

UNIVERSITY OF HAWAII LIBRARY

ARGUMENT/ADJUNCT ASYMMETRY
IN THE ACQUISITION OF INVERSION IN *WH*-QUESTIONS
BY ENGLISH-SPEAKING CHILDREN AND KOREAN LEARNERS OF ENGLISH:
FREQUENCY ACCOUNT VS. STRUCTURAL ACCOUNT

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

DECEMBER 2003

By

Sun-Young Lee

Dissertation Committee:

William O'Grady, Chairperson
Robert Bley-Vroman
Craig Chaudron
Ho-Min Sohn
Shuqiang Zhang

© Copyright 2003
by
Sun-Young Lee

ACKNOWLEDGEMENT

My life in Hawaii for last five years has been one of the most grateful moments in my life. I was blessed to be given such a precious time on a beautiful island with great teachers and friends who have affected my view of life in various ways. I would like to express my heartfelt thanks to those who were always willing to share their valuable time and thoughts with me during my graduate studies.

First of all, my deepest and endless gratitude goes to my advisor, Dr. William O'Grady, who has been a wonderful model of a great scholar and teacher. He was such a wonderful professor with a warm heart trying to support his students with such immediate and sincere responses. I was led to complete my degree by his endless encouragement and insightful advice throughout my graduate studies including the dissertation. I have to confess that he was extremely patient with me reading the various versions of the dissertation as well as other papers, sometimes shown even without having been proofread. Regardless of his busy schedule, he always welcomed me with an open mind whenever I liked to discuss my incomplete ideas and even trivial private matters. His way of studying linguistics also opened my eyes to the mysterious world of language and language acquisition as well as the beauty of the Korean language. I have always felt lucky to have such an admirable person as my advisor since I first met him five and a half years ago.

I also would like to express my gratitude and respect to Dr. Robert Bley-Vroman. He was always very supportive to my ideas with sincere and thoughtful advice. His

generosity and willingness to share his time and knowledge with me made me comfortable to ask his advice anytime without hesitation. His insightful and sharp comments on the dissertation as well as other papers always amazed me, leading me to come up with much better outcomes. It was always fun to have discussions with him in his office and in the classroom as well. Especially his lectures on corpus linguistics and second language analysis made the dissertation possible. I was so happy and proud taking his classes, being trained as an SLA researcher, which I had never imagined to be before meeting him during the first semester in Hawaii. I will never forget his warm smile and support.

I am also grateful to Dr. Ho-Min Sohn for his warm encouragement throughout my graduate studies and also for serving as one of my committee members. His lectures on Korean morphology and syntax enlightened me on Korean linguistics and taught me a new way of approaching the study of syntax. Many thanks also go to Dr. Shuqiang Zhang and Dr. Craig Chaudron for giving their valuable lectures on research methodology and statistics. They taught me what real research should be like and how to make it powerful. I especially thank Dr. Zhang for his tremendous time and patience to discuss my studies, sometimes answering my foolish questions. Without his kind help, I could not have finished my dissertation. I also remember Dr. Chaudron's frequent words of encouragement and witty jokes which made me feel happy and refreshed.

I am thankful to Dr. Hong Bae Lee, Sook Whan Cho, Young-Seok Kim and Tae-

Ok Kim for their warm welcome whenever I see them in conferences as well as at Sogang campus. Their excellent teaching of linguistics at Sogang University where I got my B.A and M.A. provided me with invaluable background knowledge about the various fields of linguistics, which I am really proud of.

I should thank the friends at Linguistics and their family for their warm friendship and hospitality. Especially Kyounkook Kim, Young-Geun Lee, and In-Kyu Park and their family were kind and nice enough to frequently invite me and my daughter to delicious meals and also to take care of my daughter while I was not able to be with her because of the classes and conferences. I am deeply indebted to them. I also thank the friends at school for their help in various ways and for sharing their time with me through happy gatherings: Sookeun Cho, Miseon Lee, Min Sun Song, Jae Yeon Kim, Mijung Lee, Insung Ko, Sang Yee Cheon, Kyung Suk Shin, Hye Kyung Hwang, Kyu Seek Hwang, another Sunyoung Lee, Marilyn Plumlee, Lan-Hsin Chang, Michiko Nakamura, Naoko Takahashi, Hsiu-Chuan Liao, Woody Mott, Paul Lasettre, Laura Sacia, Sumittra Suraratdecha, Preena Kangkun, Takako Yasuta, Mie Hiramoto, Katsura Aoyama, Hyeri Joo, Jinhwa Lee. I am deeply grateful to Hsiu-Chuan for taking care of printing the dissertation for me from a distance. My special thanks go to Siwon Park and Daniel Goldner for being such nice friends and co-workers throughout various projects we have worked on together. Their cordial friendship made my life in Hawaii cheerful and happy.

I'd like to express my deepest gratitude to Jeong Wook Hwang, Dae-Seon Kim,

Erica Sawhill and her daughter, Isabelle, and the Whittingtons (Milly & Brad) for being such wonderful roommates and friends to my daughter and me while I was struggling with my study. I am especially indebted to Milly & Brad Whittington for their warmest help and support during the writing period of the dissertation. They let me realize what real friendship is, and appreciate the hospitality of the American family. The wonderful memories they left with me including the dissertation defense party Milly threw for me will remain in my mind for ever. I also remember that Erica was nice enough to invite my daughter to various occasions and treats to keep her busy while I was studying, frequently bringing me a cup of cappuccino, and sometimes a glass of iced latte from her house.

I thank many enni's (sisters), too: Sunyoung Oh, Soo-young Kim, and Meesook Kim for showing their sincere care and encouragement for the little sister at various conference travels in the mainland of U.S.A and other countries as well. They were also willing to share my sufferings and pains and also theirs with me since my graduate school started. I also thank my friends, Sun-Kyoung Kim, Sang-Hwa Nam, Jin Sook Kim, Theresa Noh, Miran Yoon and Jeong Hwa Lee in Korea, and Cheol Back and his brother UnHyoung Back and their family in Hawaii for welcoming and encouraging me whenever I went back and forth between Korea and Hawaii. Without their continuous encouragement and love, I might not be able to have finished the program.

I am deeply indebted to my parents-in-law and grandparents-in-law for being so supportive for my study and career. I am especially grateful to my aunt-in-law and all

of her family members, particularly Bundang Grandma, for raising my daughter and taking care of my husband in their home during my absence. I also remember my Father and younger brother who would be so happy with my degree if they were alive. I'd like to give my warm gratitude to my Mother who always encouraged me in any situation.

Finally, my deepest and endless appreciation goes to my husband, Inki Lee. He supported my study physically as well as psychologically with his endless sacrifice, *patience and love for this selfish wife*. I realize how dependent I have been on him and his love since our first meeting almost seventeen years ago. I also have to thank my daughter, Eunkyong Rosa Lee for being such a wonderful child and friend of mine, *standing my absence many times but always bringing me happy thoughts and laughter*. As a small token of my gratitude and love, I'd like to dedicate this dissertation to my beloved husband and daughter.

TABLE OF CONTENTS

Acknowledgement.....	iv
Abstract.....	ix
List of Tables.....	xiii
List of Figures.....	xvi
Chapter 1: Introduction.....	1
Chapter 2: Argument/adjunct asymmetry in English-speaking children's acquisition of inversion in <i>wh</i> -questions.....	4
2.1. Structure-based generative theory.....	4
2.2. Lexical-based input frequency theory.....	10
2.3. Conclusion.....	13
Chapter 3: Argument/adjunct asymmetry in ESL learners' acquisition of inversion in <i>wh</i> -questions.....	15
3.1. Scrambled question task.....	18
3.2. Preference task.....	21
3.3. Problems.....	23
3.4. Comparative study of the first and second language development.....	25
Chapter 4: L1 Study.....	27
4.1. L1 Input Study: Corpus Analysis (Experiment 1).....	27
4.1.1. Methods.....	29
4.1.2. Results.....	37
4.1.3. Discussion.....	55
4.2. L1 Acquisition Study: Production (Experiment 2).....	60
4.2.1. Methods.....	62
4.2.2. Subjects.....	64
4.2.3. Procedure.....	64
4.2.4. Results and discussion.....	64
4.3. Comparison of Input and Acquisition.....	69
4.4. Conclusion.....	72
Chapter 5: L2 study	73
5.1. L2 Input study: Corpus Analysis (Experiment 3).....	73
5.1.1. <i>Wh</i> -questions in English textbooks in Korea.....	74
5.1.2. <i>Wh</i> -questions in scripts of American television situation comedies and movies.....	91
5.1.2.1. Scripts of television situation comedies	91
5.1.2.2. Movie scripts.....	114
5.1.3. Discussion.....	123
5.2. L2 Acquisition Study: Grammaticality judgment task (Experiment 4).....	126

5.2.1. Materials.....	129
5.2.2. Subjects.....	131
5.2.3. Methods.....	131
5.2.4. Results.....	132
5.2.5. Discussion	136
5.3 Comparison of input and acquisition.....	137
5.4 Conclusion.....	142
Chapter 6: Comparison between L1 and L2.....	144
6.1. Comparison between L1 and L2 input.....	144
6.2. Comparison between L1 and L2 acquisition.....	147
6.3. Conclusion.....	148
Chapter 7: Conclusion.....	150
7.1. Summary.....	150
7.2. Implications for input frequency theory.....	152
7.2.1 Lexical-based frequency.....	152
7.2.2 Frequency of structures.....	153
7.2.3 Negative effect of uninverted questions.....	155
7.3 Further research: Implicational relationship.....	159
7.4 Conclusion.....	162
Appendix I: Target items in Experiment 2 (Children's acquisition of <i>wh</i> -questions: an elicited production Task.....	164
Appendix II: Test items in Experiment 4 (ESL learners' acquisition of <i>wh</i> -questions: a grammaticality judgment task).....	165
References.....	167

LIST OF TABLES

<u>Table</u>	<u>Page</u>
4.1. Summary of corpora for L1 input study (selected corpora from CHILDES).....	28
4.2. Data adjustment.....	34
4.3. Frequency of <i>wh</i> -words in caregivers' utterances (individual periods).....	38
4.4. Frequency of <i>wh</i> -words in caregivers' utterances (combined periods).....	40
4.5. Frequency of <i>wh</i> -questions in caregivers' utterances (individual periods).....	42
4.6. Frequency of <i>wh</i> -questions in caregivers' utterances (combined period).....	43
4.7. <i>Wh</i> -questions in caregivers' utterances.....	45
4.8. Inversion rates for <i>wh</i> -questions in caregivers' utterances (individual periods).....	47
4.9. Inversion rates for <i>wh</i> -questions in caregivers' utterances (combined periods)	48
4.10. Frequency of formulaic expressions in caregivers' utterances.....	50
4.11. Frequency of formulaic expressions in each <i>wh</i> -question.....	51
4.12. Frequency of Non-formulaic inverted <i>wh</i> -questions and inversion rates.....	51
4.13. Mean numbers of inverted <i>wh</i> -questions.....	66
4.14. Individual data (Total number of inverted questions).....	68
4.15. Frequency of <i>wh</i> -word + <i>are you —?</i> pattern in caregivers' speech.....	71
5.1. Textbook corpus.....	75
5.2. Frequency of <i>wh</i> -words in textbooks	78

5.3.	Frequency of <i>wh</i> -questions in textbooks.....	80
5.4.	<i>Wh</i> -questions in textbooks.....	83
5.5.	Inversion rates for <i>wh</i> -questions in textbooks.....	85
5.6.	<i>Friends</i> Corpus.....	92
5.7.	Frequency of <i>wh</i> -words in <i>Friends</i>	93
5.8.	Frequency of <i>wh</i> -questions in <i>Friends</i>	95
5.9.	<i>Wh</i> -questions in <i>Friends</i>	97
5.10.	Inversion rates for <i>wh</i> -questions in <i>Friends</i>	98
5.11.	<i>Frasier</i> corpus.....	101
5.12.	Frequency of <i>wh</i> -words in <i>Frasier</i>	102
5.13.	Frequency of <i>wh</i> -questions in <i>Frasier</i>	103
5.14.	<i>Wh</i> -questions in <i>Fraiser</i>	105
5.15.	Inversion rates for <i>wh</i> -questions in <i>Fraiser</i>	106
5.16.	Movie corpus.....	115
5.17.	Frequency of <i>wh</i> -words in movies.....	115
5.18.	Frequency of <i>wh</i> -questions in movies	116
5.19.	<i>Wh</i> -questions in movies.....	118
5.20.	Inversion rates for <i>wh</i> -questions in movies.....	119
5.21.	Summary of L2 input study: Corpus Analysis.....	125
5.22.	Mean of Grammaticality Judgment Task.....	132

5.23.	Mean differences of grammaticality judgment task (inverted <i>wh</i> -questions)	134
5.24.	Mean differences of grammaticality judgment task (uninverted <i>wh</i> -questions)	135
5.25.	Frequency of [<i>wh</i> + <i>are you</i> /NP---?] (Inversion) in L2 corpus	139
5.26.	Frequency of [<i>wh</i> + <i>are you</i> /NP---?] (Inversion) excluding formulaic expressions in L2 corpus	140
5.27.	Frequency of [<i>wh</i> + <i>you</i> /NP <i>are</i> ---?] (Non-inversion) in L2 corpus	141
7.1.	Non-inversion rates for <i>wh</i> -questions in caregivers' utterances	156
7.2.	Inversion (non-formulaic) vs. non-inversion	158
7.3.	Total number of inverted questions from an elicited production task	160

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
4.1. Frequency of <i>wh</i> -words in caregivers' utterances.....	41
4.2. Frequency of <i>wh</i> -questions in caregivers' utterances.....	44
4.3. Inversion rates for <i>wh</i> -questions in caregivers' utterances.....	49
4.4. Inversion rates for <i>wh</i> -questions in caregivers' utterances (non-formulaic).....	53
4.5. Frequency of non-formulaic inverted <i>wh</i> -questions in caregivers' utterances.....	54
4.6. Frequency of non-formulaic inverted <i>wh</i> -questions in caregivers' utterances: <i>why</i> , <i>how</i> , and <i>who</i>	55
4.7. Consistency in relative frequency of <i>wh</i> -words in caregivers' utterances.....	56
4.8. Consistency in relative frequency of <i>wh</i> -questions in caregivers' utterances.....	57
4.9. Consistency in inversion rates for <i>wh</i> -questions in caregivers' utterances.....	58
4.10. Consistency in relative frequency of non-formulaic inverted <i>wh</i> -questions in caregivers' utterances.....	59
4.11. Children's production of <i>wh</i> -questions.....	66
5.1. Frequency of <i>wh</i> -words in textbooks.....	79
5.2. Frequency of <i>wh</i> -questions in textbooks.....	82
5.3. Inversion rates for <i>wh</i> -questions in textbooks.....	86
5.4. Frequency of inverted <i>wh</i> -questions in textbooks.....	88
5.5. Consistency in relative frequency of <i>wh</i> -questions in textbooks.....	89
5.6. Consistency in relative frequency of inverted <i>wh</i> -questions in textbook.....	90

