition 118
Discussion 121
Limitations of the Current Study 122
The Lexically-specific Priming Condition Facilitated the Production of Auxiliary *'is'* but not Auxiliary *'are'* 123
Children Produce the Auxiliary More Accurately in the Third Block than in the First or the Second Blocks 126
Experimental and Clinical Implications 127
Conclusion 129

CHAPTER V GENERAL DISCUSSION AND CONCLUSION 130

Summary of Main Findings 130
The Nature of Lexically-specific Constructions with Auxiliary *'is'* Varied in the Production and Priming Tasks 131
Effect of Subject Types Varied for Auxiliary and Copula *'is'* in Children with Identical Developmental Levels 133
Theory Revisited 135
Concluding Remarks 138

REFERENCES 139

APPENDIX A TARGET AND CONTRASTIVE SENTENCES OF AUXILIARY *'IS'* 147

APPENDIX B RATING OF FANTISINESS OF TARGET SENTENCES AND PICTURES 148

APPENDIX C SAMPLE PICTURES 149

APPENDIX D TARGET AND CONTRASTIVE SENTENCE FOR COPULA '*IS*' ...150

APPENDIX E PRIME-TARGET PAIRS.....155

LIST OF TABLES

Table 2.1. Frequency Distribution of Lexical Items Chosen for Target Subjects by Condition.....21

Table 2.2. Mean Number of Scorable Sentences by Condition.....26

Table 2.3. Mean Percent Correct of Auxiliary *'is'* across Children by Subject Type.....30

Table 2.4. Regression Model Showing the Likelihood of Producing auxiliary *'is'* Correctly by Subject Type.....30

Table 2.5. Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Subject Type and Tense Productivity Score.....34

Table 2.6. Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Subject Type and Tense Productivity Score.....37

Table 3.1. Predictions of the Effect of Subject Types, Predicate Types, and Frequency of Predicate Words on the Production of Copula *'is'* in Two Accounts.....66

Table 3.2. Frequency Distribution of Subjects by Subject Type.....70

Table 3.3. Number and Frequency Distribution of Predicates by Predicate Type and Frequency Category.....72

Table 3.4. Number of Target Sentences and Mean Number of Scorable Sentences by Subject Type, Predicate Type, and Predicate Frequency Category.....77

Table 3.5. Mean Sentence Length of Scorable Responses in Syllables by Subject Type, Predicate type, and Predicate Frequency Category.....80

Table 3.6. Mean Percent Correct of Target Sentences by Subject Type, Predicate Type, and Predicate Frequency Category.....81

Table 3.7. Regression Model Showing the Likelihood of Producing copula *'is'* correctly by Predicate Type and Predicate Word Frequency.....86

Table 4.1. Mean Number of Target Sentences and Number of Scorable Targets with Correctly Repeated Primes by Auxiliary BE Form and Priming Condition.....115

Table 4.2. A Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Priming Condition.....119

LIST OF FIGURES

Figure 2.1. Predicted Probability of Producing Auxiliary *'is'* Accurately by Subject Type and Tense Productivity Score 35

Figure 2.2 Predicted Probability of Producing Auxiliary *'is'* Accurately by Subject Type (Pronoun and Lexical NP Subjects) and Tense Productivity Score 38

Figure 3.1. Predicted Probability of Producing Copula *'is'* Accurately by Subject Type, Predicate type, and Tense Productivity Score 84

Figure 4.1. Mean Percent Correct of Target Sentence by Auxiliary Form and Test Order 118

Figure 4.2. Mean Percent Correct of Target Sentences by Auxiliary Form and Priming Condition.....120

CHAPTER I

INTRODUCTION

How children acquire syntactic knowledge has been a topic of intensive debate in the area of cognitive science (Bates & Goodman, 1999; Chomsky, 1956, 1965; Crain & Lillo-Martin, 1999; Elman et al., 1998; Lindner & Hohenberger, 2009; MacWhinney, 2004). The generative (or nativist) approach assumes that children are able to acquire language because they are born with innate hypotheses (e.g., rules/principles/parameters) related to structures of human languages (Chomsky, 1965; de Villiers, 2001; Lenneberg, 1967; Wexler, 1998, 2003; Yang, 2002). This assumption is based on the argument that it is impossible for children to learn the linguistic knowledge from the environment given that the input children hear is limited and full of errors and incomplete information (Chomsky, 1965). Yet, children are still able to acquire the linguistic knowledge quickly because of the guidance of innate linguistic hypotheses. Contrasting with the generative approach is the constructivist (or empiricist) approach. The constructivist approach argues that children learn linguistic knowledge from the environmental input that they are exposed to, given that the input contains rich distributional information related to linguistic structures (Christiansen & Charter, 1999; Goldberg, 2006; MacWhinney, 2004; Reali & Christiansen, 2005; Tomasello, 2005). Children figure out the linguistic structures of the ambient language from the input by means of general cognitive mechanisms, such as statistical learning, schematization, and analogy (Abbot-Smith & Tomasello, 2006; Elman et al., 1998; Gomez, 2002; Tomasello, 2003). One phenomenon in language acquisition that allows us to test these approaches is the development of tense and agreement morphemes in young children.

It has been well documented that young children acquiring English produce tense and agreement morphemes inconsistently in obligatory contexts (Brown, 1973; Lahey, Liebergott, Chasnick, Menyuk, & Adams, 1992; Legate & Yang, 2007; Pine, Conti-

Ramsden, Joseph, Lieven, & Serratrice, 2008; Wilson, 2003). These morphemes include the function words or inflections that mark for time, person, and number, such as forms of BE (e.g., *am, is, are, was, were*) and third person singular –s as in *He jumps*. Between ages two and five, young typically-developing children sometimes omit these morphemes, but when they do produce these morphemes, they tend to use them correctly (Wexler, 1994). As (1) exemplifies (Brown, 1973), for a given auxiliary or copula BE form (e.g., *am, are,* or *is*), the child may omit it or use it correctly at a given age. However, why and in what specific context children omit these morphemes is still open to debate.

(1) a. I happy. (Adam, 2;10)
b. I'm tired. (Adam, 2;10)
c. What you looking? (Adam, 2;10)
d. What're you looking? (Adam, 2,10)
e. It messy too. (Adam, 2;9)
f. It's nice outside. (Adam, 2;9)

Current accounts of typical development of tense and agreement morphemes include generative accounts (e.g., Radford, 1990; Wexler, 1998) and a constructivist account (e.g., Tomasello, 2003). The unique checking constraint (UCC) hypothesis (Wexler, 1998, 2003), one of the generative accounts, assumes that children are born with adult-like abstract representations of tense and agreement. The variable production of tense and agreement morphemes in early childhood is attributed to the presence of a maturational constraint—the UCC—in their grammar. In contrast, the usage-based account posits that children learn tense and agreement morphemes from the input in a gradual, piecemeal fashion. Children's variable production of tense and agreement morphemes is ascribed to the child's lack of adult-like abstract representations and use of lexically-specific constructions which may or may not contain these morphemes.

This thesis tests the assumptions of the UCC hypothesis and the usage-based account concerning why young children produce tense and agreement morphemes variably by examining the use of auxiliary and copula BE in three-year-olds via three experiments. Experiment 1 tested the UCC hypothesis and the usage-based approach by examining the effect of subject types on the use of auxiliary *'is'* in three-year-olds via an elicited production task. In a similar vein, Experiment 2 investigated the effect of subject types, predicate types, and predicate word frequency on the use of copula *'is.'* Experiment 3 extended Experiment 1 by exploring the degree of abstractness of representations of auxiliary BE in young children via a structural priming paradigm. While the elicited production task used one identical type of sentence to prompt the production of target sentences, the structural priming paradigm used different types of prime sentences to elicit target sentences. Thus the key manipulation in Experiments 1 and 2 was the structural and frequency properties of the target sentence. This allows us to understand the role of input on children's productions. In Experiment 3 the primary manipulation was the degree of overlap between the prompt and the target sentence with no manipulation of frequency. We can infer the abstractness of representations of auxiliary BE in young children by comparing the production of auxiliary BE in different priming conditions. Based on the findings from these experiments, the last chapter discussed the representation of auxiliary BE in three-year-olds and the potential developmental trajectory of BE forms in relation to subject types in young children.

CHAPTER II

EXPERIMENT 1: ELICITED PRODUCTION AUXILIARY *'IS'* IN YOUNG CHILDREN

Overview

It has been well documented that young children acquiring English produce tense and agreement morphemes inconsistently in obligatory contexts (Brown, 1973; Lahey, Liebergott, Chasnick, Menyuk, & Adams, 1992; Wilson, 2003; Pine, Conti-Ramsden, Joseph, Lieven, & Serratrice, 2008). Tense and agreement morphemes refer to the function words and inflections that mark for time, person, and number, such as forms of BE (e.g., *am, is, are, was, were*) and third person singular –s as in *He jumps*. Though children start to use tense and agreement morphemes early, they pass through a period in which they produce these morphemes inconsistently. One common observation of this inconsistency is that children may frequently omit tense and agreement morphemes in obligatory context, but when they do use these morphemes, they tend to use them correctly. However, why children omit these morphemes are still open to debate (Schütze & Wexler, 1996; Wexler, 1998; Wilson, 2003; Pine et al., 2008).

Current theories that account for the inconsistent use of tense and agreement morphemes in typically-developing children include generative accounts (e.g., Radford, 1990; Wexler, 1998) and the constructivist account (Tomasello, 2003). The unique checking constraint (UCC) hypothesis (Wexler, 1998, 2003), one of the generative accounts, assumes that children are born with an adult-like abstract representation of tense and agreement (or finiteness marking). The variable production of tense and agreement morphemes in early childhood is attributed to the presence of a developmental constraint—the unique checking constraint—in their grammar. In contrast, the usage-based account posits that children learn tense and agreement morphemes from the input. The variable production of tense and agreement morphemes is ascribed to the child's lack

of adult-like abstract representation and use of lexically-specific constructions, which may or may not contain these morphemes. This study tested the predictions derived from the UCC hypothesis and the usage-based approach by examining the effect of subject types (i.e., pronominal, high-frequency lexical NP, and low-frequency lexical NP subjects) on the production of auxiliary 'is' in three-year-olds through an elicited production task. In what follows, we briefly review the basic assumptions, predictions, and empirical evidence for each account. We then address the potential limitations of the existing evidence and lay out the scope of the current study.

The Unique Checking Constraint (UCC) Hypothesis

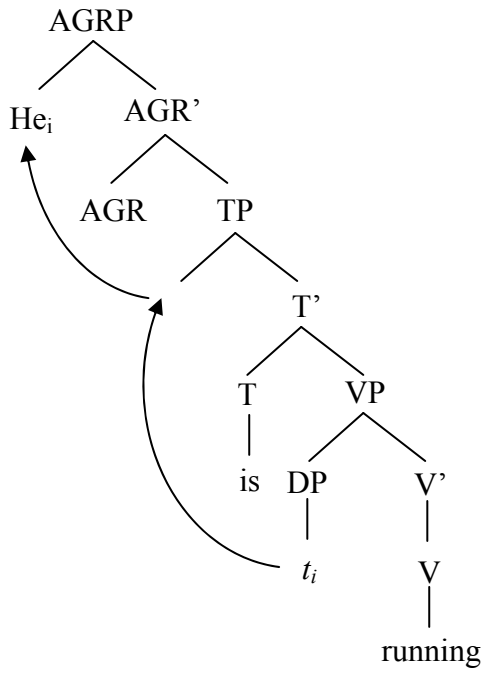
Basic Assumptions of the UCC Hypothesis

The unique checking constraint hypothesis, which was developed within the framework of minimalist syntax (see Radford, 2004), assumes that children are born with innate linguistic knowledge (Wexler, 1998). That is, the child's grammar is basically the same as the adult grammar as early as the two-word stage, which occurs around 18 months of age in typically developing children. One demonstration of this very early knowledge of grammar is that, even though young children frequently omit tense and agreement morphemes, they seldom use these morphemes in the wrong contexts (e.g., **It* <u>*are*</u> *a dog* and **Him* <u>*is*</u> *going to school*, asterisks denote ungrammatical sentences; Wexler, 1996). Omission occurs not because children have not learned the grammatical and phonological properties of these morphemes, but rather because there is a strong maturational constraint in the child's innate grammar—the unique checking constraint (UCC), which blocks the production of tense and agreement morphemes. The UCC becomes relaxed during development through maturational mechanisms that are under biological rather than experiential control.

The UCC hypothesis assumes that the syntactic representation of a sentence with tense marking (i.e., a finite sentence) has separate features of Agreement and Tense

(Avram, 2002; Adger, 2003). To produce correct tense-marking for a sentence (e.g., *He is running*), the sentence subject (e.g., *He*) must check **both** the Agreement and Tense features on the representation, as (1) schematizes (where AGRP = agreement phrase, TP = tense phrase, VP = verb phrase, DP = determiner phrase, t = trace).

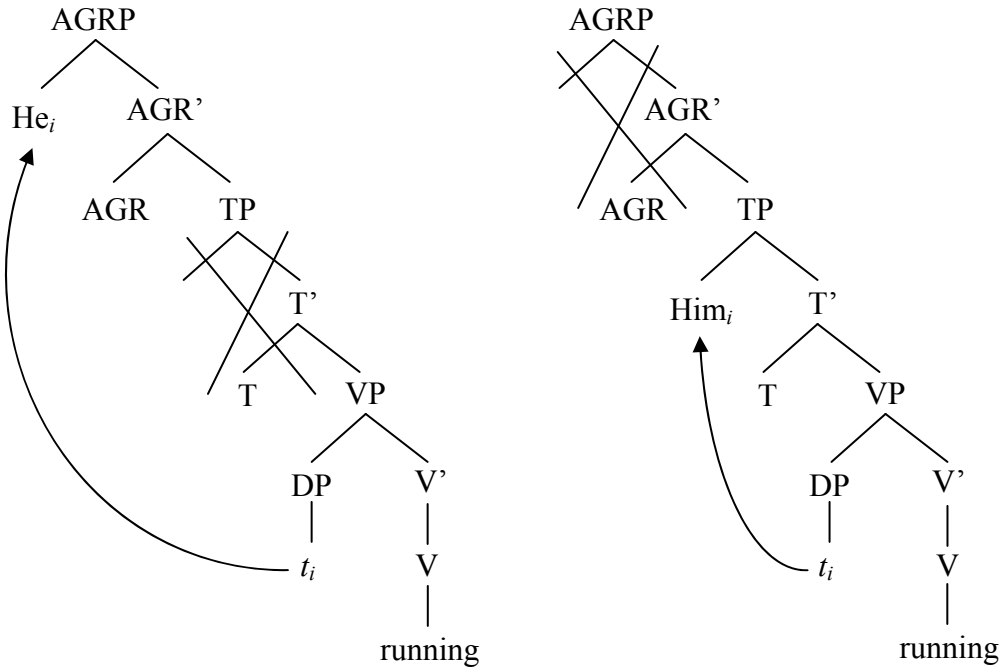
(1) A Simplified Syntactic Representation of *He is running*



However, the UCC limits this checking process in young children such that the sentence subject can check only one feature on the representation, that is, either the Agreement (AGR) feature or the Tense (TNS) feature. Because a syntactic representation with unchecked features is ruled out as ungrammatical by the grammar, the child needs to modify the syntactic representation in order to derive a grammatical sentence, given the UCC. This modification will force either the feature of AGR or TNS to be omitted from the representation. If the feature of Agreement is preserved and the feature of Tense is omitted from the representation, the child would produce a sentence

like **He running*, as (2a) illustrates. In contrast, if the feature of Tense is preserved and the feature of Agreement is omitted, the child may produce sentences like **Him running.* as (2b) demonstrates.

(2) a. AGR preserved, TNS omitted b. AGR omitted, TNS preserved

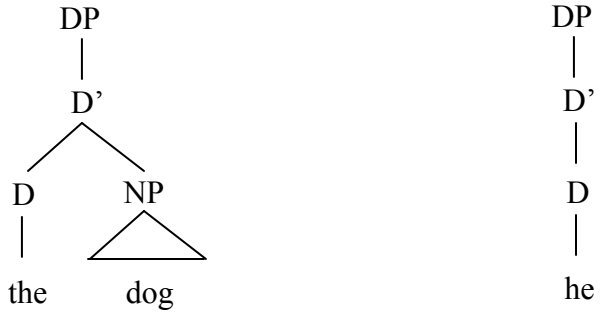


However, it should be noted that finite sentences like *He is running* are also observed in young children who omit tense and agreement morphemes. Finite sentences require both the AGR and TNS features to be preserved on the representation. A question arises here: if the UCC is present in young children's grammar, how could it be possible for them to preserve both the AGR and TNS features on the representations? Wexler (1998) adopted the notion of Minimize Violations to account for this issue. Minimize Violations basically requires the computational system of syntax to choose the derivation that violates as few grammatical properties as possible. If more than one derivation minimally violates the properties, any of them may be chosen for the output sentence. The omission of the features of AGR or TNS under the restriction of UCC

involves one violation of a grammatical property of the syntactic representation. The preservation of both the AGR and TNS features on the representation does not violate any grammatical properties but it does violate the UCC in the child grammar. Because all three of the derivations involve only one violation, any of them may be chosen as the output sentence, which results in variable use of tense and agreement morphemes.

The UCC will be developmentally relaxed as time goes by so that the child's production can eventually conform to the adult form. While the UCC is still present in the child grammar, it would constrain the sentence subject to check only one of the two required features for the derivation of tense and agreement morphemes, leading to the omission of these morphemes. Although not explicitly addressed, one prediction that follows from the UCC hypothesis is that the production/omission rate of a given form of auxiliary BE should be similar in sentences with lexical NP subjects and in those with pronominal subjects because both subject types are determiner phrases (Adger, 2003) and have the same structural properties, as (3) exemplifies (where DP= determiner phrase).

(3) a. Lexical NP subjects b. Pronominal subjects



Empirical Evidence for the UCC Hypothesis

To explore children's knowledge of finiteness, Schütze and Wexler (1996) examined the use of tense and agreement morphemes and its relation to the use of subject

(nominative) case in a corpus of three young children aged from 1; 11 to 3;1. These children used nominative (e.g., *he*) or non-nominative (e.g., *him*) subjects in a sentence. When children used nominative subjects, they sometimes omitted the tense and agreement morphemes (e.g., **He happy*) and sometimes used them correctly (e.g., *He is happy*). However, when children used non-nominative subjects, they only used tense and agreement morphemes in about 5% of these sentences (e.g., **Him is happy*). In addition, children seldom made agreement errors with nominative subjects (e.g., **He are happy*). Based on their findings, Schütze and Wexler (1996) made two claims. First, the features of AGR and TNS are represented separately so that these features can be independently omitted from the child's representation of finite sentences. Second, young children have adult-like knowledge of finiteness, but they simply view finiteness marking of sentences as optional in their grammar. This optionality in their grammar leads to the variable use of tense and agreement morphemes in obligatory contexts.

Similar patterns of use of tense and agreement morphemes were also observed in experimental studies. Wexler, Schütze, and Rice (1998) examined the use of tense and agreement morphemes in children with specific language impairment (SLI) with a mean age of 5;0 and their younger typically-developing peers matched in mean length of utterance (MLU) with a mean age of 3;0 through spontaneous data and an experimental probe. Both in the SLI and the MLU-matched groups, nominative subjects appeared together with inflected verbs, whereas non-nominative subjects almost never appeared with inflected verbs. Both groups often omitted the tense and agreement morphemes, but when they used these morphemes, they tended to use them in the correct context. These findings further support the hypothesis that the features of AGR or TNS can be optionally omitted from the syntactic representation of sentences in young children.

In a longitudinal study, Rice, Wexler, and Hershberger (1998) investigated the growth trajectory of the use of five tense and agreement morphemes in groups of children with SLI and their MLU-matched peers. Each child was tested at 6-month intervals over

3 years, for seven data points. The age range during the study was 2;6 to 6;8 for the MLU-matched group and 4;5 to 8;9 for the SLI group. In the MLU-matched groups, the growth curves for each morpheme showed similar trajectories. All the individual curves showed slow growth at the beginning, rapid growth in the middle, and an asymptote at ceiling level by the end of the study. The slow growth of tense and agreement morphemes in the beginning was interpreted as signaling the presence of a maturational constraint in the grammar (i.e., the UCC); the acceleration phase marked the relaxation of the constraint. This similarity of growth curves across tense and agreement morphemes reinforced the argument that these morphemes are controlled by a common underlying grammatical function.

In summary, the corpus and experimental studies reviewed above support the assumption of the UCC hypothesis that young children have an intact representation of tense and agreement in their grammar, as evidenced by the finding that children seldom make commission errors with these morphemes. The problem that young children have is that the features of AGR or TNS may be omitted due to the presence of the UCC in their grammar. This optional presence of AGR or TNS in the representation leads to the variable production of auxiliary BE and other tense and agreement morphemes.

However, other researchers have argued that these findings can actually be interpreted without assuming that young children have adult-like representation of finiteness. The scarcity of errors like **He **are** happy* or ***Him** is sick* in young children may be due to the low rate of such ungrammatical utterances in the input (Tomasello, 2003). In addition, if the UCC does cause the omission of tense and agreement morphemes in young children, it should affect sentences with different subject types equally. Recent studies from the usage-based approach reviewed below (e.g., Wilson, 2003; Pine et al., 2008) have shown that young children tended to produce tense and agreement morphemes more often with pronominal subjects than with lexical NP subjects.

The Usage-based Approach

Basic Assumptions of the Usage-based Approach

The usage-based approach (Tomasello, 2003; Theakston & Lieven, 2005) holds that language is composed of constructions. Constructions are conventionalized symbolic units with form and meaning pairings, which may include morphemes, words, idioms, sentence frames, and so on (Goldberg, 1995). This approach does not assume that the child has any innate abstract linguistic knowledge. Instead, it assumes that language acquisition involves learning linguistic constructions of different sizes and complexities from the input (Tomasello, 2000, 2003). The usage-based approach suggests that young children inconsistently use tense and agreement morphemes because they have not developed adult-like abstract representations and instead use constructions with fixed lexical items and open slots (i.e., lexically-specific constructions, such as *It's Verb-ing*) that may or may not contain these morphemes (Wilson, 2003; Theakston, Lieven, Pine, & Rowland, 2005; Pine et al., 2008).

In this approach, the acquisition of tense and agreement morphemes begins with gestalt learning of independent word sequences with these morphemes (e.g., *you're welcome*, *what's that?*). At some point, the child discovers the relation between these word sequences and develops a more abstract way to represent these constructions. For instance, initially the child may learn and produce sentences with forms of auxiliary BE in chunks without analyzing the elements inside, such as *He's eating* and *It's running*. After hearing a large number of constructions with similar forms of auxiliary BE, the child starts to see regularities in the input and uses some lexically-specific constructions with auxiliary BE, such as *He's Verb-ing* and *It's Verb-ing.* These constructions are lexically specific, because the child will only use the auxiliary BE together with certain specific subjects, based on how he heard the auxiliary BE used. For instance, the child might say *He's Verb-ing* but not *She's Verb-ing* because he has not learned that *he* and

she both work the same way within the sentence. That is, the child does not have abstract grammatical categories such as pronoun that would unify these. Later in development, the child will figure out from a variety of lexically-specific constructions (e.g., *Mom's Verb-ing, The deer's Verb-ing*) that there are more abstract ways to represent the present progressive construction (i.e., $NP_{subject\text{-}3sg}$*'s + Verb-ing,* where 3sg refers to third person singular) and may use auxiliary '*is'* with any subject type. Eventually, the child acquires the most abstract form $NP_{subject}$ + *Auxiliary BE + Verb-ing* and is able to produce forms of auxiliary BE flexibly and accurately. The variable production of auxiliary BE will decrease gradually as the child acquires the adult-like abstract constructions that contain auxiliary BE. It should be noted that even if children develop adult-like abstract constructions, the frequently-used lexically-specific constructions may remain in the representation and can still be adopted for language production (Bybee, 1995, 2002; Dąbrowska, 2000; Dąbrowska & Lieven, 2005). As long as children have not developed adult-like abstract constructions, they will variably produce auxiliary BE forms.

How would omission of tense and agreement morphemes occur in a child who has not developed abstract representations but only uses lexically-specific constructions of these morphemes? There are at least two possibilities. First, the child may have no or weak representations of the particular lexically-specific constructions that are required by specific discourse contexts (Wilson, 2003; Theakston et al., 2005; Theakston & Lieven, 2008). For instance, the child may only have *He's Verb-ing* and *The dog's Verb-ing* constructions in the representations but not *The sheep's Verb-ing*. However, the discourse context may require the child to say a sentence like *The sheep is crying*. Because there is no *The sheep's Verb-ing* construction in the representation and the child has not yet acquired an abstract construction like $NP_{subject\text{-}3sg}$*'s + Verb-ing*, he may just combine words together based on their semantics and produce *The sheep crying*, leading to the omission of auxiliary BE.

The second possibility is children may learn the constructions without tense and agreement morphemes directly from the input (Theakston et al., 2005; Lieven, 2008; Theakston & Lieven, 2008). For instance, the child may have learned the constructions like *He Verb-ing* and *The dog Verb-ing* due to partial processing of input utterances such as *Is he running, What is he doing*, and *I see the dog running* (Slobin, 1985; Newport, 1990; Freudenthal, Pine, Aguado-Orea, & Gobet, 2007). These constructions without auxiliary BE may compete with and win out over those with auxiliary BE when the child attempts to express the meaning of present progressive. Thus, the child may produce auxiliary BE variably even in similar contexts (e.g., *He's running* vs. *He running*) due to the competition between the correct and incorrect constructions.

Thus, we have seen that input frequency plays a crucial role in the specific constructions the child may use within the usage-based approach (Wilson, 2003; Theakston et al., 2005). For instance, children hear specific pronoun + auxiliary BE combinations (e.g., *He's eating*) more often than specific lexical NP + auxiliary BE combinations (e.g., *The deer's eating*) in the input. Because more frequent constructions will be more entrenched and easier to retrieve than less frequent constructions, children are more likely to use a particular auxiliary BE form correctly in sentences with pronominal subjects than in those with lexical NP subjects (Wilson, 2003). Therefore, frequency rather than structural properties is the driving force of acquisition of auxiliary BE, as well as other tense and agreement morphemes, within the usage-based approach.

Empirical Evidence for the Usage-based Approach

Wilson (2003) investigated the acquisition of copula BE, auxiliary BE, and third person singular *–s* (3SG-*s*) in longitudinal transcripts from five children aged from 1;6 to 3;5. The children demonstrated different types of variability in using these tense and agreement morphemes. First, the overall production rates of copula BE, auxiliary BE, and 3SG-*s* differed significantly within each child (i.e., between-morpheme variability).

Second, for the same morpheme (e.g., auxiliary *is* and *are*), the production rate varied as a function of subject types within a given child. The target morphemes were produced more accurately in sentences with closed-class subjects than in those with open-class subjects. Third, for a given BE form, the provision rates also varied in sentences with different closed-class subjects. For instance, the provision rate of auxiliary '*is*' tended to be higher in sentences with the subject *he* than in those with the subject *she*. These findings were replicated in a study by Pine and colleagues (2008), in which sentences with non-nominative subjects were included in analysis and the lexical knowledge of tense and agreement morphemes of children was controlled. The studies of Wilson (2003) and Pine and colleagues (2008) both suggested that children acquire these morphemes by learning lexically-specific constructions from the input. They argued that variable use of tense and agreement morphemes reflected inconsistent knowledge of the range of distinct constructions.

Joseph, Serratrice, and Conti-Ramsden (2002) examined the development of copula and auxiliary BE in children with SLI (aged from 3;1 to 4;8) and their peers matched on mean length of utterance (MLU) in words (aged from 1;8 to 2;4) via a corpus study. The SLI and the MLU-matched groups both showed variability in using copula and auxiliary BE. For both groups, copula BE was produced more often than auxiliary BE, and the form '*is*' was produced more often than the forms '*are*' or '*am*.' For the same BE form (e.g., copula *is*), the production rate was higher in sentences with personal-pronoun subjects (e.g., *he*) than in those with other pronominal subjects (e.g., *this*), which was in turn higher than those with lexical NP subjects. Furthermore, the distribution of BE forms in the maternal input accounted for a significant proportion of the variance in the children's use of BE forms. Joseph and colleagues (2002, p. 167) therefore argued that the variable use of copula and auxiliary BE is not "a random, across-the-board" phenomenon. The observed asymmetry between pronominal and lexical subjects is attributable to the input frequency: children hear BE forms in the former more often than

in the latter. Copula and auxiliary BE are therefore particularly entrenched in the sentence constructions with pronominal subjects and thus more likely to be produced.

Theakston and colleagues (Theakston et al., 2005) investigated the use of auxiliaries BE and HAVE in a longitudinal data set from 11 typically-developing children aged from two to three years. Although the production rates of auxiliaries BE and HAVE were positively correlated, there were significant differences in the production rates of individual forms of BE and HAVE. For auxiliary BE, children provided the form *is* (61.9%) more frequently than the form *am* (32.7%) in obligatory contexts. For auxiliary HAVE, they produced the form *has* (62.4%) more often than the form *have* (25.5%) in obligatory contexts. Theakston and colleagues suggested that the variable use of auxiliaries BE and HAVE in early childhood cannot be explained by an underlying optionality, which would be expected to function equally on different forms of the same auxiliary. The variability was also not attributable to performance limitations, in that there were no significant differences in MLU between utterances containing the target auxiliary and those in which the target auxiliary was omitted. In contrast, except for the pronominal subject *you*, the maternal input frequency was significantly correlated with the age at which a specific <u>subject + auxiliary</u> construction was acquired. Theakston and colleagues therefore argued that the input frequency and the acquisition of lexically-specific constructions may provide an adequate account for the variable use of auxiliary BE and HAVE. This variability resulted from the child's acquisition of specific <u>subject + auxiliary</u> constructions individually, depending on what he had heard from the input.

In summary, both corpus and experimental studies reviewed here support the assumption of the usage-based approach that children acquire tense and agreement morphemes from lexically-specific constructions that contain these morphemes in the environmental input. The production/omission of tense and agreement morphemes is not unified under a common underlying grammatical representation of tense marking. The assumption is supported by variation of production accuracy in different tense and

agreement morphemes (e.g., auxiliaries BE and HAVE) and in different forms of the same morpheme (e.g., *is, am,* and *are* of auxiliary BE). It is further evidenced in different production accuracy of the same form of a given tense and agreement morpheme (e.g., auxiliary *is*) in sentences with different subject types (e.g., pronominal and lexical) within the same child. The variable use of tense and agreement morphemes is attributed to young children's use of lexically-specific constructions that may or may not contain these morphemes and lack of adult-like abstract representation of these morphemes. Tense and agreement morphemes tend to be used more correctly in more frequent constructions (e.g., pronoun + auxiliary BE) than in less frequent ones (e.g., lexical NP + auxiliary BE).

However, within- and between-morpheme variability are not necessarily evidence against the notion of unified underlying representations of tense marking. Rice and colleagues (1998) showed that the growth curves were quite similar across tense and agreement morphemes in English. The similarity of individual growth curves indicated that these morphemes were unified under one grammatical function. They further argued that the variation of production rates between and within tense and agreement morphemes may result from the phonological properties of the surface form (e.g., duration) and grammatical differences across these morphemes. For instance, auxiliary DO tends to appear before negations in declarative sentences but auxiliary BE does not have this constraint. Though these factors may seem to "pull apart the [tense and agreement] morphemes over time" (Rice et al., 1998, p. 1427), they do not affect the similarity of growth curves between these morphemes.

Limitations of the Existing Evidence

Because empirical evidence supports both the UCC hypothesis and the usage-based approach, why children variably use tense and agreement morphemes remains unclear. However, although the UCC hypothesis can well explain the inconsistent use of

tense morphemes though the presence of the UCC in the grammar, it does not provide specific predictions regarding how surface and grammatical differences lead to between- and within-morpheme variability. In addition, the effect of the UCC should be equal for all forms of copula BE as well as auxiliaries BE and HAVE, regardless of subject types, a prediction challenged by evidence from the usage-based approach (Wilson, 2003; Theakston et al., 2005; Pine et al., 2008).

On the other hand, although the evidence from the usage-based approach seemed to challenge the UCC hypothesis, it may not be straightforward. A strict test of the UCC hypothesis should examine the child's production of tense and agreement morphemes at a given point of developmental time (Ambridge & Pine, 2006). However, the language samples used in the previous studies (e.g., Wilson, 2003; Pine et al., 2008) were collected over a period from 6 to 14 months. The production rate of tense and agreement morphemes by subject type was computed by collapsing the language samples together. Collapsing the language samples may have obscured this developmental change. For example, it is possible that the production rate of auxiliary *'is'* with lexical NP subjects might be equal to pronominal subjects during the later period of data collection due to the development of abstract constructions.

In addition, the determination of an obligatory context for copula and auxiliary BE in language samples may sometimes be ambiguous (Gerken, 2000). For example, the child's utterance **it there* may be coded as omission of copula '*is'* in this context. However, this utterance might actually mean *Put it there* or *I want it there*. These contexts would make it difficult to draw firm conclusions about the relationships between use of tense and agreement morphemes and specific constructions. Furthermore, previous studies have not examined whether lexical NP subjects with different frequencies would affect the production of a given form of tense and agreement morphemes to different degrees.

The Scope of the Current Study

To reduce the ambiguity of obligatory contexts and explore the effect of lexical NP subjects with different frequencies on the production of tense and agreement morphemes, this study adopted an elicited production task to investigate the child's use of auxiliary *'is'* at a given time point of development in sentences with three different subject types: pronominal subjects, high-frequency lexical NP subjects, and low-frequency lexical NP subjects. Elicited production allows for greater control of the target utterances than spontaneous samples, including factors like obligatory context, sentence length, verb type, sentence structure, and subject type. In addition, the usage-based approach stresses the role of input frequency in the acquisition of tense and agreement morphemes. Elicited production enables us to manipulate the frequency of subject types to examine this effect. In this way, we may offer more precise evidence with regard to the UCC hypothesis and the usage-based approach.