5.7.	Frequency of <i>wh</i> -words in <i>Friends</i>	93
5.8.	Frequency of <i>wh</i> -questions in <i>Friends</i>	96
5.9.	Inversion rates for <i>wh</i> -questions in <i>Friends</i>	98
5.10.	Frequency of inverted <i>wh</i> -questions in <i>Friends</i>	100
5.11.	Frequency of <i>wh</i> -words in <i>Frasier</i>	102
5.12.	Frequency of <i>wh</i> -questions in <i>Frasier</i>	104
5.13.	Inversion rate for <i>wh</i> -questions in <i>Frasier</i>	106
5.14.	Frequency of inverted <i>wh</i> -questions in <i>Frasier</i>	108
5.15.	Consistency in relative frequency of <i>wh</i> -words in <i>Friends</i>	109
5.16.	Consistency in relative frequency of <i>wh</i> -words in <i>Frasier</i>	109
5.17.	Consistency in relative frequency of <i>wh</i> -questions in <i>Friends</i>	110
5.18.	Consistency in relative frequency of <i>wh</i> -questions in <i>Frasier</i>	111
5.19.	Consistency in inversion rates for <i>wh</i> -questions in <i>Friends</i>	112
5.20.	Consistency in inversion rates for <i>wh</i> -questions in <i>Frasier</i>	112
5.21.	Consistency in relative frequency of inverted <i>wh</i> -questions in <i>Friends</i>	113
5.22.	Consistency in relative frequency of inverted <i>wh</i> -questions in <i>Frasier</i>	113
5.23.	Frequency of <i>wh</i> -words in movies	116
5.24.	Frequency of <i>wh</i> -questions in movies.....	117
5.25.	Inversion rates for <i>wh</i> -questions in movies.....	119
5.26.	Frequency of inverted <i>wh</i> -questions in movies	121
5.27.	Consistency in relative frequency of <i>wh</i> -words in movies	121

5.28.	Consistency in relative frequency of <i>wh</i> -questions in movies.....	122
5.29.	Consistency in relative frequency of inverted <i>wh</i> -questions in movies.....	123
5.30.	Grammaticality judgment task: Inverted <i>wh</i> -questions	133
5.31.	Grammaticality judgment task: Uninverted <i>wh</i> -questions	133
7.1.	Children's production of <i>wh</i> -questions.....	157

CHAPTER 1

INTRODUCTION

This dissertation will investigate the existence of a possible argument/adjunct asymmetry in the acquisition of subject-auxiliary inversion in *wh*-questions by English-speaking children and ESL (English as a Second Language) learners, testing two approaches to the study of language acquisition: the structure-based generative approach and the lexical-based input frequency approach.

One of the most frequent errors in the development of *wh*-questions in English by children and ESL learners involves a failure to perform subject-auxiliary inversion errors (Labov, W. and T. Labov 1978, Ingram and Tyack 1979, Erreich 1984, Stromswold 1990, De Villiers 1991, Rowland & Pine 2000, 2003, Van Valin 2002). As the examples in (1) illustrate, learners sometimes produce *wh*-questions without inverting the subject and auxiliary.

(1) Inversion errors in *wh*-questions

- a. What you are making?
- b. Why he is laughing?

This type of error is found in various types of *wh*-questions as shown in (1): the *what* question in (1a) and the *why* question in (1b), for example. *Wh*-questions can be divided

into two types depending on the *wh*-phrase employed: arguments (e.g., *who* and *what*) and adjuncts (e.g., *how* and *why*).

Two different theories of children's acquisition of inversion in *wh*-questions have been proposed. The structure-based generative theory claims that a UG-based difference in structural properties between arguments and adjuncts is widely manifested in the acquisition process (Stromswold 1990, De Villiers 1991), including inversion. In contrast, the lexical-based input frequency theory argues that acquisition facts can be accounted for by the relative frequency of each pattern in the children's input (Rowland and Pine 2000, 2003), unless as the indirect consequence of a frequency difference. Based on Adam's data, they argue that there is no difference between argument and adjunct *wh*-questions in terms of inversion.

By investigating children's and ESL learners' input and their acquisition of inversion in *wh*-questions, this dissertation aims to answer the following questions:

1. Is there argument/adjunct asymmetry in the acquisition of inversion in *wh*-questions by L1 and L2 learners of English?
2. Is there any similarity between L1 and L2 regarding input and/or the acquisition of inversion in *wh*-questions?
3. Which approach provides a better explanation for the acquisition of inversion in *wh*-questions?

In order to investigate these questions, various research methodologies were adopted. A corpus study was used to investigate L1 and L2 input, while a picture-aided elicited production task and a grammaticality judgment task were conducted to examine L1 and L2 acquisition, respectively. Four types of *wh*-words (*what* and *who* for argument questions, *why* and *how* for adjunct questions) were compared in the L1 and L2 input study. *Who*, *why* and *how* questions were compared in the L1 production study, while *what*, *who*, *why* and *how* questions were compared in the L2 grammaticality judgment task.

The results will show that children and L2 learners do better at inversion in argument *wh*-questions than in adjunct *wh*-questions, and that this cannot be explained by the frequency-based input theory.

This dissertation is organized as follows. Chapter 2 will review previous studies of children's acquisition of inversion in *wh*-questions, focusing on the argument/adjunct asymmetry. Chapter 3 will review the previous L2 literature regarding the possibility of an argument/adjunct distinction in the L2 acquisition of *wh*-questions. Chapter 4 presents and discusses the results from experiments on L1 input and acquisition. Chapter 5 presents and discusses the results from experiments on L2 input and acquisition. Chapter 6 compares the results from L1 and L2 input and acquisition studies. Finally, Chapter 7 will summarize the studies answering the three questions posed in this chapter.

CHAPTER 2

ARGUMENT/ADJUNCT ASYMMETRY IN ENGLISH-SPEAKING CHILDREN'S ACQUISITION OF INVERSION IN *WH*-QUESTIONS

This chapter will discuss the argument/adjunct asymmetry in English-speaking children's acquisition of inversion in *wh*-questions, focusing on two approaches to the study of language acquisition: the structure-based generative approach and the lexical-based input frequency approach. *Wh*-questions in English contain either an argument *wh*-word (e.g., *who* and *what*) or an adjunct *wh*-word (e.g., *how* and *why*), and involve subject-auxiliary inversion (e.g. *what* is she *doing*?). Regarding children's acquisition of *wh*-questions, the structure-based generative theory claims that the difference in structural properties between the argument and adjunct will be revealed in the acquisition process. On the other hand, the lexical-based input frequency theory argues that the language acquisition order can be accounted for by the relative frequency of each pattern in the children's input.

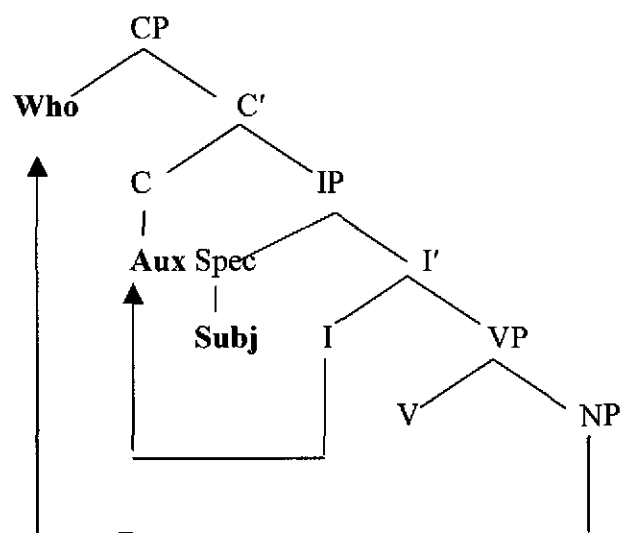
2.1 Structure-based generative theory

Many studies of the acquisition of *wh*-questions focus on the subject/object asymmetry rather than on the argument/adjunct asymmetry (Tyack and Ingram 1977, Wihelm and Hanna 1992, Stromswold 1988, 1995, Kim 1995, Yoshinaga 1996). Most of them ascribe the developmental difference between the subject question and the object

question to structural difference (e.g., depth of embedding in syntactic structure).

Relatively, there are a small number of studies on the argument/adjunct asymmetry in the acquisition of subj-aux inversion in *wh*-questions, the best known being Stromswold (1990) and De Villiers (1991). In GB theory, subj-aux inversion in English *wh*-questions involves overt movement of the *wh*-word to SPEC of CP and the auxiliary to C, resulting in subj-aux inversion as illustrated in the following tree diagram:

(1) *Wh*-questions with inversion



(e.g., *Who is she pushing?*)

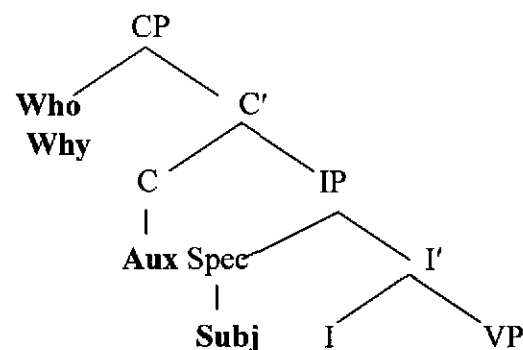
The structure in (1) clearly describes the movement of the *wh*-word, auxiliary and subject in object questions (subj-aux inversion). However, it is not clear whether subject question should be represented as the same structure as (1a) involving subj-aux inversion because the movement of inversion, if ever, is not shown overtly. That is, the

movement is vacuous (Chomsky 1986). From a learner's point of view, subject questions do not provide overt evidence for inversion rule, which means that the learners take the subject questions as uninverted *wh*-questions (Chomsky 1986). I will leave this problem open until Chapter 4 whether the subject questions should be treated as inverted ones or uninverted ones.

Based on the longitudinal data from twelve children, Stromswold found that children acquire inversion in argument *wh*-questions earlier than in adjunct *wh*-questions in the early stage of development of *wh*-questions. In her study, the overall inversion rate for argument questions was about 97 % whereas the inversion rate for adjunct questions was 85%. She attributes the difference between argument *wh*-questions and adjunct *wh*-questions to their syntactic properties based on GB theory. More specifically, the trace of the *wh*-word must be properly governed either by theta government or by antecedent government (Empty Category Principle). However, Stromswold proposes, since adjunct *wh*-words are not used to satisfy the argument requirement of the verb, they need not be associated with a trace. In contrast, this is not possible for argument *wh*-words without violating the theta criterion of the verb. Stromswold further argues that it is possible for the adjunct *wh*-word to be adjoined to the IP without leaving a trace. Those adjunct *wh*-questions in which the *wh*-word is adjoined to the IP do not trigger inversion, whereas all argument questions and those adjunct questions in which the *wh*-word is in the SPEC of CP do trigger inversion (See Stromswold 1990 for the detailed discussion of subj-aux inversion).

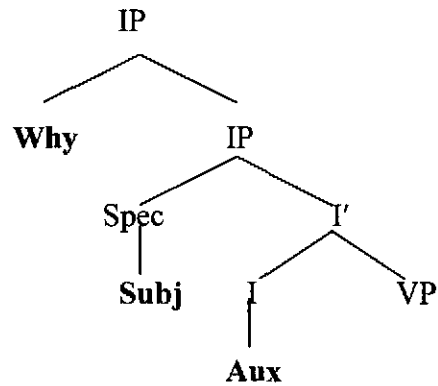
De Villiers' (1991) explanation of her findings is similar to Stromswold's in that adjunct *why* questions differ from other argument questions in terms of inversion. She found that children's acquisition of inversion with *why* questions was very late compared to other types of *wh*-questions. As a possible explanation for her finding, she proposed IP adjunction as in Stromswold (1990) (i.e., a TP, "Topic Phrase" is adjoined to IP). Thus, two structures are possible for adjunct *wh*-questions, whereas only one is possible for argument *wh*-questions in the early stage of acquisition, which causes the different inversion rate between argument questions and adjunct questions. The two different structures are shown in (2) (VP node is simplified).

(2) a. Inverted argument/adjunct *wh*-question



(e.g., *Who is she pushing?* / *Why is she smiling?*)

b. Uninverted adjunct *wh*-question



(e.g., *Why he is making bread?*)

Erreich's (1984) study shows positive evidence for a distinction between argument and adjunct in children's acquisition *wh*-questions as in Stromswold (1990) and De Villiers (1991). Her results from spontaneous and elicited production studies give the following inversion rates for each *wh*-question:

(3) Frequency of inversion by *wh*-term (from Erreich 1984, Table 4, P 588)

<i>wh</i> -term	Non-inverted			Inverted			Total
	N	/	%	N	/	%	
What	64		24	199		76	263
Where	43		28	112		72	155
Why	51		77	15		23	66
When	23		92	2		8	25
How	13		72	5		28	18
Who	1		6	15		94	16
Which one	0		0	8		100	8

As shown in (3), the inversion rates are very low in *why*, *when*, and *how* questions (23%, 8% and 28% respectively) compared to *what*, *who* and *which one* questions (76%, 94% and 100%).

However, De Villiers' claim about argument/adjunct asymmetry is not without problems. Consider Erreich's (1984) data in (3). The inversion rate of the *where* question in (3) is against the structure-based generative approach. In her experiment, *where* was used as an adjunct question but its inversion rate is very high (72%), very close to other argument questions (i.e., 76 – 100%), but very different from other adjunct questions (i.e., 8 – 28%).

In addition, consider the following data from Rowland & Pine (2000):

(4) Inverted/uninverted *wh*-word in Adam's speech (from Rowland & Pine 2000)

<u>Wh-word</u>	<u>Inverted</u>	<u>Uninverted</u>
What	55	15
Who	4	0
How	44	7
Why	3	33

There is a big difference between *how* questions and *why* questions. Both of them are adjunct questions, but Adam tends to invert only *how* questions, not *why* questions. This is not expected from the structural-based generative approach. Therefore, the structural difference between argument *wh*-questions and adjunct *wh*-

questions is not clearly shown in these data.

In conclusion, the structure-based generative approach is not convincing yet because of these conflicting data from previous studies.

2.2 Lexical-based input frequency theory

Rowland & Pine (2000, 2003) propose a lexical-based input frequency theory to account for the acquisition of inversion in *wh*-questions, arguing that children learn particular lexical-specific *wh*-word + auxiliary combinations instead of a general movement rule. Focusing on the fact that the *wh*-questions involve subj-aux inversion, they argue that children's acquisition order of subj-aux inversion in *wh*-questions can be predicted from the frequency of input from their caregivers. One of their main arguments is that children do not differentiate argument questions from adjunct questions, contrary to De Villiers (1991) and Stromswold (1990).

Based on Adam's longitudinal data, Rowland & Pine (2000) found that *what* questions are correctly inverted whereas *why* questions are not, arguing that this is due to the input frequency. That is, in Adams' mother's speech inverted *what* questions are more frequent than inverted *why* questions.

However, there are problems with Rowland & Pine's (2000) analysis. For example, even though certain types of *wh*-questions are not found in his mother's speech, the question type was correctly inverted in Adam's speech (see Table 5 on P175 in Rowland & Pine 2000). On the other hand, Adam inverted *what do* correctly with 37

instances in the input from mother, whereas he did not invert *why don't* even with 22 instances in the input. However, *why do* and *how do* are correctly inverted in Adam's speech with 13 and 14 instances of relevant input respectively. Furthermore, Adam never inverted *why don't* even with 22 instances of inverted input from mother. It is difficult to see how these facts can be explained by Rowland and Pine's input-frequency account.

Another problem of Rowland & Pine's (2000) study relates to their data. First of all, their data are not large enough to support their claim. The number of uninverted *why* questions in all periods investigated is 33 whereas that of uninverted *what* questions is 15. The number of inverted *why* questions is 3 whereas the number of inverted *what* questions is 55. At first glance, there seems to be a difference between the two, but we need more data from more children. As the data are from only one child's spontaneous speech, they are too limited to permit generalization.

A more general problem with Rowland and Pine's (2000) lexical-based input frequency approach, depending only on the frequency of certain words or types in the input, relates to the inversion rates. That is, we need to think about the inversion rate of a particular type of *wh*-question in the input as well as the frequency of inverted questions. For example, even though inverted *what* questions are very frequent in the input, as Rowland and Pine found, if the inversion rate is low, children might be confused in figuring out when inversion happens and when it does not. Notice that inversion is required with *what* in direct questions (e.g., *what is he looking for?*), but not in indirect

questions (*do you know what he is looking for?*) and exclamatives (e.g., *what a pretty doll it is!*). The other types of *wh*-words have to be taken into the same consideration. For example, *how* can be used in a direct question (e.g., *how old is she?*), an indirect question (e.g., *do you know how old she is?*) and an exclamation (e.g., *how old she is!*). *Who* can also be used in different types of sentences – a direct question (e.g., *who is he helping?*), an indirect question (*do you know who he is helping?*), and a relative clause (e.g., *this is the boy who she loves*). Therefore, in formulating the subj-aux inversion rule, children must be able to distinguish direct from indirect *wh*-questions, and even from exclamatives and relative clauses.

Finally, the concept of generalization is not clear in Rowland and Pine's input-frequency theory. If it is true that the learners draw some kind of generalization (i.e., the inversion rule) from the input frequency, the generalization is presumably based on the frequency of certain structured patterns in the input (i.e., frequency of inverted *wh*-questions) as Rowland & Pine suggested. If Rowland & Pine is right in saying that *what* questions are more correctly inverted than any other type of question because of the higher frequency of inverted *what* questions in the input, children will be better at inversion in *what* questions than any other question.

In fact, it seems plausible to think that inverted *what* questions are more frequent than inverted *who* questions in children's input considering the animacy of each *wh*-word. As *who* refers to an animate object, it's likely that *who* is used as a subject question (which does not involve subj-aux inversion). In contrast, *what* refers to an inanimate

object, so that it is likely that *what* is used as an object question (which involves subj-aux inversion).

So, the subj-aux inversion rule, which learners acquire with one of the argument *wh*-questions first (i.e., *what* questions because of the frequent inverted *what* questions in the input), may be generalized to the same type of *wh*-questions (i.e., *who* questions, another type of argument question). Therefore, even though inverted *who* questions are very low in frequency, it may not affect learners' generalization of the inversion rule from *what* to *who* once they figure out the status of *who* as an argument *wh*-question word. In this case, the lexical-based input frequency account must depend on some kind of learners' ability to generalize the syntactic property of each *wh*-question (i.e., argument vs. adjunct), regardless of frequency of each *wh*-word. On the other hand, it is not clearly mentioned in Rowland and Pine how children finally generalize inversion rule depending on the input frequency.

In conclusion, it is difficult at this point to accept Rowland & Pine's argument that children's acquisition of inversion in *wh*-questions can be accounted for by the input-frequency account, rejecting Stromswold and De Villiers' structural account.

2.3 Conclusion

I have discussed the two theories of language acquisition: how structure-based generative theory and lexical-based input frequency theory explain English-speaking children's acquisition of inversion in *wh*-questions, focusing on the possible existence of

argument/adjunct distinction. These studies from both sides are not without problems, as has been discussed so far. We have also seen that the data from various studies are not clearly supportive to either of the theories.

It is possible that the conflicting results stem from the methodology of the studies. For example, most studies (Stromswold 1990, De Villiers 1991, Rowland & Pine, 2000) used mainly children's spontaneous speech. Using spontaneous speech data to collect the inversion rate might be limited. Children might be using certain fixed expressions containing a specific *wh*-word, which will cause high inversion rate for the *wh*-word (e.g., *how do you do that?*) if the child produce only that question in the period of data collection. It is possible that the child produces *how* questions without inversion with other auxiliaries. With these types of data, we cannot be sure if the child has acquired the inversion rule per se.

Therefore, it is difficult to decide which theory provides a better explanation for children's acquisition of inversion in *wh*-questions, Rowland & Pine's (2000) lexical-based input frequency account or Stromswold (1990) and De Villiers' (1991) structure-based generative account. We are still in need of not only the input frequency data but also acquisition data from experimental studies (rather than spontaneous studies) for comparison in order to evaluate the two theories.

In conclusion, there is still a need to address the question of whether children differentiate argument questions from adjunct questions in their acquisition of *wh*-questions and subj-aux inversion, and by which theory the results can be explained better.

CHAPTER 3

ARGUMENT/ADJUNCT ASYMMETRY IN ESL LEARLERS' ACQUISITION OF INVERSION IN *WH*-QUESTIONS

In the field of second language acquisition, there seem to be no studies that specifically address the question of an asymmetry between argument *wh*-questions and adjunct *wh*-questions related with subj-aux inversion. Most studies on *wh*-questions focus on the ESL learners' developmental sequences of interrogative sentences, and the implicational relationship among the different stages. Citing Pienemann, Johnston, and Brindley's (1988) study, Spada and Lightbown (1999) propose the following developmental sequences:

(1) Developmental Stages in English Questions (Spada and Lightbown 1999)

- | | |
|---------|--|
| Stage 1 | Single words or fragments (e.g., <i>A spot on the dog?</i>) |
| Stage 2 | SVO with rising intonation (e.g., <i>A boy throw the ball?</i>) |
| Stage 3 | Fronting

<i>Do</i> -fronting (e.g., <i>Do the boy is beside the bus?</i>)

<i>Wh</i> -fronting (e.g., <i>What the boy is throwing?</i>)

Other fronting (e.g., <i>Is the boy is beside the bus?</i>) |
| Stage 4 | <i>Wh</i> -with copula BE (e.g., <i>Where is the ball?</i>)

<i>Yes-No</i> questions with aux inversion (e.g., <i>Is the boy beside
the garbage can?</i>) |
| Stage 5 | <i>Wh</i>-with auxiliary second (e.g., <i>What is the boy throwing?</i>) |

This developmental sequence has been supported by other studies (White et al. 1991, Spada & Lightbown 1999). Wode (1978b) includes two more stages involving *do*-support: *do*-support with main verbs in *wh*-questions > *do*-support with main verbs in *yes/no* questions.

On the other hand, question formation has also been studied from a typological perspective, based on Greenberg's (1966) proposal of implicational universals of question formation: (i) *wh*-inversion implies *wh*-fronting and (ii) *yes/no* inversion implies *wh*-inversion. This typological implicational relationship among the various question forms was supported by Eckman, Moravcsik and Wirth (1989).

As shown in (1), subj-aux inversion with *wh*-questions belongs to the last stage (Stage 5) of ESL learners' development of English questions. Like English-speaking children, adult ESL learners make inversion errors with *wh*-questions (Stage 3). Even though most studies reporting the learners' scores on tests of *wh*-question formation do not divide the *wh*-questions into argument *wh*-questions or adjunct *wh*-questions, some positive results for the asymmetry between argument questions and adjunct questions are still found. For example, Haznedar (2003) analyzed different types of non-subject *wh*-questions with a set of longitudinal data from one Turkish child, Erdem, who started learning English as her second language at about age four. She reports the number of inversion errors in each type of *wh*-question. The results are shown in (2), re-arranged in order to show the rates for inversion errors.

(2) Inversion errors in *wh*-questions (from Haznedar 2003, Appendix 6, P40)

<i>wh</i> -term	Inversion errors (N) /	Total	Inversion errors (%)
What	11	184	6.0
Which	2	18	11.1
Where	3	64	4.7
How	4	38	10.5
Why	3	32	9.4

As shown in (2), the rates for inversion errors are generally higher in adjunct questions (i.e., *how*, and *why*), than in argument questions (i.e., *what*). There were 18 *who* questions in the corpus (Appendix 2 Page 36 in Haznedar (2003)) and 10 of them were non-subject questions (Appendix 4 Page 38). However, the inversion errors with *who* questions are not reported (Appendix 6 Page 40) (i.e., 100 % inversion rate). Therefore, it seems that inversion errors are more frequent in adjunct *wh*-questions than in argument *wh*-questions.

However, even though the data are positive, it is not strong enough to confirm the argument/adjunct asymmetry because of the highest inversion error rate for *which* questions (an argument question) and the lowest inversion error rate for *where* questions (an adjunct question).

While Haznedar's (2003) study is based on only one, very young (four-year-old) child's naturalistic data, Spada & Lightbown (1999) provide data from an experimental study with more and older learners, reporting actual test items and the subjects' test scores in their study. Testing Pienemann's (1985) teachability hypothesis, Spada &

Lightbown (1999) took 150 francophone children (age 11-12) at stage 2 in Figure 3; and gave them focused instruction on stage 4 and 5 questions about one hour everyday for two weeks. The study included pre-tests and post-tests with various measures such as an oral production task, a preference task, and a scrambled question task. As their instruction focused on stage 4 and 5 questions, which include *wh*-questions with subj-aux inversion, their detailed reports on the pre- and post-test scores show learners' development of subj-aux inversion with specific question forms. I will reanalyze the data focusing on argument/adjunct asymmetry in subj-aux inversion. Special attention will be paid to the results of the Scrambled Question task and Preference task.

3.1 Scrambled Question task

In the Scrambled Question task, the subjects rearranged word cards to form an appropriate question. The task contained 20 cartoons accompanied by scrambled words with which to make a correct question. A part of the results on this task is presented in (3).

(3) Scrambled Questions Task: Grammatical Questions and Most Frequent Ungrammatical Questions (from Spada & Lightbown 1999:11)

Question	Percentage of Students Who Accepted	
	Pre-test	Post-test
1. What can I get for \$5.00?	40	57
1' *What I can get for \$5.00?	33	30
2. Where can I buy ice cream?	37	57
2' *Where I can buy ice cream?	47	33
3. When are we going to eat?	27	43
3' *When we are going to eat?	53	57
4. When is my mother coming home?	20	27
4' *When my mother is coming home?	63	73
5. When does the train leave?	20	33
5' *When the train does leave?	47	47
6. How can birds fly?	13	23
6' *How birds can fly?	67	60

Unfortunately, there is only one argument question in the data (*what* in 1) whereas there are five adjunct questions. Nevertheless, we can compare *what* questions with the other adjunct questions. First of all, in the pre-test, the score on the inverted *what* question (1) (which is grammatical) is higher than that on the uninverted *what* question (1') (which is

ungrammatical) (40%:33%). By contrast, the scores on inverted adjunct questions (i.e., *where*, *when* and *how*), which are grammatical, are lower than the scores on uninverted adjunct questions, which are ungrammatical, all throughout the examples ($2 - 6 < 2' - 6'$). This suggests that the learners' sensitivity to uninverted questions was different between argument questions and adjunct questions. The same tendency is found in the post-test, too. There, the percentage of students who accepted uninverted *what* questions (30% in 1') is smaller than that of students who accepted uninverted *where*, *when*, and *how* questions (33% in 2', 57% in 3', 73% in 4', 47% in 5' and 60% in 6'). Consequently, the asymmetrical results of the pre-test and post-test for *what* questions and *where*, *when* and *how* questions seem to indicate a possible argument/adjunct asymmetry in learners' production of *wh*-questions.

Secondly, even after the two weeks of focused instruction on *wh*-questions with subj-aux inversion, the learners' sensitivity toward subj-aux inversion did not improve in adjunct questions, in contrast to argument questions. The percentage of the subjects who accepted uninverted *what* questions decreased from 33% to 30% (*what* in 1') after the instruction. In contrast, the percentage of subjects who accepted uninverted adjunct questions decreased only slightly (67% > 60% *how* in 6'), remained the same (47% > 47% *when* in 5'), or even increased (53% > 57% *when* in 3' and 63% > 73% *when* in 4'). The instructional effect was high only in the *where* question (47% > 33% in 2'). Even though we cannot directly compare argument *wh*-questions and adjunct *wh*-questions with these fragmentary and uncontrolled data, there seems to be a tendency for less

improvement in inversion in adjunct *wh*-questions than in argument *wh*-questions.

3.2 Preference task

In the Preference task, the subjects judged the grammaticality of the sentences by four types of answers: (i) accept one of the sentences, (ii) accept both of the sentences, (iii) reject both, or (iv) don't know. Thirty pairs of items were tested including 22 pairs of target questions and 16 distracters. For the purpose of the study, only relevant parts of the results are presented here.