The specific question addressed is: Does production accuracy of auxiliary *'is'* in young children differ in sentences with pronominal, high-frequency lexical NP, and low-frequency lexical NP subjects? The UCC hypothesis would predict that production accuracy of auxiliary *'is'* in young children would be similar with different subject types, given that the UCC should affect sentences with pronominal and lexical NP subjects equally. In contrast, the usage-based approach would predict that production accuracy of auxiliary *'is'* in young children should be higher with pronominal subjects than with high-frequency lexical NP subjects, followed by low-frequency lexical NP subjects.

Methods

Participants

Seventy-eight children with an age range from 2;7 (year; month) to 3;5 were recruited for participation in this study. Children at this age range were recruited because they were likely to be using auxiliary *'is'* variably (Rice, Wexler, & Cleave, 1995;

Wilson, 2003; Theakston et al., 2005). Of these 78 children, 20 children (11 girls) with a mean age at 3;0 (ranged from 2;8 to 3;4) were included in the analysis reported below. Fifty-eight of the 78 children recruited were excluded for analysis because they had a history of speech and language problems (n = 6), did not produce at least three scorable items (n = 4) or at least one correct usage of auxiliary *'is'* (n = 5), had reached ceiling level (n = 24), or did not finish the study (n = 19).

To be included for analysis, the participants had to be monolingual native English speakers and did not have a history in speech and/or language problems. They had to perform below the ceiling level (i.e., 90% correct; Brown, 1973) of auxiliary 'is' so that sources of variability could be legitimately examined. A child would be included as long as he/she produced below 90% correct of auxiliary *'is'* in one or more of the three conditions. In addition, the participants had to produce at least 3 scorable items for each condition (Balason & Dollaghan, 2002) and produce at least one correct usage of auxiliary *'is'* in the task. This last requirement is because it has been argued that testing the UCC hypothesis is only valid when the child has shown the target morpheme in his/her productive lexical inventory (Schütze, 2001; Pine et al., 2008).

All of the 20 participants were typically developing as documented by parent report, performance above 10th percentile on a standardized language test, hearing within normal limits per ASHA standards (American Speech-Language-Hearing Association, 1997), and no history of cognitive or motor disorders. The standardized language tests were *Preschool Language Scale-Fourth Edition* (*PLS-4*, Zimmerman, Steiner, & Pond, 2002) for children who were younger than 3;0 and *Structured Photographic Expressive Language Test-Preschool* (*SPELT-P*, Werner & Kresheck, 1983) for those who were at or older than 3;0.[1] We administered the SPELT-P for those who were 3;0 or older

[1] The sensitivity and specificity are 83.3% and 90.5% for SPELT-P (Plante & Vance, 1995) and 80% and 88% for PLS-4, respectively (Spalding, Plante, & Farinella, 2006).

because we only tested three-year-olds at the beginning stage of this study. The PLS-4 was added when the age range was extended to 2;7. As a component of assessment, we collected a conversational language sample for each child. All children demonstrated an MLU within the typical range with reference to Miller's (1981) norm.

Stimuli

The target items consisted of simple declarative transitive sentences that required the use of auxiliary *'is'* and varied with regard to subject types (i.e., pronominal, high-frequency lexical NP, and low-frequency lexical NP subjects). Each condition had 10 target sentences. To ensure that the words used in the target sentences were likely to be familiar to typically developing children around 3 years of age, the selection of lexical items for the subjects and verbs in the target sentences was primarily based on Part 1 of *The MacArthur-Bates Communicative Development Inventory: Words and Sentences* (*MB-CDI: WS*, Fenson et al., 1993).

To select subject words with different frequencies, we used the frequency list of Moe, Hopkins, and Rush (1982) to determine the token frequency of 183 nouns from *MB-CDI: WS* and 16 additional nouns that were considered child-friendly. Nouns with frequencies above the median (i.e., 29) for all 199 nouns were considered high frequency and selected as subjects for the high-frequency lexical NP subject condition. Pronouns that were under consideration for the pronominal subject condition included *he, she, it, this*, and *that*. We did not choose *this* and *that* as the subjects because they did not sound felicitous in the transitive present progressive sentences (e.g., ?*This is/That's eating a cake*). Because all the rest of the three pronouns had the frequencies above 2700, which was much higher than the frequency (i.e., 860) of the most frequent noun among the 199 nouns, we did not include a low-frequency pronominal subject condition in this study.

In total, there were 3 pronominal subjects (i.e., *he, she, it*), 6 high-frequency lexical NP subjects (i.e., *cat, dog, duck, goat, mom, pig*) and 5 low-frequency lexical NP

subjects (i.e., *ant, deer, frog, queen, sheep*) for the target sentences. Because each condition had 10 sentences, all of these lexical items were repeated as subjects across target sentences, except *duck* and *Mom*. Each pronominal subject appeared in no more than 4 target sentences, while each lexical NP subject appeared in no more than 2 sentences. We repeated the lexical NP subjects in target sentences for two reasons. First, there were not many choices of nouns available under the constraints mentioned above, especially for low-frequency lexical NP subjects. Second, because repeating the pronominal subjects for target sentences was inevitable, repeating the lexical NP subjects would make the target sentences more similar across conditions.

The median and range of raw frequency and mean log frequency of lexical items selected for target subjects are tabulated in Table 2.1 by condition.

Table 2.1. Frequency Distribution of Lexical Items Chosen for Target Subjects by Condition

	Median	Range	Log Mean (SD)
Pronominal Subject (e.g., *he*)	5408.0	2753-7583	3.69 (0.18)
High-frequency NP Subject (e.g., *the dog*)	448.0	167-606	2.54 (0.23)
Low-frequency NP Subject (e.g., *the deer*)	10.0	5-27	0.99 (0.29)

Note: SD = standard deviation

A one-way ANOVA showed that the mean log frequency of lexical items chosen for target subjects differed significantly across sentences with different subject types, F (2, 27) = 320.03, p = .000, η_p^2 = .96. *Post hoc* Tukey test at .05 level showed that the mean log frequency of the selected pronouns was higher than the high-frequency lexical NPs. In turn, the mean log frequency of high-frequency lexical NPs are higher than the selected low-frequency lexical NPs.

The frequency comparison above was based on lexical frequency. However, the prediction of the usage-based approach is actually based on the **co-occurrence** frequency of the subject and auxiliary *'is.'* What remains unknown is whether high-frequency nouns also co-occur with auxiliary *'is'* more often than low-frequency nouns. To answer this question, we counted the co-occurrence frequency of each lexical item chosen as a target subject and auxiliary *'is'* in the parental input of the American English corpora in CHILDES (MacWhinney, 2000) and then computed the correlation of log frequency of each lexical item and the log frequency of co-occurrence. The log frequencies were significantly correlated, $r = .779$, $p < .01$. This high correlation validates the use of lexical frequency to choose the subject words for the target sentences in this study.

All the subject words denoted an animal or person so that they could serve as the agent of the transitive verb selected the target sentence. The lexical NP subject words were also required to be monosyllabic to control for sentence length and end with a vowel or non-sibilant consonant to create contexts that allow the contraction of auxiliary *'is.'* Contexts for contraction were important because the usage-based approach suggests that children may learn auxiliary *'is'* from the specific subject + contracted auxiliary *'is'* constructions (e.g., *he's, Mommy's*) in the input (Wilson, 2003). Thus, contexts for contracted auxiliary *'is'* allowed us to test the prediction of the usage-based approach on the effect of subject types more precisely.

A set of 10 transitive verbs from MB-CDI: WS was used repeatedly across three conditions in order to control for verb frequency. The length of each sentence was controlled at 6 syllables (e.g., *He's eating a cookie, The goat's driving the car*). The target sentences with auxiliary '*is'* were listed in Appendix A.

We used pictures to elicit target sentences in this study. Because the target sentences/pictures involved fantasy events (e.g., *the goat's driving the car*), we asked 21 adults to rate the fantisiness of the target sentences together with their corresponding pictures though the effect of fantisiness on the production of tense and agreement

morphemes remains unknown. The procedure of conducting the rating was specified in Appendix B. Kruskal-Wallis H tests showed that there was no significant difference in fantasiness of target sentences/pictures across subject types, $\chi^2 = 2.81$, $df = 2$, $p = 0.25$.

Procedures

Each child was tested individually by a trained examiner. The examiners were eight undergraduate students and two graduate students at the University of Iowa, who were all monolingual English speakers. The task began with a warm-up comprehension activity, in which children were required to point to one of four pictures at each trial after hearing the instruction *"Show me the _____"* (e.g., *Show me* <u>*the dog*</u>) from the examiner. All of the subject lexical NPs in the target sentences were tested in this activity. This was to familiarize the child with these NPs and the child's performance did not affect inclusion/exclusion for analysis. If the child made errors, the examiner named all four pictures for the child and asked the child to point to the target picture again.

Next, the child was invited to play a game to talk about pictures. We adopted a parallel structure method to elicit the production of the target sentences. Target sentences were presented in drawings that depicted the event described by the target sentences (e.g., *The deer's eating a cake*) together with a contrastive event (e.g., *The horses are wiping the floor*). The contrastive events always used plural-form subjects and auxiliary '*are.*' The contrastive sentences are listed in Appendix A. Sample pictures of target and contrastive sentences are illustrated in Appendix C. To eliminate order effects, all the target sentences were randomized into one list in the task. Five practice items were used to familiarize the child with the experiment.

Each trial began with the examiner pointing to the picture and prompting the child by describing the contrastive event. For all items, we included the verb phrase in the prompt (e.g., *Eating a cake.*) in order to reduce the task demands. To maximize the use of target subject words, we prompted the child slightly differently for sentences with

pronominal subjects (see 4a) or lexical NP subjects (see 4b). To elicit responses with pronominal subjects, the subject words in the contrast events were always pronouns. We also prompted the target pronominal subject in the last sentence of the instruction. To elicit responses with lexical NP subjects, the subject words in the contrast events were always lexical NPs. We did not prompt the lexical subject in the last sentence of the prompt because children were more likely to use lexical NP subjects in sentences if they did not hear the NP in the previous discourse (Matthews, Lieven, Theakston, & Tomasello, 2006).

(4) a. Target: He's eating a cookie.

Prompt: Look! Making a cake. Eating a cookie. **They are** making a cookie. What's happening **to him?**

b. Target: The deer's eating a cake.

Prompt: Look! Wiping the floor. Eating a cake. **The horses are** wiping the floor. What's happening **in this picture?**

At the end of the prompt, the examiner pointed to the target picture and waited for the child's response. When the child responded with a non-target subject in the sentence, the examiner only prompted if the child substituted a pronoun with a noun or vice versa. Under these circumstances or if the child only produced the verb phrase, the examiner prompted the child by saying "Can you start by saying Target Subject Word?" (e.g., *Can you start by saying he/the deer?*). The examiner maximally used three additional prompts for each item.

Transcription and Coding

All responses were audiotaped for further analysis. The children's responses were transcribed by the examiner and coded as correct, incorrect, or unscorable. If there was

more than one response for a given item, we typically treated the last response as the target response. The rules of coding are described below.

1. Correct productions

A. Target production (245 responses): All elements in the target sentences were present, with auxiliary *'is'* being contracted (e.g., *The dog's eating a cake*) or uncontracted (e.g., *The dog is eating a cake*). Production with missing elements unrelated to tense marking or with non-target verb or object NP were also counted as target productions (e.g., *The dog's* <u>*chewing food*</u>).

B. Production with an alternative subject (66 responses): The target subject was replaced by another subject. Although correct, the response was counted as a correct production of the alternative subject type. If the responded lexical NP subjects changed frequency, they were reclassified as high-/low-frequency based on the median frequency of 29.

C. Production with a non-nominative subject case (7 responses): The auxiliary and verb inflection were correctly used but the case of the pronominal subject form was incorrect, such as <u>*Him*</u> *is driving an airplane* instead of *He's driving an airplane*. In this case, the response was still counted as a correct production of pronominal subject type. It is worth noting that the use of non-nominative subject case with a tense and agreement morpheme is not predicted by either the UCC hypothesis (Schütze & Wexler, 1996) or the usage-based approach (Tomasello, 2003). Nevertheless, because the study's focus was on auxiliary production, we treated this response as correct to avoid underestimating the child's ability to use auxiliary *'is.'*

2. Incorrect Production

A. Omission of auxiliary (199 responses): Auxiliary *'is'* was omitted (e.g., **The dog eating a cake*). Like the correct productions, we also recorded whether the

child used target sentences, non-target subjects, or non-nominative subjects in the omission errors. In total, there were 148 omission errors with target sentences, 36 with alternative subjects, and 15 with non-nominative subjects.

B. Agreement error (0 responses): A third person plural auxiliary '*are'* was used with third person singular subject (e.g., **The dog are eating a cake*).

3. Unscorable Production

A. Use of an unrelated structure (16 responses): A grammatical sentence was produced but was unrelated to the target structure (e.g., *The dog eats the cake*, *The dog's eaten a cake,* or *The dog is naughty*).

B. Omission of subject (61 responses): The subject was omitted (e.g., *Eating a cake* or *Is eating a cake*). Given that no subject was included, we were unable to code the subject type for this response.

C. No response (6 responses): the child did not respond to the stimuli.

Recall that, to be included, each child had to produce at least 3 scorable sentences in each condition. This criterion of 3 scorable sentences was chosen because this number would allow us to reliably test to the child's performance in each condition without excluding too many participants from analysis (Lahey et al., 1992; Balason & Dollaghan, 2002). Table 2.2 summarizes the mean number of scorable sentences by condition.

Table 2.2. Mean Number of Scorable Sentences by Condition

Condition	Mean	SD	Range
Pronominal subjects	8.70	2.42	3-14
High-frequency lexical NP subjects	8.90	3.49	3-16
Low-frequency lexical NP subjects	8.20	1.91	5-11

Note: The maximum number of scorable items of each condition was greater than 10 because some children replaced the target subject types with alternative subject types.

Reliability

To check the transcription and coding reliability in this study, we randomly sampled 20% of the participants (n = 4). These 4 samples were re-transcribed and re-coded by the first author. The mean reliability was 92.03% in identifying the target sentences, 96.39% in transcribing content words in the target sentences, 92.06% in transcribing grammatical morphemes, and 91.46% in coding correctness of using auxiliary *'is.'*

Statistical Analysis

The dependent measure was accuracy of auxiliary *'is'* of each scorable response for each child. Because of the dichotomous (i.e., correct or incorrect) nature of the dependent measure, a binominal logistic regression model was adopted to evaluate the effect of subject types on the production of auxiliary *'is*.' Like linear regression, the logistic model relates one or more predictor variables to the dependent variable. The predictor variables may be either categorical (e.g., subject types) or continuous. When a categorical variable has more than two levels, one of the levels is selected as reference category/condition, as (5) exemplifies (where $SUBJ_{Pron}$ refers to the pronominal subject condition; $SUBJ_{Low}$ refers to the low-frequency lexical NP subject condition).

(5) $\text{Logit}(Y_i) = \beta_0 + \beta_1 SUBJ_{Pron} + \beta_2 SUBJ_{Low}$

In (5), the model predicts the effect of subject types on the accuracy of auxiliary *'is*' and the high-frequency lexical NP subject condition is selected as the reference category. The β-coefficients can be used to compute the logit associated with each condition. The logit is the natural logarithm of predicted odds for a given event, such as the log odds of producing auxiliary *'is'* accurately in the pronominal subject condition.

The β-coefficients associated with each term (except the intercept term) also reflect the log odds ratios between the category/condition of interest and the reference category/condition. For instance, β_1 in (5) reflects the log odds ratio of producing correct auxiliary *'is'* when the pronominal subject condition is compared to the high-frequency lexical NP subject condition. A positive β-coefficient indicates an increased likelihood to produce auxiliary *'is'* accurately in the condition of interest (e.g., pronominal subjects) as compared to the reference condtion (e.g., high-frequency lexical NP subjects). The *p*-values of the β-coefficients reported in the model indicate whether the β-coefficient is significantly different than 0. For instance, if β_1 in (5) is positive and reaches a significant level, this would suggest that children are more likely to produce auxiliary *'is'* correctly with pronominal subjects than with high-frequency lexical NP subjects. Log odds ratios can be transformed into odds ratios (OR) to ease interpretation.[2]

We chose this analysis rather than using proportion data in a repeated-measures ANOVA model to avoid distortion of statistics, violation of the assumption of normal distribution, and examination of the random effects in separate (by-item, by-subject) ANOVAs (Baayen, Davidson, & Bates, 2008; Dixon, 2008; Jaeger, 2008; Quené & van den Bergh, 2008). Inclusion of random factors within the model also weights the responses via the number of data points within each level of an independent variable (Quené & van den Bergh, 2008). Thus a level with more scorable responses (e.g., temporary-adjectival predicates among predicate types) is treated as more reliable and comparisons associated with this level are more likely to reach significance.

---

[2] OR larger than 1 means that children are **more** likely to produce auxiliary *'is'* (or *'are'*) in the condition of interest than in the reference condition. In contrast, OR less than 1 means that children are **less** likely to produce auxiliary *'is'* in the condition of interest than in the reference condition. For instance, if the β-coefficient associated with pronominal subjects is 0.97, this means that the log odds ratio of producing auxiliary *'is'* correctly between the pronominal subject condition and the reference condition (i.e., high-frequency lexical NP subjects) is 0.97 (OR = exp(0.97) = 2.64). It means that children are 2.64 times more likely to use auxiliary *'is'* correctly with pronominal subjects than with high-frequency lexical NP subjects.

Results

Sentence Length of Target Responses

Because children omitted and/or changed sentence elements, the sentence length of the target responses could have changed across subject conditions. The difference in sentence length could confound the effect of subject types on the production of auxiliary *'is'*. To examine this possibility, we computed the length of target responses (excluding auxiliary *'is'*) in syllables for each item produced by each child and compared sentence length across subject conditions (Theakston, et al., 2005).

The mean sentence length of target responses excluding auxiliary *'is'* was 5.76 syllables (SD = 0.39) for the pronominal subject condition, 5.82 syllables (SD = 0.54) for the high-frequency lexical NP subject condition, and 5.77 syllables (SD = 0.33) for the low-frequency lexical NP subject condition. A one-way repeated-measures ANOVA showed that mean sentence length did not differ significantly across subject conditions, $F(2, 38) = 0.153$, $p = .86$. Thus we see that the sentence length remained similar across different subject types even if the child omitted and/or changed sentence elements in the target responses and can be ignored in all future analyses. To be especially cautious, we also considered sentence length as a predictor of production accuracy in each of the mixed model binomial logistic regressions reported below. It was not a significant predictor (all ps > 0.14), and did not significantly improve model fit. For these reasons sentence length was dropped from all models and coefficients are not reported.

Effect of Subject Types on the Production of Auxiliary *'is'*

Table 2.3 presents the actual proportion correct of auxiliary *'is'* across children by subject type. To examine the effect of subject types on the production of auxiliary *'is,'* a mixed model binomial logistic regression was performed, with child and item treated as random factors and subject type treated as a fixed factor. Table 2.4 presents the results of the model.

Table 2.3. Mean proportion Correct of Auxiliary *'is'* across Children by Subject Type

	Mean	SD
Pronominal Subjects	0.62	0.22
High-frequency Lexical NP Subjects	0.54	0.31
Low-frequency Lexical NP Subjects	0.66	0.33

Note: SD = standard deviation

Table 2.4. Regression Model Showing the Likelihood of Producing auxiliary *'is'* Correctly by Subject Type

	Variance	SD		
Random Factors				
Item	0.02	0.16		
Child	1.30	1.14		
	Coefficient	**SE**	**OR**	**p-value**
Fixed Factors				
Intercept	0.32	0.31	1.38	0.31
Subject Type (reference condition = high-frequency lexical NP subjects)				
Pronominal Subjects	0.30	0.26	1.35	0.24
Low-frequency Lexical NP Subjects	0.47	0.26	1.60	0.07

Note: SE = standard error; OR = odds ratios

The model indicated that the likelihood of producing auxiliary *'is'* correctly in the high-frequency lexical NP subject condition did not differ significantly from the pronominal subject (Pronominal/HighNP OR = 1.35) or the low-frequency lexical NP subject conditions (LowNP/HighNP OR = 1.60). However, this model did not directly reveal whether the pronominal and low-frequency lexical NP subject conditions differed

from each other. To examine the difference between these conditions, we re-ran the model with pronominal subjects as the reference condition (Jaccard, 2001). The results make it clear that there was also no significant difference between the pronominal and low-frequency lexical NP subject conditions (LowNP/Pronominal OR = 1.19, $p = 0.53$).

Comparison with Other Studies

As can be seen in Table 2.4, the likelihood of producing auxiliary *'is'* accurately did not differ across subject types in young children; a finding which is inconsistent with the previous corpus studies (e.g., Wilson, 2003; Pine et al., 2008). Because the previous studies did not divide lexical NP subjects by frequency, we combined sentences with high-frequency and low-frequency lexical NP subjects together in a mixed model logistic regression to allow a more direct comparison between our data and the results from these studies. Even with this modification, the likelihood of producing auxiliary *'is'* correctly did not differ across subject types (Pronominal/Lexical OR = 1.11, $p = 0.65$).

Discrepant findings between the current study and previous research could result from the differences in the design. First, the previous studies used language samples collected over a period of time to explore the child's production of auxiliary *'is'* and other tense and agreement morphemes. In contrast, the current study adopted an experimental paradigm to examine children's use of auxiliary *'is'* at a given point of developmental time. One possible reason for the discrepant results could be that we observed children's performance using different-sized time windows.

Another possible reason could be that we examined children at different developmental stages. The age of children at the beginning of the study in Wilson (2003) was from 1;8 to 2;8 and in Pine and colleagues (2008) was from 1;8 to 2;0, both of which were younger than the current study. The reason these studies showed a high-frequency advantage might have been because their participants were young and still relied heavily on highly frequent/ lexically-specific constructions to produce tense and agreement

morphemes. The presence of older children (ages 2;8-3;4) with more abstract representation in the current study may have obscured this finding. Furthermore, it is possible that children in the current study were at various developmental levels. Children at higher developmental levels might have developed certain abstract constructions, allowing them to produce auxiliary *'is'* more flexibly without being limited to only lexically-specific constructions. In contrast, children at lower developmental levels might still rely heavily on highly frequent/ lexically-specific constructions due to the item-based nature of their representations. The facilitative effects of high-frequency subjects may have been obscured because we mixed children with different developmental levels together. Based on this logic, we may observe the high-frequency advantage in children with lower developmental levels of tense marking.

To test this possibility, we considered the child's developmental level of tense and agreement morphemes in the analysis and examined whether the effect of subject types on the production of auxiliary *'is'* would vary as a function of tense development, that is, interaction between subject types and the child's tense development. Two indices of tense development were considered and will be discussed in turn: the finite verb morphology composite (Leonard, Miller, & Gerber, 1999) and the tense productivity score (Hadley & Short, 2005).

The first index was the finite verb morphology composite (FVMC; Leonard et al., 1999). This index is based on the child's percent correct use of four tense and agreement morphemes in obligatory contexts in a language sample: copula BE, auxiliary BE, third person singular *–s*, and past tense *–ed*. The mean FVMC for the children in our sample was 76.9% (SD = 15.8%, range = 33.3% - 96.8%). All children fell within the normal range for their age as reported in Goffman and Leonard (2001).

To examine whether there was an interaction effect of FVMC and subject types on the use of auxiliary *'is'*, we added FVMC into the model as a main effect and as an interaction term. Two mixed models of binomial logistic regression were performed: one

without interaction terms and one with interaction terms. In both models, each child and item were treated as random factors, and subject types and the child's FVMC were treated as fixed factors. We compared the log-likelihood values of these two models with the model with subject type alone (i.e., the reduced model) using a chi-square test for goodness of fit to examine whether adding FVMC improved prediction of the production of auxiliary *'is'* (Kleinbaum, Kupper, Muller, & Nizam, 1998). The tests showed that neither the model with subject type and FVMC as main effects ($\chi^2 = 1.12$, $df = 1$, $p = 0.28$) nor the model with interaction between these two factors ($\chi^2 = 1.46$, $df = 3$, $p = 0.69$) significantly improved model fit, as compared to the reduced model. That is, adding the child's FVMC as a main effect or an interaction term did not improve the model fit in terms of predicting the production accuracy of auxiliary *'is.'*

The second index we considered was tense productivity scores (Hadley & Short, 2005). This index is computed via counting the use of the following 5 tense and agreement morphemes in the language sample: copula BE, auxiliary BE, third person singular *–s*, past tense *–ed*, and auxiliary DO. The child receives a point for each sufficiently different use of these morphemes up to 5 points for each morpheme. To be counted as a sufficiently different use, verb inflections must appear on different lexical verbs and copula BE, auxiliary BE, and auxiliary DO must occur with different sentence subjects in declarative or interrogative sentences. To avoid overestimating the child's tense development, all the forms contracted with pronominal subjects (e.g., *he's, it's*) are excluded for analysis. The total score for a child can range from 0 to 25. The mean tense productivity score for the children in this study was 8.75 (SD = 4.13, range = 2 -15).

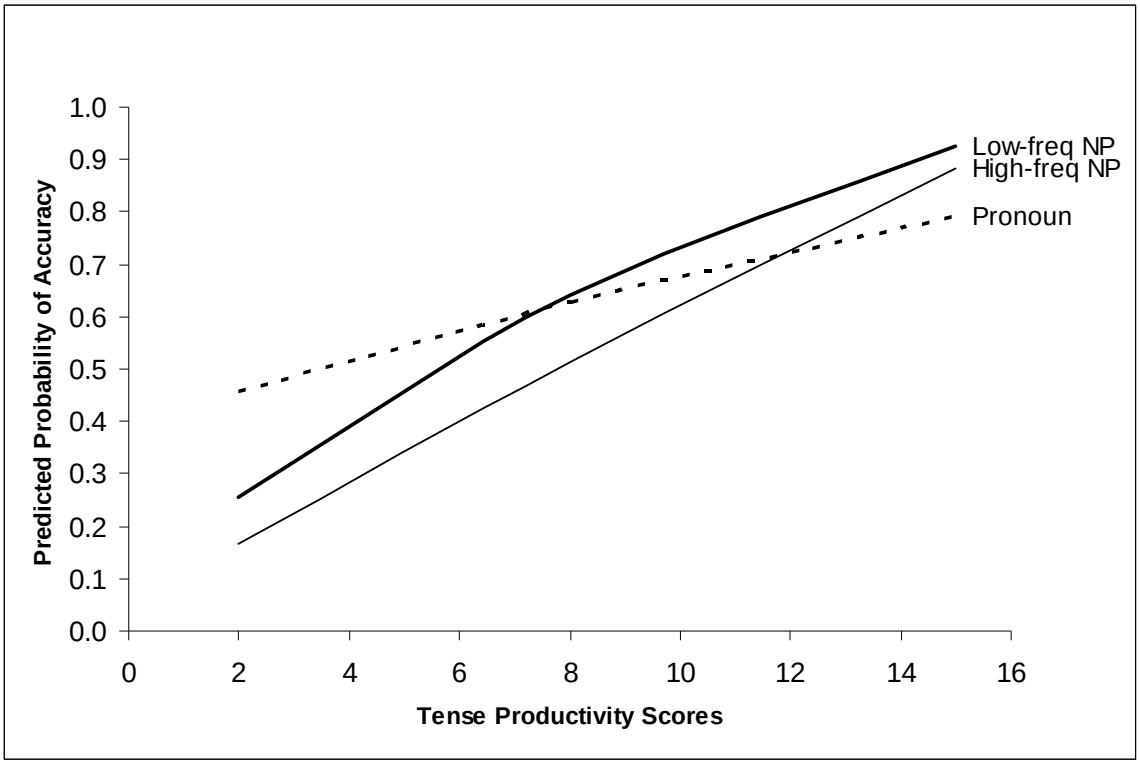
To assess whether there was an interaction effect of the child's tense productivity score and subject types on the production of auxiliary *'is,'* comparable models like those above were completed, with the child's FVMC replaced by tense productivity score. The goodness-of-fit test showed that the model with subject type and tense productivity score as main effects significantly increased model fit as compared to the model with subject

type alone (χ^2 = 12.28, *df* = 1, *p* < 0.001), meaning that the child's tense productivity score was a significant predictor of the production accuracy of auxiliary *'is.'*. The model with in the interaction between subject type and tense productivity, in turn, significantly improved data likelihood as compared to the model with the main effects only; χ^2 = 7.54, *df* = 2, *p* = 0.02. That is, subject types had different effects on the production accuracy of auxiliary *'is,'* depending on the child's tense productivity scores. The results of the model with the interaction between subject type and tense productivity score are reported in Table 2.5. To ease interpretation, a plot of the model in probability space is shown in Figure 2.1.

Table 2.5. Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Subject Type and Tense Productivity Score

	Variance	SD		
Random Factors				
Item	0.06	0.24		
Child	0.64	0.80		
	Coefficient	**SE**	**OR**	**p-value**
Fixed Factors				
Intercept	-2.18	0.66	0.11	< 0.01
Subject Type (reference condition = high-frequency lexical NP subjects)				
Pronominal Subjects	1.78	0.65	5.93	0.01
Low-frequency Lexical NP Subjects	0.57	0.67	1.77	0.39
TnsProd	0.28	0.07	1.32	< 0.01
Subject Type*TnsProd (reference condition = high-frequency lexical NP subjects)				
Pronominal Subjects*TnsProd	-0.16	0.07	0.85	**0.01**
Low-frequency Lexical NP Subjects*TnsProd	-0.01	0.07	0.99	0.94

Figure 2.1. Predicted Probability of Producing Auxiliary *'is'* Accurately by Subject Type and Tense Productivity Score



As can be seen in Figure 2.1, the relationship between tense productivity and rate of use of auxiliary *'is'* is affected by subject type. Pronominal subjects increase in accuracy more slowly as tense productivity increases than lexical NP subjects[3].

To confirm whether the difference between the pronominal and low-frequency lexical NP subject conditions in predicting auxiliary *'is'* production had a similar

[3] The predicted probability in the plot was computed via the β-coefficients in the model. For instance, to compute the predicated production accuracy of auxiliary *'is'* for pronominal subjects in children with a tense productivity score of 2, we first computed its logistic odds as follows: (-2.18) + (1.78)*1 + (0.57)*0 + (-0.16)*1*2 + (-0.01)*0*2 = -0.16. We then take exponential of this value to convert the log odds to odds. Taking the exponential of -0.16 yields 0.84, that is, exp(-0.16) = 0.84. We can convert the odds to a probability by this formula: probability = odds / (1 + odds). In this example, probability is 0.84/(1+0.84) = 0.45.

relationship to that of the pronominal and high-frequency lexical NP conditions (β= -0.16, p = 0.01), we re-ran the model with low-frequency lexical NP subjects as the reference condition. The model showed that the odds of using auxiliary *'is'* correctly increased at a slower rate with pronominal subjects than with the low-frequency lexical NP subjects (β= -0.16, p = 0.02). Likewise, this model reconfirmed that probability of producing auxiliary *'is'* correctly in the high- and low-frequency lexical subject conditions did not differ in children with higher or lower tense productivity scores.

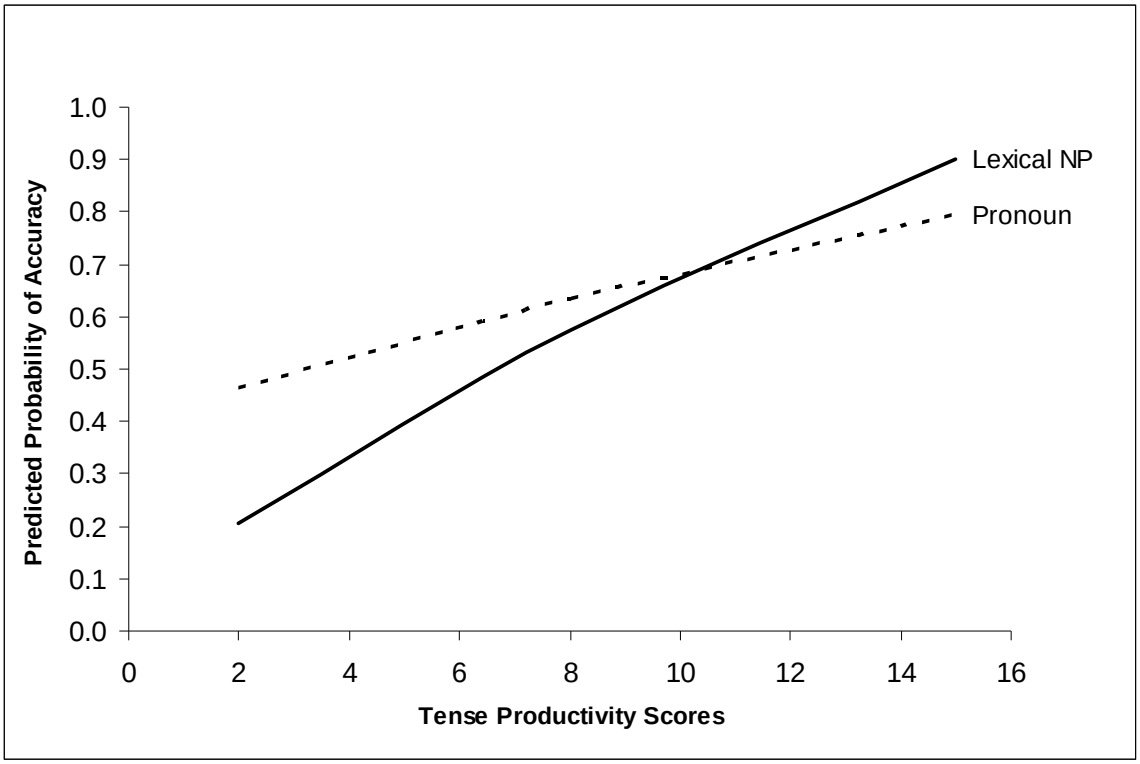
Once again, to directly compare with the previous corpus studies, we collapsed the two lexical NP conditions and repeated the goodness-of-fit tests above. The model with subject type and tense productivity score as main effects significantly improved model fit as compared to the model with subject type alone (χ^2 = 12.28, df = 1, p < 0.001). The model with the interaction between subject type and tense productivity, in turn, significantly improved model fit as compared to the model with main effects only (χ^2 = 7.21, df = 1, p < 0.001). This result indicated that the pronominal and lexical NP subjects had different effects on the production accuracy of auxiliary *'is*,' depending on the child's tense productivity score. Table 2.6 presents the model with the interaction between subject type and tense productivity score and Figure 2.2 plots the predicted accuracy of auxiliary *'is'* in probability space.