(4) Preference Task: Rate of Acceptance of Each Question (Grammatical and ungrammatical) (Adopted from Spada & Lightbown 1999:12)

Percentage of Students Who Accepted

<u>Question</u>	<u>Pre-test</u>	<u>Post-test</u>
1. What is your brother doing?	27	47
1' *What your brother is doing?	60	57
2. What can we watch on TV tonight?	53	70
2' *What we can watch on TV tonight?	67	67
3. When are you going to eat breakfast?	83	87
3' *When you are going to eat breakfast?	30	37
4. Where can I buy a bicycle?	70	80
4' *Where I can buy a bicycle?	57	47
5. When can you visit your uncle?	57	83

5' *When you can visit your uncle?	53	50
6. Where are your parents working?	50	57
6' *Where your parents are working?	57	70
7. Where is the teacher going?	43	63
7' *Where the teacher is going?	67	93
8. Why do children like McDonalds?	33	43
8' *Why children like McDonalds?	83	70
9. Why is he at home today?	33	57
9' *Why he is at home today?	73	73
10. Why can fish live in water?	17	30
10' *Why fish can live in water?	83	80

The results of the Preference task are similar to those of the scrambled question task in terms of argument/adjunct asymmetry. The learners are more likely to accept uninverted adjunct *wh*-questions than uninverted argument *wh*-questions. For example, in the post-test after instruction, the percentages of the students that accepted uninverted adjunct *where* and *why* questions (e.g., 70% in 6', 70% in 8', 73% in 9', and 80% in 10') are higher than those of the students that accepted uninverted *what* questions (e.g., 57% in 1' and 67% in 2'). It is noticeable, though, that among *where* questions, the uninverted "argument" *where* question (7') (i.e., argument of the verb, *go*), was accepted grammatical by more students than uninverted "adjunct" *where* questions (4') and (6') (i.e., 93% vs. 47% and 70% respectively).

In sum, the results of the Preference task as well as the Scrambled Question task seem to show argument/adjunct asymmetry in the inversion in *wh*-questions.

However, we still need data from more controlled experiments. Since the data reviewed here were from Spada & Lightbown (1999), whose aim and purpose of the study was different from ours, the results cannot provide clear answers for our research question. Nevertheless, their study reminds us of some possible problems for us to control in the present study.

3.3 Problems

First of all, the learners in Spada & Lightbown's (1999) study were children aged 11-12, not adults. Many SLA studies have found that there is little difference between children and adults in terms of their developmental sequence (including the acquisition of question formation). Nevertheless, we are not sure of adult learners' behavior particularly related to the argument/adjunct asymmetry in their acquisition of subj-aux inversion in *wh*-questions.

Second, the learners' L1 is French in Spada & Lightbown's (1999) study. In French there are several ways to form *wh*-questions. *Wh*-phrases can be *in situ* as in (5a), or by movement as in (5b) and (5c). It is also possible to leave subjects and verbs in uninverted position as in (5b,c) or invert them (5d). Inversion is not obligatory with pronouns but is not permitted with noun phrases as in (5e).

- (5) a. Marie demeure où?
 Marie live where
 'Where does Mary live?'
- b. Où est-ce que Marie demeure?
 Where is-it that Mary live
 'Where does Mary live?'
- c. Où c'est qu'elle demeure?
 Where it's that she live
 'Where does she live?'
- d. Où demeure-t-elle?
 Where live she
 'Where does she live?'
- e. *Où demeure Marie?
 where live Mary
 'Where does she live?' (Examples from White et al.1991: 419)

As French allows various forms of *wh*-question formation as illustrated in (5), including optional inversion, it may have affected the learners' acquisition of *wh*-questions of English. For example, Spada & Lightbown (1999) found that the subj-aux inversion was more frequent in *wh*-questions with a pronoun subject than in those with a noun phrase subject (P.16). This seems to be due to L1 transfer, because in French, *wh*-questions with a noun phrase subject do not allow inversion as in (5e). Therefore, it would be better to have subjects whose L1 does not have influential factors. For example, languages such as Korean and Japanese are *wh-in-situ* languages and do not allow subj-aux inversion.

Third, we need more controlled data balancing the number of argument questions

and adjunct questions for better comparison.

Fourth, as mentioned above, it has been shown that the form of the subject influences inversion in *wh*-question formation. This might be because of an L1 transfer effect as Spada & Lightbown (1999) pointed out. However, it might also be an interlanguage characteristic itself regardless of L1 transfer. Therefore, we need to control the form of the subject in the experiment of subj-aux inversion.

Fifth, as Spada & Lightbown (1999) also pointed out, it is possible for the learners to use the knowledge of fixed expressions involving *wh*-questions. For example, one of the *what* questions in Spada & Lightbown (1999), *what can I get for—?* might be such a fixed expression. With these kinds of examples, we cannot exactly measure the learners' knowledge about subj-aux inversion.

Finally, we need input frequency data to compare the learners' behavior of inversion and the frequency of inversion in their input. The L2 learners might get different kinds of input from L1 learners in terms of inversion in *wh*-questions. We are not sure if the learners' behavior (i.e., argument/adjunct distinction) might be co-related with *argument/adjunct distinction of input frequency*, for which we have no data yet.

3.4 Comparative study of first and second language development: the role of input

There have been findings showing that first language learners and second language learners seem to go through similar developmental stages for particular structures such as relative clauses and negation, and so on (Larsen-Freeman and Long

1991). It has also been found that L2 learners follow a similar sequence regardless of their learning situation (i.e. instructed or uninstructed learning). The information about the similarities is important to explain the learning mechanism of language acquisition whether it is a language-specific acquisition device or a more general problem-solving device based on input frequency. The present study, which investigates similarities between L1 and L2 development by comparing the learners' input and acquisition of inversion in *wh*-questions will try to show that language development is more likely controlled by a faculty-specific language acquisition device than by learners' input frequency.

CHAPTER 4

L1 STUDY

This chapter consists of two parts. The first investigates children's input from their caregivers. The second involves children's production. Four types of *wh*-questions were investigated in the input study: *what*, *who*, *how*, and *why* questions. Based on the input study, *who*, *how* and *why* questions were selected for the production study. *What* was excluded because of the enormously high frequency of *what* questions compared with other types of questions and also because of a practical reason to be explained later. *Where* was also excluded from both studies because it sometimes functions as an argument and sometimes as an adjunct.

4.1 L1 Input study: corpus analysis (Experiment I)

In order to find whether there is any similarity among caregivers' speech in terms of the relative frequencies of *wh*-questions, caregivers' speech in the CHILDES database was analyzed (MacWhinney 1995, 1998). If there is a similarity among caregivers in terms of the relative frequency of inverted *wh*-questions among the different *wh*-words, we may assume that children receive similar kinds of input from their caregivers.

The speech of the caregivers of four children (Peter, Ross, Shem and Hass) was analyzed. In some files in the database, there are several people who are interacting with the child (e.g., mother, father, investigator, siblings and friends). However, only

the main interlocutor was chosen for our analysis, based on the number of utterances (i.e., the person who is talking to the child more frequently than others — usually the mother and the investigator). Another reason for excluding the other participants' utterances in the data is to control the data to generalize our results (i.e., children's input from their "caregivers", not from "anybody") even though utterances of other people can serve as input to children). In addition, the number of utterances of other participants in each data was very small, except Peter's. In Peter's data, utterances from three people (i.e., mother, and investigators, Pat and Loi) were included because of the similar number of utterances among them. A summary of the CHILDES corpus investigated in this study is presented in the following table.

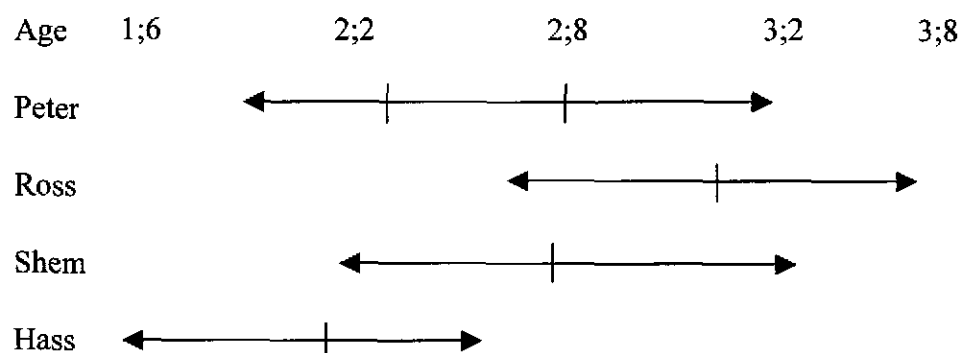
Table 4.1 Summary of corpora for L1 input study (selected corpora from CHILDES)

Folder name	File name	Total no. of files	Child's age	Caregivers	Word count
Bloom 70	Peter 1-9	10	1;9-2;2	MOT 1, PAT 1, LOI 1	161,764
	Peter 10-16	7	2;3-2;7	MOT 2, PAT 2, LOI 2	173,781
	Peter 17-20	4	2;8-3;1	MOT 3, PAT 3, LOI 3	100,990
Machwin	Ross 20-31	12	2;6-2;11	FAT 1	52,800
	Ross 32-40	9	3;0-3;5	FAT 2	70,744
Clark	Shem1-19	19	2;2-2;7	INV 1	134,941
	Shem20-40	28	2.8-3;2	INV 2	136,960
Cornell	Hass1917-1923	6	1;6-1;11	MOT 1	15,907
	Hass1924-1929	4	2;2-2;5	MOT 2	13,203
Total		99			861,090

- MOT: mother, PAT: Pat, LOI: Loi, FAT: father, INV: investigator, (Pat and Loi are investigators)

The data of each child are divided into two or three periods according to his or her age in order (i) to find any possible differences in input depending on the children's age and (ii) to include various age ranges from different children's data. The following age ranges from the children's data were investigated in this study.

(1) Age ranges investigated from each child's data



- The periods are divided into two or three according to the different ages (See Table 4.1 for detailed information).

4.1.1. Methods

All the caregivers' utterances that included *wh*-words were sought out with a search program, MonoConc Pro, and analyzed manually. As the coding of CHILDES is arranged by line, and one line includes only one utterance by one speaker, it is possible to find a particular speaker's utterances with a particular word using the search program.

Consider the following transcript from Peter's data.

(2) *MOT: what's this?

*PET: a dog.

Using MonoConc's "text search" with the search string, *MOT @ what (@ is a wild card), we can call up all the lines that include *MOT and *what* from a designated corpus.

However, it is important to adjust the range of the wild card depending on the search words, and manual analysis is necessary after the search for the sake of accuracy. Special attention needs to be paid to the wild card in this study. The following factors must be taken into consideration. First, a *wh*-word usually appears at the beginning of the utterance, but it is also possible for *what* to appear in different positions in the utterance. Compare the following two utterances:

(3) a. *MOT: what are you doing?

b. *MOT: Peter, I don't know what you are saying.

In (3a), *what* appears in the first position of the utterance, whereas it appears in the fifth position in (3b). In other words, the number of intervening words between *MOT and *what* is zero in (3a), whereas it is four in (3b) (i.e., *Peter*, *I*, *don't*, and *know*). Therefore, in order to include utterances like (3a), the range of the wild card should start from zero. And in order to include utterances like (3b), the upper range must be bigger than four.

Still, it is possible to miss lines that do not fall within this range.

Second, it is possible to catch two consecutive lines at the same time as follows.

- (4) *MOT: what?
 *PET: umm
 *MOT: what did you say?

Since there are four words between the first **MOT* and the second *what*, the command treats these three lines as one line and regards it as one line in the frequency count.

Therefore, a manual search must follow the automatic search to include the missing items.

On the other hand, the search command can also find lines in which the *wh*-word appears in the other speaker's utterance:

- (5) *MOT: did you eat?
 *PET: what?

In (4), there are four words intervening between **MOT* and *what* (i.e., *did*, *you*, *eat*? and **PET*) but the *wh*-word *what* belongs to Peter's utterance, not the mother's. These kinds of cases must be excluded during the manual search.

Third, it is also important to consider the particular characteristics of the utterances in the CHILDES database, which is a dyad between a caregiver and a very young child. The utterances are quite short and simple, which means that we do not

need to have a very large range for the wild card.

Considering these influential factors, the range of the wild card was set from zero to five in this study and it worked quite well at calling up most of the lines with *wh*-words. The search command included not only *what*, *who*, *how* and *why* but also *what's*, *who's*, *how's* and *why's*, because the transcriptions in the CHILDES database sometimes used these contracted forms instead of separate forms (i.e., *what is*, *who is*, *how is*, and *why is* respectively).

Various combinations of *wh*-words and subjects and auxiliaries were found and analyzed. In the analysis, each sentence was coded manually in the following manner.

(6) Coding

Inverted questions:	i	(e.g., <i>What are you doing?</i>)
Uninverted questions:	u	(e.g., <i>Do you know what I am doing?</i>)
Other questions:	o	(e.g., <i>What?</i> , <i>What about it?</i>)
Subject questions:	s	(e.g., <i>Who is going?</i>)
Relative clauses:	r	(e.g., <i>Do you know the little boy who is crying all the time?</i>)
Exclamatives:	e	(e.g., <i>What a good boy!</i>)

Multiple coding was done for those lines that have more than two relevant examples of the particular *wh*-words as in (3). On the other hand, lines in which a *wh*-word belonged

to a different speaker were marked “x” and excluded from the frequency count. As the frequency of *wh*-words is enormous in some cases (e.g., there were over 1,000 tokens of *what* in Shem’s Investigator’s utterances), it was thought impractical to code every single line in those cases. In those cases, the coding was done for randomly chosen lines (e.g., every two lines, or every 10 lines). To estimate the total for each coding the number in the sample was later multiplied by the appropriate number. For example, if the coding was done for every second line, the numbers for each coding were multiplied by two to arrive at an estimate for the total. The total number of *wh*-words and the total number of *wh*-questions are calculated based on the total numbers for each coding. The following table shows which questions in which data were obtained by this estimation method.

Table 4.2 Data adjustment

Cargivers	Total <i>wh</i> -word			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-MOT 1	15	17	5	62 x 2
P-MOT 2	31	42	33	103 x 2
P-MOT 3	10	9	6	37
P-PAT 1	55	76	61	121 x 5
P-PAT 2	37	88	75	111 x 5
P-PAT 3	31	64	36	88 x 4
P-LOI 1	44	60	89	118 x 5
P-LOI 2	38	48	45	105 x 5
P-LOI 3	39	69	42	75 x 4
R-FAT 1	112	102	51	86 x 5
R-FAT 2	68 x 2	71 x 2	74	101 x 5
S-INV 1	64 x 2	81 x 2	81	72 x 20
S-INV 2	61 x 3	46 x 5	90	60 x 20
H-MOT 1	19	7	15	165 x 5
H-MOT 2	38	14	25	69 x 3

In Table 4.2, the numbers without an “x” mark are the counts of the sample and the number after the “x” mark show the sampling rate. For example, 62 x 2 in P-MOT 1 means that there were originally 124 cases of *what* questions in Peter’s mother’s utterances in Period 1 but every second example of the *what* questions was analyzed. After that, the numbers for each coding were multiplied by 2 to estimate the total number. This estimation technique was used where the number of lines was greater than 100. Estimation was made mainly for *what* questions and three cases of *how* and *why*

questions, providing us with estimated numbers. Except for these cases, other cases are original numbers (not estimated). It would be possible to reduce the size of the corpus (i.e., the number of files) instead of estimating based on the whole corpus. However, as the relative frequency of *wh*-words seems to be sensitive to the activities in which caregivers and children are involved at the time of recording (i.e., each file), it seemed more reasonable to include all the files to reduce the variability as much as possible.