Table 2.6. Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Subject Type and Tense Productivity Score

	Variance	SD		
Random Factors				
Item	0.09	0.30		
Child	0.63	0.79		
	Coefficient	**SE**	**OR**	**p-value**
Fixed Factors				
Intercept	-1.89	0.56	0.15	< 0.01
Subject Type (reference condition = lexical NP subjects)				
Pronominal Subjects	1.51	0.56	4.53	0.01
TnsProd	0.27	0.06	1.31	< 0.01
Subject Type*TnsProd (reference condition = lexical NP subjects)				
Pronominal Subjects*TnsProd	-0.16	0.06	0.85	< 0.01

Figure 2.2 shows that children with lower tense productivity were more likely to produce auxiliary *'is'* accurately with pronominal subjects than with lexical NP subjects. This is consistent with Wilson's (2003) and Pine and colleagues' (2008) work, using spontaneous language sample. The mean percent correct of auxiliary BE in Wilson's study was 48% for pronominal subjects and 21% for lexical NP subjects, which was close to the performance of children with a tense productivity score of 2 in the current study. In contrast, children with higher tense productivity did not appear to use auxiliary *'is'* differently with pronominal or lexical NP subjects, but the overall accuracy level of the lexical NP subject condition in the current study (i.e., 0.69) also exceeded those reported in the other corpus studies (e.g., 0.47 in Pine et al., 2008), suggesting that we are indeed looking at children functioning at a higher developmental level.

Figure 2.2 Predicted Probability of Producing Auxiliary *'is'* Accurately by Subject Type (Pronoun and Lexical NP Subjects) and Tense Productivity Score



Discussion

The study tested the unique checking constraint (UCC) hypothesis and the usage-based approach by examining the effect of subject types on the production of auxiliary *'is.'* When we only considered the effect of subject types, we found that the odds of producing auxiliary *'is'* accurately in young children were similar in sentences with pronominal subjects or lexical NP subjects. This finding was inconsistent with the predictions of the usage-based approach. We speculated that this inconsistency could have resulted from ignoring children's developmental levels. To test this hypothesis, we included children's levels of tense development into the model and found that the effect of subject types on the production of auxiliary *'is'* did indeed vary in children with

different levels of tense development. Children with lower tense productivity scores (Hadley & Short, 2005) produced auxiliary *'is'* more accurately with pronominal subjects than with high- or low-frequency lexical NP subjects, but accuracy grew with tense productivity at a slower rate for pronominal subjects than for lexical NP subjects . This suggests that, as children's representations become more abstract, the influence of lexical frequency and frozen forms on accuracy changes. However, such an observation is only possible when an appropriately sensitive measure of abstraction/developmental level is used. When we documented children's tense development with the finite verb morphology composite (Leonard et al., 1999), we were unable to find any effect of tense development on young children's production of auxiliary *'is'* in sentences with different subject types.

Limitations of the Current Study

Before we discuss the results in detail, we should consider two limitations of this study. First, we excluded a large number of potential participants from the final data analysis. Fifty-eight out of 78 participants were excluded from analyses for a variety of reasons. At first glance, this could suggest that the experimental task may be too challenging for three-year-olds. However, recall that 24 children—or approximately half of the excluded participants—were excluded because of ceiling level performance, suggesting that the task was age-appropriate. A second concern is that the children changed the lexical item used in the subject for approximately 21% (124/600) of the items. Due to these changes, it is possible that we may not have had enough items to reliably measure some children's performance for any given category, particularly if the majority of these changes occurred in one condition. We attempted to address this problem in three ways: 1) by excluding children from analysis who produced fewer than 3 items in any of the conditions, 2) by reclassifying the lexical items the children actually

produced into an appropriate category to maximize retention of data, and 3) by our choice of statistical approaches.

Effect of Subject Types on the Production of Auxiliary *'is'* Varies in Children with Different Developmental Levels

This study confirmed that subject types affect the production accuracy of auxiliary *'is'* in young children, but these effects varied across different developmental levels as documented by children's tense productivity scores. Children with lower tense productivity used auxiliary *'is'* more accurately with pronominal subjects than with either high- or low-frequency lexical NP subjects or with lexical NP subjects in general. The finding that pronominal subjects facilitated the production of auxiliary *'is'* more than lexical NP subjects was expected given results from previous corpus studies (e.g., Joseph et al., 2002; Wilson, 2003; Pine et al., 2008). Recall that in these studies, children tended to produce tense and agreement morphemes (e.g., auxiliary and copula BE) more accurately with pronominal subjects than with lexical NP subjects.

In contrast, children with higher tense productivity did not use auxiliary *'is'* differently with pronominal subjects or lexical NP subjects. It is worth considering why we might have obtained different findings than those described above (Joseph et al., 2002; Wilson, 2003; Pine et al., 2008). First, these children were older than those who participated in the corpus studies and may have developed a more abstract knowledge of auxiliary *'is'* than the younger participants. Abstract knowledge of '*is'* would allow these children to produce the auxiliary without being constrained by lexically-specific constructions. In addition to differences in age, the current study adopted an elicited production task while prior work relied primarily on longitudinal spontaneous language samples. Elicited production tasks may encourage the children to use abstract constructions that they have already acquired but do not yet use fluently in spontaneous speech. One advantage of the elicited production paradigm is that it allowed us to

examine children's performance at a given developmental time point. The earlier corpus studies examined children's production by collapsing language samples collected over a more extended period of time. Collapsing across time was necessary in order to obtain a sufficient number of utterances for evaluation in those studies. It was possible that some children in the corpus studies may have produced auxiliary BE similarly with pronominal subjects and with lexical NP subjects at some points of the later developmental stage but, by collapsing language samples over a period of several months, this pattern was not readily observable.

Testing the UCC Hypothesis

With regard to the two language learning hypotheses we were testing, the results are mixed. The asymmetries observed between lexical and pronominal subject types does not support the predictions of the UCC hypothesis, which states that it is the failure to check the features of Agreement or Tense that leads to the omission of tense and agreement morphemes (Wexler, 1998). Because both pronominal and lexical subjects have similar structural properties (i.e., both are determiner phrases), they should be affected equally by the UCC. The odds of producing auxiliary *'is'* accurately should be similar across these two subject types, which was not confirmed by the current study.

One could argue that the different production rates of auxiliary *'is'* in sentences with pronominal subjects or lexical NP subjects may result from the differences in sentence length. Indeed, it was possible that the sentences with pronominal subjects were shorter than those with lexical NP subjects so that children were more likely to produce auxiliary *'is'* in sentences with pronominal subjects. The sentence length analysis, however, ruled out this possibility and showed that the three sets of sentences were well-matched with regard to length. Furthermore, sentence length was actually not a significant factor alone or in combination with other predictors in predicting the likelihood of producing auxiliary *'is.'*

Testing the Usage-based Approach in Children
with Lower Tense Productivity

Turning now to the usage based approach, we see that the finding that auxiliary *'is'* was used more accurately with pronominal subjects than with lexical NP subjects as observed in the children with lower overall performance is consistent with their predictions. This stronger facilitation effect could be because the pronominal constructions are strengthened by the frequency with which they co-occur with the auxiliary form (Wilson, 2003; Lieven, 2008). However, this also raises the question of why auxiliary *'is'* production was variable even if we restrict our examination to sentences with pronominal subjects. One would expect that the pronoun + auxiliary *'is'* constructions would be highly entrenched due to their frequency. One explanation is that even though specific pronoun + auxiliary *'is'* constructions may have stronger representations as compared to the specific lexical NP + auxiliary *'is'* constructions, the representations for both subject types must be gradually built up from the input. This means that until the child's representations of these frequent constructions reach a robust or adult-like level, the retrieval of the specific pronoun + auxiliary *'is'* may still fail sometimes, leading to the inconsistent productions that we observed.

What about the lexical NP subject types? If input frequency does play a role, we would expect to see differences in the production rates of auxiliary *'is'* across sentences with high- and low-frequency lexical NP subjects in these same children, but this was not the case. One explanation is that although specific high-frequency lexical NP + auxiliary *'is'* constructions may be relatively more frequent than specific low-frequency lexical NP + auxiliary *'is'* constructions, the instantiations of both types of constructions are not sufficiently frequent in the input to have an effect on children's productions when they are at the early stages of abstraction. The difference in the strength of representations of these constructions may have not yet reached a level that would influence their productions. In addition, to select age-appropriate nouns for the subjects in target

sentences, we picked the nouns primarily based on the word list in *MB-CDI: WS* (Fenson et al. 1993). The nouns in *MB-CDI: WS* are actually frequent words in the input. Though low-frequency lexical NP subjects are relatively low in frequency, they may be still commonly heard in the input, which makes it even more difficult to observe the frequency effect between the high- and low-frequency lexical NP subject conditions. However, this explanation is only tentative and future work with very young children and/or children with low levels of auxiliary production are needed to verify these results.

Testing the Usage-based Approach in Children
with Higher Tense Productivity

If we look at children at the higher end of tense productivity, it seems that these children produced auxiliary *'is'* more accurately with lexical NP subjects than with pronominal subjects though the current statistical method did not allow us to evaluate whether this difference reached a significant level. A finding that lexical NP subjects and pronominal subjects lead to similar accuracy levels is not, in fact, incompatible with the usage-based approach. Within this approach, children will eventually develop an adult-like abstract representations of auxiliary BE (e.g., *NP + BE + Verb-ing* or $NP_{subject\text{-}3sg}$*'s + Verb-ing*). It is possible that children with higher tense productivity are approaching adult-like abstract representations of auxiliary BE and thus differences in accuracy levels are gradually diminished.

Even if the difference were to be significant, this result would not be a complete surprise. In a study by Theakston and colleagues (2005), children tended to produce auxiliary BE less accurately in sentences with the pronominal subjects *I* and *You* than in sentences with other subjects (e.g., *he, it*, lexical NPs, proper names) though the constructions *I'm* and *You're* were more frequent than other specific subject + auxiliary BE constructions in the input.

One explanation of this finding is that increased accuracy with low-frequency lexical NPs may be attributed to the competition of *ungrammatical* constructions, such as *He Verb-ing* and *It Verb-ing* (Theakston et al., 2005; Lieven, 2008). Although these higher functioning children may have developed abstract constructions like $NP_{subject\text{-}3sg}$*'s + Verb-ing* to express the meaning of present progressive but lexically-specific constructions that were learned earlier might still remain in the representation (Bybee, 1995; 2002; Dąbrowska, 2000; Dąbrowska & Lieven, 2005; Lieven, 2008). Such lexically-specific constructions may compete with the abstract construction when the child attempts to produce a sentence with the meaning of present progressive. Recall that lexically-specific constructions can be grammatical (e.g., *He's Verb-ing, It's Verb-ing*) or ungrammatical (e.g., (is) *he Verb-ing, (I saw) it Verb-ing*). The ungrammatical *pronoun + Verb-ing* sequences could have strong representations because children at this developmental level may have accumulated a number of instantiations of these constructions. Because of the strength of the underlying representation, these constructions may sometimes triumph, leading to the omission of auxiliary *'is'* (Lieven, 2008; Theakston & Lieven, 2008). In other words, sentences with pronominal subjects are more vulnerable to the omission of auxiliary *'is'* than those with lexical NP subjects as children's representations become more abstract because the former are more likely to encounter competition from the ungrammatical lexically-specific constructions than the later (Lieven, 2008).

If the ungrammatical lexically-specific constructions remain in the representation, how do they stop being strong competitors to the grammatical constructions, given that the omission of auxiliary *'is'* becomes rare in typically-developing children even a few months later? Perhaps as children's capacity of sentence processing expands, they become able to process utterances completely (instead of partial processing) and figure out that these ungrammatical constructions are actually part of longer constructions (e.g., *Auxiliary BE +* ***pronoun + Verb-ing***) (Newport, 1990; Freundenthal et al, 2007).

As long as these ungrammatical constructions are contained within longer constructions, children's use of these constructions to express the meaning of present progressive will gradually decrease over time. Thus, the ungrammatical truncated constructions will cease to be strong competitors for the grammatical constructions due to the declining rate at which children store truncated sentences.

In general, this study supported the predictions of the usage-based approach that the frequency of subject types play a role in auxiliary *'is'* production. However, the effects of subject types are not identical in children at different developmental levels. Without considering children's developmental levels, the effect of subject type would have been obscured. It should also be noted that, although we only reported the results from children who were variable, we tested 78 children with an age range from 2;7 to 3;5. While some children did not even produce one correct instance of auxiliary '*is*,' 24 had reached a ceiling level of performance. This variation among individuals suggests that even children within a narrow age range can have substantially different developmental levels of auxiliary *'is.'* Taken together, these findings stress the importance of measuring children's developmental level when investigating the effects of frequency on use of grammatical morphemes.

Tense Productivity Scores are More Sensitive to Children's Use of Auxiliary 'is' than Finite Verb Morphology Composites

Up to this point, we have been discussing our results in the context of using tense productivity scores in the model to examine the effect to subject types. In this study, we also considered the finite verb morphology composite (FVMC, Leonard et al., 1999) in the model. However, including FVMC in the model did not improve model fit. That is, FVMC was not a significant predictor of the child's use of auxiliary *'is'* in the current study.

One reason FVMC was not a significant predicator of the child's use of auxiliary *'is'* is because this index might overestimate the child's tense development. FVMC is computed using the percent correct of tense and agreement morphemes in *all* obligatory contexts, including forms of BE contracted to the pronominal subjects (e.g., *he's, it's*). According to the usage-based approach and work by Hadley and colleagues, these pronoun + contracted BE combinations could be rote, unanalyzed chunks and may not be generated through syntactic processing (Tomasello, 2003; Hadley & Short, 2005; Rispoli, Hadley, & Holt, 2009). Including these combinations in the computation potentially overestimates the child's knowledge of tense and agreement morphemes. For instance, the child HP only produced one correct item each for the pronominal subject and low-frequency lexical NP subject conditions in the task but had a FVMC of 87%. Because of this discrepancy, we inspected her language sample. This child used constructions like *That's + NP* and *What's that?* frequently in the language sample but did not produce a variety of tense and agreement morphemes in other contexts. It is possible that HP had not yet developed an abstract or productive knowledge of tense and agreement morphemes. Unfortunately this lack of abstract knowledge was not reflected in her FVMC. The limited ability of FVMC to account for children's use of auxiliary 'is' in sentences with different subject types in the current study may result from such an overestimation.

Clinical and Experimental Implications

The current findings have clinical and experimental implications. First, this study stresses the importance of incorporating knowledge of a child's developmental level into the experimental hypotheses. In this study, we observed that the facilitative effect of pronominal subjects changed in children with different levels of tense productivity. The use of precise measures for this purpose are not only of importance when researchers investigate the effect of frequency on the production of tense and agreement morphemes,

but also when they undertake other studies of language learning and development. A finding that developmental level is important is not uncommon when a developmental process approach is taken (Spencer, Blumberg, McMurray, Robinson, Samuelson, & Tomblin, in press) and has been observed in areas of language other than grammar. For instance, children aged between 17 to 30 months tended to generalize names by shape more often with solid stimuli than with non-rigid or non-solid stimuli, but this effect was limited to children with advanced noun vocabulary development (Samuelson & Smith, 1999). Similarly, 14-month-olds who had larger vocabulary sizes were better able to learn phonetically similar words (e.g, *bih* and *dih*) than those who had smaller vocabulary sizes (Werker, Fennell, Corcoran, & Stager, 2002). Clinically this would suggest that a speech language pathologist should clearly understand her client's level of functioning prior to selecting a particular intervention approach or the linguistic materials to be used. Attention to such details (at a level not previously anticipated) may enhance intervention efficacy.

The results of this study may also generate hypotheses for intervention techniques that could be used with children with specific language impairment (SLI), a population that has documented problems in using tense and agreement morphemes. For instance, if a clinician wished to employ errorless learning as a therapy technique (Fillingham, Hodgson, Sage, & Ralph, 2003), she may initially choose highly frequent subjects (e.g., pronouns) to maximize the production of auxiliary *'is'* and other related tense and agreement morphemes for children at lower levels of production of these morphemes. From a more theoretical perspective, if children with higher tense productivity in this study did use auxiliary *'is'* more accurately with lexical NP subjects than with pronominal subjects and if the same patterns of shifting accuracy from low to high tense productivity are also observed in children with SLI, one could argue that Leonard's (2007) hypotheses about the underlying mechanisms of SLI might be accurate. Specifically, because of processing capacity limitations (Leonard, 1998), children with SLI may

continue to process sentences partially while typically-developing peers are able to process sentences completely. The ungrammatical constructions (e.g., *(Is)* *he running*) learned from partial processing may remain in the representation of children with SLI to compete with the grammatical constructions for a protracted period of time as compared to typically-developing children, leading to lower-than-typical production rates of tense and agreement morphemes in children with SLI (Freudenthal et al., 2007). That is, the difficulty using tense and agreement morphemes in children with SLI may actually result from the interaction between linguistic input and processing capacity limitations.

Conclusion

This study tested the unique checking constraint (UCC) hypothesis and the usage-based approach by examining the effect of subject types on the production of auxiliary *'is'* in three-year-olds via an elicited production task. The results showed that subject types affected the production of auxiliary *'is'* but the effect varied in children at different developmental levels. These findings may generally be accounted for by the notion of input frequency, which supports the assumption of the usage-based approach that children learn tense and agreement morphemes from the input in a gradual piecemeal fashion. In contrast, the findings did not conform to the predictions of the UCC hypothesis. Finally, we are not claiming that frequency is the only factor that accounts for the variable production of auxiliary *'is'* and other tense and agreement morphemes in young children. Rather, it is an important factor to consider when we investigate why young children tend to omit tense and agreement morphemes.

CHAPTER III

EXPERIMENT 2: ELICITED PRODUCTION OF COPULA *'IS'* IN YOUNG CHILDREN

Overview

How children come to use tense and agreement morphemes accurately is a central issue in language acquisition research (Brown, 1973; Kuczaj, 1981; Lahey, Liebergott, Chasnick, Menyuk, & Adams, 1992; Schütze & Wexler, 1996; Wilson, 2003; Legate & Yang, 2007). Tense and agreement morphemes include both function words (e.g., forms of BE, such as *The dog is big*) and inflections (e.g., third person singular –s, such as *The dog runs*) that mark for person, number, and time. Although young children use tense and agreement morphemes as soon as they produce two-word utterances, they do not achieve adult-like production for several years. Most commonly, young children omit these morphemes (e.g., **He happy*; asterisks refer to ungrammatical sentences), but when they do use these morphemes, they tend to use them correctly (e.g., *He is happy*) and errors of commission (e.g., **He are happy*) are rarely observed (Wexler, 1994; Schütze & Wexler, 1996). One question that must be addressed by any acquisition theory is what causes the variable use of tense and agreement morphemes in young children and how children move from variable to stable, accurate productions (e.g., Radford, 1990; Wexler, 1998; Yang, 2002; Tomasello, 2003).

Current theories that address this issue include generative accounts (e.g., Radford, 1990; Wexler, 1998) and constructivist accounts (e.g., Tomasello, 2003). One generative account, the unique checking constraint (UCC) hypothesis, posits that children have very early knowledge of tense and agreement (Wexler, 1998). Inconsistent production occurs because of the presence of a maturational constraint—the unique checking constraint—in their grammar. In contrast, the usage-based approach, a constructivist account, holds that children are not born with knowledge of tense and agreement. Rather, they learn tense

and agreement morphemes from the input in a gradual, piecemeal fashion. Young children use these morphemes variably because they have not yet developed abstract representations and instead use lexically-specific constructions that may or may not contain these morphemes.

This study tested the UCC hypothesis and the usage-based approach by examining the use of copula '*is*' in typically-developing three-year-olds via an elicited production task. These two theories make different predictions about how subject types (e.g., *He/The dog is happy*) and predicate types (e.g., *He is a king/happy/at home*) influence the production of copula BE. In the following, we briefly review the basic assumptions and predictions of, and the existing empirical evidence for, these two theories as they relate to production of copula BE.

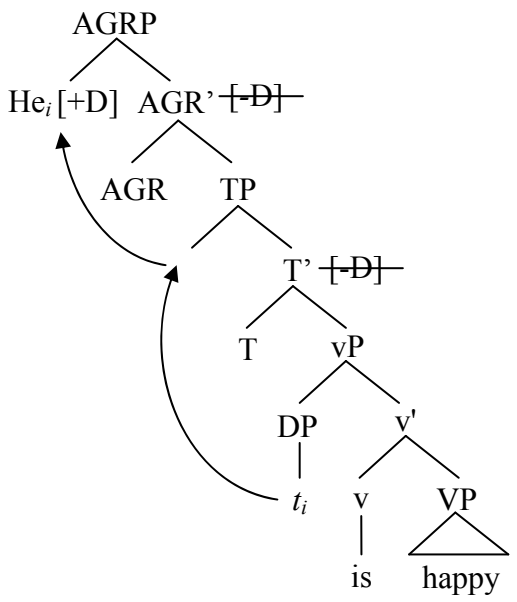
The Unique Checking Constraint Hypothesis

Grounded within the framework of minimalist syntax (see Radford, 2004), the unique checking constraint hypothesis posits that children are born with innate linguistic knowledge (Wexler, 1998, 2003). Young children's knowledge of tense and agreement is basically adult-like as early as the onset of multiword utterances, which occurs around 18 months of age in typically developing children. This early knowledge is evidenced in the observation that young children rarely use tense and agreement morphemes in unacceptable contexts, such as **That are a dog* and **Her is sad* (Wexler, 1996).

Following the minimalist program, the UCC hypothesis assumes that the syntactic representation of a sentence with tense marking in English has separate functional categories for Agreement (AGR) and Tense (TNS), as (1) shows. Both AGR and TNS have an uninterpretable D-feature (i.e., a morphosyntactic feature related to determiner phrases). To produce a grammatical sentence, these uninterpretable features must be eliminated / checked off by their interpretable counterparts (Avram, 2002; Adger, 2003). Thus, the subject determiner phrase (i.e., DP), which has the interpretable D-feature, must

raise to check off the uninterpretable D-features **both** on AGR and TNS. That is, the subject DP needs to check **twice** to derive a grammatical finite sentence.

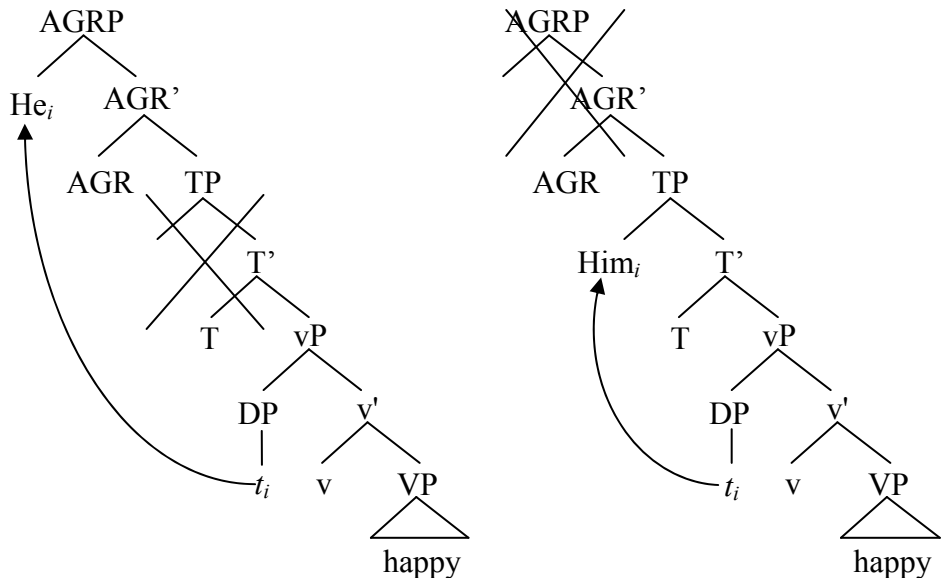
(1) A simplified syntactic representation: *He is happy*; where AGRP = agreement phrase, TP = tense (TNS) phrase, VP = verb phrase, t = trace; [+D] = interpretable D-feature; [-D] = uninterpretable D-feature



However, the UCC hypothesis claims that initially children can only check **once** due to the presence of the maturational constraint. This limits children's ability to check off both of the D-features on AGR and TNS. Thus either AGR or TNS, whichever has an unchecked D-feature, will be dropped from the representation. When only AGR or TNS is preserved on the representation, generation of tense and agreement morphemes will be blocked. If AGR is preserved and TNS is dropped, a sentence like **He happy* may be produced, as (2a) shows. In contrast, if TNS is kept and AGR is omitted, a sentence like **Him happy* may be produced, as (2b) shows. It should be noted that the form of the subject DP in (2a) is in the correct nominative case (i.e., *He*) but that in (2b) is not (i.e., *Him*) because the nominative case is assigned by AGR. Because the use of tense and

agreement morphemes implies the presence of both AGR and TNS in the grammar, these morphemes should only occur with nominative subjects.

(2) a. AGR preserved, TNS omitted b. AGR omitted, TNS preserved



To demonstrate that children have early knowledge of tense and agreement, Schütze and Wexler (1996) examined the production of tense and agreement morphemes in relation to the production of the subject form in sentences of three young children between the ages of 1;11 and 3;1. When these children used nominative subjects (e.g., *she*), both accurate use and omission of tense and agreement morphemes were observed. However, when children used non-nominative subjects, they used tense and agreement morphemes in approximately 5% of these sentences (e.g., **Her is quiet*). Similar patterns of use of tense and agreement morphemes were also observed in an experimental study, which included children with and without language impairment (Wexler, Schütze, & Rice, 1998). Because omission of tense and agreement morphemes may occur with nominative or non-nominative subjects, Schütze and colleagues (Schütze & Wexler, 1996; Wexler et al., 1998) argued that AGR and TNS are represented separately and can be independently

dropped. In addition, the scarcity of the co-occurrence of non-nominative subjects with tense and agreement morphemes reveals that young children know that the presence of AGR requires the nominative case to be assigned to the subject, implying that they have early knowledge of tense and agreement.

Given that the UCC causes AGR or TNS to be dropped from the representation, we are left asking why young children still produce tense and agreement morphemes correctly in some sentences. Wexler (1998) adopts the notion of minimal violation to explain this issue. During the process of sentence generation, a number of candidate sentences are derived. The computational system of syntax prefers the candidate sentence that violates as few grammatical properties as possible. If more than one candidate sentence minimally violates the properties, any one of them may be chosen as the output sentence. Dropping AGR or TNS under the restriction of UCC involves one violation of a grammatical property. The preservation of both AGR and TNS on the representation also violates a grammatical property, namely the UCC, due to the fact that the D-features of AGR and TNS are both checked. Because the candidate sentence that preserves both AGR and TNS (e.g., *He is happy*) and the candidate sentences that drop either of these two functional categories (e.g., **He happy* and **Him happy*) involve one violation each, any one of them may be chosen as the output sentence.

Over time the UCC relaxes so that the child's use of tense and agreement morphemes will gradually approach adult levels of performance. As long as the UCC is still present in the child's grammar, it will limit the child's ability to check AGR or TNS, resulting in the omission of tense and agreement morphemes.

The presence and relaxation of the UCC can be observed via the growth trajectory of tense and agreement morphemes. Rice, Wexler, and Hershberger (1998) investigated the growth trajectory of the use of five tense and agreement morphemes in children with SLI (age range: 4;5-8;9) and their typically-developing peers (age range: 2;6-6;8) who were matched on mean length of utterance (MLU). Language sample collection and

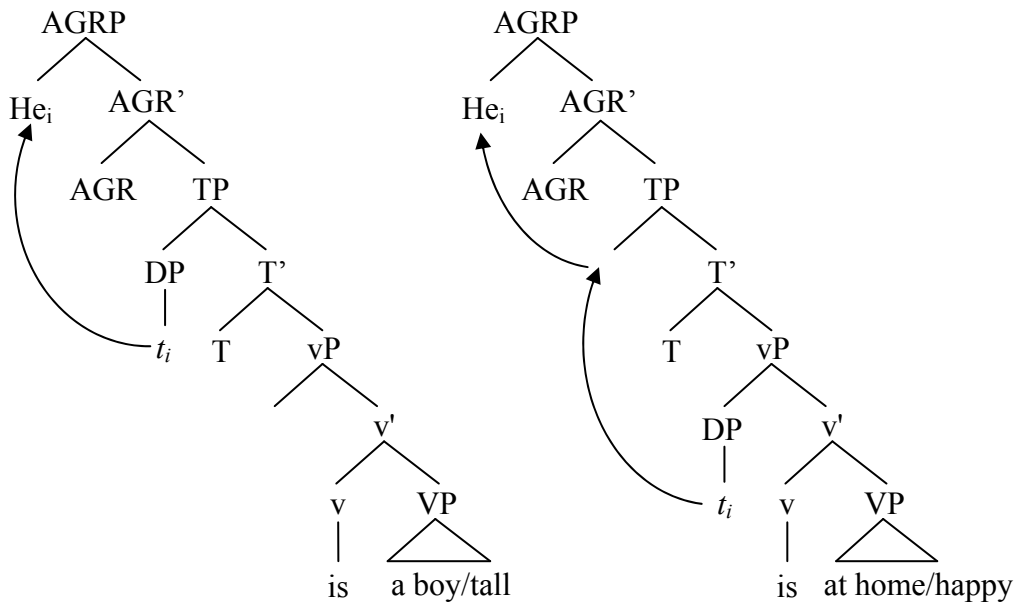
experimental probes were conducted for each child at 6-month intervals over 3 years. For typically-developing children, the curve of each individual morpheme exhibited flat growth at the onset, accelerated growth in the middle, and an asymptote at ceiling level by the end of the study. Rice and colleagues (1998) suggested that the similarity of growth trajectories across morphemes supports the assumption that tense and agreement morphemes are unified under a common grammatical function (i.e., AGR and TNS). The flat growth trajectory at the onset signaled the presence of the UCC in the grammar whereas an acceleration phase indicated the withering of the constraint.

Several predictions of the UCC hypothesis fall out from the grammatical assumptions associated with the minimalist syntax. For instance, pronominal (e.g., *it*) and lexical noun phrase (lexical NP, e.g., *the cat*) subjects are both determiner phrases and have similar structural properties within the representation (Adger, 2003). One prediction that can be derived from this proposition is that the production rate of a given form of copula BE (e.g., *is*) should be similar in sentences with different subject types (i.e., pronominal and lexical NP subjects) because the UCC should function similarly for elements with similar structural properties.

In contrast to subject types, Wexler (2000) argued that the production rate of a given form of copula BE (e.g., *is*) should vary in sentences with different predicate types because of the difference in syntactic representations. Copula BE may occur with individual-level or stage-level predicates. Individual-level predicates refer to those which express a temporally unbounded property, such as nominal predicates (e.g., *He is a boy*) and permanent-type adjectives (e.g., *He is tall*), whereas stage-level predicates are those which express a temporally bounded property, such as locative predicates (e.g., *He is at home*) and temporary-type adjectives (e.g., *He is hungry*). Wexler argued that subject DPs are generated outside of the vP shell (i.e., on the specifier position of AGRP or TP) in individual-level predicates and within the vP shell in stage-level predicates, as (3) illustrates. Those subjects that are generated within the vP shell must check twice in the

derivation and risk the copula being omitted under the UCC. In contrast, those subjects that are generated outside of the vP shell must check only once, which makes the copula more likely to be retained and produced correctly.

(3) a. Individual-level predicates b. Stage-level predicates



Two predictions about copula *'is'* production in sentences with different predicate types can be derived from the UCC hypothesis. First, the production rate of copula *'is'* form should vary across sentences with different predicate types. Specifically, children should produce copula *'is'* more often with nominal or permanent-adjectival predicates (both are individual-level predicates) than with locative or temporary-adjectival predicates (both are stage-level predicates). Second, the production rate of copula *'is'* should be similar in sentences with different subtypes of the same predicate type for a given young child. For example, both locative predicates (e.g., *he is <u>at home</u>*) and temporary-type adjectives (e.g., *he is <u>happy</u>*) are stage-level predicates. The production rate of copula *'is'* in sentences with these sub-types of stage-level predicates should be similar because they are similar in syntactic representation.

In a series of studies, Becker (2000a, 2000b, 2004) investigated the effect of subject types and predicate types on the use of copula BE in longitudinal corpora of four children collected between the ages of 2;0 and 3;4. She found that although two children consistently produced copula BE more accurately with pronominal subjects than with lexical NP subjects, the other two children did not. In contrast, these four children consistently used copula BE more accurately with individual-level predicates than with stage-level predicates. Children were more likely to produce copula BE correctly with nominal predicates (76.3%) than with locative predicates (18.8%). Similarly, children tended to use copula BE more accurately with permanent-type adjectival predicates (62.2%) than with temporary-type adjectival predicates (47.0%). Based on these findings, Becker suggested that predicate types, rather than subject types, were the crucial factor affecting young children's production of copula BE.

In summary, the corpus and experimental studies reviewed above argued that young children have an adult-like representation of tense and agreement in their grammar and that young children are inconsistent in producing tense and agreement morphemes because of the presence of the UCC in their grammar. Production accuracy of copula *'is'* should be similar with pronominal and lexical NP subjects because the UCC should affect these subject types equally. In contrast, production accuracy of copula *'is'* should be higher with individual-level predicates than with stage-level predicates because of the difference of the position of subjects within the representations. Furthermore, production accuracy of copula *'is'* should be similar in sentences with different subtypes of the same predicate types.

Although these studies appear to account for the variable use of tense and agreement morphemes demonstrated by young children, other researchers have argued that these findings can actually be interpreted without positing that young children have an intact, abstract representation of finiteness. In the next section, we review studies from the usage-based approach that argue against the notion of intact, abstract knowledge

of finiteness in young children via the variable production of tense and agreement morphemes with different subject types (e.g., Wilson, 2003; Pine et al., 2008).

The Usage-based Approach

The usage-based approach (Tomasello, 2003) holds that language consists of constructions. Constructions are conventionalized symbolic units with form-meaning pairings, which may include morphemes, words, phrases, idioms, sentence frames, and so on (Goldberg, 1995). Instead of assuming innate abstract linguistic knowledge, the usage-based approach posits that language acquisition involves learning linguistic constructions of different sizes and complexity from the input (Lieven, Pine, & Baldwin, 1997; Tomasello, 2000; Dąbrowska & Lieven, 2005; Bannard & Matthews, 2008). At the beginning, the constructions that children learn may be unanalyzed chunks (e.g., *What's that?*) or constructions with fixed lexical items and open slots (i.e., lexically-specific constructions, such as *That's a* ____). Children gradually develop abstract constructions from chunks or lexically-specific constructions.

This approach hypothesizes that young children inconsistently use tense and agreement morphemes because they have not yet developed sufficiently adult-like abstract constructions and instead rely primarily on lexically-specific constructions that may or may not contain these morphemes (Wilson, 2003; Theakston, Lieven, Pine, & Rowland, 2005; Pine et al., 2008). For these children, the omission of these morphemes in sentences may result from two possible mechanisms. First, the child may have limited representations of the particular constructions required to support the use of tense and agreement morphemes in a specific discourse context (Ambridge, Rowland, Theakston, & Tomasello, 2006; Theakston & Lieven, 2008). For instance, the child may have developed *It's Adjective Phrase* (AdjP) and *He's AdjP* constructions but not *The frog's AdjP* in the representation. However, the discourse context may require the child to produce a sentence such as *The frog's happy*. Because the child does not have the

particular *The frog's AdjP* construction and has not yet developed the abstract construction $NP_{subject\text{-}3sg}$*'s + AdjP* (where 3sg refers to third person singular) to support his production, he may just combine the words based on semantics and produce a sentence like **The frog happy*, resulting in the omission of copula *'is.'*

Second, the child may use ungrammatical constructions without tense and agreement morphemes because he has observed these constructions in the input (Theakston et al., 2005; Lieven, 2008; Theakston & Lieven, 2008). For instance, the child may have learned the ungrammatical constructions like **He AdjP*, **Him AdjP*, and **The dog AP* due to partial processing of input utterances such as *Is he happy*, *You make him mad*, and *Why is the dog sad* (Slobin, 1985; Newport, 1990, Elman, 1993; Freudenthal, Pine, Aguado-Orea, & Gobet, 2007). Depending on their representational strength, these ungrammatical constructions may compete with and triumph over a comparable construction that contains copula BE, yielding a sentence like **He happy*.

Later in development, the child may figure out from a variety of lexically-specific constructions that there are more abstract ways to represent these constructions (e.g., $NP_{subject\text{-}3sg}$*'s + AdjP*) and may use copula '*is'* flexibly with any subject type. Eventually, the child develops the most abstract form (e.g., $NP_{subject}$ + *Copula BE + AdjP*) and is able to produce forms of copula BE flexibly and accurately. The inconsistent production of copula BE will decrease gradually as the child develops sufficiently adult-like abstract constructions that contain copula BE. It should be noted that even if the individual develops adult-like abstract constructions, the frequently-used lexically-specific constructions may remain in the representation and continue to be used for language production even into adulthood (Bybee, 1995, 2002; Dąbrowska, 2000; Dąbrowska & Lieven, 2005; Theakston & Lieven, 2008). As long as the child has not developed adult-like abstract constructions, he will inconsistently produce copula BE forms. The production of copula BE will reflect the constructions that have been learned and their associated strength (Theakston & Lieven, 2008).

The representational strength of constructions partly depends on their frequency of use in the input (Bybee, 1995, 2002; Theakston et al., 2005; Ambridge et al., 2006). In general, high-frequency constructions tend to have stronger representations than low-frequency constructions. For instance, specific pronoun + copula BE combinations (e.g., *He's*) are more frequent in the input and thus have stronger representations than specific lexical NP + copula BE combinations (e.g., *The deer's*). Because more frequent constructions will be more entrenched and easier to retrieve than less frequent constructions, children are more likely to use a particular copula BE form correctly with pronominal subjects than with lexical NP subjects (Lieven, 2008).

Empirical evidence has shown that there is a link between the production of tense and agreement morphemes and the input that children receive. Joseph, Serratrice, and Conti-Ramsden (2002) explored the use of copula and auxiliary BE in children with SLI (aged from 3;1 to 4;8) and their peers who had matched on MLU in words (aged from 1;8 to 2;4) via longitudinal corpora. Overall, the production of copula and auxiliary BE in these children can be accounted for by the distribution of use of BE in the maternal input. The SLI and the MLU-matched groups both produced copula BE more accurately than auxiliary BE, and used the form '*is'* more correctly than the forms '*are'* or '*am.'* These children also produced a given form of BE (e.g., copula *is*) more often with personal-pronoun subjects (e.g., *he*) than with other pronominal subjects (e.g., *this*), followed by lexical NP subjects. Joseph and colleagues (2002) therefore suggested that BE and other tense and agreement morphemes were learned independently from the input in a piecemeal fashion, but not governed by a unified underlying representation. Constructions with the form *'is'* or personal-pronoun subjects were produced with higher accuracy than other with other forms of BE or other subjects because the former were heard more frequently and thus learned earlier.

Additional evidence demonstrates frequency effects on the production rate of tense and agreement morphemes. Wilson (2003) examined the effect of subject types on

the production of copula and auxiliary BE in longitudinal transcripts of five children collected between the ages of 1;6 and 3;5. For a given form of BE, the production accuracy was higher with pronominal subjects than with lexical NP subjects within each child. This finding was replicated in a study by Pine and colleagues (2008), in which sentences with non-nominative subjects were included in analysis and children's lexical knowledge of BE was controlled. Similarly, Theakston and colleagues (2005) investigated the use of auxiliaries BE and HAVE in longitudinal samples from 11 typically-developing children between the ages of 1;8 and 2;0 over a year. Children were more likely to produce BE and HAVE correctly with pronominal subjects *He* or *It* than with lexical NP subjects. In addition, the age of acquisition of specific subject + auxiliary constructions tended to correlate with the input frequency. These corpus studies suggested that children used BE and HAVE more accurately with pronominal subjects than with lexical NP subjects because BE and HAVE co-occurred with pronominal subjects more often in the input and were more entrenched in the representation.

However, this advantage of pronominal subjects may not always hold as the child's representation changes. Indeed, our previous work (Guo, Owen, & Tomblin, submitted) tested the effect of pronominal and lexical NP subjects on the use of auxiliary *'is'* in children between the ages of 2;7 and 3;5 and found that effect of subject types changed with children's developmental level of tense, as measured by tense productivity scores (Hadley & Short, 2005). Children with lower developmental levels of tense produced auxiliary *'is'* more often with pronominal subjects than with lexical NP subjects. In contrast, children with higher developmental levels produced auxiliary *'is'* similarly with pronominal and lexical NP subjects. We attributed the disappearance of an advantage for pronominal subjects in children with higher tense productivity to the development of adult-like abstract representations of auxiliary BE. This finding suggests that the developmental level and the dynamic nature of representations in children need to be incorporated into analysis when researchers test acquisition theories.

Although the usage-based approach is most explicit with regard to the role of subject types, frequency of predicate types may also influence the use of copula BE. Recall that copula BE can occur with nominal, permanent-adjectival, locative, or temporal-adjectival predicates. Due to the item-based nature of early syntactic development (Tomasello, 2000), children may learn copula BE independently for constructions that contain different predicate types. Because constructions with one predicate type may be more frequent than those with the other type, the representational strength of constructions with different predicate types might also vary.

Wilson (2003) also investigated the effect of predicate types and their interaction with subject types on the production of copula BE. He found that, when subject types were not controlled, 3 out of 5 children produced copula BE with nominal predicates (e.g., *He's* <u>*a boy*</u>) more accurately than with adjectival predicates (e.g., *He's* <u>*tall/happy*</u>), followed by locative predicates (e.g., *He's* <u>*at home*</u>). This pattern was consistent with Becker's predictions described earlier. However, this pattern did not always hold when the sentences were broken down according to subject types. For sentences with lexical NP subjects, this pattern of copula BE use was observed in 4 out of 5 children; for those with pronominal subjects, in only 1 out of 5 children. Wilson (2003) further investigated the effect of subject types on copula *'is'* production when the predicate type was held constant. Although production accuracy of copula *'is'* did not differ across subject types in sentences with nominal predicates, children tended to produce copula *'is'* more accurately with pronominal subjects than with lexical NP subjects in sentences with adjectival or locative predicates. Because the effect of subject types of copula *'is'* production was consistent when predicate types were controlled but not vice versa, Wilson concluded that subject types, rather than predicate types, were the major predictor of copula BE production.

However, because Wilson did support Becker's (2004) prediction that the production accuracy of copula *'is'* varied with predicate types in sentences with lexical

NP subjects, perhaps a better interpretation is that the effect of predicate types was moderated by the effect subject types. Furthermore, it should be possible to distinguish between the UCC hypothesis and the usage-based approach by examining the production accuracy of copula *'is'* in sentences with different subtypes of the same predicate type (e.g., nominal and permanent-adjectival predicates as subtypes of individual-level predicates), which Wilson (2003) did not do directly. The usage-based approach would predict that there should be variation in production accuracy of copula *'is'* in sentences with different subtypes of the same predicate type whereas Becker/ the UCC hypothesis should predict that the production accuracy should be similar.

To the best of our knowledge, there are no studies that have explicitly examined how frequency of predicate types may influence the production of copula BE. To make specific predictions concerning the effect of predicate types on the production of copula *'is'* for the current study, we counted the co-occurrence frequency of copula *'is'* and each predicate type in the parental input from the American corpora in CHILDES (MacWhinney, 2000). We limited our counting to the complete and intelligible utterances containing the particular sentence subjects that were used in this study (see the Method section). The co-occurrence frequency is 8550 for nominal predicates, 3107 for permanent-adjectival predicates, 1303 for temporary-adjectival predicates, and 840 for locative predicates. Based on this frequency counting, we predict sentences with nominal or permanent-adjectival predicates (both are individual-level predicates) should have higher production accuracy of copula *'is'* than those with locative or temporary-adjectival predicates (both are stage-level predicates). In addition, the production rate of copula *'is'* should vary in sentences with different subtypes of the same predicate type in children. Specifically, children should use copula *'is'* more correctly with nominal predicates than with permanent-adjectival predicates and produce copula *'is'* more often with temporary-adjectival predicates than with locative predicates.

Furthermore, the frequency of predicate words may also play a role in the production accuracy of copula '*is'* in young children. Development of an abstract predicate slot in constructions with copula BE (e.g., *It's* *NP*) requires the child to hear and learn a variety of instantiations (e.g., *It's* *a dog*, *It's* *a cat*) and derive the abstract predicate slot from these instantiations (Bybee, 2002). Before the child derives the abstract slot, he may rely on the specific instantiations that he learned to support his production. Because the child hears copula '*is'* co-occurring with high-frequency predicate words more often than with low-frequency predicate words in the input, he should use copula *'is'* more often with high-frequency predicate words (e.g., *It's* *happy*) than with low-frequency predicate words (e.g., *It's* *itchy*). The effect of predicate word frequency on the use of copula *'is,'* however, may vary with predicate types. Recall that constructions with one predicate type may be acquired earlier than the other type due to the difference in frequency. The predicate types that are acquired later may have less abstract representations than those that are acquired earlier. Children are more likely to use highly frequent/ lexically-specific representations to support the production of those constructions that are acquired later. Thus, the effect of predicate word frequency on the use of copula *'is'* may vary in sentences with different predicate types.

In summary, the corpus studies support the assumption of the usage-based approach that the production of tense and agreement morphemes in young children depends on the lexically-specific constructions that children have learned from the input, and not on the maturation of a unified underlying grammatical function. This is supported by variability in production accuracy between different tense and agreement morphemes. Even stronger evidence comes from the finding that children use a given form of BE and HAVE more accurately with pronominal subjects than with lexical NP subjects, which reflects the distribution of usage in the parental input.

Although the corpus studies based on the usage-based approach tend to challenge the UCC hypothesis, the evidence is not that straightforward. First, because these studies

used language samples, the determination of obligatory contexts for copula BE and other tense and agreement morphemes may sometimes be ambiguous (Gerken, 2000). For instance, the utterance *The book there* can be coded as omission of copula *'is'* but it could actually denote *Put the book there* or *I want the book there*.

Second, a strict test of the UCC hypothesis should examine the child's performance at a single point of developmental time (Ambridge & Pine, 2006). The previous corpus studies tended to collapse across language samples that were collected over a period of 6 to 14 months (e.g., Joseph et al., 2002; Pine et al., 2008). While this was probably necessary to obtain sufficient utterances to draw reliable conclusions, the developmental change in morpheme use might obscure performance that is consistent with the UCC hypothesis. In addition, our previous study (Guo et al., submitted) found that the effect of subject types on the production of auxiliary *'is'* varied in children with different developmental levels. This further stresses the importance of incorporating the variation in children's developmental levels into the analysis when researchers explore the frequency effect of sentence elements on the use of tense and agreement morphemes.

Third, though the asymmetry between subject types in the production accuracy of BE and HAVE can be attributed to the frequency effect, it can also result from difference in internal structures between these subject types. Though both pronominal and lexical NP subjects are DPs, there is internal branching for lexical NP subjects (e.g., $[_{DP}$ $[_{D}$ *The*$]$ $[_{NP}$ *dog*$]]$) but not for pronominal subjects (e.g., $[_{DP}$ $[_{D}$ *He*$]]$). Given that the internal structure of lexical NP subjects is more complex than that of pronominal subjects, producing sentences with lexical NP subjects could be relatively more demanding. A direct way to test this alternative is to investigate whether lexical NP subjects with different frequencies facilitate the use of tense and agreement morphemes to different degrees, which has not been done in the existing corpus studies. In addition, like predicate word frequency, the effect of subject types may also vary with predicate types

because some predicate types may have more abstract representations than the other. The interaction between subject and predicate types remain unclear in the literature.

Fourth, to what extent predicate types, predicate word frequency, and their interaction may influence the production accuracy of copula *'is'* remains unclear.

Summary of Research Questions

This study explores the effect of subject types, predicate types, and frequency of predicate words on children's production of copula *'is'* at a given point of developmental time via an elicited production task. Elicited production not only eliminates ambiguity in identifying obligatory contexts, but also allows for greater control over target sentences than spontaneous language sample, such as sentence length and sentence structure. It also allows us to manipulate the frequency of subclass of subject types, predicate types, and frequency of predicate words and to obtain sufficient data from children at a given point of developmental time, which generates data to test the UCC hypothesis directly. In addition, we explored how children's developmental levels of tense, as documented by the tense productivity score (Hadley & Short, 2005), may refine the effect of subjects/ predicates on the use of copula *'is.'* Specifically, we asked the following questions:

1) Would the production accuracy of copula *'is'* differ with pronominal, high-frequency lexical NP, or low-frequency lexical NP subjects in three-year-olds?

2) Would the production accuracy of copula *'is'* vary in sentences with different types of predicates (i.e., individual-level vs. stage-level predicates), subject types of predicates, and/or predicate word frequency?

3) Are there interaction effects between subject types and predicate types or between predicate word frequency and predicate types on the use of copula *'is*?'

4) Would the effect of each of the factors and/or interactions between these factors change with children's developmental levels of tense?

The specific predictions of the UCC hypothesis and the usage-based approach are listed in Table 3.1.

Table 3.1. Predictions of the Effect of Subject Types, Predicate Types, and Frequency of Predicate Words on the Production of Copula *'is'* in Two Accounts

Independent Variables	The UCC Hypothesis	The Usage-based Approach
Subject types	Pronominal = High-frequency Lexical NP = Low-frequency Lexical NP	Pronominal > High-frequency Lexical NP > Low-frequency Lexical NP
Predicate types	Individual-level > Stage-level	Individual-level > Stage-level
Subtypes of predicates	1. Nominal = Permanent-type adjectival predicates 2. Locative = Temporary-type adjectival predicates	1. Nominal > Permanent-type adjectival predicates 2. Temporary-type adjectival > locative predicates
Frequency of predicate words	High-frequency = Low-frequency	High-frequency > Low-frequency
Interaction between the main factors with developmental levels	No	Yes

Methods

Participants

Seventy-eight children with an age range from 2;7 (year; month) to 3;5 were recruited for the experiment and data from 12 children were retained for analysis. Eight of these 12 children also participated in Experiment 1. Children at this age range were recruited because they were likely to be producing copula *'is'* inconsistently (Rice, Wexler, & Cleave, 1995; Wilson, 2003; Pine et al., 2008). Sixty-six of the 78 children

recruited were excluded for analysis because they had reached ceiling levels of using copula *'is'* (38 children), were unable to fully complete the task (11), dropped out for personal reasons (9), had a history of speech and/or language problems (6), or did not produce at least one correct usage of copula *'is'* (2).

The 12 children (7 girls) included in the analysis had a mean age of 3;0 (ranging from 2;10 to 3;3). They had to perform below ceiling level (i.e., 90% correct, Brown, 1973) of using copula *'is'* for one or more levels of each of the three independent variables so that the sources of variability could be legitimately examined. In addition, the participants had to produce at least 3 scorable responses for each level of all three independent variables (Balason & Dollaghan, 2002) and produce at least one correct usage of copula *'is'* in the task. This last requirement is to ensure that the participants have copula '*is'* in their productive lexical inventory (Schütze, 2001).

All of these 12 children were typically-developing monolingual English speakers as documented by parent report and performance above 10th percentile on one standardized test. The standardized language tests were *Preschool Language Scale-Fourth Edition* (*PLS-4*, Zimmerman, Steiner, & Pond, 2002) for children who were younger than 3;0 and *Structured Photographic Expressive Language Test-Preschool* (*SPELT-P*, Werner & Kresheck, 1983) for those who were at or older than 3;0. We used the SPELT-P for those who were at or older than 3;0 because we only saw three-year-olds at the beginning stage of this study. The PLS-4 was added when the age was extended to 2;7. All children also had hearing within normal limits per ASHA standards (American Speech-Language-Hearing Association, 1997) and did not have a history of cognitive or motor problems. As a component of assessment, we also collected a conversational language sample for each child. All children demonstrated a mean length of utterance within the typical range with reference to the Miller norms (1981).

From the language sample, we documented the child's developmental level of tense with the tense productivity score (Hadley & Short, 2005) for statistical analyses.

The index is computed via counting the use of copula BE, auxiliaries BE and DO, third person singular *–s*, and past tense *–ed* in the language sample. The child receives a point for each sufficiently different use of these morphemes up to 5 points for each morpheme, for a possible total of 25. To be considered as a sufficiently different use, *-s* and *–ed* must occur with different lexical verbs, and BE and DO must occur with different subjects in declarative or interrogative sentences. To avoid overestimating the child's tense productivity, all the BE forms contracted with pronominal subjects are excluded for analysis because they may be learned as unanalyzed chunks (Rispoli, Hadley, & Holt, 2009; Tomasello, 2003). Scores can range from 0 to 25. The mean tense productivity score for the children in this study was 6.17 (SD = 2.66, range = 2-12).

Stimuli

The stimuli consisted of simple declarative sentences that required the use of copula *'is'* and varied in terms of subject types (i.e., pronominal, high-frequency lexical NP, and low-frequency lexical NP subjects), predicate types (i.e., nominal, permanent-adjectival, locative, temporary-adjectival predicates) and frequency of predicate words (i.e., high and low). Each target sentence was a combination of these three variables. For example, *The dog's hot* is a sentence with a high-frequency lexical NP subject and a high-frequency temporary adjective. Each combination had 5 target sentences, for a total of 120 sentences (3 subject types x 4 predicate types x 2 predicate frequencies x 5 sentences per combination). The target sentences are reported in Appendix D.

To ensure that the target sentences were composed of words that were likely to be familiar to typically-developing 3-year-olds, the selection of subject and predicate words was primarily based on Part 1 of T*he MacArthur-Bates Communicative Development Inventory: Words and Sentences* (*MB-CDI: WS*, Fenson et al., 1993). A total of 183 nouns, 63 adjectives, and 23 locative prepositions were selected from MB-CDI: WS; 16 additional nouns, 19 additional adjectives, and 1 additional locative preposition that were

considered child-friendly were also chosen. To manipulate the frequency of subject and predicate words, we used the frequency list of Moe, Hopkins, and Rush (1982) to determine the token frequency of each selected word. Because the frequency distributions of nouns, adjectives, and locative prepositions were different, we used a median split for each syntactic category to select high- and low-frequency words for subjects and predicates. The median frequency is 29 for common nouns, 30 for adjectives, and 479 for locative prepositions. Words with the frequency above the median are considered high frequency.

The lexical frequency of Moe and colleagues (1982) were based on the spoken discourse of first-graders. It is not clear whether this frequency count would reflect input frequency of children-directed speech. Moreover, the predictions of the usage-based approach are based on the **co-occurrence** frequency of subject + copula *'is'* and copula *'is'* + predicate. It is unknown from literature whether high-frequency lexical items also tend to co-occur with copula *'is'* more often than low-frequency ones. To answer these two questions, we counted the co-occurrence of each lexical item chosen as a target subject or predicate and copula *'is'* in the parental input of the American English corpora in CHILDES (MacWhinney, 2000). The correlation of log frequency of lexical items from Moe and colleagues (1982) and log frequency of co-occurrence from CHILDES was then computed. The log frequency of subject words was significantly correlated with the log co-occurrence frequency of subject + copula *'is'*, $r = 0.73$, $p < 0.01$. The log frequency of predicate words was significantly correlated with the log co-occurrence frequency of copula *'is'* + predicate for nominal predicates ($r = 0.76$, $p < 0.01$) and for adjectival predicates ($r = 0.75$, $p < 0.01$). Because only 5 locative prepositions were used in the target sentences (see below), we directly compared the co-occurrence frequency of copula '*is'* and each preposition for locative predicates and found that copula *'is'* co-occurred with high-frequency locative prepositions (i.e., *in, on, at*) more often than with low-frequency locative predicates (i.e., *near, under*). The trends in correlation and direct

comparison validate the use of lexical frequency to select subject and predicate words for the target sentences in this study.

Vocabulary Items for Subjects

In total, there were 4 pronominal subjects, 16 high-frequency lexical NP subjects, and 17 low-frequency subjects. The pronouns that were considered for the pronominal subjects included *he, she, it, that*, and *this*. We avoided *this* in the target sentences because we wanted all subjects to end with a vowel or non-sibilant consonant to allow for the contraction of copula *'is.'* Contexts for contraction were created because previous studies suggest that young children may learn specific subject + contracted copula *'is'* combinations (e.g., *it's*, *Daddy's*) but not subject + *un*contracted copula *'is'* as chunks to support their use of copula *'is*,' especially for pronominal subjects (Wilson, 2003).

Because there were 120 target sentences in total, some subjects were repeated across the target sentences (see Table 3.2). We repeated the lexical NP subjects in target sentences to make the target sentences more similar across subject conditions because repeating the pronominal subjects for target sentences was inevitable. The median and range of the raw frequency and mean log frequency of lexical items selected for target subjects are tabulated in Table 3.2 by subject type.

Table 3.2. Frequency Distribution of Subjects by Subject Type

	N (n of rep.)	Median	Range	Log Mean (SD)
Pronominal Subject (e.g., *he*)	4 (4)	5408.0	2753-7583	3.71 (0.15)
High-freq. NP Subject (e.g., *the pig*)	8 (16)	430.0	45-717	2.47 (0.36)
Low-freq. NP Subject (e.g., *the ant*)	10 (17)	10.0	1-27	0.99 (0.33)

Note: N refers to the total number of predicates that are used in target sentences; n of rep. refers to the number of predicate words that were repeated; freq. refers to frequency

A one-way ANOVA verified that the mean log frequency of subjects differed as expected across sentences with different subject types, F (2, 117) = 1039.40, $p < 0.001$, $\eta_p^2 = 0.95$. *Post hoc* Tukey test at the 0.05 level showed that all three types were significantly different from each other.

Vocabulary Items for Predicates

With regard to predicate words, we selected 16 nouns for nominal predicates, 17 adjectives for permanent-adjectival predicate, 5 prepositions for locative predicates, and 13 adjectives for temporary-adjectival predicates. Some of the predicate words were repeated across sentences with different subject types (e.g., *He's / The deer's a fireman*) because there were not many appropriate choices available (see Table 3.3). For instance, in order to maintain the permanent characteristics, meet the frequency requirement, and choose words that were familiar enough to 3-year-olds that they would produce the word in an elicited production task, the nominal predicate tended to be a noun representing an occupation or a character (e.g., *He's a fireman* / *Dad's an alien*). In addition, we avoided adjectival predicates with initial /s/ sounds to directly follow copula *'is'* by adding intervening intensifiers (e.g., *He's very sad*) for the ease of scoring.

The distributional information about the frequency of the predicate words are listed in Table 3.3 by predicate type and frequency. One-way ANOVAs showed that there were significant differences in mean log frequency between high- and low-frequency predicate words within each predicate type (all ps < 0.001, $\eta_p^2 > 0.72$). Regardless of predicate types, mean log frequency was higher in high-frequency predicate words than in low-frequency ones, F (1, 118) = 172.25, $p < 0.001$, $\eta_p^2 = 0.59$.

We chose to use a categorical variable with a median split within predicate types rather than a continuous variable (i.e., log frequency of predicate words) to examine the effect of predicate word frequency because the frequency of locative predicates is much higher than the others. If we use a continuous frequency of predicate words and obtain

frequency effects, we are not sure whether the effects are from high-frequency predicate words in general or from locative predicates only (Goodman, Dale, & Li, 2008).

Table 3.3. Number and Frequency Distribution of Predicates by Predicate Type and Frequency Category

Predicate type/frequency category	N (n of rep.)	Median	Range	Log Mean (SD)
Nominal				
High frequency (e.g., *It's a teacher*)	7 (5)	80	45-516	2.11 (0.43)
Low frequency (e.g., *It's a zebra*)	9 (4)	6	2-13	0.74 (0.23)
Permanent-adjectival				
High frequency (e.g., *It's really big*)	10 (4)	148	89-689	2.18 (0.24)
Low frequency (e.g., *It's very sweet*)	7 (6)	12	4-19	0.91 (0.29)
Locative				
High frequency (e.g., *It's on the table*)	3 (3)	3213	860-3710	3.38 (0.28)
Low frequency (e.g., *It's under the table*)	2 (2)	98	4-98	1.44 (0.70)
Temporary-adjectival				
High frequency (e.g., *It's very hot*)	7 (5)	86	42-172	1.97 (0.21)
Low frequency (e.g., *It's very dirty*)	6 (6)	9	1-17	0.81 (0.48)

Although sentence length is not considered a significant factor in the use of tense and agreement morphemes in the UCC hypothesis (Becker, 2004) or the usage-based hypothesis (Theakston et al., 2005), it has traditionally been a factor related to accurate production of morphemes (Bloom, 1990). The average sentence length between levels of each variable was controlled in terms of number of syllables. Because lexical NP subjects (e.g., *the dog*) tend to have one more syllable than pronominal subjects (e.g., *he*), the predicate words used in the target sentences were selected with this in mind.

For nominal predicates, we tended to select disyllabic nouns (e.g., *doctor*) for sentences with pronominal subjects (e.g., *he*) and monosyllabic lexical NP subject (e.g., *Mom*) but monosyllabic nouns (e.g., *nurse*) for sentences with other lexical NP subjects

than *mom* and *dad* (e.g., *the dog*), with a few exceptions. For permanent-type and temporary-type adjectival predicates, we used the intensifiers (i.e., *very, really*) and adjectives with different numbers of syllables (e.g., *tall, beautiful*) to control sentence length. For low-frequency locative predicates, we used the prepositions *under* and *near* for monosyllabic and disyllabic subject NPs, respectively. One-way ANOVAs indicated that sentence length did not differ significantly between different subject types, F (2, 117) = 1.24, $p = 0.29$, $\eta_p^2 = 0.02$; or between high- and low-frequency predicates, F (1, 118) = 2.04, $p = 0.16$, $\eta_p^2 = 0.02$. However, length differed between different predicate types, F (3, 116) = 8.09, $p < 0.001$, $\eta_p^2 = 0.17$. *Post hoc* Tukey tests showed that sentences with locative predicates were longer than sentences with other types of predicates, which did not differ from each other. Although we used intensifiers for adjectival predicates to increase the sentence length, we were not able to do so for nominal predicates. Because it was difficult for us to find multisyllabic nouns that named characters familiar to three-year-olds, we did not increase the sentence length for nominal, permanent-type adjectival, and temporary-type adjectival predicates, and instead, the effect of sentence length was considered in the statistical analysis.

Procedure

To eliminate order (or practice) effects, the target sentences were randomized and divided into four blocks, with 30 sentences in each block. Each block was tested in one or two separate sessions, depending on the progress of the child in the task. The order of blocks was counterbalanced across participants.

Each child was tested individually by a trained examiner. The examiners were eight undergraduate and two graduate students at the University of Iowa, who were all monolingual English speakers. Each session began with a warm-up comprehension activity, in which the child was asked to point to one of the four pictures after hearing the examiner say "*Show me the* _____" (e.g., *Show me* <u>*the cat*</u>). The pictures were the

subject lexical NPs in the target sentences tested in that session. This activity was to familiarize the child with these NPs and the child's performance did not affect being included/excluded for analysis. If the child made errors on a given trial, the examiner named all four pictures and asked the child to point to the target picture again.

After the warm-up activity, the elicitation task began. We adopted a parallel structure technique to elicit target sentences from the child. Target sentences were presented in black-and-white drawings that depicted the entity or event described by the target sentence (e.g., *The cat's very sad*) and beside the target picture was the contrast picture (e.g., *The mice are very happy*). The contrastive entities/events, listed in Appendix E, always involved plural subjects and copula *'are.'* Five practice items were used to familiarize the child with the experiment.

Each trial began with the examiner pointing to the picture and prompting the child by describing the contrastive picture. For all items, we included the target predicate words in the prompt (e.g., *very sad*) in order to facilitate the child's production of target predicate words in the response and to reduce the task demands. To maximize the child's production of target subject words, we prompted the child slightly differently for sentences with pronominal subjects (see 4a) or lexical NP subjects (see 4b).

(4) a. Target: It's really big.

Prompt: Look! Really small. Really big. **They are** really small. Can you tell me about **it?**

b. Target: The deer's under the tree.

Prompt: Look! On the flower. Under the tree. **The butterflies are** on the flower. Can you tell me about **this picture?**

For targets with pronominal subjects, the subject words in the contrastive sentences were always pronouns and the target pronominal subject was prompted in the last sentence of

the instruction. To elicit responses with lexical NP subjects, the subject words in the contrastive sentences were always lexical NPs. We did not prompt the target lexical NP subject in the last sentence of the instruction because children were more likely to use lexical NP subjects in the sentence if they did not hear the same lexical NP in the previous discourse context (Matthews, Lieven, Theakston, & Tomasello, 2006).

At the end of the prompt, the examiner pointed to the target picture and waited for the child's response. When the child responded with a non-target subject in the sentence, the examiner gave the child more prompts if the child substituted a pronominal subject with a lexical NP subject or vice versa. Under these circumstances or if the child only produced the predicate, the examiner prompted the child by saying "Can you start by saying Target Subject Word?" (e.g., *Can you start by saying it/the deer?*). Similarly, the examiner also prompted if the child substituted a predicate type with another predicate type or if the child only produced the subject word. Under these circumstances, the examiner prompted the child by repeating the original instruction (e.g., *Remember? On the flower. Under the tree. The butterflies are on the flower. Can you tell me about this picture?*). The examiner maximally used three additional prompts for each item.

Transcription and Coding

All responses were audiotaped for later analysis. The child's responses were transcribed by the examiner and coded as correct, incorrect, or unscorable. The rules of coding correct, incorrect, and unscorable responses are described below.

1. Correct Production

A. Target production (558 responses): All elements in the target sentence were present, with copula *'is'* being contracted (e.g., *He's very mad*) or uncontracted (*He is very mad*). Productions of missing elements unrelated to copula *'is'* (e.g., were also counted as target productions (e.g., *Cat is sad*)

B. Production with an alternative subject and/or predicate (261 responses): The target subject/predicate word was substituted with another target subject/predicate word (e.g., *The cat's sad* → *She's sad* or *The cat's small*). Although correct, the response was coded as a correct production of the alternative subject type, predicate type, and/or predicate word frequency.