Three frequency searches were mainly conducted and compared. First, the total of all six codings (i, u, o, s, r, and e) were counted as the total number of *wh*-words. Second the total of four codings (i, u, o, and s) were counted as the total number of *wh*-questions. Third, each *wh*-question was separated into one of the three groups: inverted (coding “i”), uninverted (coding “u”) and other questions (coding “s” and “o”). In the analysis, the “total number of *wh*-words” indicates the number of occurrences of each *wh*-word in each caregiver’s utterances. The “total number of *wh*-questions” indicates the number of *wh*-questions (which is smaller than the number of *wh*-words). “Inverted” indicates the number of questions with the word order, [*wh*-word+ Aux+Subj] (coded as “i”). “Uninverted” indicates the number of questions with the word order, [*wh*-word + Subj + Aux] (coded as “u”). Uninverted questions include mainly indirect questions (e.g., *I do not know what you want*). “Other questions” includes two types of questions: (i) those *wh*-questions that do not contain any verbs (e.g., *who?* or *what about* –?, coded as “o”) and (ii) *what* and *who* subject questions (e.g., *what makes you think so?*, *who is going?*, coded as “s”). *What* and *who* subject questions are included in this

category because they do not belong to [*wh*-word + Aux + Subj] or [*wh*-word + Subj + Aux].

It is possible that *what* and *who* subject questions can be treated as uninverted *wh*-questions if we follow the idea of vacuous movement (Chomsky 1986b). Children construct the grammar based on overt evidence. However, subject *wh*-questions do not give the learners any overt evidence for movement of *wh*-word and auxiliary. In addition, auxiliary insertion is not possible in subject *wh*-questions (i.e., **who did help Mary?* vs. *who helped Mary?*). Therefore, it is plausible to categorize subject *wh*-questions as uninverted questions. Nevertheless, they are categorized into “other questions” in this study, taking into account both Chomsky’s and Rowland & Pine’s point of view, and particularly learners’ point of view. Rowland and Pine (2000) counted [*wh* + aux] combinations as input for inverted *wh*-questions. According to their counting, *who* subject questions involving a [*wh* + aux] combination (e.g., *who will go?*) in caregivers’ utterances should be counted as input for inverted questions. However, as discussed above, *who* subject questions do not provide very clear evidence for subj-aux inversion, compared to other types of questions involving the same combination [*wh* + aux] (e.g., *how do you do that?*). *How* and *why* questions which involve either [*wh*-word + Aux + Subj] or [*wh*-word + Subj + Aux] show a direct contrast involving subj-aux inversion whereas *who* subject questions involving [*who* + aux] do not. However, non-subject *who* questions involving [*wh* + aux + subj] (e.g., *who are you waiting for?*) provides as clear evidence for inversion in *wh*-questions as *how* and *why* questions.

Considering this learner's point of view, it was decided that *who* subject questions do provide clear evidence for either inverted questions or uninverted questions. Therefore, they were categorized as "other questions".

In the analysis of each *wh*-question, only one-word *wh*-word phrases (e.g., *how?*), not multi-word *wh*-phrases (e.g., *how many people?*), were counted as *wh*-elements because of differences in their grammatical status. For example, in the question, *how many do you have?*, the *how* phrase is not an adjunct, but an argument phrase. In the question, *what time do you leave?*, the *what* phrase is not an argument, but an adjunct phrase. The category "Subj" includes both pronominals and regular noun phrases. For the purpose of preliminary analysis, "Aux" includes copula (e.g., *who is she?* where *is* is not an auxiliary verb but a main verb) as well as auxiliary verbs (e.g., *who is he looking at?*). After the initial compilation, inversion rates for each *wh*-question are calculated and compared. Finally, the relative frequency of each type of inverted *wh*-questions was compared.

4.1.2 Results

The total number of each *wh*-word is reported in the following table:

Table 4.3 Frequency of *wh*-words in caregivers' utterances (individual periods)

Caregivers	Total number of <i>wh</i> -word					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
P-MOT 1	15	17	5	124	161	9%	11%	3%	77%	100%
P-MOT 2	31	42	33	206	312	10%	13%	11%	66%	100%
P-MOT 3	10	9	6	37	62	16%	15%	10%	60%	100%
Total	56	68	44	367	535	10%	13%	8%	69%	100%
P-PAT 1	55	76	61	605	797	7%	10%	8%	76%	100%
P-PAT 2	37	88	75	555	755	5%	12%	10%	74%	100%
P-PAT 3	31	64	36	352	483	6%	13%	7%	73%	100%
Total	123	228	172	1,512	2,035	6%	11%	8%	74%	100%
P-LOI 1	44	60	89	590	783	6%	8%	11%	75%	100%
P-LOI 2	38	48	45	525	656	6%	7%	7%	80%	100%
P-LOI 3	39	69	42	300	450	9%	15%	9%	67%	100%
Total	121	177	176	1,415	1,889	6%	9%	9%	75%	100%
R-FAT 1	112	102	51	430	695	16%	15%	7%	62%	100%
R-FAT 2	136	142	74	505	857	16%	17%	9%	59%	100%
Total	248	244	125	935	1,552	16%	16%	8%	60%	100%
S-INV 1	128	162	81	1,440	1,811	7%	9%	4%	80%	100%
S-INV 2	183	230	90	1,200	1,703	11%	14%	5%	70%	100%
Total	311	392	171	2,640	3,514	9%	11%	5%	75%	100%
H-MOT 1	19	7	15	825	866	2%	1%	2%	95%	100%
H-MOT 2	38	14	25	207	284	13%	5%	9%	73%	100%
Total	57	21	40	1,032	1,150	5%	2%	3%	90%	100%

Table 4.3 shows that the frequency of *what* is much higher than that of the other *wh*-words (i.e., *who*, *how*, and *why*) in all caregivers' utterances. About 60 – 90 % of *wh*-words in caregivers' utterances involve *what*, whereas only 2 – 16 % of *wh*-words involve the other three types. The big difference in frequency between *what* and the

other question types is found regardless of the children's age. For example, the frequency rates (9% in *how*, 11% in *why*, 3% in *who* vs 77% in *what*) in Peter's mother's utterance in period 1 (i.e., P-MOT1) are similar to those in her utterances in period 2 (i.e., P-MOT2: 10 % in *how*, 13% in *why*, 11% in *who* vs. 66% in *what*) and period 3 (i.e., P-MOT3: 16% in *how*, 15% in *why*, 10% in *who*, vs. 60% in *what*). A similar tendency is found in all the caregivers.

However, there seems to be a difference among the different periods involving the frequency of *wh*-words. The relative frequency of *what* seems to decrease whereas the relative frequency of other *wh*-words seems to increase. For example, the percentage of *what* in Peter's mother's utterances decrease from Period 1 to Period 3 (i.e., 77% in Period 1 > 66% in Period 2 > 60 % in Period 3), whereas the percentage of the other three *wh*-words increases (i.e., 9% > 10% > 16% in *how*, 11% > 13% > 15% in *why* and 3% > 11% in *who*). A similar tendency is found for other caregivers in Table 4.3. This seems to indicate that the types of caregivers' questions become more varied as the children's cognitive abilities develop (from *what* question to other types of questions such as *how*, *why*, and *who*).

The next table shows the integrated data from each period for the comparison of the overall data.

Table 4.4 Frequency of *wh*-words in caregivers' utterances (combined periods)

	Total number of <i>wh</i> -words					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
P-MOT	56	68	44	367	535	10%	13%	8%	69%	100%
P-PAT	123	228	172	1,512	2,035	6%	11%	8%	74%	100%
P-LOI	121	177	176	1,415	1,889	6%	9%	9%	75%	100%
R-FAT	248	244	125	935	1,552	16%	16%	8%	60%	100%
S-INV	311	392	171	2,640	3,514	9%	11%	5%	75%	100%
H-MOT	57	21	40	1,032	1,150	5%	2%	3%	90%	100%
Total	916	1,130	728	7,901	10,675	9%	11%	7%	74%	100%
Mean	152.7	188.3	121.3	1,315.8	1,782.5	9%	11%	7%	74%	100%

Overall, *what* is more frequent than any other *wh*-word. Among the other *wh*-words, *why* is more frequent than the other two (*how* and *who*) except in R-FAT (*how* > *why*) and H-MOT (*how*, *who* > *why*). Comparing *how* and *who*, *how* is more frequent than *who* except in two cases (P-PAT and P-LOI). As PAT and LOI are involved in the same activity with the same child at the same time, it seems that they show a similar tendency in this frequency count. Overall, the relative frequency of *wh*-words can be stated as follows:

(7) The order of relative frequency of *wh*-words in caregivers' utterance

what > *why* > *how* > *who*

The following graph helps us to see the relative frequency order easily.

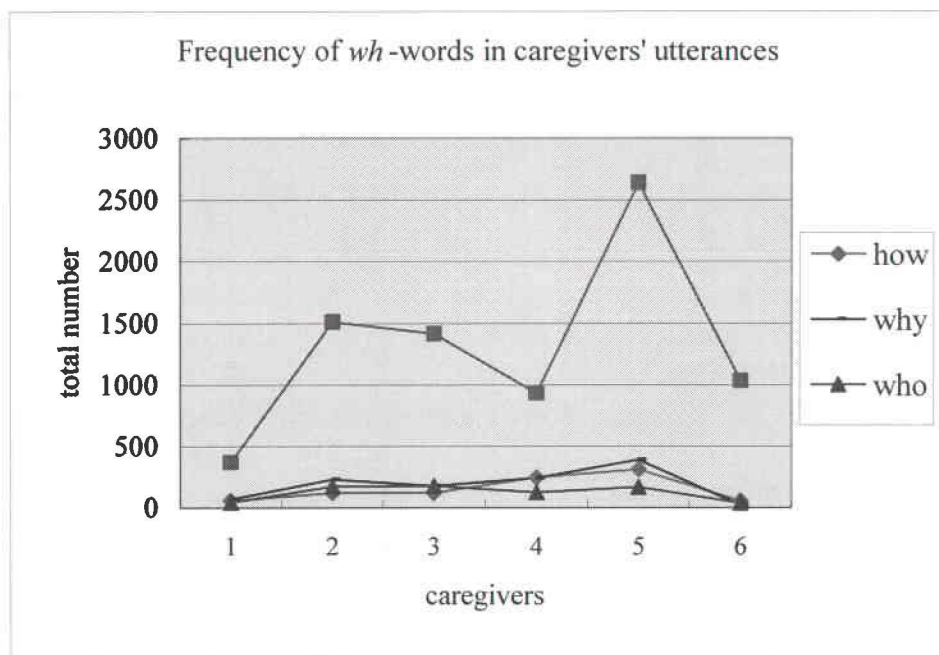


Figure 4.1 Frequency of *wh*-words in caregivers' utterances

Second, the frequency of *wh*-questions was calculated excluding exclamatives, relative pronouns and multi-word *wh*-phrase questions (e.g., *how many?*). The results are presented in the following table:

Table 4.5 Frequency of *wh*-questions in caregivers' utterances (individual periods)

	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
P-MOT 1	8	17	5	118	148	5%	11%	3%	80%	100%
P-MOT 2	20	41	33	176	270	7%	15%	12%	65%	100%
P-MOT 3	7	9	6	30	52	13%	17%	12%	58%	100%
Total	35	67	44	324	470	7%	14%	9%	69%	100%
P-PAT 1	24	76	60	535	695	3%	11%	9%	77%	100%
P-PAT 2	13	86	75	460	634	2%	14%	12%	73%	100%
P-PAT 3	15	63	35	272	385	4%	16%	9%	71%	100%
Total	52	225	170	1,267	1,714	3%	13%	10%	74%	100%
P-LOI 1	15	60	89	550	714	2%	8%	12%	77%	100%
P-LOI 2	20	48	44	475	587	3%	8%	7%	81%	100%
P-LOI 3	17	68	42	260	387	4%	18%	11%	67%	100%
Total	52	176	175	1,285	1,688	3%	10%	10%	76%	100%
R-FAT 1	48	99	50	405	602	8%	16%	8%	67%	100%
R-FAT 2	84	142	67	460	753	11%	19%	9%	61%	100%
Total	132	241	117	865	1,355	10%	18%	9%	64%	100%
S-INV 1	94	156	78	1,420	1,748	5%	9%	4%	81%	100%
S-INV 2	114	220	82	1,040	1,456	8%	15%	6%	71%	100%
Total	208	376	160	2,460	3,204	6%	12%	5%	77%	100%
H-MOT 1	5	6	15	785	811	1%	1%	2%	97%	100%
H-MOT 2	10	14	25	174	223	4%	6%	11%	78%	100%
Total	15	20	40	959	1,034	1%	2%	4%	93%	100%

There is not much difference between the frequency of *wh*-words and the frequency of *wh*-questions, as shown in Table 4.3 and Table 4.5, which compare the percentage of each item. This tells us that relative pronouns and exclamations are very rare compared

to the *wh*-questions in mother and child dyads. *What* questions are much more frequent than any other type of question (64-93%). Among the other three, *why* questions are more frequent than *how* and *who* questions except for one case (H-MOT1, *how* = *why*). *Who* questions are more frequent than *how* questions except in five cases (i.e., P-MOT1, P-MOT3, R-FAT2, S-INV1 and S-INV2). However, the difference of percentages among the three types of *wh*-questions is very small compared to the percentage of *what* questions. The percentage of *what* questions decreases as children's age increases. In contrast, the percentage of the other three types of questions increases as children's age increases. This tendency is found in all the children in Table 4.5.

The number of *wh*-questions in each period was combined and the results are shown in the following table.