C. Production with a non-nominative subject case (22 responses): The case of the subject was incorrect but copula *'is'* was correct (e.g., *He's very mad* → **Him is very mad*). In this case, this response was counted as a correct production with pronominal subjects. It should be noted that the use of non-nominative subject case together with a tense and agreement morpheme is not predicted by either the UCC hypothesis (Schütze & Wexler, 1996; Wexler, 1998) or the usage-based approach (Tomasello, 2003). However, because the current focus was on the production of copula *'is,'* we counted this type of production as correct to avoid underestimating the child's ability to use copula *'is.'*

2. Incorrect Production

A. Omission of copula (452 responses): Copula *'is'* was omitted (e.g., **He very mad*) in the target response. Like the correct productions, we also recorded whether the child used target sentences, non-target subjects and/or predicates, or non-nominative subjects in the omission errors. In total, there were 300 omission errors with target sentences, 129 with alternative subjects and/or predicates, and 23 with non-nominative subjects.

B. Agreement error (3 responses): A third person plural copula *'are'* was used with third person singular subject (e.g., **Pig are dirty*).

3. Unscorable Production

A. Use of unrelated structure (54 responses): A grammatical but non-target sentence was produced (e.g., *The lion's running fast, He has the crown*).

B. Omission of subject or predicate (56 responses): The subject (e.g., *Very mad* or *Is very mad*) or the predicate was omitted (e.g., *The deer*). Given that no subject or predicate was included, we were unable to code the subject type or predicate type/frequency for this response.

C. No response or unintelligible response (24 responses): The child did not respond or responded with an unintelligible utterance.

Table 3.4 summarizes the number of responses produced by the children. It should be noted because children used alternative subject and/or predicate words, the maximum numbers may exceed the upper limits of the number of targeted items.

Table 3.4. Number of Target Sentences and Mean Number of Scorable Sentences by Subject Type, Predicate Type, and Predicate Frequency Category

	N of Targets	Mean	SD	Range
Subject types				
Pronominal subjects	40	32.75	7.78	13-42
High-frequency lexical NP subjects	40	40.08	4.52	33-51
Low-frequency lexical NP subjects	40	36.00	2.70	31-41
Predicate types				
Nominal predicates	30	27.25	1.76	24-30
Permanent-adjectival predicates	30	25.92	2.43	21-30
Locative predicates	30	27.00	1.48	25-30
Temporary-adjectival predicates	30	28.67	2.35	26-35
Predicate frequency				
High-frequency predicate words	60	58.58	4.66	51-68
Low-frequency predicate words	60	50.33	5.29	42-58

Reliability

To check the transcription and coding reliability in this study, we randomly sampled 20% of the participants (n = 3). These 3 samples were re-transcribed and re-coded by the first author. The mean reliability was 95.19% in transcribing content words in the sentences, 92.89% in transcribing grammatical morphemes, and 96.28% in coding accuracy of using copula *'is.'*

Statistical Analysis

The dependent measure was the accuracy of copula *'is'* of each scorable response. Because of the dichotomous nature of the dependent measure, we chose a binominal logistic regression model to explore the effect of subject types, predicate types, and predicate word frequency on the production accuracy of copula *'is.'* Like linear regression, logistic regression relates one or more predictor variables to the dependent variable. The predictor variable can be either continuous or categorical (e.g., subject types, predicate types). When a model involves a categorical variable, one level of the categorical variable is selected as reference category/condition, as (5) exemplifies (where $PRED_{NOM}$ refers to nominal predicates; $PRED_{PERM}$ refers to permanent-adjectival predicates, and $PRED_{TEMP}$ refers to temporary-adjectival predicates).

$$(5)\ \text{Logit}\ (Y_i) = \beta_0 + \beta_1 PRED_{NOM} + \beta_2 PRED_{PERM} + \beta_3 PRED_{TEMP}$$

The model in (5) predicts the effect of predicate types on the production accuracy of copula *'is'* and the locative predicate was selected as the reference category. The β values can be used to compute logit associated with each condition. The logit is the natural logarithm of predicted odds for a given event, such as the predicted log odds of producing copula *'is'* accurately in the nominal predicate condition.

The β-coefficient associated with each term (except the intercept term) also reflects the log odds ratios between the category of interest and the reference category. For instance, β_1 in (5) reflects the log odds ratio of producing copula *'is'* correctly when the nominal predicate condition is compared to the locative predicate condition. A positive β-coefficient indicates an increased likelihood of producing copula *'is'* correctly in the condition of interest (e.g., the nominal predicate condition) as compared to the reference condition (e.g., the locative predicate condition). The *p*-value of the β-coefficient reveals whether the log odds ratio between conditions is significantly different from 0. For instance, if β_1 in (5) is positive and reaches a significant level, this suggests that children are more likely to produce copula *'is'* correctly with nominal predicates than with locative predicates. Log odds ratios can be transformed into odds ratios (OR) for the ease of interpretation.[4]

Results

Sentence Length of Target Responses

Differences in sentence length could confound the effects of interest on the production of copula *'is'* (Bloom, 1989, 1990). Not only because sentences with locative predicates were longer but also because children omitted and/or substituted words in the response, the length of target sentences was likely to be unequal across levels of independent variables. To explore this possibility, we measured the length of scorable responses (excluding copula *'is'*) in syllables for each item produced by each child and

---

[4] OR larger than 1 means that children are **more** likely to produce copula *'is'* in the condition of interest than in the reference condition. In contrast, OR smaller than 1 means that children are **less** likely to produce copula *'is'* in the condition of interests than in the reference condition. For instance, if the β-coefficient associated with the nominal predicate is 1.1035, this means that the log odds ratio of producing copula *'is'* correctly between the nominal predicate and the reference category (i.e., the locative predicate) is 1.1035 (OR = exp(1.1035) = 3.01). It means that children are 3.01 times more likely to use copula *'is'* correctly with nominal predicates than with locative predicates.

computed mean sentence length across subject types, predicate types, and predicate frequency categories (Theakston et al., 2005), as is shown in Table 3.5.

Table 3.5. Mean Sentence Length of Scorable Responses in Syllables by Subject Type, Predicate type, and Predicate Frequency Category

	Mean	SD
Subject types		
Pronominal subjects	3.99	0.18
High-frequency lexical NP subjects	4.27	0.34
Low-frequency lexical NP subjects	4.24	0.25
Predicate types		
Nominal predicates	4.01	0.30
Permanent-adjectival predicates	3.93	0.25
Locative predicates	4.75	0.45
Temporary-adjectival predicates	4.01	0.23
Predicate frequency		
High-frequency predicate words	4.06	0.25
Low-frequency predicate words	4.31	0.28

Given that sentence length of target responses varied across independent variables [subject types, $F(2, 22) = 8.78$, $p < 0.01$, $\eta_p^2 = .44$; predicate types, $F(3, 33) = 26.33$, $p < 0.01$, $\eta_p^2 = .71$; and predicate frequency categories, $F(1, 11) = 5.34$, $p = 0.04$, $\eta_p^2 = .33$], we examined whether sentence length affected the production accuracy of copula *'is.'* A mixed model binomial logistic regression was performed, with item and child treated as random factors and sentence length as a fixed factor. The model indicated that production accuracy of copula *'is'* decreased as sentence length increased (OR = 0.78, $p < 0.01$). Because this may have influenced the factor of interests, we added sentence length as an interaction term in each of the mixed model binomial logistic regressions reported below. The interaction terms with sentence length did not significantly improve

model fit (all *p*s > 0.14). For this reason, sentence length was treated as a covariate in each of the models reported below (Kleinbaum, Kupper, Muller, & Nizam, 1998) and its interaction with other variables was dropped from all models.

Effects Subject Types, Predicate Types, and Predicate Word Frequency

Table 3.6 presents mean percent correct of target sentences across children by subject type, predicate type, and predicate frequency category. In the subsections below, we first focus on the effect subject types and then the effect of predicate types and predicate word frequency. Within each subsection, we examine the main effect of each factor to test the predictions of the UCC hypothesis and the usage-based approach, and then explore the interaction between these factors as well as children's tense productivity scores to build a model that best predicts the production of copula *'is'* in young children.

Table 3.6. Mean Proportion Correct of Target Sentences by Subject Type, Predicate Type, and Predicate Frequency Category

	Mean	SD
Subject types		
Pronominal subjects	0.65	0.21
High-frequency lexical NP subjects	0.64	0.22
Low-frequency lexical NP subjects	0.67	0.22
Predicate types		
Nominal predicates	0.77	0.18
Permanent-adjectival predicates	0.71	0.23
Locative predicates	0.46	0.31
Temporary-adjectival predicates	0.67	0.24
Predicate frequency		
High-frequency predicate words	0.63	0.20
Low-frequency predicate words	0.68	0.21

Effect of Subject Types and Their Interaction with Other Factors

To test the hypothesis that subject types influence the production of copula *'is*,' we put the factor of subject types into the model, with child and item treated as random factors, sentence length treated as a covariate, and subject types treated as a fixed factor (i.e., full model). We compared the log-likelihood values of the full model and the reduced model (i.e., the model with sentence length only) using a chi-square test for goodness of fit (Kleinbaum et al., 1998). The model with subject types and sentence length did not significantly improve model fit as compared to the model with sentence length only ($\chi^2 = 0.78$, $df = 2$, $p = 0.68$), indicating that subject types were not a significant predictor of production accuracy of copula *'is'* over and above sentence length alone. This result was at odds with the previous corpus studies (e.g., Wilson, 2003; Pine et al., 2008). Because the previous studies did not separate lexical NP subjects by frequency, we combined the high- and low-frequency lexical NP subjects in a mixed model logistic regression to allow a more direct comparison between our data and the previous findings. Even with this modification, subject types remained insignificant in predicting the production accuracy of copula *'is'* ($\chi^2 = 0.03$, $df = 1$, $p = 0.86$). This finding seems inconsistent with the usage-based approach, which predicts that pronominal subjects facilitated the use of copula *'is'* more than lexical subjects. Instead, the finding favors the UCC hypothesis, which predicts no differences between subject types.

Because our previous work has shown that children's developmental levels affect the production of tense and agreement morphemes, we examined whether children's developmental levels of tense would refine the effect of subject types on the production accuracy of copula *'is.'* We added children's tense productivity scores into the model with subject types. Two models were completed: one with the interaction between subject types and children's tense productivity scores and one without the interaction.

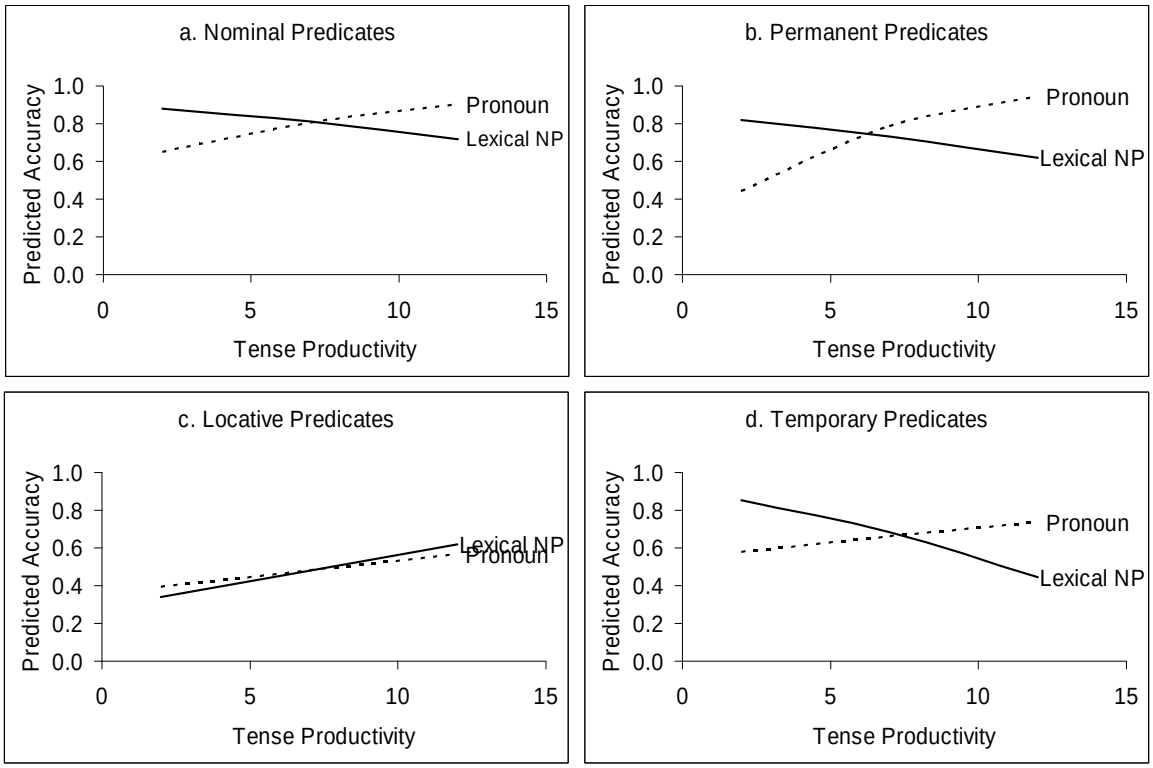
The model with the interaction significantly improved model fit as compared the model without the interaction when high- and low-frequency lexical NP subjects were separated (χ^2 = 13.69, *df* = 2, *p* = 0.001) or combined (χ^2 = 11.96, *df* = 1, *p* < 0.001). That is, as expected, subject types affected children's production accuracy of copula *'is,'* differently, depending on the children's developmental levels of tense.

Because the effect of subject types may also vary with predicate types, we added predicate types into the model to examine whether the interaction between subject types and tense productivity scores would be further refined with the addition of predicate types in the model (i.e., three-way interaction). The model with the three-way interaction between subject types, predicate types, and tense productivity scores significantly improved model fit as compared the model with only two-way interactions when high- and low-frequency lexical NP subjects were combined (χ^2 = 8.18, *df* = 3, *p* = 0.04) but not when they were separated (χ^2 = 9.34, *df* = 6, *p* = 0.16). That is, after sentence length was controlled, subject types had different effects on the production accuracy of copula *'is,'* depending on predicate types and the child's tense productivity. The model with all interactions is best depicted visually in probability space for the ease of interpretation, as Figure 3.1 illustrates.

The model with the three-way interaction indicates that subject types influenced the production of copula *'is'* differently in children with different levels of tense productivity in sentences with nominal (β = 0.26, *p* = 0.03), permanent-adjectival (β = 0.41, *p* = 0.003), or temporary predicates (β = 0.27, *p* = 0.02), but not in those with locative predicates (β = -0.04, *p* = 0.68). For nominal, permanent-adjectival, or temporary-adjectival predicates, children with lower tense productivity scores tended to use copula '*is'* with lexical NP subjects more accurately than with pronominal subjects after sentence length was controlled. In contrast, children with higher tense productivity scores were more likely to produce copula *'is'* correctly with pronominal subjects than with lexical NP subjects, which is consistent with the previous corpus studies (e.g.,

Wilson 2003; Pine et al., 2008). For locative predicates, the production accuracy of copula *'is'* did not vary with the child's tense productivity. This model further reveals that the production accuracy of copula *'is'* did not differ with pronominal or lexical subjects (Pronominal/Lexical OR = 1.37, $p = 0.68$) in sentences with locative predicates, regardless of the child's tense productivity.

Figure 3.1. Predicted Probability of Producing Copula *'is'* Accurately by Subject Type, Predicate type, and Tense Productivity Score (Sentence Length = 4 syllables)



Note: The predicted probability is computed from the β coefficients in the full model

Effect of Predicate Types and Predicate Word Frequency
and Their Interaction with Other Factors

Because one goal of this study was to directly test the predictions of the UCC hypothesis and the usage-based approach with regard to the role of predicate types and predicate word frequency on the use of copula *'is'* and neither of these accounts explicitly addresses potential interactions between these factors and developmental levels or subject types, we return to a simpler for the purpose of hypothesis testing. A mixed model binomial logistic regression was performed, with child and item treated as random factors, sentence length treated as a covariate, and predicate types and predicate word frequency treated as fixed factors (i.e., full model). The full model significantly improved model fit as compared to the model with sentence length alone ($\chi^2 = 60.79$, $df = 4$, $p < 0.001$), indicating that predicate types and predicate word frequency were significant predictors of the production accuracy of copula *'is.'*

Table 3.7 presents the full model. The full model indicated that, after predicate word frequency was controlled, children were more likely to produce copula *'is'* correctly with nominal, permanent-adjectival, or temporary-adjectival predicates than with locative predicates (ORs > 2.43, *p*s < 0.01). However, this model did not directly test the differences between nominal, permanent-adjectival, and temporary-adjectival predicates. To further examine the differences between these predicate types, we re-ran the model with different reference conditions (Jaccard, 2001). The model showed that children produced copula *'is'* more accurately with nominal predicates than with temporary-adjectival predicates (Nominal/Temporary OR = 1.82, $p < 0.01$). The production accuracy of copula *'is'* was marginally higher with nominal predicates than with permanent-adjectival predicates (Nominal/Permanent OR = 1.42, $p = 0.09$). In addition, the production accuracy of copula *'is'* did not differ significantly with permanent- or temporary-adjectival predicates (Permanent/Temporary OR = 1.31, $p = 0.20$). That is, the production accuracy of copula *'is'* was highest with nominal

predicates, followed by permanent-adjectival and temporary-adjectival predicates, which was in turn higher than locative predicates. This finding was consistent with Wilson (2003). Moreover, the full model showed that children were more likely to produce copula *'is'* accurately with low-frequency predicate words than with high-frequency ones after predicate types were controlled (Low/High OR = 1.32, $p = 0.04$).

Table 3.7. Regression Model Showing the Likelihood of Producing copula *'is'* correctly by Predicate Type and Predicate Word Frequency

	Variance	SD		
Random Factors				
Item	0.06	0.25		
Child	1.02	1.01		
	Coefficient	**SE**	**OR**	**p-value**
Fixed Factors				
Intercept	0.42	0.44	1.52	0.32
Length	-0.17	0.06	0.84	0.008
Predicate Type (reference condition = locative predicates)				
Nominal Predicates	1.50	0.21	4.48	< 0.001
Permanent-adjectival Predicates	1.19	0.20	3.28	< 0.001
Temporary-adjectival Predicates	0.89	0.19	2.43	< 0.001
Predicate Word Frequency (reference condition = high-frequency predicate words)				
Low-frequency Predicate Words	0.28	0.06	1.32	0.04

Note: SD = standard deviation; SE = standard error; OR = odds ratios

Because the effect of predicate word frequency may vary with predicate types on the production accuracy of copula *'is'* and children's developmental levels of tense may refine the effect of predicate types and predicate word frequency, we then explored whether there were any two-way or three-way interaction effects between these factors on

the use of copula *'is'* through identical procedures above. The goodness-of-fit tests indicated that there were no two-way or three-way interactions between predicate types, predicate word frequency, and tense productivity scores (χ^2s < 5.50, *p*s >0.51), except that the effect of predicate types on the production accuracy of copula *'is'* varied with children's tense productivity scores ($\chi^2 = 7.83$, $df = 3$, $p = 0.04$). Given that the interaction between predicate types and children's tense productivity scores was further qualified by considering subject types (see Figure 3.1) in the model, the details of this two-way interaction was not reported here.

Discussion

The study tested the unique checking constraint (UCC) hypothesis and the usage-based approach by examining the role of subject types, predicate types, and predicate word frequency in the production accuracy of copula *'is.'* When we did not take the child's tense development into account, we found that children were more likely to produce copula *'is'* correctly with nominal predicates than with permanent- or temporary-adjectival predicates, followed by locative predicates. Regardless of predicate types, children tended to produce copula *'is'* more accurately with low-frequency predicate words than with high-frequency predicate words. However, the production accuracy of copula *'is'* did not differ with pronominal or lexical NP subjects, which, at the first glance, supported the UCC hypothesis. However, after we took predicate types and the child's tense development into account, we found that subject types did affect the production of copula *'is,'* and its effect varied with these factors.

Limitations of the Current Study

Before we discuss the results in detail, we should consider two limitations of this study. First, sentence length was not equal across levels of each independent variable because of the constraints in designing the sentences and the children's substitution/ omission of sentence elements in the task. Given that sentence length could confound the

effect of factors of interests (Bloom, 1990), we attempted to address this problem by treating sentence length as a covariate in the analysis. Thus, sentence length was statistically controlled when we inspected the effect of independent variables on the use of copula *'is.'* A second concern is that we had a small number of children in the statistical analysis because many children we tested had reached ceiling levels of using copula *'is,'* thus limiting the power of the analysis to find differences between conditions. Nevertheless, we were still able to find the effect of subject types, predicate types, and predicate word frequency on the use of copula *'is,'* indicating that this is an robust finding.

Predicate Types and Predicate Word Frequency Affect the Production Accuracy of Copula *'is'*

This study confirmed that predicate types and predicate word frequency affected the production accuracy of copula *'is.'* When we did not consider children's developmental levels, children tended to use copula *'is'* more correctly with nominal predicates than with permanent-adjectival or temporary-adjectival predicates, followed by locative predicates after predicate word frequency was controlled. This finding was consistent with the study of Wilson (2003), in which production accuracy of copula *'is'* was highest with nominal predicates, followed by adjectival predicates and locative predicates in that order. Our result was also partially consistent with the studies of Becker (2000a, 2000b, 2004). Recall that in these studies, production accuracy of copula *'is'* was higher with nominal predicates than with locative predicates, and higher with permanent-adjectival predicates than with temporary-adjectival predicates. For the effect of predicate word frequencies, children tended to produce copula *'is'* more accurately with low-frequency than with high-frequency predicates after predicate types were controlled.

Testing the UCC hypothesis

With regard to the two language learning hypotheses that we were testing, the results are mixed. The UCC hypothesis predicted that children should produce copula *'is'* more accurately with individual-level predicates (i.e., nominal and permanent-adjectival predicates) than with stage-level predicates (i.e., locative and temporary-adjectival predicates). This prediction was generally supported by the current study though there was no significant difference between permanent- and temporary-adjectival predicate conditions.

However, the variation in production accuracy of copula *'i*s' between subtypes of the same predicate types (e.g., locative and temporary-adjectival predicates) and between high- and low-frequency predicate words does not support the predictions of the UCC hypothesis, which states that it is the inability to check the functional categories of agreement or tense that results in the omission of tense and agreement morphemes (Wexler, 1998). Thus, subtypes of the same predicate types or predicate words with different lexical frequency should be equally (in)vulnerable to the UCC, which was not confirmed by the current study.

One might argue that the different production accuracy of copula *'is'* between subtypes of the same predicate types may result from the differences in sentence length. This possibility, however, was ruled out because the difference between subtypes of the same predicate type was robust over and above differences due to sentence length alone. In addition, sentence length was not able to account for the differences between the nominal and permanent-adjectival predicates, given that these two sets of sentences were similar with regard to length, even after children's changes to the sentence elements were considered.

Testing the Usage-based Approach

Turning now to the usage-based approach, we see that children tended to produce copula *'is'* more correctly with individual-level predicates than with stage-level predicates. More importantly, the production accuracy of copula *'is'* also varied between subtypes within individual-level or within stage-level predicates. These findings confirmed the prediction of the usage-based approach that predicate types with higher co-occurrence frequency with copula *'is'* facilitated the production of copula *'is'* more than those with lower co-occurrence frequency. This is because constructions with higher frequency would have stronger representations and be easier to retrieve than those with lower frequency (Wilson, 2003; Theakston & Lieven, 2008).

However, if frequency really plays a role in the production of copula *'is,'* why was there no difference between the permanent-adjectival and the temporary-adjectival predicates? One possibility that children may have gradually developed shared representations (e.g., *He's AdjP*, *NP's AdjP*) for permanent-adjectival and temporary-adjectival predicates, given that these predicates are structurally and functionally similar (Bybee, 2002; Theakston & Lieven, 2008; Tomasello, 2003). The shared representations may be used to support the child's production and the difference between both types of adjectival predicates is minimized.

A even greater challenge to the usage-based approach is that children use copula *'is'* more correctly with low-frequency predicate words than with high-frequency predicate words regardless of predicate types, a finding which is in opposite direction of that predicted by the theory. One possible explanation is that lower production accuracy of copula *'is'* with high-frequency predicate words than with low-frequency ones may be attributed to the competition of ungrammatical constructions that contain high-frequency predicate words, such as *NP big, NP mad*, and *NP in NP* (Lieven, 2008; MacWhinney, 2004). Take, the adjectival predicates, for example. Children in this study may have developed constructions with abstract predicate slots like *He's AdjP* and *NP's AdjP* to

describe the properties of an object or event but the constructions with specific predicate words that were learned earlier may remain in the representation (Bybee, 2002; Dąbrowska, 2000; Dąbrowska & Lieven, 2005; Lieven, 2008). Such constructions with specific predicate words may compete with the constructions with abstract predicate slots when the child attempts to describe the property of objects. Recall that these specific constructions may be grammatical (e.g., *NP's big, NP's mad, It's big, He's mad*) or ungrammatical (e.g., *(is) NP big, (is) it big, (I make) NP mad, (I make) him mad*) (Freudenthal et al., 2007). The ungrammatical constructions might have strong representations because children in the current study may have accumulated a number of instantiations of these constructions. Due to the strength of representations, these ungrammatical constructions may sometimes triumph, leading to the omission of copula *'is'* (Lieven, 2008; MacWhinney, 2004). In other words, sentences with high-frequency predicate words are relatively more vulnerable to the omission of copula *'is'* than those with low-frequency ones as children develop the abstract predicate slot because the former are more likely to encounter competition from the ungrammatical constructions.

Effect of Subject Types on the Production of Copula *'is'* Varied with Predicate Types and Tense Productivity

In addition to predicates, subject types affect the production accuracy of copula *'is,'* but only when predicate types and tense productivity are taken into account. For nominal, permanent-adjectival, and temporal-adjectival predicates, children with lower tense productivity produced copula *'is'* more accurately with lexical NP subjects than with pronominal subjects. In contrast, children with higher tense productivity produced copula *'is'* more correctly with pronominal subjects than with lexical NP subjects. The finding that pronominal subjects facilitated the production accuracy of copula *'is'* more than lexical NP subjects was expected given the results from the previous studies (e.g., Joseph et al., 2002; Wilson, 2003; Pine et al., 2008). For locative predicates, the

production accuracy of copula *'is'* with pronominal or lexical subjects did not differ significantly, regardless of the child's tense productivity. This finding is at odds with the study of Wilson (2003), in which children tended to use copula BE more often with pronominal subjects than with lexical NP subjects in sentences with locative predicates.

Testing the UCC hypothesis

The initial finding that the production accuracy of copula *'is'* did not vary with subject types was consistent with the prediction of the UCC hypothesis, which predicted that production accuracy of copula *'is'* should be similar with pronominal and lexical subjects due to the similarity in their structural properties. However, the asymmetries between subject types in sentences with nominal or adjectival predicates were not consistent with the UCC hypothesis. One could argue that children with higher tense productivity produced copula *'is'* more accurately with pronominal subjects than with lexical NP subjects because of the difference in internal structures between these subject types. Children with higher tense productivity may therefore omit copula *'is'* more often with lexical NP subjects than with pronominal subjects because lexical NP subjects are structurally more complex than pronominal subjects. However, this structure-processing account is not able to explain the reverse pattern observed in children with lower tense productivity or the lack of differences between subject types in sentences with locative predicates.

Testing the Usage-based Approach

In contrast, the finding that children with higher tense productivity produced copula *'is'* more accurately with pronominal subjects than with lexical subjects in sentences with nominal or adjectival predicates was consistent with the prediction of the usage-based approach. Pronominal subjects have a stronger facilitative effect because they co-occur with copula *'is'* more frequently in the input (Joseph et al., 2002; Kemmer & Barlow, 2000; Lieven, 2008).

However, this advantage of pronominal subjects did not hold for all children. Recall that children with lower tense productivity produced copula *'is'* more correctly with lexical subjects than with pronominal subjects in sentences with nominal or adjectival predicates. In fact, this result was not a complete surprise. Theakston and colleagues (2005) found that children tended to use auxiliary BE less correctly with pronominal subjects *I* and *You* than with other subjects (e.g., *he, it*, lexical NPs) though the constructions *I'm* and *You're* were more frequent than other specific subject + auxiliary BE constructions in the input.

Like the effect of predicate words discussed above, the finding that children with lower tense productivity used copula *'is'* more correctly with lexical subjects than with pronominal subjects can be attributed to the competition of ungrammatical constructions with pronominal subjects, such as *It NP*, *He AdjP*, and *Her AdjP*. Children with lower tense productivity may have developed grammatical constructions like *NP's NP, He's NP, NP's AdjP*, and *It's AdjP*, but the acquired ungrammatical constructions may compete with the grammatical constructions, risking the omission of copula *'is'* in sentences with pronominal subjects (Freudenthal et al., 2007; Lieven, 2008; Theakston & Lieven, 2008). In contrast, children with higher tense productivity used copula *'is'* more accurately with pronominal subjects than with lexical NP subjects. This may be because they gradually figure out that these ungrammatical sequences are part of longer utterances. Their use of these ungrammatical constructions will gradually decrease over time and cease to be strong competitors (Freudenthal et al., 2007).

Though the notion of competition of ungrammatical constructions may explain the dynamic relation between pronominal and lexical subjects, one discrepancy between the current and our previous (Guo et al., submitted) studies remains to be resolved. In the current study, children with lower tense productivity used copula *'is'* more accurately with lexical NP subjects than pronominal subjects. In our previous study, however, children with lower tense productivity used auxiliary *'is'* more often with pronominal

subjects than with lexical NP subjects (Guo et al., submitted). This inconsistency may occur because copula and auxiliary BE have different developmental trajectories. It has been documented that children acquire copula BE earlier than auxiliary BE (Brown, 1973; Joseph et al., 2002; Wilson, 2003), which suggests that although these verbs share the same surface forms, they are actually unique categories that the child must acquire. For children with a tense productivity score of 2, the production accuracy was approximately 0.70 for copula *'is'* in the current study and 0.30 for auxiliary *'is'* in our previous study.

It is possible that children with lower tense productivity have accumulated more instantiations of both grammatical and ungrammatical constructions involving copula *'is'* and pronominal subjects. Competition between the two forms may lead to lower rates of use with pronouns initially. Lexical NP subjects may have fewer ungrammatical competitors available. In contrast, children with lower tense productivity may not have accumulated enough instantiations of ungrammatical constructions involving auxiliary *'is'* with pronominal subjects. Thus, the ungrammatical constructions are still not strong enough to compete with the grammatical constructions with specific pronominal subjects that contain auxiliary *'is'* (e.g., *It's + Verb-ing*). In addition, these children may have not learned enough grammatical constructions with lexical NP subjects. Children with lower tense productivity scores therefore produced auxiliary *'is'* more accurately with pronominal subjects than with lexical NP subjects but used copula *'is'* more correctly with lexical NP subjects than with pronominal subjects. However, children in these two studies did not totally overlap. This interpretation is thus only tentative and longitudinal studies involving children with different levels of tense productivity producing both copula and auxiliary BE are needed to verify this interpretation. A corpus study examining the relative rates of copula and auxiliary constructions in simple and complex sentences would also provide additional evidence to verify this hypothesis.

However, if the frequency of subject types does affect the use of copula *'is,'* why was there no difference between pronominal and lexical NP subjects in sentences with

locative predicates? One possibility is that children were at floor levels of using copula *'is'* with locative predicates. This explanation does not seem valid because the mean accuracy of copula *'is'* was approximately 47% with locative predicates and 5 out 12 children scored above 60% correct. The other possibility is that although specific Pronoun*'s* + PP constructions (e.g., *He's PP*, *It's PP*) may be relatively more frequent than the NP*'s* + PP construction in the input, the difference may not be strong enough to be reflected in the child's production. To test this possibility, we counted the frequency of these constructions in the parental input of the American Corpora in CHILDES. We limited our counting on prepositions *'at'* and *'under.'* In terms of the preposition *'at,'* the frequency is 61 for specific Pronoun*'s* + PP constructions and 62 for the NP*'s* + PP construction; in terms of the preposition *'under,'* the frequency is 46 for specific Pronoun*'s* + PP constructions and 7 for the NP*'s* + PP construction. Because of the small difference in input frequency, the difference in the representational strength between specific Pronoun*'s* + PP constructions and the NP*'s* + PP construction may not be strong enough to be easily reflected in the child's production.

Experimental Implications

This study has experimental implications for future studies. First, based on the finding that the effects of subject types on the production of copula *'is'* changed with the child's tense productivity, we stress the importance of incorporating the child's developmental level into the experimental hypotheses when researchers investigate factors affecting language development. The importance of considering the child's developmental level has been observed in areas of language other than grammar. For instance, children between the ages of 17 months to 30 months were more likely to generalize names by shape for solid stimuli than for non-solid stimuli, but this effect was limited to children with larger vocabulary sizes (Samuelson & Smith, 1999). Similarly, first-graders with low word decoding scores achieved greater growth in word decoding in

classrooms that spent more time in teacher-managed explicit decoding instruction. In contrast, the time spent in teacher-managed explicit decoding instruction did not play a significant role in children with high decoding scores (Connor, 2009; Connor, Morrison, & Katch, 2004). Taken together, these studies further support that considering children's developmental levels in contrasting experimental hypotheses will lead to a better understanding of cues or instructions that facilitate language learning in typical and atypical populations.

In addition, the results in this study may generate hypotheses for the underlying mechanisms of children with specific language impairment (SLI), a group of children who have difficulty using tense and agreement morphemes. Leonard (2007) proposed that, because children with SLI have limited processing capacity, they may continue to partially process the sentences while their typically-developing peers are able to completely process the sentences. As a result, the ungrammatical constructions (e.g., *(I see) him running*) that are learned from partial processing may continue to be stored in the representation and remain to be strong competitors to grammatical constructions in children with SLI for a protracted period of time, leading to lower-than-typical production rates of tense and agreement morphemes (Freudenthal et al., 2007; MacWhinney, 2004). If future studies find that children with SLI tend to use copula *'is'* more often with lexical NP subjects than with pronominal subjects, and/or more accurately with low-frequency predicate words than with high-frequency predicate words for an extended period of time as compared to typically-developing children, this might suggest that Leonard's hypothesis is correct. Furthermore, recent studies have found that children with SLI were less efficient in inhibiting competition from phonologically similar vocabulary items than their age- and vocabulary-matched peers (Coady, 2009; Mainela-Anold, Evans, & Coady, 2008). The lower-than-typical production rate of tense and agreement morphemes in children with SLI may partly result from their difficulty inhibiting the competition from the ungrammatical constructions. Further studies are

needed to understand how the interaction between linguistic input, processing capacity, and inhibition ability may account for the difficulty using tense and agreement morphemes in children with SLI.

Conclusion

This study tested the unique checking constraint (UCC) hypothesis and the usage-based approach by exploring the effects of subject types, predicate types, and predicate word frequency on the production of copula *'is'* in three-year-olds via an elicited production task. The results indicated that these factors play a role in the production accuracy of copula *'is.'* The role of subject types was evident only when we considered its interaction with predicate types and the child's developmental levels. Overall, these effects may be explained by the notion of input frequency. Taken together, these findings tend to support the assumption of the usage-based approach that tense and agreement morphemes are learned from the input in a gradual, piecemeal fashion. Children use tense and agreement variably because they have not acquired adult-like abstract constructions and instead use the constructions that may or may not these morphemes. In contrast, the findings did not confirm the UCC hypothesis, which posits that children were born with principles and parameters related to tense and agreement and that children use tense and agreement morphemes variably because the presence of the UCC in the grammar. Though the results may be accounted for input frequency, we are not claiming that input frequency is the only determinant that account for the variable use of tense and agreement morphemes in young children. Furthermore, in order to fully account for the frequency effect, more mechanisms such as competitions from ungrammatical constructions and children's developmental level need to be taken into consideration.

CHAPTER IV

EXPERIMENT 3: REPRESENTATION OF AUXILIARY BE IN YOUNG CHILDREN

Overview

Whether young children have abstract syntactic representations is an open question in the field of language acquisition (Fisher, 2002; Naigles, 2002; Tomasello, 2000; Tomasello & Abbot-Smith, 2002; Tomasello & Akhtar, 2003). Tense and agreement morphemes, which include both function words (e.g., forms of BE, such as *The cat is running*) and inflections (e.g., third person singular *–s*, such as *The cat runs*) that mark for time, person, and number, have been one area of intensive study concerning this question (see Avram, 2000; Wilson, 2003). Although current acquisition theories generally agree that typically-developing children eventually acquire adult-like abstract representations of these morphemes, even theories like the usage-based approach that assume gradual acquisition of language are not clear with regard to *when* children reach this point (Tomasello, 2003). This study tested whether three-year-olds have an adult-like abstract representation of auxiliary BE using a structural priming paradigm (Bock & Griffin, 2000; Shimpi, Gámez, Huttenlocher, & Vasilyeva, 2007). In the following, we briefly review the usage-based view of representations of tense and agreement morphemes, the rationale behind structural priming paradigm, and the relevant studies that explored the abstractness of (morpho-)syntactic representations in children.

Representations of Tense and Agreement Morphemes in the Usage-based Approach

Though typically-developing children may start to use tense and agreement morphemes as early as they start to combine words together, they experience a period in which they use these morphemes variably (Brown, 1973; Schütze & Wexler, 1996; Wilson, 2003). Young children may sometimes omit tense and agreement morphemes

and use them correctly at other times. Researchers holding the usage-based view have attributed the variable use of these morphemes to the lack of abstract representations (Joseph, Serratrice, & Conti-Ramsden, 2002; Theakston, Lieven, Pine, & Rowland, 2005).

The usage-based approach, which is grounded in functional-cognitive linguistics (Goldberg, 1995; Langacker, 2000), assumes that language acquisition involves learning linguistic constructions, such as words, idioms, sentence frames, and so on, from the input (Tomasello, 2003). At the beginning, the constructions that children learn may be unanalyzed chunks (e.g., *What's that?*) or constructions with fixed lexical items and open slots (i.e., lexically-specific constructions, such as *He's* <u>*Verb*</u>*-ing*). High-frequency chunks or lexically-specific constructions tend to be learned earlier than the low-frequency ones. It is assumed that abstraction occurs in a piecemeal fashion as children accumulate sufficient exemplars to abstract over. Young children use tense and agreement morphemes variably because they have not yet learned the abstract constructions of tense and agreement morphemes (e.g., *NP BE Verb-ing*, where NP refers to noun phrase) and primarily operate with lexically-specific constructions that they have learned (e.g., *It's Verb-ing*). This view is supported by findings from corpus studies (Joseph, et al., 2001; Wilson, 2003) that young children tend to use tense and agreement morphemes (e.g., auxiliary *'is'*) more accurately when the sentence contains certain classes of lexical items, like pronominal subjects (e.g., *He*), rather than lexical NP subjects (e.g., *The dog*).

However, differences in percent correct of auxiliary *'is'* between pronominal and lexical subjects could actually be attributed to increased processing demands rather than the lack of an abstract representation (Bloom, 1990). Lexical subjects tend to be longer and have more complex phrase structure than pronominal subjects (e.g., [$_{DP}$ [$_{D}$ *The*] [$_{NP}$ *dog*]] and [$_{DP}$ [$_{D}$ *He*]]; where DP refers to determiner phrase). Processing sentences with lexical subjects may make more processing demands than those with pronominal subjects, leading to higher omission rates of auxiliary *'is'* with the former. Thus, the abstractness

of representations of tense and agreement morphemes in young children remains an open question. The structural priming paradigm allows us to address this question by assessing the abstractness of representations of tense and agreement morphemes.

The Structural Priming Paradigm

The priming paradigm has been widely used to explore the nature of cognitive and linguistic representations (Bock & Griffin, 2000; Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995; Shimpi et al., 2008). Structural (or syntactic) priming refers to a phenomenon in which a preceding sentence of a particular form (e.g., the transitive construction) increases the likelihood of producing another sentence with the same form even when these sentences have virtually no words or morphemes in common (Huttenlocher, Vasilyeva, & Shimpi, 2004; Miller & Deevy, 2006; Shimpi et al., 2007). In a typical structural priming paradigm, the participant first listens to and repeats the prime (e.g., *The dog is biting a ball*). Then, he/she is asked to describe a picture that is compatible with more than one syntactic form (e.g., *The cat is eating a cookie* or *The cat is eating*). The participant is more likely to describe the picture with a transitive sentence if he/she is primed with the same structure. If the prime is related to the alternative structure, the participant is likely to produce the alterative structure.

The relationship between the prime sentence and the participant's response allows the experimenter to infer whether the person has an abstract representation of a particular syntactic structure (Branigan et al., 1995; Savage, Lieven, Theakston, Tomasello, 2003). If a priming effect occurs, that is, a prime with a particular structure increases the rate of producing the same structure in the target response though the prime and target do not share lexical items in common, we can infer that the two syntactic forms (i.e., prime-target) are related to each other within psychological reality, implying an abstract representation of the syntactic form under examination.

Critical to structural priming is that there is no overlap of lexical items between the prime and target. Nevertheless, the priming effect is stronger when the prime and target share specific lexical items (e.g., lexical verbs) than when they do not (Pickering & Branigan, 1998), suggesting that some of the observed facilitation effects could result from lexical overlap between the prime and target. Based on this finding, recent studies (Savage et al., 2003; Savage, Lieven, Theakston, & Tomasello, 2006) have adapted the structural priming paradigm by having some lexical overlap between the prime and target, such as the subject pronoun (e.g., Prime: *It's running*. Target: *It's jumping*). This adaptation is called lexically-specific priming and has been used to investigate the nature of syntactic representations in young children. Presumably, if the likelihood of producing the syntactic construction of interest increases with lexically-specific priming, but not with structural priming, as compared to a baseline or no priming condition, the child's representation of that construction is likely to contain specific lexical items (e.g., *It's Verb-ing*) and has not yet reached the adult-like level of abstractness (e.g., *NP BE Verb-ing*) that would be revealed by structural priming.

Structural Priming and Sentence Constructions

Savage and colleagues (2003) explored the representation of active/passive constructions in three-, four-, and six-year-olds via lexically-specific and structural primes. Six-year-olds increased the use of active or passive constructions after the lexically-specific (e.g., prime: *It got pushed by it*; target: *It got kicked by it*) and structural primes (e.g., prime: *The bricks got pushed by the digger*; target: *The cat got kicked by the mouse*) whereas three- and four-year-olds increased the use of these constructions only after the lexically-specific primes. Thus, Savage and colleagues suggested that six-year-olds have abstract representations of active/passive constructions whereas three- and four-year-olds only have lexically-specific representations of these constructions (e.g., *It is Verb-ing it*, *It got Verb-ed by it*). In a subsequent study, Savage and colleagues (2006)

further confirmed that lexically-specific primes enhanced the production of passive sentences in 4-year-olds. In addition, the priming effect still occurred a month after the priming event. This is consistent with work on adults (Bock, Dell, Chang, & Onishi, 2007; Bock & Griffin, 2000), suggesting that priming can also function as linguistic input and strengthen the underlying linguistic representations to facilitate the production of target constructions.

Huttenlocher, Vasilyeva, and Shimpi (2004) extended the work of Savage and colleagues (2003) to other sentence constructions by examining the use of transitive (active/passive) and dative constructions (double objects: e.g., *The boy is feeding the dog a bone* /preposition: e.g., *The boy is feeding a bone to the dog*) in children between 4;1 and 5;8. As a whole, children within this age range increased their use of transitive or dative constructions after they repeated related prime sentences. Following a similar experimental design, Shimpi and colleagues (2007) further examined priming effects in three- and four-year-olds. Four-year-olds were able to be primed to produce transitive or dative constructions after either listening to or repeating related prime sentences. However, three-year-olds could only be primed after they repeated the prime sentences. Shimpi and colleagues concluded that three-year-olds possess abstract syntactic representations of transitive and dative constructions but access to these representations is affected by the task demands. Repeating the prime reduced task demands (e.g., attention, memory) and helped young children access the abstract syntactic representations required for the target sentences.

The studies reviewed above consistently show that children at the age of five or older possess abstract representations of sentence constructions. However, the results for three- and four-year-olds are mixed. The discrepant findings between studies could result from differences in task demands. Children as young as three years of age could have abstract representations of sentence constructions, but these representations may not be readily accessible for language production due to processing factors (Shimpi et al., 2007).

Thus, priming effects would be relatively observable for three-year-olds if we could minimize processing demands, thereby facilitating these children's access to the syntactic representations under examination.

Structural Priming and Tense and Agreement Morphemes

In addition to examining the representations of sentence constructions, the structural priming paradigm has also been employed to examine the use of tense and agreement morphemes. Leonard, Miller, Deevy, et al. (2002) investigated the effect of repeating priming sentences that included auxiliary '*are*,' past tense *–ed*, or a nonfinite structure on the use of auxiliary '*is'* in the target sentence. The participants were children with specific language impairment (SLI group) between 4;3 and 6;10 and their younger peers matched on finite verb morphology composite scores (ND group), ranging in age from 2;8 to 4;0. The SLI and the ND groups were more likely to use auxiliary '*is'* in the target sentence (e.g., *The goose is chasing the cat*) after they repeated prime sentences with auxiliary '*are'* (e.g., *The dogs are chewing the sock*) than after they repeated those with past tense (e.g., *The doctor smiled*). Use of auxiliary *'is'* was least likely when the prime involved a nonfinite structure (e.g., *We see the horse eating the grass*). Leonard, Miller, Deevy, et al. (2002) suggested that auxiliary '*are'* increased the use of auxiliary *'is'* because the prime and the target sentences in this case shared the same syntactic, and perhaps prosodic, structures. The structures needed for the target sentence had already been activated after the child repeated the prime, and was thus more accessible for the target sentence. This reduced the processing demands and allowed the child to allocate more resources to retrieve auxiliary *'is.'* The nonfinite present progressive was least likely to facilitate the use of auxiliary *'is'* because it might have activated a sentence frame without any finite markers.

Although the focus of the studies was on children with SLI, the findings of Leonard and colleagues (Leonard, Miller, Grela, et al., 2000; Leonard, Miller, Deevy, et

al., 2002) may also offer evidence for the representation of tense and agreement morphemes in young typically-developing children. One possibility is that children within the age range under examination have fully abstract representations of auxiliary BE so that repeating prime sentences with auxiliary *'are'* facilitates the production of auxiliary '*is*' in the target sentence. However, because the typically-developing children were selected to match children with SLI on a language measure, the age range of the typically-developing children was wide and the results cannot be generalized to younger children unambiguously. Four-year-olds may have different representations of auxiliary BE from three-year-olds (Savage et al., 2006). It is possible that the priming effect of auxiliary *'are'* occurred mostly in children who were close to four years old but not in those close to three. Because of this possibility, we cannot be sure that a typically-developing three-year-old has an abstract representation of auxiliary BE.

To better understand the nature of representations of auxiliary BE in children close to three years old, Theakston and Lieven (2008) adopted a structural priming paradigm to investigate the production of auxiliary BE after questions or declaratives in children between the ages of 2;5 and 2;10. It was hypothesized that children may omit auxiliary BE more often after they hear questions as primes (e.g., *Are* <u>*the giraffes flying*</u>) than declaratives (e.g., *The giraffes are flying*) because the word order in questions may prime children to use the ungrammatical constructions that they have learned, such as **NP Verb-ing*. As predicted, children produced auxiliary BE more often with prime declaratives than with prime questions. However, because there was no baseline condition for comparison, it was not clear whether declaratives had a facilitative effect, questions had a detrimental effect, or both. In addition, because children produced/omitted auxiliary BE in target sentences throughout the experiment, it was not clear whether the use of auxiliary BE was actually influenced by the prime sentences from the examiner, by the prior responses from the children themselves, or both.

To answer these questions, Theakston and Lieven (2008) further examined how the previous five utterances influenced the use of auxiliary BE in the target utterance in longitudinal language samples collected from a child between the ages of 2;8 to 3;2. As compared to the baseline (i.e., no prior context), prior parental questions and prior declaratives in which the child omitted auxiliary BE had negative effects on the child's use of auxiliary BE in subsequent declaratives. In addition, the prior use of auxiliary BE in the child's declaratives, but not in the parental input, facilitated the production of auxiliary BE in the subsequent declaratives. This facilitative effect was restricted to the target form of BE only and did not extend to different forms of BE. The effect was most prominent when the prior and subsequent declaratives shared identical sentence subjects and BE forms. Theakston and Lieven (2008) thus suggested that the representations of auxiliary BE may be lexically-specific in three-year-olds, given that the priming effects occurred only when the child's prior and subsequent declaratives had the same specific <u>subject + auxiliary BE</u> combination.

Taken together, the studies of Leonard and colleagues (Leonard, Miller, Grela, et al., 2000; Leonard, Miller, Deevy, et al., 2002) and Theakston and Lieven (2008) indicate a developmental change such that four-year-olds have abstract representations of auxiliary BE while three-year-olds may possess only lexically-specific representations. These findings seem to support the assumption of the usage-based approach that young children may not have abstract but lexically-specific constructions.

However, the suggestion that three-year-olds may only possess lexically-specific representations of auxiliary BE needs more verification for two reasons. First, the observation that the child's prior use of different forms of BE did not promote the use of BE in subsequent declaratives was based on only seven obligatory contexts in the language samples from 2;8 to 2;10, which is not sufficient for reliable conclusions. At the same time, the child's prior use of different forms of BE did have similar, although lower, facilitative effects as the same form of BE in the language samples from 2;11 to

3;2. Second, the absence of a priming effect from the parent's productions may result from increased processing demands rather than a lack of abstract representations. Recall that three-year-olds in the study of Shimpi and colleagues (2007) were able to be primed to produce target sentences only when children repeated the primes. It is possible that repeating the prime helps children to retain the primed structure to a greater degree than simply listening to the prime. Because the child in the study of Theakston and Lieven (2008) typically did not repeat the parental utterances with auxiliary BE and there were intervening utterances between the parental utterance with BE and the child's subsequent use of BE, the priming effects may have been compromised (Shimpi et al., 2007).

Research Questions and Predictions

Given the concerns of wide age range of participants, insufficient number of obligatory contexts for target morphemes, and changing task demands, we cannot conclude whether three-year-olds have abstract representations of auxiliary BE. To address these concerns, this study adopted a structural priming paradigm to examine the abstractness of the representation of auxiliary BE in three-year-olds. This paradigm allows us to elicit a sufficient number of obligatory contexts to reliably evaluate the child's use of auxiliary BE under different priming conditions. In addition, to eliminate processing factors that could compromise the priming effects, we asked children to repeat the prime sentences in the experiment. The primes were limited to declaratives or lists of letters, numbers, or words. The letter/number/word lists, rather than questions or nonfinite structures, were used in the baseline condition because questions and nonfinite structures may actively interfere with access to the representation of auxiliary BE and thus decrease the use of auxiliary BE, rather than serving as a neutral baseline (Leonard, Miller, Deevy, et al., 2002; Theakston & Lieven, 2008).

Specifically, we examined whether the likelihood of using auxiliary *'is'* and *'are'* would change in three priming conditions: no priming condition (e.g., prime: *One, two,*

three, four; target: *The frog's flying the kite*), lexically-specific priming condition (e.g., prime: *The bear's eating the pizza*; target: *The bear's flying the kite*), and structural priming condition (e.g., prime: *The deer's chasing the cow*; target: *The cats are fixing the car*). The no priming condition served as a baseline to measure the priming effects in the other conditions.

We computed the likelihood of producing auxiliary *'is'* and *'are'* separately for each condition because recent studies indicated that children may learn these auxiliary BE forms individually (Joseph et al., 2001; Wilson, 2003). It is possible that three-year-olds have learned a variety of lexically-specific constructions with auxiliary *'is'* but only a few lexically-specific constructions with auxiliary *'are.'* The facilitative effect of the lexically-specific priming condition may only be observable for auxiliary *'is'* but not for auxiliary *'are'* because it is more likely to have specific subject + auxiliary *'is'* combinations in the prime that match the child's lexically-specific representations than specific subject + auxiliary *'are'* combinations. Combining auxiliary *'is'* and *'are'* in the computation would obscure the facilitative effect of lexically-specific priming condition for auxiliary *'is.'* At the same time, if the priming effects of lexically-specific and/or structural priming conditions did occur, separating these morphemes in the analysis would allow for a more precise inspection of these priming effects.

The research question addressed is: Does the likelihood of using auxiliary *'is'* and/or *'are'* in three-year-olds vary in different priming conditions? If three-year-olds have an abstract representation of auxiliary BE, then both the lexically-specific priming and structural priming conditions should enhance the child's production of auxiliary *'is'* and '*are'* above baseline (i.e., the no priming condition) because they both prime the child with abstract linguistic structures used in the target sentence. The likelihood of using auxiliary *'is'* and *'are'* should be higher in the lexically-specific priming than in the structural priming condition (i.e., lexically-specific priming > structural priming > no priming) because the former primes the child not only with the structure but also with the

specific lexical items (i.e., subject and auxiliary BE form) in the target, which reduces processing demands in language production (Pickering & Branigan, 1998).

In contrast, if three-year-olds only have lexically-specific representations of auxiliary BE, only the lexically-specific priming condition would significantly enhance the production of auxiliary *'is'* and/or *'are'* above the baseline. The structural priming condition will not have a strong priming effect (i.e., lexically-specific priming > structural priming = no priming). The structural priming condition should not be an effective prime if the child does not have an abstract representation of auxiliary BE and hence cannot see the relation between prime-target pairs that share the linguistic structure in common but differ in lexical items.

Method

Participants

Twenty-one children with an age range of 2;7 (year;month) to 3;5 were recruited to examine the priming effects on the production of auxiliary *'is'* and *'are,'* and data from 13 children were retained for analysis. Ten of the 13 children also participated in Experiment 1 and nine of the 13 children participated in Experiment 2. Children in this age range were recruited because they were likely to be producing auxiliary *'is'* and *'are'* inconsistently (Rice, Wexler, & Cleave, 1995; Wilson, 2003), allowing us to evaluate the priming effects across different conditions. To be retained for analysis, each child had to perform below ceiling (i.e., 90% correct, Brown, 1973) on the use of auxiliary *'is'* and/or *'are'* for one or more priming conditions so that the priming effect could be legitimately examined. In addition, the participants had to produce at least 3 scorable items for each condition and produce at least one correct usage of auxiliary *'is'* and/or *'are'* in the task. This last requirement was to ensure that participants had auxiliary *'is'* and/or *'are'* in their productive lexical inventory (Schütze, 2001). Eight of the 21 children recruited were excluded for analysis because they had reached ceiling levels of using both

auxiliaries *'is'* and *'are'* (5 children), or did not produce at least one correct use of auxiliaries *'is'* or *'are'* (3 children) in the task.

The 13 children (7 girls) included in the analysis had a mean age of 3;0 (ranging from 2;7 to 3;4). Among these 13 children, nine children were included for both the analyses of auxiliaries *'is'* and *'are*' because they met all the criteria mentioned above; three children were included for the analysis of auxiliary *'is'* only because they did not produce at least one correct use of auxiliary *'are*;' and one child was included for the analysis of auxiliary *'are'* only because he had reached ceiling level of using auxiliary *'is*.' In total, there were 12 children included for auxiliary *'is'* and 10 children for auxiliary *'are*.'

All of these 13 children were typically-developing monolingual English speakers as documented by parent report and performance above the 10th percentile on a norm-referenced standardized test. The standardized language tests were the *Preschool Language Scale-Fourth Edition* (*PLS-4*, Zimmerman, Steiner, & Pond, 2002) for children who were younger than 3;0 and the *Structured Photographic Expressive Language Test-Preschool* (*SPELT-P*, Werner & Kresheck, 1983) for those who were 3;0 or older. We administered the SPELT-P for those who were 3;0 or older because we only tested three-year-olds at the beginning stage of this study. The PLS-4 was added when the age range was extended to 2;7. As a component of assessment, we also collected a conversational language sample for each child. All children demonstrated an MLU within the typical range with reference to the norms of Miller (1981). They also had hearing acuity within normal limits per ASHA standards (American Speech-Language-Hearing Association, 1997) and did not have a history of cognitive or motor problems.

Stimuli

The experiment consisted of three conditions: a no priming condition, a structural priming condition, and a lexically-specific priming condition. Prime-target pairs were

created for each condition. In each condition, the target sentences were sixteen simple, transitive, present progressive declarative sentences, of which half involved obligatory use of auxiliary '*is'* (e.g., *The frog's flying the kite*) and the other half auxiliary '*are'* (e.g., *The lions are sweeping the floor*). A set of 16 lexical NP subjects and 16 verbs were used in the target sentences across conditions to control the frequency of subjects and verbs between conditions. Each lexical NP subject co-occurred with a different verb in each of the conditions. The frequency of object NPs and the sentence length (i.e., number of syllables) of the target sentences were also controlled across conditions. One-way ANOVAs showed that target sentences across conditions did not differ significantly in the frequency of object NPs, F (2, 45) = .23, p = .79, η_p^2 = 0.01; or in sentence length, F (2, 45) = .04, p = .96, η_p^2 = 0.002.

The key difference in the three priming conditions was the relationship between the prime and target sentences. In the **no priming condition**, the primes were utterances involving counting or naming a list of objects and did not include the use of auxiliary BE. For instance, the utterance *One, two, three, four* was used to prime the target *The frog's flying the kite.* This condition served as baseline against which we can measure the priming effect of the other priming conditions (Shimpi et al., 2007). Presumably, repeating words, letters, or numbers maintains a similar task procedure but neither facilitates nor inhibits any sentence structure. In the **structural priming condition**, prime sentences contained different subjects and auxiliary BE forms than the target sentences to test whether children had abstracted away from the lexically-specific representation of the auxiliary BE. For example, the sentence *The sheep's drawing the fish* was used to prime *The dogs are pushing the brain*. In the **lexically-specific priming condition**, prime sentences contained both the subject NP and the auxiliary BE form identical to target sentences. For example, the sentence *The bear's eating the pizza* was used to prime *The bear's flying the kite*. Both the subject and the auxiliary BE form, instead of just the auxiliary BE form, overlapped in the prime and target sentences

because recent studies suggest that the representations of auxiliary BE in young children are likely to be specific subject + auxiliary BE combinations, such as *He's Verb-ing, The dog's Verb-ing* (Joseph et al., 2001; Pine et al., 2008; Theakston & Lieven, 2008). It is assumed that the likelihood of using auxiliary BE in the target would increase because the specific subject + auxiliary BE combinations in the prime would naturally match the child's lexically-specific representations. This is the most facilitative condition with the lowest processing load. The prime-target pairs in each condition are listed in Appendix E.

Procedure

All target-prime pairs were divided into three sets, which were blocked by conditions. That is, one set had no-priming utterances; the other, lexically-specific priming sentences; and still another, structural-priming sentences. Within each set, the order of prime-target pairs were randomized so that production of targets with auxiliary *'is'* or *'are'* were intermixed. Each condition was administered on a separate day to reduce the interference between prime types and decrease fatigue (Bock & Griffin, 2000; Bock et al., 2007; Savage et al., 2006). The order of presentation of the three sets of priming sentences was counterbalanced across children who were tested to eliminate order effects. However, because we excluded some children from analyses afterwards, the order of priming conditions was not completely counterbalanced and thus we considered the effect of order in the statistical analysis later.

Each child was tested individually by a trained examiner. The examiners were four undergraduate students at the University of Iowa, who were all monolingual English speakers. Each session began with a warm-up comprehension activity to familiarize the child with subject NPs that were used in the experiment. The child was asked to identify the target from four pictures in each trial after hearing the examiner say "Show me the _____" (e.g., "Show me the bear"). The child's performance did not affect being

included or excluded for analysis. If the child made errors on a given trial, the examiner named all four pictures and asked the child to point to the target picture again.

After the warm-up activity, the priming task began. Five practice prime-target pairs were offered to familiarize the child with the task. The child was invited to play in a game in which he needed to repeat exactly what the examiner said first and then tell the examiner about a picture (Miller & Deevy, 2006). In order to ensure that priming has occurred (Leonard, Miller, Deevy, et al., 2002) and to reduce the task demands (Shimpi et al., 2007), the child was asked to repeat the prime sentence.

Each trial began with the examiner reading the prime sentence. If the child did not imitate the prime, the examiner said "Say exactly what I say" or "Let's play copy cat" and then read the prime again. It should be noted that some children did not always repeat the auxiliary BE form in the prime sentence accurately (e.g., omission errors). We did not ask them to repeat the prime again when they made errors so that all children would repeat the prime sentence just once for each item. The accuracy of repeating prime sentences will be considered in the statistical analysis.

After the child repeated the prime, the examiner presented the target picture, said "Tell me about this picture," and waited for a response from the child. If the child did not respond at all, the examiner prompted the child with the subject NP by saying "Tell me about the <u>subject NP</u>" (e.g., "Tell me about <u>the cats</u>"; Target: *The cats are fixing the car*). If the child responded with a pronominal subject and contracted auxiliary BE (e.g., *They're fixing the car*), the examiner prompted the child by saying "Can you start by say the <u>subject NP</u>." This decision was made because those <u>pronoun + contracted BE</u> combinations could be unanalyzed chunks and could be produced without activating the syntactic representations related to tense and agreement morphemes (Rispoli, Hadley, & Holt, 2009). To avoid overestimating the child's abstractness of representations of these morphemes, we tried to eliminate these unanalyzed chunks from the responses. The examiner maximally used three additional prompts for each item.

Transcription and Coding

All responses to the primes and targets were audio-recorded for later analysis. The child's responses to the targets were transcribed by the examiner and coded as correct, incorrect, or unscorable, based on the use of auxiliary *'is'* or *'are.'* The rules of coding are described below.

1. **Correct Production** (Auxiliary *'is'*: 161 responses; Auxiliary *'are'*: 93 responses)

Correct productions included the responses that matched the target sentence (e.g., *The ducks are kicking the ball.*) or those with alternative, missing, and/or additional elements unrelated to auxiliary BE (e.g., *The <u>birds</u> are kicking <u>ball</u> <u>like that</u>*). However, if the subject NP was changed in the lexically-specific priming condition, we recoded this as a target response as for the structural priming condition because this alternative response had the same structure but a different subject NP from the prime sentence.

2. **Incorrect Production**
   A. Omission of auxiliary BE (*'is'*: 112 responses; *'are'*: 135 responses): Auxiliary BE was omitted (e.g., **The ducks biting the gloves*).
   B. Agreement error (*'is'*: 0 responses; *'are'*: 5 responses): A wrong form of auxiliary BE was used (e.g., **The ducks <u>is</u> biting the gloves*).
3. **Unscorable Production**
   A. Use of an unrelated structure (*'is'*: 6 responses; *'are'*: 2 responses): A grammatical but non-target sentence was produced (e.g., *The cats bite the gloves*, *The cats are happy*).
   B. Use of a verb phrase (*'is'*: 7 responses; *'are'*: 8 responses): Both the subject and auxiliary BE were omitted (e.g., *Kicking the ball*).
   C. No response or unintelligible response (*'is'*: 2 responses; *'are'*: 2 responses): The child did not respond or responded with an unintelligible utterance.

The child's repetitions of the prime sentences were also transcribed. Recall that the prime sentences in the structural and lexically-specific priming conditions also contained auxiliary *'is'* or *'are.'* To check whether priming had occurred, we also coded whether the repetition of auxiliary *'is'* or *'are'* in the prime was accurate for the scorable (i.e., correct and incorrect) target responses in these conditions based on the rules above. Seven children produced errors of repeating auxiliary *'is*;' 6 children, auxiliary *'are*.' In the structural priming condition, 64 out of 93 primes with auxiliary *'is'* and 56 out of 76 primes with auxiliary *'are'* were repeated correctly. In the lexically-specific priming condition, 69 out of 92 primes with auxiliary *'is'* and 56 out of 73 primes with auxiliary *'are'* were repeated correctly. All of the primes that were not repeated correctly had omission errors, except 1 repetition with an unrelated structure (e.g., a verb phrase) for auxiliary *'is'* and 2 for auxiliary *'are.'*

Given that some of the primes in the structural and lexically-specific priming conditions were not repeated correctly, we could not determine whether priming had occurred in these items. It might be interesting to investigate whether primes repeated without auxiliary BE had a detrimental effect on the targets. However, not enough children produced at least three primes with omitted *'is'* or *'are,'* so we were not able to reliably examine the effect of these incorrect primes. Thus, we excluded those scorable target responses that were paired with incorrect or unrelated repetitions of primes for further analysis. After we excluded all these responses, 3 more children did not meet the criterion of having at least three scorable target responses in the structural or lexically-specific priming conditions for auxiliary *'is'*; 2 more children did not meet this criterion for auxiliary *'are.'* These children were therefore excluded. In total, there were 9 children left for the statistical analysis of auxiliary *'is'* and 8 children for auxiliary *'are*.'

Table 4.1 summarizes the mean number of scorable responses that were paired with correctly repeated primes by auxiliary BE form and condition.

Table 4.1. Mean Number of Target Sentences and Number of Scorable Targets with Correctly Repeated Primes by Auxiliary BE Form and Priming Condition

	N of Targets	Mean	SD	Range
Auxiliary *'is'* (9 children)				
No priming condition	8	7.56	0.53	7-8
Structural priming condition	8	5.89	1.54	3-8
Lexically-specific priming condition	8	7.89	1.83	4-10
Auxiliary *'are'* (8 children)				
No priming condition	8	7.88	0.35	7-8
Structural priming condition	8	7.38	1.30	5-9
Lexically-specific priming condition	8	5.88	1.96	3-8

Note: The maximum number of scorable items of the structural priming condition was greater than 8 because some children used alternative subject NPs in the targets of lexically-specific priming condition and these sentences were recoded as target responses for the structural priming condition.

Reliability

To check the transcription and coding reliability in this study, we randomly sampled 20% of the participants (n = 3). These 3 samples were re-transcribed and re-coded by the first author. The mean reliability was 94.78% in transcribing content words in the sentences, 98.78% in transcribing grammatical morphemes, and 94.69% in coding accuracy of using auxiliary BE.

Statistical Analysis

The dependent measure was the accuracy of the auxiliary for each scorable response in which the child had accurately repeated the prime. Because of the dichotomous nature of the dependent measure, we adopted a multi-level logistic regression model to examine the effect of priming conditions on the production accuracy of auxiliary *'is'* and *'are.'* Similar to linear regression, logistic regression relates one or

more predictor variables to the dependent measure. When the predictor variable is categorical (e.g., priming conditions), one level of the categorical variable is selected as the reference category/condition, as (1) exemplifies (where $PRIME_{STRUCTURAL}$ refers to the structural priming condition and $PRIME_{LEXICAL}$ refers to the lexically-specific priming condition).

$$(1)\ \text{Logit}\ (Y_i) = \beta_0 + \beta_1 PRIME_{STRUCTURAL} + \beta_2 PRIME_{LEXICAL}$$

The model in (1) predicts the effect of priming conditions on the production accuracy of the auxiliary and the no-priming condition was selected as the reference condition. The β-values generated from the model can be used to compute the logit associated with each condition. The logit is the natural logarithm of predicted odds for a given event, such as the predicted log odds of producing auxiliary *'is'* accurately in the lexically-specific priming condition.