Table 4.6 Frequency of *wh*-questions in caregivers' utterances (combined period)

	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
P-MOT	35	67	44	324	470	7%	14%	9%	69%	100%
P-PAT	52	225	170	1,267	1,714	3%	13%	10%	74%	100%
P-LOI	52	176	175	1,285	1,688	3%	10%	10%	76%	100%
R-FAT	132	241	117	865	1,355	10%	18%	9%	64%	100%
S-INV	208	376	160	2,460	3,204	6%	12%	5%	77%	100%
H-MOT	15	20	40	959	1,034	1%	2%	4%	93%	100%
Total	494	1,105	706	7,160	9,465	5%	12%	8%	76%	100%
Mean	82.3	184.2	117.7	1,193.3	1,577.5	5%	12%	8%	76%	100%

The results in Table 4.6 are little different from those in Table 4.4, which shows the frequency of *wh*-words. The relative order of *how* and *who* questions has switched

because some of the multi-word *how* questions (e.g., *how old*, *how many*) were eliminated in this count. Four of the caregivers show the *who* > *how* order, but the other two caregivers show the opposite *how* > *who* order. The difference of frequency is very small in each case. On the other hand, *why* questions seem to be consistently more frequent than *how* and *who* questions except in one case (H-MOT). Therefore the relative order of frequency of *wh*-questions can be ranked as follows.

(8) The relative order of frequency of *wh*-question

what > why > how, who

The following figure also helps us to see the relative frequency of *wh*-questions given in Table 4.6.

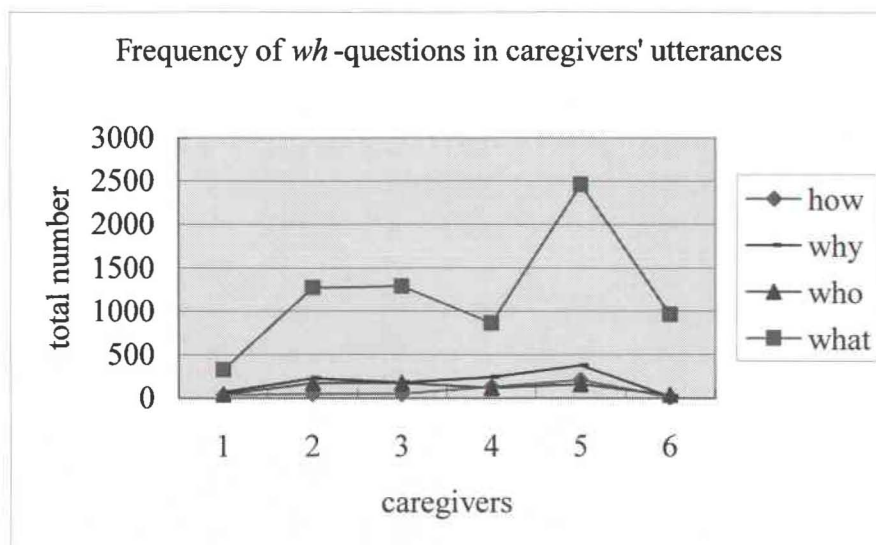


Figure 4.2 Frequency of *wh*-questions in caregivers' utterances

Third, the frequency of inverted, uninverted and other questions was counted and the results are shown in the following table:

Table 4.7 *Wh*-questions in caregivers' utterances

	Inverted (<i>Wh</i> +Aux+Subj)				Uninverted (<i>Wh</i> +Subj+Aux)				Other questions			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-M1	4	15	2	70	0	1	1	14	4	1	2	34
P-M2	9	33	4	60	4	1	0	20	7	7	29	96
P-M3	4	8	0	12	0	0	0	2	3	1	6	16
Total	17	56	6	142	4	2	1	36	14	9	37	146
P-P1	8	72	19	230	4	1	2	25	12	3	39	280
P-P2	7	62	24	215	1	0	2	20	5	24	49	225
P-P3	9	38	3	72	2	1	0	16	4	24	32	184
Total	24	172	46	517	7	2	4	61	21	51	120	689
P-L1	7	58	29	210	6	1	0	55	2	1	60	285
P-L2	11	44	13	275	2	0	0	25	7	4	31	175
P-L3	8	42	9	104	1	4	0	20	8	22	33	136
Total	26	144	51	589	9	5	0	100	17	27	124	596
R-F1	37	41	11	285	3	2	0	5	8	56	39	115
R-F2	48	68	12	285	2	6	2	15	34	68	53	160
Total	85	109	23	570	5	8	2	20	42	124	92	275
S-I1	50	100	41	600	22	2	6	80	22	54	31	740
S-I2	66	180	35	600	21	10	5	120	27	30	42	320
Total	116	280	76	1,200	43	12	11	200	49	84	73	1,060
H-M1	4	3	10	625	0	0	0	110	1	3	5	50
H-M2	2	7	5	99	6	0	0	6	2	7	20	69
Total	6	10	15	724	6	0	0	116	3	10	25	119

Overall, all the *wh*-question types (i.e., *how*, *why*, *who*, and *what*) are used both as inverted and uninverted questions by most of the caregivers except for P-L and H-M. Uninverted *who* questions are not found in P-L and H-M, and uninverted *why* questions are not found in H-M. Inverted questions are more frequent than uninverted questions, which means that most of the caregivers' questions involve direct questions rather than indirect questions. There is not much difference between the different periods in each parent and among the caregivers. As the most relevant input for the subject and auxiliary inversion consists of the inverted questions, the inversion rates for each *wh*-question were calculated and the results are shown in the following table.

Table 4.8 Inversion rates for *wh*-questions in caregivers' utterances (individual periods)

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> +Aux+Subj)				Inversion rates			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-M1	8	17	5	118	4	15	2	70	50%	88%	40%	59%
P-M2	20	41	33	176	9	33	4	60	45%	80%	12%	34%
P-M3	7	9	6	30	4	8	0	12	57%	89%	0%	40%
Total	35	67	44	324	17	56	6	142	49%	84%	14%	44%
P-P1	24	76	60	535	8	72	19	230	33%	95%	32%	43%
P-P2	13	86	75	460	7	62	24	215	54%	72%	32%	47%
P-P3	15	63	35	272	9	38	3	72	60%	60%	9%	26%
Total	52	225	170	1,267	24	172	46	517	46%	76%	27%	41%
P-L1	15	60	89	550	7	58	29	210	47%	97%	33%	38%
P-L2	20	48	44	475	11	44	13	275	55%	92%	30%	58%
P-L3	17	68	42	260	8	42	9	104	47%	62%	21%	40%
Total	52	176	175	1,285	26	144	51	589	50%	82%	29%	46%
R-F1	48	99	50	405	37	41	11	285	77%	41%	22%	70%
R-F2	84	142	67	460	48	68	12	285	57%	48%	18%	62%
Total	132	241	117	865	85	109	23	570	64%	45%	20%	66%
S-I1	94	156	78	1,420	50	100	41	600	53%	64%	53%	42%
S-I2	114	220	82	1,040	66	180	35	600	58%	82%	43%	58%
Total	208	376	160	2,460	116	280	76	1,200	56%	74%	48%	49%
H-M1	5	6	15	785	4	3	10	625	80%	50%	67%	80%
H-M2	10	14	25	174	2	7	5	99	20%	50%	20%	57%
Total	15	20	40	959	6	10	15	724	40%	50%	38%	75%

- The inversion rates the ratio of the number of inverted questions to the number of total questions of each type.

There doesn't seem to be any consistent difference in inversion rates among different types of *wh*-questions for different ages. Only the inversion rates for *who* questions decrease from the early period to the later period in every caregiver. The reason for this

seems to be related to the increase of other *who* questions, especially *who* subject questions (362 out of 471 other questions), which will be discussed in detail later. The high inversion rates for *why* questions are mainly due to the *why don't you* pattern (332 out of 771 inverted *why* questions). Comparing *how* and *who*, the inversion rates for *how* questions are always higher than those for *who* questions in all the cases except for two cases out of fifteen (S-INV1 and H-MOT2), where, the inversion rates for *who* and *how* are the same.

For easier comparison, caregivers' data from different periods were combined across age samples and shown in the following table:

Table 4.9 Inversion rates for *wh*-questions in caregivers' utterances (Combined periods)

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> +Aux+Subj)				Inversion rate			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-M	35	67	44	324	17	56	6	142	49%	84%	14%	44%
P-P	52	225	170	1,267	24	172	46	517	46%	76%	27%	41%
P-L	52	176	175	1,285	26	144	51	589	50%	82%	29%	46%
R-F	132	241	117	865	85	109	23	570	64%	45%	20%	66%
S-I	208	376	160	2,460	116	280	76	1,200	56%	74%	48%	49%
H-M	15	20	40	959	6	10	15	724	40%	50%	38%	75%
Total	494	1,105	706	7,160	274	771	217	3,742	51%	69%	29%	54%
Mean	82.3	184.2	117.7	1,193	45.7	128.5	36.2	623.7	51%	69%	29%	54%

It seems clear that there are differences in inversion rates for different *wh*-questions.

The inversion rate of *what* questions is higher than that of *who* questions in all caregivers' utterances, and the inversion rate of *how* questions is higher than that of *who* questions in

all caregivers' utterances. On the other hand, the inversion rate for *why* questions is higher than that of *what* questions except in two cases (R-FAT and H-MOT). In sum, the following rank order of inversion rates for *wh*-questions is found:

(9) Order of inversion rates for *wh*-questions in caregivers' utterances

why > what, how > who

The relative inversion rates among *wh*-questions are shown in the following figure.

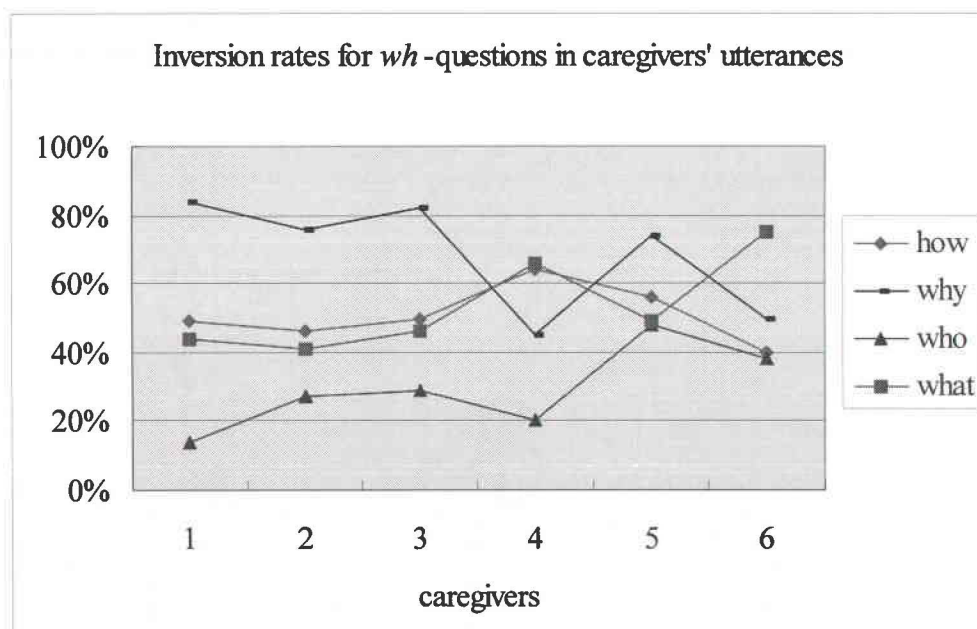


Figure 4.3 Inversion rates for *wh*-questions in caregivers' utterances

Before we conclude the frequency search, we need to consider the huge number of formulaic expressions such as *how are you?*, *why don't you—?*, *who's that?* and *what's that?*, which do not seem to be segmented into their component parts by the child as [*wh*-

+ Aux + Subj]. For example, high inversion rates for *why* questions (average 69%) are mainly due to the *why don't you—?* pattern (332 out of 771). It seems desirable to eliminate these expressions from our analysis in order to get more reliable frequency data from the child's point of view regarding generalizable forms.

To find out what proportion of the inverted *wh*-questions are formulaic expressions, an additional search was conducted. First, formulaic expressions (which involve inversion) were searched in the same corpus. The search items and their frequencies are presented in the following table.

Table 4.10 Frequency of formulaic expressions in caregivers' utterances

Care-givers	<i>How are you (doing)?</i>	<i>How's it/that —?</i>	<i>Why don't you—?</i>	<i>Who's that—?</i>	<i>Who's this—?</i>	<i>What's that—?</i>	<i>What's this—?</i>	Total
P-M	1	1	34	1	1	19	9	66
P-P	3	0	132	29	0	115	38	317
P-L	3	1	116	25	4	80	28	257
R-F	1	2	12	3	2	18	7	45
S-I	9	0	23	10	18	56	95	211
H-M	1	0	9	5	5	59	51	130
Total	18	4	326	73	30	347	228	1,026
Mean	3.0	0.7	54.3	12.2	5.0	57.8	38.0	171.0

It was decided that an expression is formulaic if (i) it is semantically non-compositional (e.g., *why don't you—?*, *how are you (doing)?*) or (ii) its component parts are fixed, allowing for little or no variation in the way that they are used (e.g., *who's that*, not *who was that?*, *which man is that?* or even *who is that?*). Table 4.10 shows that a great

number of formulaic expressions were included in our first search for inverted *wh*-questions. The following table gives the total number of formulaic expressions for each type of *wh*-question.

Table 4.11 Frequency of formulaic expressions in each *wh*-question in caregivers' utterances

Caregivers	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
P-M	2	34	2	28	66
P-P	3	132	29	153	317
P-L	4	116	29	108	257
R-F	3	12	5	25	45
S-I	9	23	28	151	211
H-M	1	9	10	110	130
Total	22	326	103	575	1,026
Mean	3.6	54.3	17.2	95.8	171.0

Eliminating these formulaic expressions from the number of inverted *wh*-questions (in Table 4.9) results in new inversion rates for each type of *wh*-question as follows:

Table 4.12 Frequency of non-formulaic inverted *wh*-questions and inversion rates

	Total number of <i>wh</i> -questions				Non-formulaic inverted <i>wh</i> -questions (<i>Wh</i> +Aux+Subj)				Inversion rate			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-M	35	67	44	324	15	22	4	114	43%	33%	9%	35%
P-P	52	225	170	1,267	21	40	17	364	40%	18%	10%	29%
P-L	52	176	175	1,285	22	28	22	481	42%	16%	13%	37%
R-F	132	241	117	865	82	97	18	545	62%	40%	15%	63%
S-I	208	376	160	2,460	107	257	48	1,049	51%	68%	30%	43%
H-M	15	20	40	959	5	1	5	614	33%	5%	13%	64%
Total	494	1,105	706	7,160	252	445	114	3,167	45%	30%	15%	45%
Mean	82.3	184.2	117.7	1,193.3	42.0	74.2	19.0	527.8	45%	30%	15%	45%

The new data for each question results in the following relative order of inversion rates:

(10) The order of inversion rates for *wh*-question in caregivers' utterances

what, how > why > who

Comparing the inversion rates in (10) with our previous one (9), which included formulaic expressions, *why* (69%) > *what* (54%), *how* (51%) > *who* (29%), the relative order of inversion rates among *wh*-questions is the same except for *why* question.