The β-coefficient associated with each term (except the intercept term) in the model also reflects the log odds ratios between the condition of interest and the reference condition. For instance, β_1 in (1) reflects the log odds ratio of producing auxiliary *'is'* correctly when the lexically-specific priming condition is compared to the no-priming condition. A positive β-coefficient indicates an increased likelihood of producing auxiliary *'is'* correctly in the condition of interest (e.g., the lexically-specific priming condition) as compared to the reference condition (e.g., the no priming condition). The *p*-value of the β-coefficient reveals whether the log odds ratio between the condition of interest and the reference condition is significantly different from 0. For instance, if β_1 in (1) is positive and reaches a significant level, this suggests that children are more likely to produce auxiliary *'is'* correctly in the lexically-specific priming condition than in the no

priming condition. Log odds ratios can be transformed into odds ratios (OR) for the ease of interpretation.[5]

Results

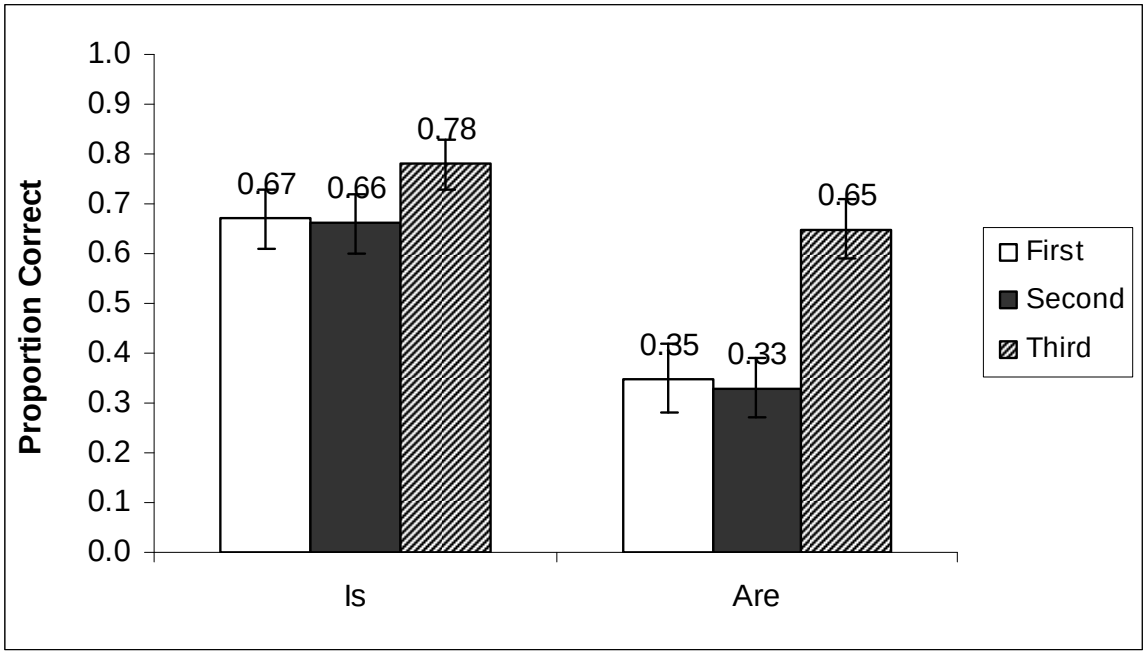
Effect of Test Order

Because some children were excluded from analysis after we finished data collection, the order of the priming conditions was not evenly counterbalanced. Figure 4.1 shows the actual percent correct of target sentences with auxiliary *'is'* or *'are'* across children by test order. To explore whether the order of the priming conditions would affect the accuracy of using the auxiliary, two mixed model binominal logistic regressions were performed for auxiliary *'is'* and *'are'* respectively, with item and child treated as random factors, and test order as a fixed factor. The models indicated that, regardless of priming conditions, children produced the auxiliary more accurately in the third block than in the first (OR = 3.29, $p = 0.02$ for *'is'*; OR = 7.43, $p < 0.001$ for *'are'*) or second blocks (OR = 2.61, $p = 0.07$ for *'is'*; OR = 10.04, $p < 0.001$ for *'are'*). The likelihood of producing the auxiliary correctly did not differ significantly between the first and the second blocks (OR = 1.26, $p = 0.64$ for *'is'*; OR = 0.74, $p = 0.59$ for *'are'*).

---

[5] OR larger than 1 means that children are **more** likely to produce auxiliary *'is'* (or *'are'*) in the condition of interest than in the reference condition. In contrast, OR smaller than 1 means that children are **less** likely to produce auxiliary *'is'* (or *'are'*) in the condition of interest than in the reference condition. For instance, if the β-coefficient associated with the lexically-specific priming condition is 1.2804, this means that the log odds ratio of producing auxiliary *'is'* correctly between the lexically-specific priming condition and the reference condition (i.e., no priming condition) is 1.2084 (OR = exp(1.2084) = 3.59). It means that children are 3.59 times more likely to use auxiliary *'is'* correctly in the lexically-specific priming condition than in the no priming condition.

Figure 4.1. Mean Proportion Correct of Target Sentence by Auxiliary Form and Test Order



Note: The error bar is standard error.

Given that this order effect may have influenced the effect of priming conditions, we added test order in each of the mixed model binomial logistic regressions reported below. The interaction terms with test orders did not significantly improve model fit (*p*s > 0.18). For this reason, test order was treated as a covariate in each of the models reported below (Kleinbaum, Kupper, Muller, & Nizam, 1998) and its interaction with priming conditions was dropped from all models.

Effect of Priming Condition

To explore the effect of priming conditions on the production accuracy of auxiliary *'is,'* we added the factor of priming conditions into the model, with child and item treated as random factors, test order treated as a covariate, and priming condition treated as a fixed factor (i.e., full model). We compared the log-likelihood values of the

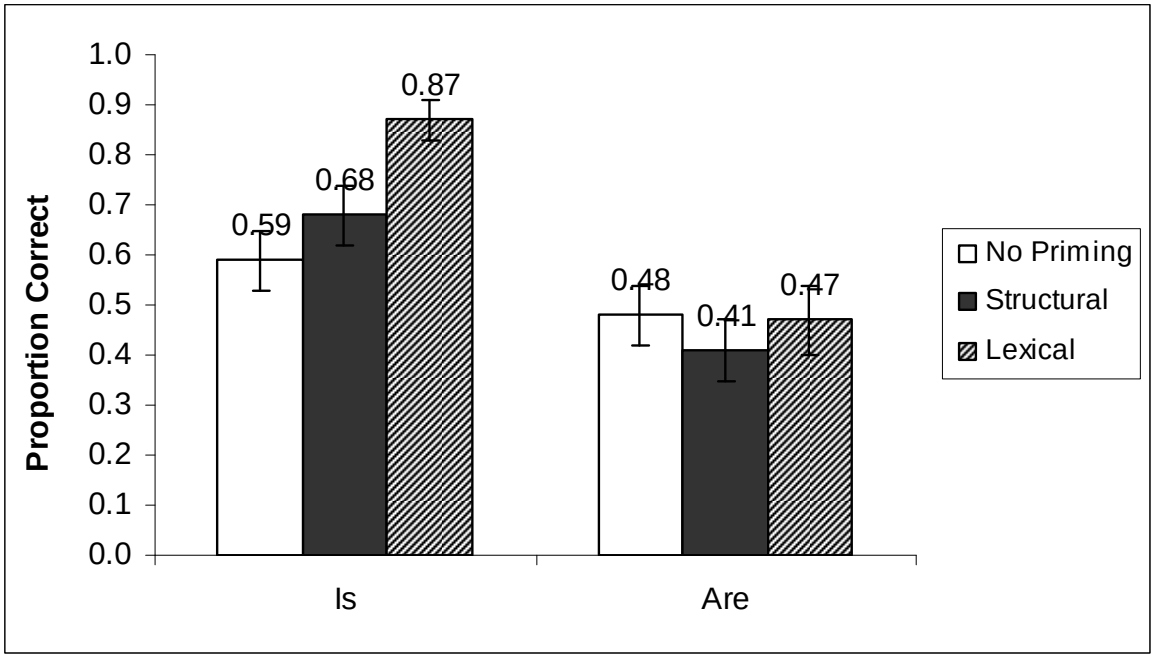
full model with the model with test order alone (i.e., the reduced model) using a chi-square test for goodness of fit (Kleinbaum et al., 1998). The test showed that the model with priming condition and test order significantly improved model fit as compared to the model with test order only ($\chi^2 = 11.99$, $df = 2$, $p = 0.002$), indicating that priming conditions were a significant predictor of the production accuracy of auxiliary *'is'* over and above test order alone. Table 4.2 presents the results from the full model and Figure 4.2 presents actual mean proportion correct of target sentences by priming conditions for auxiliary *'is'* and *'are.'*

Table 4.2. A Regression Model Showing the Likelihood of Producing Auxiliary *'is'* Correctly by Priming Condition

	Variance	SD		
Random Factors				
Item	0.00	0.00		
Child	2.94	1.72		
	Coefficient	**SE**	**OR**	**p-value**
Fixed Factors				
Intercept	2.24	0.86	9.39	0.01
Block (reference condition = first block)				
Second Block	0.26	0.56	1.30	0.63
Third Block	0.94	0.54	2.56	0.08
Priming Condition (reference condition = no priming condition)				
Structural Priming Condition	0.59	0.48	1.80	0.21
Lexically-specific Priming Condition	2.16	0.62	8.67	<0.001

Note: SD = standard deviation; SE = standard error; OR = odds ratios

Figure 4.2. Mean Proportion Correct of Target Sentences by Auxiliary Form and Priming Condition



After test order was controlled, children were more likely to use auxiliary *'is'* correctly in the lexically-specific priming condition than in the no priming condition (OR = 8.67, $p < 0.001$). The likelihood of producing auxiliary *'is'* correctly did not differ significantly between the structural priming and the no priming conditions (OR = 1.80, $p = 0.21$). However, this full model did not directly test the contrast between the structural and lexically-specific priming conditions. To examine the difference between the structural priming and lexically-specific priming conditions, we re-ran the model with the structural priming condition as the reference condition (Jaccard, 2001). The model indicated that children produced auxiliary *'is'* more accurately in the lexically-specific priming than in the structural priming conditions (OR = 4.83, $p = 0.01$). That is, children were more likely to use auxiliary *'is'* accurately in the lexically-specific priming

condition than in the no priming or structural priming conditions, which did not differ from each other.

The effect of priming conditions on the production of auxiliary *'are'* was examined next. The full and reduced models were identical to those used for the analysis for auxiliary *'is.'* The goodness of fit test showed that the model with priming condition and test order did not significantly improve model fit as compared to the model with test order only ($\chi^2 = 0.38$, $df = 2$, $p = 0.83$), meaning that priming conditions were not a significant predictor of production accuracy of auxiliary *'are'* over and above test order alone, or that neither prime condition promoted the production accuracy of auxiliary *'are'* above the baseline.

Discussion

This study examined the abstractness of the representation of auxiliary BE in three-year-olds by using a structural priming paradigm. In order to reduce processing demands (Shimpi et al., 2007) and ensure that priming had occurred, we asked children to repeat the primes and only the target responses paired with correctly repeated primes were included for analysis. For the target sentences containing *'is,'* children were more likely to produce the auxiliary accurately in the lexically-specific priming condition than in the no priming or the structural priming conditions after test orders were controlled. The likelihood of producing auxiliary *'is'* accurately did not increase significantly above the baseline in the structural priming condition. For target sentences with *'are*,' the likelihood of producing the auxiliary correctly did not vary across the three priming conditions. In other words, children do not seem to have an abstract representation of auxiliary BE, but do seem to have lexically-specific representations of auxiliary *'is'* and *'are'* although with different degrees of strength. Regardless of priming conditions or the target auxiliary involved, children tended to produce the auxiliary more accurately in the

third block than in the first or second blocks, suggesting that participation improved performance.

Limitations of the Current Study

Before we discuss the results in detail, we should consider two limitations of this study. First, this study had a limited number of children in the analysis, limiting both statistical power and generalizability. This limited number resulted from excluding children who had reached a ceiling level of using auxiliary BE, did not produce the target auxiliary BE form, or did not produce at least three scorable target items. These children were excluded because ceiling or floor performance limits our ability to legitimately examine priming. For instance, when a child produced with 100% correct for each condition, we were unable to determine whether the structural and lexically-specific priming conditions had facilitated the production by comparing these conditions with the baseline condition. On the other hand, when the child did not produce at least one correct target BE form, we were not sure whether this floor effects result from the lack of representation of BE or the lack of the target BE form in his lexical inventory (Schütze, 2001). Nevertheless, even with a limited number of children, we were still able to find the facilitative effect of the lexically-specific priming condition on the use of auxiliary *'is,'* indicating that statistical power was sufficient for determining differences between two extreme conditions. However, inspection of the results suggests that the effect of structural priming for auxiliary *'is'* may have been observed with more children. In contrast to *'is*,' the null finding for *'are'* appears legitimate because the production accuracy of the auxiliary across the three priming conditions was close to each other. Given that the null results in this study generally did not result from limited statistical power, we can infer that our results are generalizable to a wider population as well.

The second concern is that the test order of priming conditions was not evenly counterbalanced after we excluded some children from the statistical analysis. Given that

test order could confound the effect of priming conditions (Feldt, 1993), we attempted to address this issue by treating test order as a covariate in the analysis. Thus, test order was statistically controlled when we examined the effect of priming conditions on the production of auxiliary BE.

The Lexically-specific Priming Condition Facilitated the Production of Auxiliary *'is'* but not Auxiliary *'are'*

The current findings confirmed the prediction that only the lexically-specific priming condition, but not the structural priming condition, would enhance the production of auxiliary BE above the no priming condition in young children. Children were more likely to use auxiliary *'is'* accurately in the targets (e.g., *The bird is watching TV*) after repeating the primes that had the same sentence structure and specific subject + auxiliary *'is'* combinations (e.g., Prime: *The bird is chasing the worm*) in comparison to a baseline condition (Prime: e.g., *One, two, three, four*). The likelihood of producing auxiliary *'is'* accurately in the targets did not increase significantly above the baseline when children repeated the primes that only shared the same sentence structure but did not contain lexical overlap with the target (e.g., Prime: *The wolves are drinking the juice*). These findings are consistent with the case study reported by Theakston and Lieven (2008). In their study, the child was more likely to use auxiliary BE accurately in the target utterance when he had produced declaratives with the same subject + auxiliary BE combinations in the prior five utterances than when he had used unrelated structures in prior utterances. The prior declaratives with a different form of auxiliary BE did not significantly enhance the child's use of auxiliary BE in the target utterances, as compared to prior utterances with unrelated structures.

The current finding suggests that young children, or at least three-year-olds who use auxiliary *'is'* variably, may have only lexically-specific representations of auxiliary BE (Joseph et al., 2002; Theakston et al., 2005; Theakston & Lieven, 2008; Wilson,

2003). However, recall that we excluded some children from analysis because they had reached a ceiling level of using auxiliary *'is'* and *'are.'* It seems that some three-year-olds can have an abstract representation of auxiliary BE. These findings, in fact, are not incompatible. Given individual differences, some three-year-olds may learn language faster than the other. It is possible that these advanced three-year-olds may have developed an abstract representation of auxiliary BE from the lexically-specific constructions with auxiliary BE. Had we tested these children at a younger age, we might have been able to find the lexically-specific priming effect on the use of the auxiliary in them too. It is also possible that these advanced children have strong/ wide ranging lexically-specific constructions stored in the representations. These children may have extended capacity of memory to learn a wide range of lexically-specific constructions related to auxiliary *'is'* and/or *'are'* but have not figured out the relationship between these lexically-specific constructions to form an abstract representations of auxiliary BE. Even so, the strong/wide ranging lexically-specific constructions are sufficient to support their productions to a ceiling level.

However, if young children do have lexically-specific representations of auxiliary BE, why did we find the facilitative effect of the lexically-specific condition only on auxiliary *'is'* but not on *'are?'* One explanation is that children may have stronger/ wider ranging lexically-specific representations of auxiliary *'is'* than *'are,'* and thus the facilitative effect of the lexically-specific priming was relatively easier to observe for auxiliary *'is'* than *'are.'* Recent research has shown that different forms of auxiliary BE are learned individually at the beginning (Joseph et al., 2002; Pine, Conti-Ramsden, Joseph, Liebergott, & Serratrice, 2008; Theakston et al., 2005). Auxiliary *'is'* tend to be acquired relatively earlier than auxiliary *'are'* partly due to the frequency of use. Children may have learned more/ stronger specific <u>*subject + is +Verb-ing*</u> constructions (e.g., *The bird is Verb-ing*) than specific <u>*subject + are +Verb-ing*</u> constructions (e.g., *The ducks are Verb-ing*) in their representations. It is possible the lexically-specific primes in

the current study might not contain specific *subject + are +Verb-ing* constructions that matches the lexically-specific constructions in child's representations so that the priming effect was not observable (Theakston & Lieven, 2008). It is also possible that the prime sentences may match specific *subject + are +Verb-ing* constructions in the representations, but these constructions are not robust enough to reflect the priming effect by adding one or two exemplars during the task (i.e., the examiner's prime and the child's correct repetition). Constructions are not represented in an all-or-none but rather gradient fashion and can have different levels of strength (Abbot-Smith & Tomasello, 2006; Bybee, 1995; Bybee & McClelland, 2006; Savage et al., 2006). The success of one task may require the child's representation to reach a certain level of strength (Tomasello & Akhtar, 2003). In this study, correct repetition of primes and the reflection of priming effects may require different levels of representational strength. Recall that some children did not even repeat the prime sentence with auxiliary *'are'* accurately, suggesting that their representation was not strong enough to support the correct repetition of auxiliary *'are'* in the prime even when they had just heard an exemplar in the prime. Similarly, some children may have the representational strength to support the correct repetitions of auxiliary *'are'* but not the reflection of the priming effect.

In summary, three-year-olds may have only lexically-specific representations of auxiliary BE, in which the subject and auxiliary forms are fixed. Overall, auxiliary *'is'* has stronger/ wider ranging lexically-specific representations than *'are.'* It should be noted that while three-year-olds may have only the constructions with specified subject and auxiliary BE form, we are not able to rule out the possibility that these children may also have *NP + is +Verb-ing* or *NP + are +Verb-ing* constructions with different degrees of strength in the representation, which include an abstract subject slot but a specified BE form (Theakston & Lieven, 2008). It is possible that the facilitative effect of the lexically-specific priming condition for auxiliary *'is'* in the current study may have resulted from the overlap of auxiliary *'is'* between the prime and the *NP + is +Verb-ing*

construction in the representation. Future studies that include a condition in which the prime and target overlap in auxiliary BE form but not subject NP are needed to test this possibility.

One might argue that the absence of the structural priming effect on producing auxiliary *'is'* may result from task demands rather than the lack of abstract representations in young children. Because Shimpi and colleagues (2007) suggested that repeating primes would reduce processing demands particularly for three-year-olds and thus facilitate their access to syntactic representations, we followed a similar procedure and avoided questions and nonfinite structures as the primes. We also limited our analysis to the target responses paired with correctly repeated prime sentences to ensure that priming had occurred. Yet, we were unable to observe the facilitative effect of the structural priming condition on the production of auxiliary *'is'* even in these ideal circumstances. Furthermore, if processing demands had played a role, the lexically-specific priming condition should have enhanced the production of auxiliary *'are'* in the target as compared to the no priming condition because both sentence structures and lexical items would have been activated in the lexically-specific priming condition. This was not what we found in this study. It seems that processing demands may not be a crucial factor for the current findings.

Children Produce the Auxiliary More Accurately in the Third Block than in the First or the Second Blocks

An additional finding in this study is that children were more likely to produce auxiliary *'is'* and *'are'* correctly in the third block than in the first or second blocks, regardless of priming condition. Similarly, Savage and colleagues (2003) found that children used passive constructions more often at trial five than at trial one within the same block in a priming task.

One interpretation of this additional finding is that children became more skilled in responding the targets over the blocks due to practice (Feldt, 1993). The other interpretation is that the prime sentences offered by the examiners served as linguistic input and strengthened children's representations, thereby facilitating the use of the auxiliary BE over time (Chang, Dell, & Bock, 2006). Previous studies adopting the structural priming paradigm have shown that listening to/repeating a particular syntactic structure in the prime sentence tended to facilitate the production of that structure in adults as well as children, as least over a short period of time (Bock et al., 2007; Savage et al., 2006; Vasilyeva, Huttenlocher, & Waterfall, 2006). Bock and colleagues (Bock et al., 2007; Chang et al., 2006) argued that the persistence of the priming effect occurred because processing the prime sentence actually enhanced the strength of representations that supported its use. Thus, structural priming is regarded by Bock and colleagues as a form of implicit learning. It is possible that cumulative exposure to the examiner's primes with auxiliary *'is'* or *'are'* (and perhaps the children's own correct repetitions of primes and responses to targets) in the current study strengthened the children's representations of constructions with auxiliary *'is'* or *'are.'* Given this cumulative learning effect, the representations of constructions with auxiliary *'is'* or *'are'* may thus be stronger and easier to retrieve in the third block than in the first or second blocks, leading to higher likelihood of producing the auxiliary accurately in the third block. However, whether the improvement of production accuracy of the auxiliary over blocks resulted from the practice or the effects of implicit learning is a question beyond the scope of the current study. Future studies may include conditions with and without prime sentences and compare the magnitude of change over time to answer this question.

Experimental and Clinical Implications

Though this study did not directly test how different formats of the structural priming paradigm may influence priming effects due to the difference in processing

demands, we stress the importance of considering this issue in future studies that examine the syntactic representations in young children, given that this population is susceptible to processing factors (Bloom, 1990). In the extant literature, the structural priming paradigm has been administered with a variety of formats. For instance, the prime sentences were either presented in blocks (Shimpi et al., 2007) or alternated with responses to target (Huttenlocher et al., 2004), were either repeated (Bock & Griffin, 2000; Leonard, Miller, Deevy, et al., 2002; Savage et al., 2003) or not repeated (Bock et al., 2007; Shimpi et al., 2007 in Experiments 1 & 2), and were accompanied with (Savage et al., 2003) or without (Bock & Griffin, 2000) corresponding pictures. The number of prime sentences also varied across studies (e.g., Savage et al., 2003; Shimpi et al., 2007). These different formats may require different amount of processing resources (e.g., attention, memory) during the task. In addition, given that structural priming indeed offers linguistic input and may change the child's linguistic representations (Bock et al., 2007), these different task formats actually offer different amount of linguistic input to the child and thus strengthen the child's representation to different degrees. To better understand the nature of syntactic representations in young children by using the structural priming paradigm, future studies need to test, or at least consider, the influence of different task formats.

Furthermore, the finding that the likelihood of producing auxiliary BE correctly increased over the blocks suggests that the structural priming paradigm may be used to facilitate (morpho-)syntactic learning in atypical populations. Researchers have shown that the structural priming paradigm can be modified to facilitate the acquisition of syntactic structures in preschoolers (Savage et al., 2006; Vasilyeva et al., 2006). To the extent that atypical populations can also acquire language structure implicitly, investigations of which formats of the priming task would best facilitate the learning of (morpho-)syntactic structures for atypical populations at various developmental levels may be a productive avenue to pursue.

Conclusion

This study examined the nature of the representation of auxiliary BE in three-year-olds via a structural priming paradigm, in which we asked children to repeat prime sentences to facilitate their access to the representations. Only the lexically-specific condition, but not the structural priming condition, enhanced the production of auxiliary *'is'* above the baseline condition. The likelihood of producing auxiliary *'are'* accurately, however, did not differ across priming conditions. Based on these findings, we suggest that young children, at least three-year-olds included in the analysis, do not have an abstract but rather a lexically-specific representation of auxiliary BE. More studies are needed to investigate how children's representations of auxiliary BE change from being lexically-specific to abstract and how the representations might be instantiated during the process of abstraction.

CHAPTER V

GENERAL DISCUSSION AND CONCLUSION

Summary of Main Findings

This study tested the UCC hypothesis and the usage-based account concerning why young children produced tense and agreement morphemes variably via three experiments. Experiment 1 investigated whether subject types influenced the production accuracy of auxiliary *'is'* in three-year-olds by using an elicited production task. Children with lower tense productivity were more likely to use auxiliary *'is'* accurately with pronominal subjects than with lexical NP subjects. The rate of use of axuiliariy 'is' increased more rapidly as tense productivity increased for lexical NP subjects but grew at a slower rate with pronominal subjects.

To determine whether predicates also affected the production accuracy of tense agreement morpheme, Experiment 2 further examined the role of subject types, predicate types, and predicate word frequency on the use of copula *'is'* in three-year-olds via an elicited production task. Regardless of children's tense productivity, the likelihood of using copula *'is'* correctly was higher with nominal predicates than with permanent- or temporary-adjectival predicates, followed by locative predicates. Children were also more likely to produce copula *'is'* accurately with low-frequency predicate words than with high-frequency predicate words. Similar to the findings in Experiment 1, the effect of subject types on the use of copula *'is'* varied with children's tense productivity but the pattern not only differed but also depended on predicate types. For sentences with nominal, permanent-adjectival, or temporary-adjectival predicates, children with **lower** tense productivity were more likely to use copula *'is'* correctly with lexical NP subjects than with pronominal subjects. In contrast, children with **higher** tense productivity used copula *'is'* more accurately with pronominal subjects than with lexical NP subjects. For

locative predicates, the likelihood of producing copula *'is'* correctly did not differ with pronominal or lexical subjects, regardless of the child's tense productivity.

Experiment 3 extended Experiment 1 by exploring the degree of abstractness of representations of auxiliary BE via a structural priming task. The production accuracy of auxiliary *'is'* in three-year-olds increased above the baseline when the prime and target shared the same structure and subject + auxiliary *'is'* combinations, but not when the prime and target only shared the same structure. However, the production accuracy of auxiliary *'are'* did not change with the types of primes that children repeated.

Taken together, Experiments 1 and 2 indicated that the frequency, rather than the structural properties, of sentence elements influenced the production accuracy of auxiliary and copula *'is'* in young children. However, the frequency effect depended on the children's developmental levels of tense. In addition, Experiments 1 and 3 suggested that young children, or at least three-year-olds who used auxiliary *'is'* variably, have only lexically-specific representations of auxiliary BE.

However, there are also some discrepancies between these experiments. First, to what extent the constructions with auxiliary *'is'* were lexically specified in young children, especially in the subject position, varied between the elicitation production task (Experiment 1) and the priming task (Experiment 3). Second, the effect of subject types varied for auxiliary and copula *'is'* in the children with the similar developmental levels of tense.

The Nature of Lexically-specific Constructions with Auxiliary *'is'* Varied in the Production and Priming Tasks

Both the elicited production task and the priming task indicated that three-year-olds may have only lexically-specific representations of auxiliary *'is'* but these tasks seems to suggest different levels of abstractness of these constructions. Given that the production accuracy of auxiliary *'is'* did not differ with high- or low-frequency lexical

NP subjects in three-year-olds, the elicited production task suggested that these children may have the *NP**'s** + Verb-ing* construction although the representational strength varied with children's developmental levels of tense. In this construction, the auxiliary BE form is specified whereas the subject position is an abstract, open slot. In contrast, the priming task suggested that three-year-olds may have only lexically-specific constructions with auxiliary *'is,'* in which both the subject and the auxiliary BE form were specified (e.g., ***The dog's** + Verb-ing*). This is because only the lexically-specific condition, in which the prime-target pair shared the same subject and BE form, facilitated the production of auxiliary *'is'* above the baseline.

One possible explanation is that the current design of the priming task did not allow us to infer whether three-year-olds had developed the *NP**'s** + Verb-ing* construction. To examine whether these children had developed the *NP**'s** + Verb-ing* construction, we would need to include a condition, in which the prime-target pair shares the auxiliary BE form but not the subject (e.g., Prime: *The dog**'s** running*. Target: *The cat**'s** eating*). If this condition significantly increased the likelihood of producing auxiliary *'is'* correctly above the baseline, we could infer that children have the *NP**'s** + Verb-ing* construction.

From the results of the elicited production and the priming tasks, we suggest that three-year-olds who still use auxiliary *'is'* variably do not seem to have the *NP + BE + Verb-ing* construction (i.e., without specified BE form) but do seem to have the *NP**'s** + Verb-ing* construction although the representational strength of the *NP**'s** + Verb-ing* construction varies with children's developmental levels of tense. However, it should be noted that even if children have developed this construction with an abstract subject slot, highly frequent constructions with specific subject and auxiliary *'is'* that they have learned can remain in the representation to support children's production (Bybee, 1995, 2002; Dąbrowska, 2000; Dąbrowska & Lieven, 2005).

Effect of Subject Types Varied for Auxiliary and Copula *'is'* in Children with Identical Developmental Levels

The other discrepancy between the experiments is that the effect of subject types on the production of auxiliary and copula *'is'* changed as a function of children's tense productivity, but the patterns were different for these morphemes. The different patterns between auxiliary and copula *'is'* may occur because of the different developmental trajectories between these morphemes.

Previous studies have shown that copula BE tend to be acquired earlier than auxiliary BE partly because of the difference in the frequency of use (Brown, 1973; Cameron-Faulkner, Lieven, & Tomasello, 2003; Joseph et al., 2001; Wilson, 2003). In the current study, the production accuracy was approximately 0.70 for copula *'is'* and 0.30 for auxiliary *'is'* in children with a tense productivity score of 2. It is possible that, children with lower tense productivity may have accumulated enough instantiations of ungrammatical constructions with specific pronominal subjects that lack copula *'is'* from the input (e.g., **(Is)He NP/AP*). These ungrammatical constructions may then become strong competitors to the grammatical constructions with specific pronominal subjects that contain copula *'is'* (e.g., *It's + NP/AP*). In contrast, these children may have not accumulated enough instantiations of ungrammatical constructions with specific pronominal subjects that do not have auxiliary *'is'* (e.g., **(I see)It Verb-ing*). Thus, these ungrammatical constructions are still not strong enough to compete with the grammatical constructions with specific pronominal subjects that contain auxiliary *'is'* (e.g., *It's + Verb-ing*). Children with lower tense productivity scores therefore produced auxiliary *'is'* more accurately with pronominal subjects than with lexical NP subjects but used copula *'is'* more correctly with lexical NP subjects than with pronominal subjects.

The competition from ungrammatical constructions changes in children with higher tense productivity. Children at this developmental level may figure out that the ungrammatical constructions with specific pronominal subjects that lack copula *'is'* are

actually part of longer constructions due to increased of processing capacity (Freudenthal et al., 2007). The use of these ungrammatical constructions with pronominal subjects thus decreases over time. These ungrammatical truncated constructions will cease to be strong competitors for the grammatical constructions with pronominal subjects that contain copula *'is'* due to the declining rate at which children store truncated sentences (MacWhinney, 2004). Similarly, we did not observe strong competition from ungrammatical constructions with specific pronominal subjects that lack auxiliary *'is'* to the grammatical constructions with specific pronominal subjects and auxiliary *'is'* in children with higher tense productivity. It seems that, because of the increased processing capacity, children with higher tense productivity did not encounter strong competitions from ungrammatical constructions with pronominal subjects for both copula and auxiliary *'is.'* Thus, the production accuracy of copula or auxiliary *'is'* was not higher with lexical NP subjects than with pronominal subjects.

However, recall that children at higher end of tense productivity did show a trend to use auxiliary *'is'* more accurately with lexical NP subjects than with pronominal subjects though the statistical method we employed did not allow us to directly evaluate whether this difference was significant. If the difference did reach a significant level, a possible interpretation is that children at higher tense productivity may just accumulate a sufficient number of ungrammatical constructions with specific pronominal subjects that lack auxiliary *'is'* to compete with the grammatical constructions with specific pronominal subjects and auxiliary *'is.'*

In addition, if children with higher tense productivity did use auxiliary *'is'* more accurately with lexical subjects than with pronominal subjects, we can speculate that the developmental trajectory of BE forms in relation to subject types is as follows: For a given BE form, initially children use the BE form more correctly with pronominal subjects than with lexically-specific constructions because specific <u>pronoun + BE</u> combinations are more frequent than specific <u>NP + BE</u> combinations. Later on, the

ungrammatical constructions without the BE form start to compete with the grammatical constructions with pronominal subjects and BE forms. The production accuracy of BE is thus lower with pronominal subjects than with lexical NP subjects. Eventually, children will figure out that these ungrammatical constructions were actually part of larger constructions and start to decrease the use of these truncated, ungrammatical constructions. The production accuracy of BE is then higher with pronominal subjects than with lexical NP subjects again.

However, we do not rule out the possibility that children may also learn ungrammatical abstract constructions from the input, such as **NP AdjP* (e.g., *(*Is) the dog happy*) and *NP Verb-ing* (e.g., *(I see) the dog running*), to compete with their grammatical counterparts when the sentence involves the use of lexical NP subjects. When and how these ungrammatical abstract constructions may affect the production of BE in sentences with lexical NP subjects are beyond the scope this thesis. Future studies adopting polynomial regressions are needed to illuminate how the competition from different types of ungrammatical constructions influences the developmental trajectory of BE forms in relation to subject types. Computational models may also be enlightening with regard to the speculated developmental trajectories.

Theory Revisited

This thesis tested the assumptions of the UCC hypothesis (Wexler, 1998) and the usage-based approach (Tomasello, 2003) concerning why young children produced tense and agreement morphemes variably. The UCC hypothesis posits that young children possess abstract representations of tense and agreement morphemes from the time that they begin to combine words. They use these morphemes variably because of the presence of the unique checking constraint in their grammar (Wexler, 1998). However, evidence from the priming study seems to suggest that young children have only

lexically-specific representations of auxiliary BE, a finding which argues against the assumptions of the UCC.