Because we eliminated a huge number of *why don't you—?* patterns, the inversion rate for *why* decreased to 30%. Therefore, the inversion rate for *why* questions is lower than that for *what* questions and *how* questions but higher than that for *who* questions. In addition, this order seems to be consistent among all caregivers except one (68% in S-INV). The relative order of inversion rates among *what*, *how* and *who* remains the same as the previous order. However, comparing *how* with *what*, the inversion rate for *what* questions seems to be lower than that for *how* questions if we take into account the general tendency among caregivers. The inversion rate for *how* questions is higher than that for *what* questions in all caregivers except one, H-MOT, (64% in *what* vs. 33% in *how*). It therefore seems plausible to say that the inversion rate is higher in *how* questions than in *what* questions. Therefore, we came up with the following order of inversion rates:

(11) Relative order of inversion rates for *wh*-question in caregivers' utterances

how > what > why > who

The relative order in (11) seems to be consistent among the caregivers, as depicted in Figure 4.4.

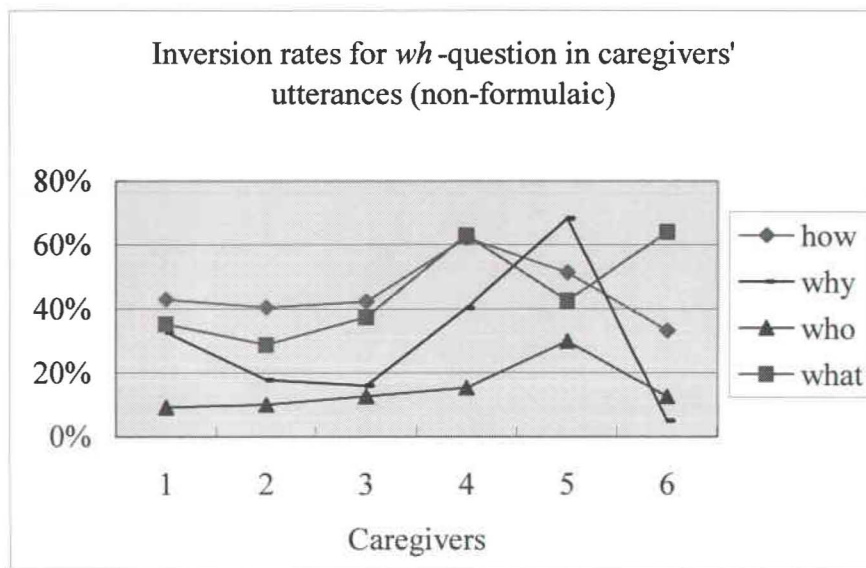


Figure 4.4 Inversion rates for *wh*-questions in caregivers' utterances (non-formulaic)

Different frequency and inversion rates among *wh*-questions result in different frequencies of inverted *wh*-questions. The frequency of inverted questions for each *wh*-question is shown in Table 4.12. Even after eliminating the formulaic expressions, the frequency of inverted *what* questions is highest because of the enormous number of *what* questions. The inverted *why* question is the next frequent pattern even after eliminating

a great number of formulaic *why don't you—?* patterns. Inverted *how* questions are more frequent than inverted *who* questions in all caregivers except for two (P-L and H-M), for whom the frequencies of the two patterns are the same. In sum, we find the following rank order of relative frequency of inverted *wh*-questions:

- (12) Relative order of frequency of non-formulaic inverted *wh*-questions
 what > why > how > who

The following figure helps us see the relative frequency among them:

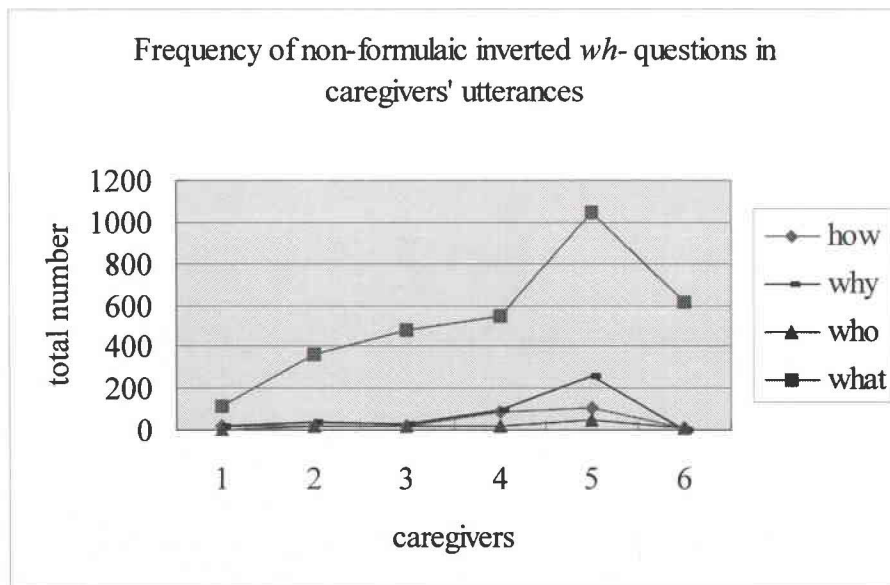


Figure 4.5 Frequency of non-formulaic inverted *wh*-questions in caregivers' utterances.

Figure 4.5 shows an enormously high frequency of *what* questions compared to the other

three types of *wh*-questions. Another figure is presented with only the other three patterns in order to help show their relative frequency in all caregivers (i.e., *why* > *how* > *who*):

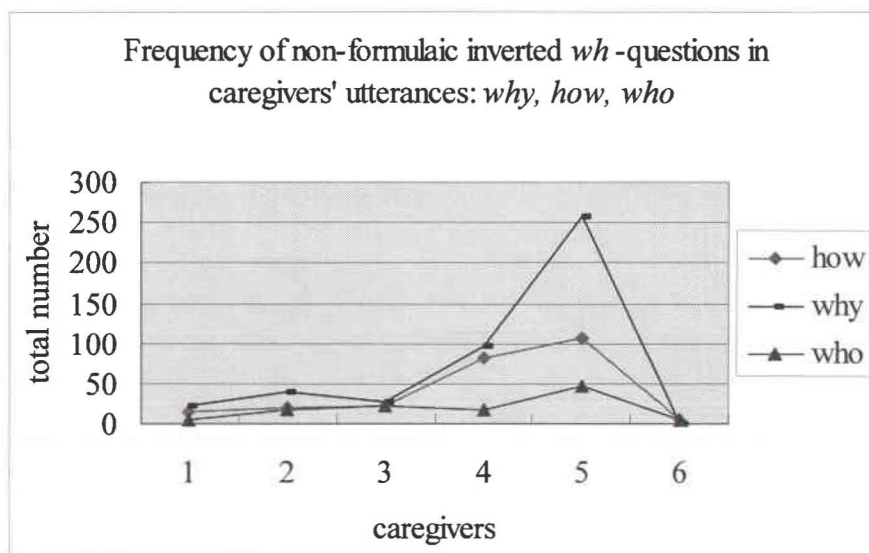


Figure 4.6 Frequency of non-formulaic inverted *wh*-questions in caregivers' utterances: *why*, *how*, and *who*

4.1.3 Discussion

The corpus analysis of *wh*-questions in caregivers' utterances in caregiver-child dyads seems to show a consistency in the input that children receive in four ways: (i) the frequency of *wh*-words, (ii) the frequency of *wh*-questions, (ii) the inversion rates for the *wh*-questions, and (iv) the frequency of inverted *wh*-questions. There seems to be an order among different *wh*-questions in these four aspects:

(13) *Wh*-questions in caregivers' utterances

- i) Frequency of *wh*-words: what > why > how > who
- ii) Frequency of *wh*-questions: what > why > how, who
- iii) Inversion rates: how > what > why > who
- iv) Frequency of inverted *wh*-questions: what > why > how > who

These orders are quite consistent among different caregivers. The following graph helps us see the consistency of the order of relative frequency of *wh*-words and *wh*-questions:

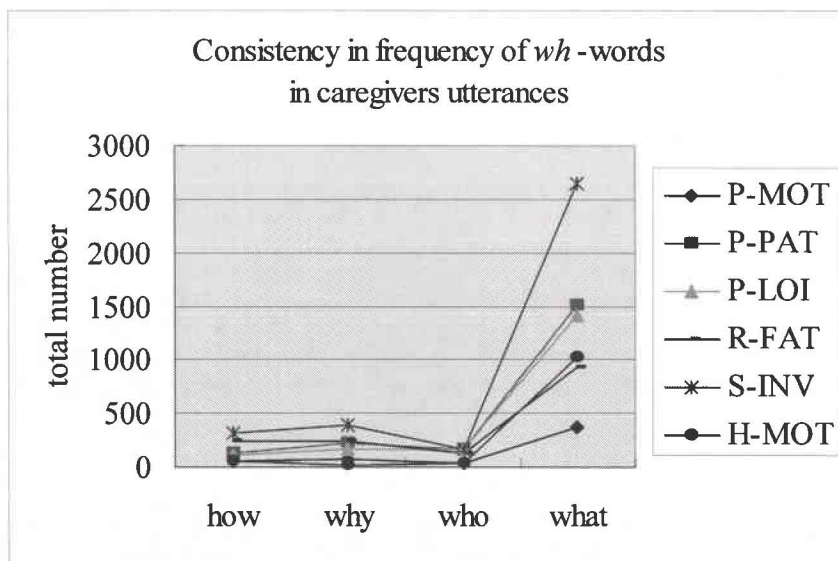


Figure 4.7 Consistency in relative frequency of *wh*-words in caregivers' utterances

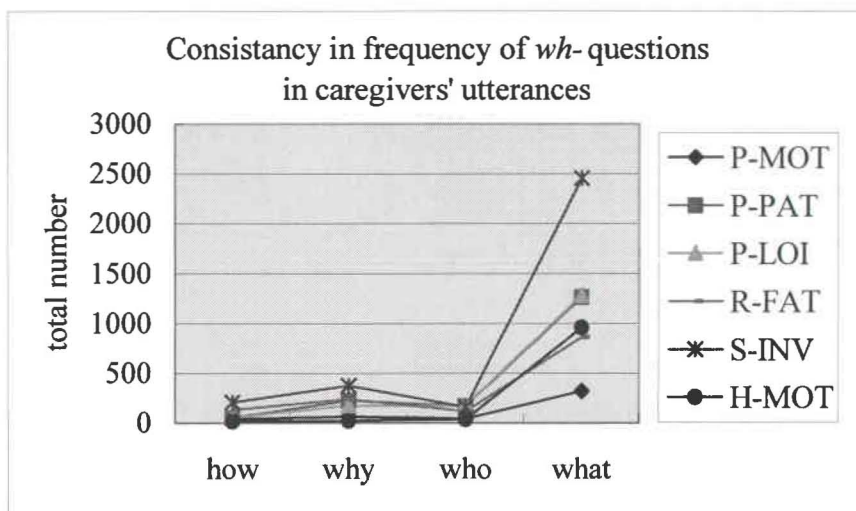


Figure 4.8 Consistency in relative frequency of *wh*-questions in caregivers' utterances

The shapes of the curves are similar among six caregivers in the figures, which means that there is a similarity among the six caregivers in terms of the relative frequency of *wh*-words and *wh*-questions in their speech to their child. The following graph shows the consistency in the relative order of inversion rates for *wh*-questions (non-formulaic):

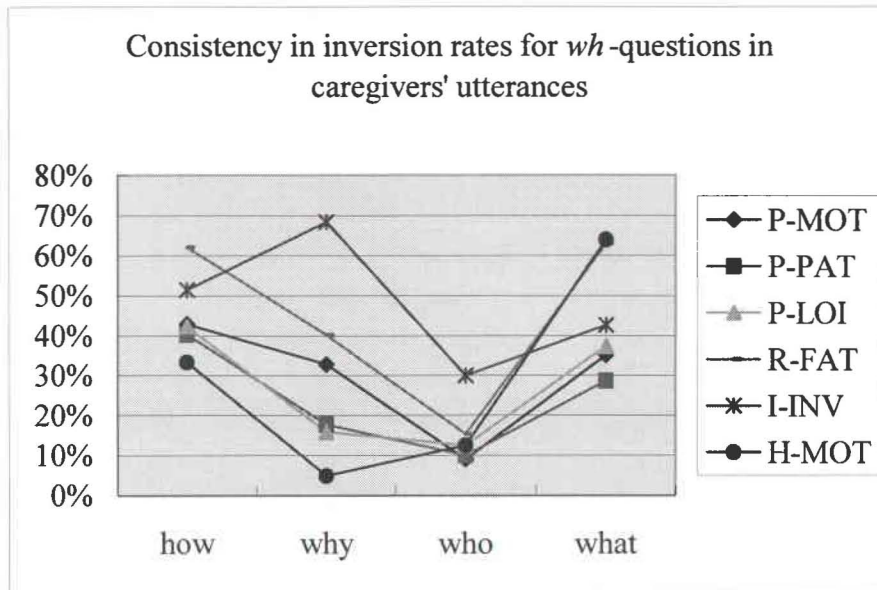


Figure 4.9 Consistency in inversion rates for *wh*-question in caregivers' utterances

The shape of of curves is similar among *how*, *who* and *what* (i.e., *how* > *who*, and *what* > *who*), which means that the relative rank order of inversion rates among these questions are similar among different caregivers.

Finally, the following graph shows the consistency of the relative frequency of non inverted *wh*-questions among different caregivers:

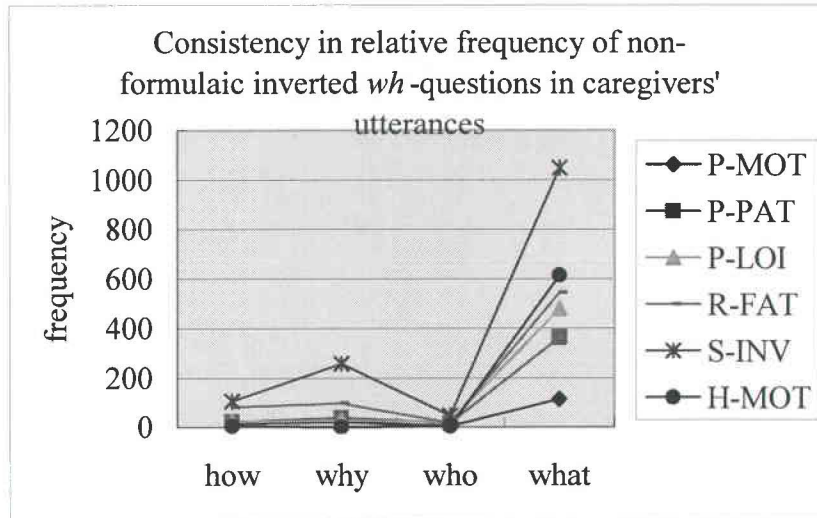


Figure 4.10 Consistency in relative frequency of non-formulaic inverted *wh*-questions in caregivers' utterances

Again, the shape of all the curves are very similar (especially among *how*, *who* and *what*), which means that caregivers' utterances are very similar to each other in terms of the relative frequency of inverted *wh*-questions in their conversation with their child.