The UCC hypothesis also explicitly predicts that young children should produce copula *'is'* more accurately with permanent-type predicates than with temporary-type predicates due to structural differences in the syntactic derivation (Becker, 2004; Wexler, 2000). This prediction is also made by the usage-based approach, but for reasons associated with frequency. Indeed Experiment 2 did show that permanent-type predicates were associated with higher rates of use of copula *'is'* than temporary-type predicates. However, there were findings that structural properties were not able to account for. For instance, structural properties cannot explain the asymmetry between pronominal and lexical subjects on the production of copula and auxiliary *'is'* given that both of the pronominal and lexical subjects are both determiner phrases, which should be equally (un)affected by the UCC. Based on the predictions of the UCC with regard to movement and feature checking, there is no reason to infer that the UCC would affect subtypes of the same predicate types differently (e.g., locative and temporary-adjectival predicates). It might be possible to explain the observed asymmetry/variation in subjects and predicate subtypes by attributing this difference to differences in processing demands. For instance, lexical subjects are longer and have more complex internal structure, potentially making more demands than shorter and less complex pronominal subjects. These processing demands could lead to higher omission rates of copula/auxiliary *'is'* in sentences with lexical subjects. Thus, the UCC hypothesis might be correct, but a mechanism to account for why the UCC applies to these cases would need to be incorporated into the theoretical framework to account for the variable use of tense and agreement morphemes in young children. Given the preciseness of the UCC and the implicit rejection of changes in syntactic processing as a factor, the addition of such mechanisms seems necessarily post-hoc and strains the test of parsimony associated with the theory.

Contrary to the UCC hypothesis, the usage-based approach assumes that young children use tense and agreement morphemes variably because they have not developed adult-like abstract representation and possess only lexically-specific representations of these morphemes. Evidence from the priming study seems to support this assumption. In addition, the finding that the frequency of subject types influenced the use of copula/auxiliary 'is' further supported this assumption. However, recall that the role of frequency seemed to change with the morpheme being elicited and the child's developmental level in non-intuitive ways. These findings contradicts the predictions of the usage-based approach. Although we speculated that the results can be attributed to competition of ungrammatical constructions, this hypothesis needs further support from corpus analyses and computational simulation before the mechanism can be clearly defined.

Though not tested in the current study, one account that could explain the variable use of tense and agreement morphemes is the implicit realization account (Levelt, 1989; Owen & Goffman, 2007). In this account, young children have some knowledge of tense and agreement morphemes. Due to articulatory difficulties or insufficiently strong representations of the morpheme, these children may not always overtly realize the phonetic forms of these morphemes (e.g., *'s* in *He's*) in a way that is perceptible to normal adult listeners. However, given that children have the knowledge of tense morphemes that they are not able to articulate, they might phonetically realize these morphemes by lengthening the surrounding segments, such as vowels, or otherwise leaving a timing slot for the morphological form (Carter & Gerken, 2004).

One ideal way to test this account is to measure the duration of particular segments in minimal-pair sentences from children. For instance, if the child produces sentences *Joe playing*, *Joe's playing,* and *Is Joe playing*, we can compare the duration of <u>*Joe*</u> in *I saw* <u>*Joe*</u> *playing* and <u>*Joe*</u> *playing* and <u>*Joe's*</u> in *Joe's playing*. If the child has knowledge of tense and agreement morphemes, we would expect differences between

nonfinite complement and simple sentence with an omission, but not between the two simple sentences, even though these are in contrast to the phonetic realization. Although we have many exemplars of correct and incorrect productions, we were not able to directly test this account because the children did not produce minimal-pair sentences. Furthermore, even if an implicit realization can be observed in young children, this account is limited in its ability to distinguish between the two accounts we were testing in this study. It is unclear whether the articulatory realization is due to weakly stored representations, consistent with the usage-based approach or due to limited motor skills, a result consistent with the UCC. Furthermore, subphonemic changes as an explanation of variability would not explain findings such as the variation between subtypes of the same predicate types.

Concluding Remarks

This thesis confirmed that young children had only lexically-specific representations of BE and the use of BE depended on the co-occurrence frequency between BE and other sentence elements (e.g., subjects, predicates). Overall, the findings supported the usage-based approach in specific and the empiricist approach in general that children learned language structures from the input. However, current theories within the empiricist approach remains unclear as to how children make use of the general cognitive mechanisms to create abstract constructions out of concrete utterances, how and when the process of abstraction occurs, and how the representations of constructions change along the process of abstraction. Uncovering the process of abstracting linguistic constructions in future studies would inform us not only how we could promote language learning in typical and atypical populations but also how human mind works.

REFERENCES

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23*, 275-290

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Ambridge, B., & Pine, J. (2006). Testing the Agreement/Tense Omission Model using an elicited imitation paradigm. *Journal of Child Language*, 33, 879-898.

Ambridge, B., Theakston, A., Lieven, E., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21*, 174-193.

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*. Rockville, MD: Author.

Avram, L. (2002). *An introduction to language acquisition from a generative perspective*. Bucharest: Editura Universitatii din Bucuresti.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

Balason, D, & Dollaghan, C. (2002). Grammatical morpheme production in 4-year-old children. *Journal of Speech, Language, and Hearing Research*, *45*, 961-969.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Review, 19*(3), 241-248.

Bates, E., & Goodman, J. (1999). On the emergence of Grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language* (pp. 29-79). New Jersey: Mahwah.

Becker, M. (2000a). The acquisition of the English copula. In S. C. Howell (Ed.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 104-115). Somerville, MA: Cascadilla Press.

Becker, M. (2000b). *The development of copula in child English: The lightness of be.* Unpublished doctoral dissertation, University of California, Los Angeles.

Becker, M. (2004). Copula omission is a grammatical reflex. *Language Acquisition, 12*(2), 157-167.

Bloom, P. (1989). Why do children omit subject? *Papers and Reports on Child Language Development, 28*, 57-64.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 24*, 491-504.

Bock, K., Dell, G., Chang, F., & Onishi, K. (2007). Persistent structural priming from language comprehension to language production. *Cognition, 104*, 437-458.

Bock, K., & Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General, 129*(2), 177-192.

Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research, 24*(6), 489-506.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425-455.

Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givon & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109-134). Amsterdam: John Benjamins.

Bybee, J., & McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*, 381-410.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*, 843-873.

Carter, A. & Gerken, L. (2004). Do children's omissions leave traces?. *Journal of Child Language, 31,* 561-586.

Chang, F., Dell, G., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234-272.

Chomsky, N. (1956). *Syntactic structure*. The Hague: Mouton & CO.

Chomsky, N. (1965). Three models for the description of language. In R. Luce, R. Bush & E. Galanter (Eds.), *Readings in mathematical psychology* (Vol. II, pp. 105-124). New York: John Wiley and Sons.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*(2), 157-205.

Coady, J. (2009). Rapid naming by children with and without specific language impairments. Poster presented at *The 30th Symposium on Research in Child Language Disorders,* University of Wisconsin at Madison, Madison

Connor, C. M. (2009). The effect of children's language and literacy skills on the efficacy of early reading instruction: Causal implications of child characteristics by instruction interactions. Presented at *The 30th Symposium on Research in Child Language Disorders*, University of Wisconsin at Madison, Madison.

Connor, C. M., Morrison, F., & Katch, L. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions of growth in early reading. *Scientific Studies of Reading, 8*(4), 305-336.

Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Malden, MA: Blackwell.

Dąbrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics*, *11*(1/2), 83-102.

Dąbrowska, E, & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437-474.

de Villers, J. (2001). Continuity and modularity in language acquisition and research. *Annual Review of Language Acquisition, 1*, 1-64.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59, 447-456.

Elman, J. L. (1993). Learning and development n neural networks: The importance of starting small. *Cognition, 48*, 71-99.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Feldt, L. (1993). *Design and analysis of experiments in the behavioral sciences*. Iowa City, IA: The University of Iowa.

Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., et al. (1993). *The MacArthur Communicative Development Inventories.* San Diego: Singular Publishing.

Fillingham, J., Hodgson, C., Sage, K., & Ralph, M. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation, 13*(3), 337-363.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition, 82*, 259-278.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science, 31*, 311-341.

Gerken, L. (2000). Examining young children's morphosyntactic development through elicited production. In L. Menn & N. B. Ratner (Eds.), *Methods for studying language production* (pp. 45-67). Mahwah, NJ: Lawrence Erlbaum Associates.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implication for assessment and intervention. *American Journal of Speech-Language Pathology, 9*, 151-161.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Goodman, J., Dale, P., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35, 515-531.

Guo, L., Owen, A., & Tomblin, J. (submitted). Effect of subject types on the production of auxiliary 'is.' *Journal of Speech, Language, and Hearing Research*.

Hadley, P., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1344-1362.

Hottenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*, 182-195.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language* (pp. vii-xxviii). Standford, CA: CSLI Publications.

Kleinbaum, D., Kupper, L., Muller, K., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods.* Pacific Grove, CA: Duxbury Press.

Kuczaj, S. (1981). The acquisition of copula and auxiliary BE forms. *Papers and Reports on Child Language Development, 20*, 78-83.

Jaccard, J. (2001). *Interaction effects in logistic regression*. Thousand Oaks, CA: Sage.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.

Joseph, K., Serratrice, L., & Conti-Ramsden, G. (2002). Development of copula and auxiliary BE in children with specific language impairment and younger unaffected controls. *First Language*, 22, 137-172.

Lahey, M., Liebergott, J., Chesnick, M., Menyuk, P., & Adams, J. (1992). Variability in children's use of grammatical morphemes. *Applied Psycholinguistics, 13*, 373-398.

Langacker, R. (2000). A dynamic usage-based model. In S. Kemmer & M. Barlow (Eds.), *Usage-based models of language* (pp. 1-64). Standford, CA: CSLI Publications.

Legate, J. & Yang, C. (2007). Morphosyntactic learning and the development of tense. Language Acquisition, *14*(3), 315-344.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Leonard, L. (1998). *Children with specific language impairment.* Cambridge, MA: MIT Press.

Leonard, L. (2007). Accounting for the uneven morphosyntactic profile in specific language impairment: A crosslinguistic perspective. Presented at The 28th Symposium on Research in Child Language Disorders, University of Wisconsin at Madison, Madison.

Leonard, L. B., Miller, C. A., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 678-689.

Leonard, L. B., Miller, C. A., Deevy, P., Rauf, L., Gerber, E., & Charest, M. (2002). Production operations and the use of nonfinite verbs by children with language impairment. *Journal of Speech, Language, and Hearing Research, 45*, 744-758.

Leonard, L. B., Miller, C. A., Grela, B., Holland, A. L., Gerber, E., & Petucci, M. (2000). Production operations contribute to the grammatical limitations of children with specific language impairment. *Journal of Memory and Language, 43*, 362-378.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lieven, E., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*, 187-219.

Lieven, E. (2008). Learning the English auxiliary. In H. Behrens (Ed.), *Corpora in Language Acquisition Research* (pp. 61-98). Philadelphia: John Benjamins Publishing.

Lindner, K., & Hohenberger, A. (2009). Introduction: Concepts of development, learning, and acquisition. *Linguistics, 47*(2), 211-239.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Earlbaum.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31*, 883-914.

Mainela-Arnold, E., Evans, J., & Coady, J. (2008). Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task. *Journal of Speech, Language, and Hearing Research, 51*, 381-393.

Matthews, D., Lieven, E., Theakston, A. L., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied psycholinguistics, 27*, 403-422.

Miller, C. A., & Deevy, P. (2006). Structural priming in children with and without specific language impairment. *Clinical Linguistics & Phonetics, 20*(5), 387-399.

Miller, J. F. (1981). *Assessing language production in children: Experimental procedures*. Baltimore: University Park Press.

Moe, A., Hopkins, C., & Rush, R. T. (1982). *The vocabulary of first-grade children*. Springfiled, IL: Charles C Thomas.

Naigles, L. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition*, *86*, 157-199.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

Owen, A., & Goffman, L. (2007). Acoustic correlate of inflectional morphology in the speech of children with specific language impairment and their typically developing peers. *Clinical Linguistics & Phonetics, 21*(7), 501-522.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633-651.

Pine, J. M., Conti-Ramsden, G., Joseph, K. L., Liebergott, J., & Serratrice, L. (2008). Tense over time: Test the Agreement/Tense Omission Model as an account of the pattern of tense -marking provision in early child English. *Journal of Child Language, 35*, 55-75

Quené, H. & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413-425.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Cambridge, MA: Basil Blackwell.

Radford, A. (2004). *Minimalist syntax: Exploring the structure of English.* Cambridge: Cambridge University Press.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science, 29*, 1007-1028.

Rice, M., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extend optional infinitive. *Journal of Speech and Hearing Research, 38*, 850-863.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1412-1431.

Rispoli, M., Hadley, P., & Holt, J. (2009). The growth of tense productivity. *Journal of Speech, Language, and Hearing Research, 52*, 930-944.

Samuelson, L., & Smith, L. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73, 1-33.

Savage, C., Lieven, E., Theakston, A. L., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science, 6*(5), 557-567.

Savage, C., Lieven, E., Theakston, A. L., & Tomasello, M. (2006). Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development, 2*(1), 27-49.

Schütze, C. T. (2001). Productive inventory and case/agreement contingencies: A methodological note to Rispoli (1999). *Journal of Child Language, 28*, 507-515.

Schütze, C. T., & Wexler, K. (1996). Subject case licensing and English root infinitives. In A. Stringfellow, D. Cahan-Amitay, E. Hughes & A. Zukowski (Eds.), *Proceedings of the 20th Boston University Conference of Language Development* (pp. 670-681). Boston, MA: MIT Press.

Shimpi, P., Gámez, P., Huttenlocher, J., & Vasilyeva, M. (2007). Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology, 43*(6), 1334-1346.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition Volume 2: Theoretical issues*. Hillsdale, NJ: Lawrence Erlbaum associates.

Spaulding, T., Plante, E., & Farinella, K. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, *37*, 61-72.

Spencer, J., Blumberg, M., McMurray, B., Robinson, S., Samuelson, L., &Tomblin, J.B. (in press). Short arms and talking eggs: The inconvenience of understanding process. In *Child Development Perspectives*.

Theakston, A. L., & Lieven, E. V. (2005). The acquisition of auxiliaries BE and HAVE: An elicitation study. *Journal of Child Language, 32*, 587-616.

Theakston, A. L., & Lieven, E. (2008). The influence of discourse context on children's provision of auxiliary BE. *Journal of Child Language, 35*, 129-158.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16*(1), 247-277.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209-253.

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review, 22*, 183-197.

Tomasello, M., & Abbot-Smith, K. (2002). A tale of two theories: response to Fisher. *Cognition, 83*, 207-214.

Tomasello, M., & Akhtar, N. (2003). What paradox? A response to Naigles (2002). *Cognition, 88*, 317-323.

Vasilyeva, M., Huttenlocher, J., & Waterfall, H. (2006). Effects of language intervention on syntactic skill levels in preschoolers. *Developmental Psychology, 42*(1), 164-174.

Werker, J., Fennell, C., Corcoran, K., & Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1-30.

Werner, E., & Kresheck, J. D. (1983). *Structured Photographic Expressive Language Test–Preschool (SPELT-P)*. Sandwich, IL: Janelle Publications.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 305-350). Cambridge: Cambridge University Press.

Wexler, K. (1996). The development of inflection in a biologically based theory of language acquisition. In M. Rice (Ed.), *Toward a genetics of language* (pp. 113-144). New Jersey: Lawrence Erlbaum Associates.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 103*, 23-79.

Wexler, K. (2000). Three problems in the theory of the optional infinitive stage: Stage/individual predicates, eventive verbs and finite null-subjects. In R. Billerery & D. Lillehaugen (Eds.), *Proceedings of the 19th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press.

Wexler, K. (2003). Lenneberg's dream: Learning, normal language development, and specific language impairment. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wexler, K. Schütze, C., & Rice, M. (1998). Subject case in children with SLI and unaffected controls: Evidence for the Agr/Tns Omission Model. *Language Acquisition, 7*, 311-344.

Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language, 30*, 75-115.

Yang, C. (2002). *Knowledge and learning in natural language.* New York: Oxford University Press.

Zimmerman, I., Steiner, V., & Pond, R. (2002). *Preschool Language Scale-Fourth Edition (PLS-4)*. San Antonio, Tx: Pearson.

APPENDIX A

TARGET AND CONTRASTIVE SENTENCES OF AUXILIARY *'IS'*

Contrastive Sentences	**Target Sentences**
Pronominal Subjects	
1. They're making a cake.	1. He's eating a cookie.
2. They're cleaning the table.	2. She's drinking the water.
3. They're walking the dog.	3. It's licking the doctor.
4. They're cooking the chicken.	4. He's drawing a flower.
5. They're painting the door.	5. She's hitting the TV.
6. They're wiping the window.	6. It's chasing the bunny.
7. They're touching the ball.	7. He's driving the airplane.
8. They're climbing the wall.	8. She's feeding the rabbit.
9. They're throwing the ball.	9. It's kissing the princess.
10. They're moving the orange.	10. He's cutting the apple.
High-frequency Lexical NP Subjects	
11. The foxes are moving the pumpkin.	11. The dog's eating a cake.
12. The chickens are making a pie.	12. The goat's drinking the juice.
13. The angels are cleaning the chair.	13. The dog's licking the boy.
14. The boys are walking the cat.	14. The duck's drawing a star.
15. The wolves are wiping the floor.	15. The pig's hitting the tree.
16. The witches are painting the wall.	16. The cat's chasing the mouse.
17. The mice are climbing the ladder.	17. The goat's driving the car.
18. Tom and Jane are watering the flower.	18. Mom's feeding the turtle.
19. The bunnies are throwing the apple.	19. The cat's kissing the girl.
20. The penguins are touching the table.	20. The pig's cutting the cake.
Low-frequency Lexical NP Subjects	
21. The horses are wiping the floor.	21. The deer's eating a cake.
22. The wolves are painting the chair.	22. The queen's drinking the tea.
23. The foxes are climbing the ladder.	23. The sheep's licking the boy.
24. The chickens are throwing the ball.	24. The ant's drawing a tree.
25. The angels are walking the dog..	25. The frog's hitting the boat.
26. The elephants are moving the TV.	26. The sheep's chasing the duck.
27. The lions are watering the tree.	27. The frog's riding a bike.
28. The witches are making the soup.	28. The queen's feeding the bird.
29. The monkeys are cleaning the table.	29. The deer's kissing the prince.
30. The mice are touching the cheese.	30. The ant's cutting the bread.

APPENDIX B

RATING OF FANTISINESS OF TARGET SENTENCES AND PICTURES

To evaluate the realness/fantasiness of the target sentences/pictures, we recruited twenty-one subjects to judge the pictures. The pictures were divided into three parts in a powerpoint slide show. Part I included the pictures for copula BE production; Part II, those for auxiliary BE production; and Part III, those for priming study. The order of the pictures within each part was randomized. In total, there were 198 pictures. The participants were instructed to judge whether the event that the picture depicted existed in real world. A picture was judged as fantasiful if they felt the picture describing an event that was not likely to happen in real world. In contrast, a picture was judged as real if the event depicted by the picture was likely to happen in real world. Each picture judged as fantasiful was assigned a score of 2 whereas that judged as real, a score of 1. Higher scores means more fantasiful.

The scores were tallied for each item in the three tasks. Because the scores of realness/fantasiness in this rating were ordinal *per se*, we performed Kruskal-Wallis H tests to examine the differences in fantasiness of pictures by level of each independent variable in each task. For the pictures in the auxiliary BE production task, there was no significant differences in fantasiness of pictures across subject types, $\chi^2 = 2.806$, $df = 2$, $p = 0.246$. For the pictures in the copula BE production task, there were no significant differences in fantasiness of pictures across subject types, $\chi^2 = 1.269$, $df = 2$, $p = 0.530$; across predicate types, $\chi^2 = 5.974$, $df = 3$, $p = 0.113$; or across predicate frequency, $\chi^2 = 1.347$, $df = 1$, $p = 0.246$. For those in the priming task, there was no significant differences in fantasiness of pictures across priming conditions, $\chi^2 = 0.036$, $df = 2$, $p = 0.982$.

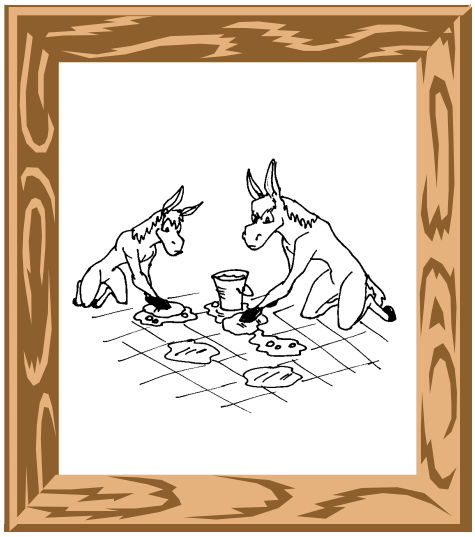
APPENDIX C

SAMPLE PICTURES

1. Contrastive: They're making a cake.

Target: He's eating a cookie.

2. Contrastive: The horses are wiping the floor.



Target: The deer's eating a cake.

APPENDIX D

TARGET AND CONTRASTIVE SENTENCE FOR COPULA '*IS*'

Contrastive Sentence	**Target Sentence**
Pronominal Subjects, High-frequency Nominal Predicates	
1. They're firemen.	1. He's a policeman. (5)[6]
2. Those are boys.	2. That's a girl. (3)
3. They're students.	3. It's a teacher. (4)
4. Those are nurses.	4. That's a doctor. (4)
5. They're foxes.	5. It's a chicken. (4)
Pronominal Subjects, High-frequency Nominal Predicates	
6. They're policemen.	6. He's a fireman. (4)
7. They're muffins.	7. It's a donut. (4)
8. They're elephants.	8. It's a zebra. (4)
9. Those are doctors.	9. That's a nurse. (3)
10. They're princes.	10. She's a princess. (4)
High-frequency Lexical NP Subjects, High-frequency Nominal Predicates	
11. The women are nurses.	11. The man's a doctor. (5)
12. The ducks are students.	12. The bear's teacher. (5)
13. The bunnies are girls.	13. The pig's a boy. (4)
14. The lions are queens.	14. The dog's a king. (4)
15. The monkeys are boys.	15. The cat's a girl. (4)
High-frequency Lexical NP Subjects, Low-frequency Nominal Predicates	
16. The cows are princesses.	16. The dog's a prince. (4)
17. Johnny and Bobby are humans.	17. Dad's an alien. (4)
18. The kids are students.	18. The man's a soldier. (5)
19. The mice are doctors.	19. The bear's a nurse. (4)
20. The lions are kings.	20. The cat's a queen. (4)
Low-frequency Lexical NP Subjects, High-frequency Nominal Predicates	
21. The birds are nurses.	21. The deer's a doctor. (5)
22. The cats are queens.	22. The ant's a king. (4)
23. The bears are girls.	23. The frog's a boy. (4)
24. The dogs are students.	24. The owl's a teacher. (5)
25. The dogs are boys.	25. The sheep's a girl. (4)

[6] The parenthesized number represents the number of syllables of each sentence.

Contrastive Sentence	**Target Sentence**
Low-frequency Lexical NP Subjects, Low-frequency Nominal Predicates	
26. The cows are princesses.	26. The frog's a prince. (4)
27. The birds are policemen.	27. The deer's a fireman. (5)
28. The mice are doctors.	28. The sheep's a nurse. (4)
29. The lions are kings.	29. The squirrel's a queen. (4)
30. The alligators are princes.	30. The ant's a prince. (4)
Pronominal Subjects, High-frequency Permanent-adjectival Predicates	
31. Those are very short.	31. That's very long. (4)
32. They're really small.	32. It's really big. (5)
33. They're really soft.	33. It's really hard. (5)
34. Those are pink.	34. That's yellow. (3)
35. They're ugly.	35. She's pretty. (3)
Pronominal Subjects, Low-frequency Permanent-adjectival Predicates	
36. Those are beautiful.	36. That's ugly. (3)
37. They're very sour.	37. It's very sweet. (4)
38. They're very tall.	38. He's very short. (4)
39. They're very short	39. That's very tall. (4)
40. They're very sweet.	40. It's very spicy. (5)
High-frequency Lexical NP Subjects, High-frequency Permanent-adjectival Predicates	
41. The chickens are very small.	41. The bear's very big. (5)
42. The cakes are very soft.	42. The rock's very hard. (5)
43. The bananas are yellow.	43. The apple's red. (4)
44. The cars are short.	44. The train's long. (3)
45. The doves are white.	45. The cat's blue. (3)
High-frequency Lexical NP Subjects, Low-frequency Permanent-adjectival Predicates	
46. The boys are ugly.	46. The girl's beautiful. (5)
47. The pigs are very fat.	47. The monkey's very thin. (5)
48. The bunnies are pretty.	48. The pig's ugly. (4)
49. The giraffes are very tall.	49. The duck's very short. (5)
50. The flowers are short.	50. The tree's tall. (3)
Low-frequency Lexical NP Subjects, High-frequency Permanent-adjectival Predicates	
51. The cars are red.	51. The balloon's yellow. (5)
52. The cookies are very small.	52. The pie's very big. (5)
53. The crabs are red.	53. The snowman's white. (4)
54. The rabbits are white.	54. The ant's black. (3)
55. The pigs are pink.	55. The frog's green. (3)

Contrastive Sentence	**Target Sentence**
Low-frequency Lexical NP Subjects, Low-frequency Permanent-adjectival Predicates	
56. The sharks are ugly.	56. The deer's beautiful. (5)
57. The elephants are fat.	57. The ant's thin. (3)
58. The butterflies are pretty.	58. The worm's ugly. (4)
59. The lemons are very sour.	59. The pie's very sweet. (4)
60. The rabbits are short.	60. The zebra's tall. (4)
Pronominal Subjects, High-frequency Locative Predicates	
61. They're on the street.	61. He's in the kitchen. (5)
62. They're under the chair.	62. It's on the table. (5)
63. They're at home.	63. She's in the park. (4)
64. Those are under the table.	64. That's on the door. (4)
65. They're outside the house.	65. He's in the house. (4)
Pronominal Subjects, Low-frequency Locative Predicates	
66. They're on the chair.	66. It's under the table. (6)
67. They're in the tree.	67. It's under the flower. (6)
68. They're on the flower.	68. He's under the swing. (5)
69. Those are on the table.	69. That's under the chair. (5)
70. They're near the cup.	70. She's near the bed. (4)
High-frequency Lexical NP Subjects, High-frequency Locative Predicates	
71. The ducks are at school.	71. The dog's in the bedroom. (6)
72. The wolves are outside the house	72. The pig's in the house. (5)
73. The moose are in the hospital.	73. The bear's at church. (4)
74. The giraffes are in the zoo.	74. The cat's at school. (4)
75. Johnny and Bobby are at home.	75. Dad's in the park. (4)
High-frequency Lexical NP Subjects, Low-frequency Locative Predicates	
76. The turtles are on the bed.	76. The bird's under the bench. (6)
77. The foxes are on the table.	77. The cat's under the bed. (6)
78. The mice are near the cheese.	78. The bear's near the bed. (5)
79. Tom and John are on the swing.	79. Dad's under the slide. (5)
80. Johnny and Bobby are near the flower.	80. Mom's near the car. (4)

Contrastive Sentence	**Target Sentence**
Low-frequency Lexical NP Subjects, High-frequency Locative Predicates	
81. The cups are under the chair.	81. The fork's on the table. (5)
82. The butterflies are on the hat.	82. The child's in the box. (5)
83. The cats are under the table.	83. The worm's on the chair. (5)
84. The bears are in the park.	84. The sheep's at church. (4)
85. The penguins are in the playground.	85. The deer's at school. (4)
Low-frequency Lexical NP Subjects, Low-frequency Locative Predicates	
86. The shoes are on the table.	86. The sock's under the chair. (6)
87. The butterflies are on the flower.	87. The deer's under the tree. (6)
88. The chickens are near the stroller.	88. The frog's near the bed. (5)
89. The mice are near hamburger.	89. The ant's near the cake. (5)
90. The spoons are near the bowl.	90. The pen's near the book. (5)
Pronominal Subjects, High-frequency Temporary-adjectival Predicates	
91. They're very happy.	91. She's very scared. (4)
92. They're very slow.	92. He's very fast. (4)
93. They're very cold.	93. It's very hot. (4)
94. Those are very hot.	94. That's very cold. (4)
95. They're very happy.	95. He's very mad. (4)
Pronominal Subjects, Low-frequency Temporary-adjectival Predicates	
96. They're very clean.	96. It's very dirty. (5)
97. They're very happy.	97. He's very sad. (4)
98. They're very noisy.	98. She's very quiet. (5)
99. Those are very dry.	99. That's very wet. (4)
100. They are fine.	100. He's itchy. (3)
High-frequency Lexical NP Subjects, High-frequency Temporary-adjectival Predicates	
101. The kids are very happy.	101. The man's very mad. (5)
102. The tigers are very slow.	102. The lion's very fast. (5)
103. Jack and David are very hot.	103. Mom's very cold. (4)
104. The mice are cold.	104. The dog's hot. (3)
105. The tigers are fine.	105. The bear's hurt. (3)

Contrastive Sentence	**Target Sentence**
High-frequency Lexical NP Subjects, Low-frequency Temporary-adjectival Predicates	
106. The elephants are very happy.	106. The cat's very sad. (5)
107. The horses are clean.	107. The pig's dirty. (4)
108. The zebras are fine.	108. The dog's itchy (4)
109. The cows are quiet.	109. The bear's noisy. (4)
110. The rabbits are dry.	110. The duck's wet. (3)
Low-frequency Lexical NP Subjects, High-frequency Temporary-adjectival Predicates	
111. The monkeys are awake.	111. The zebra's tired. (4)
112. The mice are very happy.	112. The frog's very mad. (5)
113. The foxes are fine.	113. The squirrel's hurt. (4)
114. The men are very hot.	114. The child's very cold. (5)
115. The drinks are cold.	115. The tea's hot. (3)
Low-frequency Lexical NP Subjects, Low-frequency Temporary-adjectival Predicates	
116. The dinosaurs are very happy.	116. The deer's very sad.(5)
117. The chickens are noisy.	117. The sheep's quiet. (4)
118. The kings are fine.	118. The queen's itchy. (4)
119. The lions are quiet.	119. The frog's noisy. (4)
120. The shirts are dry.	120. The sock's wet. (3)

APPENDIX E

PRIME-TARGET PAIRS

1. No Priming Condition

	Prime	Target
Singular	1. A, B, C, D 2. One, two, three, four 3. Square, circle, triangle 4. Winnie the Pooh 5. Count three apples 6. Ernie, Elmo, Bambi 7. Grapes, strawberries 8. Paper, scissors, rock	1. The crab's flying the kite. 2. The bear's reading the book. 3. The zebra's holding the doll. 4. The squirrel's washing the car. 5. The bird's closing the door. 6. The cow's petting the bunny. 7. The monkey's watching TV 8. The child's riding the bike.
Plural	9. Egg, bread, pancake 10. A, B, C, D 11. Five, six, seven 12. Eye, ear, nose 13. Sleeping Beauty, Snow White 14. Count two horses 15. Count three books. 16. Superman, Mickey Mouse	9. The lions are sweeping the floor. 10. The firemen are digging a hole. 11. The frogs are pulling the boat. 12. The cats are building the house 13. The ducks are kicking the ball. 14. The turtles are biting the gloves. 15. The dogs are fixing the car. 16. The ants are pushing the cake.

2. Structural Priming Condition

	Prime	Target
Singular	1. The chickens are eating the bread. 2. The wolves are drinking the juice. 3. The horses are licking the ice cream. 4. The witches are drawing the fish. 5. The boys are moving the box. 6. The bees are chasing the prince. 7. The girls are feeding the goat. 8. The donkeys are cutting the cake.	1. The monkey's flying the kite 2. The bird's reading the book. 3. The bear's holding the apple. 4. The crab's washing the chair. 5. The squirrel's closing the door. 6. The zebra's petting the cat. 7. The child's watching TV. 8. The cow's riding the bike.
Plural	9. The tiger's eating the soup. 10. The turkey's drinking the juice. 11. The giraffe's licking the apple. 12. The alien's cutting the carrot. 13. The pig's moving the desk. 14. The king's feeding the bird. 15. The deer's chasing the cow. 16. The sheep's drawing the fish.	9. The frogs are sweeping the floor. 10. The ants are digging a hole. 11. The firemen are pulling the car. 12. The turtles are building the house. 13. The lions are kicking the tree. 14. The ducks are biting the shoe. 15. The cats are fixing the truck. 16. The dogs are pushing the train.

3. Lexically-specific Priming Condition

	Prime	Target
Singular	1. The cow's eating the pizza. 2. The squirrels drinking the juice. 3. The monkey's licking the cheese. 4. The zebra's drawing the ball. 5. The crab's moving the block. 6. The child's feeding the goat. 7. The bird's chasing the worm. 8. The bear's cutting the cake.	1. The cow's flying the kite. 2. The squirrel's reading the book. 3. The monkey's holding the ball. 4. The zebra's washing the chair. 5. The crab's closing the door. 6. The child's petting the dog. 7. The bird's watching TV. 8. The bear's riding the bike.
plural	9. The cats are chasing the mouse. 10. The lions are feeding the bear. 11. The ducks are drawing the tree. 12. The ants are moving the jar. 13. The firemen are drinking the milk. 14. The dogs are licking the pie. 15. The frogs are cutting the bread. 16. The turtles are eating the grass.	9. The cats are sweeping the floor. 10. The lions are digging a hole. 11. The ducks are pulling the boat. 12. The ants are building the house. 13. The firemen are kicking the ball. 14. The dogs are biting the sock. 15. The frogs are fixing the bike. 16. The turtles are pushing the car.