In sum, we can draw the following conclusion:

- i) The caregivers' utterances are very similar to each other in terms of (i) the frequency of *wh*-words, (ii) the frequency of *wh*-questions, (iii) inversion rates for the *wh*-questions and (iv) the frequency of inverted *wh*-questions. In other words, children are exposed to very similar input regardless of their interlocutors in these respects.
- ii) This consistency makes it possible for us to assume that the children who are involved in an experimental study will have been exposed to similar kinds of input in these four respects. That is, production data from other

children in an experimental situation can be compared with the input data of the children investigated in this study to compare any differences between children's input and production.

- iii) The conclusions in (i) and (ii) enable us to compare children's input and their acquisition, testing the lexical-based input frequency hypothesis. If acquisition order follows frequency order, we will expect the production study to yield a result similar to the input-based findings in (12), that is, *what* > *why* > *how* > *who*. Accordingly, we will not expect any argument and adjunct distinction because the two argument questions are separated by two adjunct questions in the rank of input frequency.

4.2 L1 Acquisition study: Production (Experiment 2)

The results of the first part of this study enable us to test the input frequency account versus the structural account at the same time. Even though we cannot directly compare the input data obtained from the children of our input study with the acquisition data obtained from different children of other studies, we can at least generalize our results, that is, most children receive similar kinds of input regarding the frequency of inverted *wh*-questions (that is, *what* > *why* > *how* > *who*). Based on this generalization, if the input theory is right, children should invert subject and auxiliary in *what* questions better than they do in *why*, *how* and *who* questions. Among the other three, *why*, *how* and *who*, children will be better at inversion with *why* questions than with *how* and *who*

questions, because inverted *why* questions are more frequent than inverted *how* and *who* questions. Between *how* and *who* questions, children will be better at inversion with *how* questions than with *who* questions. On the other hand, if the structure-based generative account is right, there should be some similarity between *what* questions and *who* questions on the one hand, and between *how* questions and *why* questions on the other in terms of children's inversion in *wh*-questions, even though the frequencies of the inverted questions are different among these four in the input. Moreover, both *how* and *why* questions should show a different pattern from *what* and *who* questions if the structural difference between argument (*who* and *what*) and adjunct (*how* and *why*) is reflected in language acquisition. The second part of this chapter investigates children's production of *wh*-questions with an experiment involving an elicited production task to test two theories.

An experiment was designed to elicit *wh*-questions from children. Although children's longitudinal data are available in the CHILDES database, there are some problems with using these data for our study as mentioned in Chapter 2 and Chapter 3. In addition, the children's production data in CHILDES are different from one another in terms of their size, the actual number of each type of *wh*-questions, and the children's ages. In addition, the kind and the number of the auxiliaries involved in *wh*-questions were different throughout the data. Therefore, in order to obtain more controlled data, an elicited production task was conducted using pictures.

4.2.1. Methods

The same basic technique was used as in Kim (1995) and Yoshinaga (1996) to elicit *wh*-questions from children. A child is given a picture which is partly covered with colored paper. The child is told to address a question to the animal in the picture to figure out what is hidden under the color paper, pretending the animal is alive. Children like to talk to the animals in the pictures. The whole session of the experiment goes smoothly, like a picture game, without any artificial feeling. An example of the protocol to elicit each type of question is given below (the *why*-question is given with its relevant picture)

(14) Protocol to elicit *wh*-questions

a. *Who are you* –? questions

Experimenter: [Showing a picture] This is an elephant. The elephant is pushing somebody, but I don't know who.
[Pointing to the elephant in the picture] Ask him/her who.
Child: *Who are you pushing?*
Experimenter: [Uncovering the hidden part of the picture] Pinocchio.

b. *How are you* –? questions

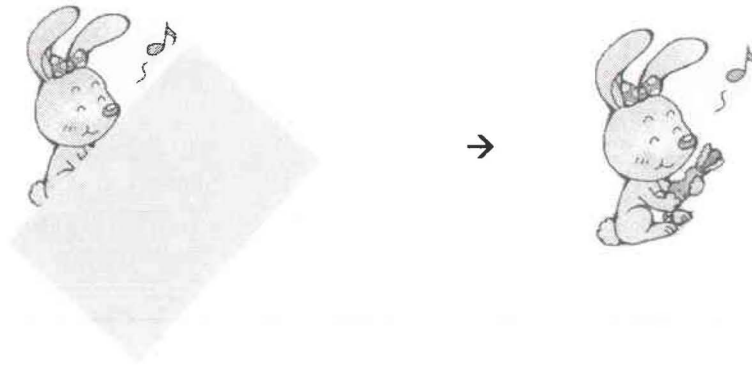
Experimenter: [Showing a picture] This is a witch. The witch is flying in some way. I don't know how. [Pointing to the witch in the picture] Ask her how.
Child: *How are you flying?*
Experimenter: [Uncovering the hidden part of the picture] On my broom.

c. *Why are you* –? questions

Experimenter: [Showing a picture] This is a little bunny. The little bunny is smiling for some reason. I don't know why.
[Pointing to the bunny in the picture] Ask her why.

Child: *Why are you smiling?*

Experimenter: [Uncovering the hidden part of the picture] Because I have a carrot!



Five tokens for three question types and two warm-up questions (a total of 17 questions with corresponding pictures) were prepared for the whole experiment. The 15 target questions were randomized and presented after the two warm-up questions. Only the combination of *who/how/why* + *are* + *you* + *V-ing?* was used as target items to control the data, the purpose of which will be discussed in detail later. The *what* question was excluded because of the enormous input (found in the input study in the previous section), which might influence the children's production of the target items. In other words, children might be used to *what are you V-ing?* (as a fixed expression) so that they might apply the fixed pattern during the experiment. In addition, the input frequency account and the generative account predict the same results (i.e., *what* >

adjunct *wh*-questions). Therefore it seemed unnecessary to include *what* questions.

Also, including *what* questions increases the number of target items, which will lengthen the time of the task, possibly distracting the children.

4.2.2 Subjects

A total of seventeen preschool children (age range 3;5 – 4;6) participated in the experiment. Seven of the children were enrolled in the University of Hawai‘i Children’s Center and ten of the children were enrolled in Motherese Preschool. Both preschools are located near the University of Hawai‘i at Manoa in Honolulu, Hawai‘i. All the subjects were monolingual English-speaking children whose parents were native speakers of English.

4.2.3 Procedure

The task was done individually in a quiet classroom in each preschool during playtime. Each child completed the task pleasantly with the experimenter. All the conversations during the experiment session were audio-taped and transcribed later by the researcher.

4.2.4. Results and discussion

First of all, all the tokens from seventeen children were combined and presented according to the types in Figure 4.11. Notice that only the target subj-aux combination

was considered for each inverted/uninverted question. That is to say, the target questions involve a specific combinations of *wh*-word + Aux + Subj, which is *who/why/how* + *are you V-ing?*. Therefore, only responses involving *who/why/how* + *are you*+ *V-ing?* are treated as inverted questions and only responses involving *who/why/how* + *you are*+ *V-ing?* are treated as uninverted questions. The other types of answers were categorized into “others”, which include questions that involve inverted or uninverted questions with different combinations of Subj + Aux or copular *be* (e.g., *how do you do that?*, *why are you cold?*, *who did you call?*, etc.). This will be discussed in detail later.

Some of the children produced only inverted questions, which is correct, but others produced both inverted and uninverted questions (see Table 4.14 for individual data). Figure 4.11 shows the difference between *who* questions and *why/how* questions in terms of the subj-aux inversion rates.

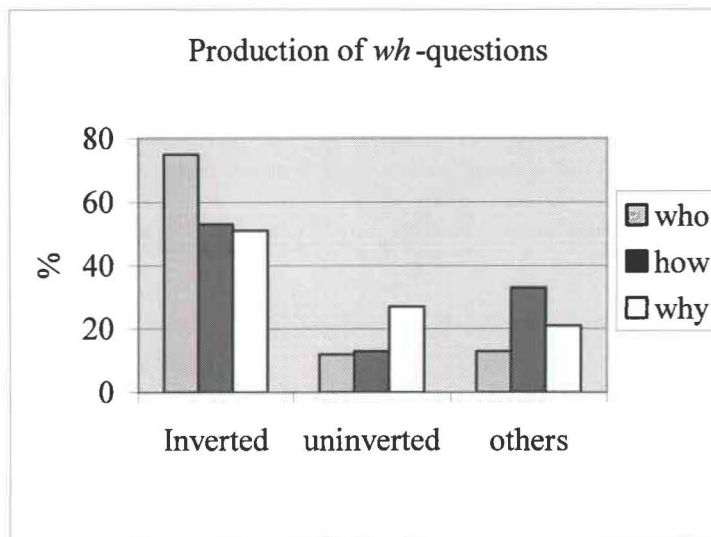


Figure 4.11 Childrens' production of *wh*-questions

- "Others" include such errors as *who* subject questions, double auxiliary questions, one-word questions, and auxiliary missing questions.

The results show that inversion was more frequent with *who* questions (75%) than with *how* and *why* questions (53% and 51% respectively). Only half of the *how* and *why* questions were correctly inverted whereas three-fourths of the *who* questions were correctly inverted. The mean (maximum 5.0) difference among the three types of questions is presented in Table 4.13.

Table 4.13 Mean numbers of inverted *wh*-questions

	<i>who</i>	<i>how</i>	<i>why</i>
Mean	3.88	2.94	2.82

$P < 0.05$

A one-way repeated measures ANOVA was conducted and it was found that the effect of the *wh*-words was found significant ($F(2, 32) = 5.30^*, p = 0.0103$).

Subsequent Tukey's tests were conducted. The results showed that the mean difference between *who* and *how*, and that between *who* and *why* were significant, but that the mean difference between *why* and *how* was not significant (critical significant difference = 0.8758, 2-tailed, $\alpha = 0.05$).

The results of the experiment show that children differentiate argument questions from adjunct questions in terms of subj-aux inversion. As the inversion with *who* questions is more frequent than that with *how* and *why* questions, children seem to do better at inversion with argument *wh*-questions than with adjunct *wh*-questions.

In order to show the difference among the children, the individual data are presented in Table 4.14. This seems to show relative developmental stages among children (not within a child) in the acquisition of *wh*-questions.

Table 4.14 Individual data (Total number of inverted questions)

Subject	<i>who</i>	<i>how</i>	<i>why</i>	STAGE
21	5	5	5	All inversion
29	5	5	5	
19	5	5	4	<i>Who</i> inversion
20	5	5	4	
23	5	5	3	
26	5	5	2	
30	5	4	4	
17	5	4	2	
25	5	2	4	
33	5	0	3	
27	5	1	1	
18*	4	3	5	Inversion trial
24	4	4	4	
31	3	2	2	
14	0	0	0	Non-inversion
16	0	0	0	
28	0	0	0	
Total (17)	66	50	48	
Mean	3.88	2.94	2.82	

The scores for *who* questions are always higher than (or the same as, at least) those for *how* questions and *why* questions in the table, except for one child (subject 18). This seems to show that children do better at inversion with *who* questions than with *how* or *why* questions. Furthermore, there seems to be an implicational developmental sequence between argument questions and adjunct questions in terms of subj-aux inversion. Every child that produced *how* and *why* questions with inversion 100% of the time also produced *who* questions with inversion 100 % of the time (subject 21, 29), but not vice versa (subject 30, 17, 25, 33, and 27). Therefore, generally speaking, subj-aux

inversion with adjunct *wh*-questions implies subj-aux inversion with argument *wh*-questions.

Let's consider developmental stages among the children. Subjects 21 and 29 seem to have mastered subj-aux inversion in *wh*-questions, since they produce all types of questions with inversion, whereas Subjects 14, 16, and 28 did not ever use inversion. Among those who are in the middle of learning the inversion rule, Subjects 19, 20, 23, 26, 30, 17, 25, 33, and 27 seem to be in the *who* inversion stage (or more generally, argument question inversion stage), where they have mastered the inversion rule with *who* questions but not with *how* and *why* questions. On the other hand, Subjects 24 and 31 did not produce inversion 100% correctly with any *wh*-question words. Therefore, the data shows the following sequence in the development of subj-aux inversion in *wh*-questions.

(15) Development of subj-aux inversion in *wh*-questions

- Stage I: Non-inversion
- Stage II: First attempts at inversion
- Stage III: Inversion with argument questions (*who*)
- Stage IV: Inversion with adjunct questions (*how* and *why*)

4.3 Comparison of input and acquisition

The results of the children's input and production studies enable us to evaluate

the structure-based generative theory and the lexical-based input frequency theory. In the input study, inverted *how* questions were found more frequent than inverted *who* questions in all the caregivers' utterances. By contrast, in the production study, it was found that children did better at inversion with *who* questions than with *how* questions. In addition, the inversion rates for *how* and *why* questions were similar to each other, but significantly different from *who* questions. These findings support the structure-based generative approach, not the lexical-based input frequency approach to account for children's acquisition of *wh*-questions.

Interestingly, inversion seems to occur in *who* questions once the children start to produce *who* questions. In other words, children's error involving subj-aux inversion is very rare in *who* questions. Fourteen children out of seventeen produced *who* questions with inversion either 100% or 0% of the time. On the other hand, the inversion scores for *how* and *why* questions vary among the children. This seems to support Stromswold and De Villiers' proposal that at the early stage of language acquisition, argument *wh*-questions have only one structure with the *wh*-word in Spec of CP, while adjunct *wh*-questions have two structures – one with IP adjoined *wh*-word and one with *wh*-word in Spec of CP as shown in (2) in Chapter 2.

Before concluding this chapter, we need to consider another possibility from the perspective of input theory more specifically. The present experiment involves children's production of *who are you V-ing?*, *how are you V-ing?*, and *why are you V-ing?*. It is possible to think that children produce more inverted *who* questions than inverted

how or *why* questions in this experiment because they are exposed to the input for the specific combination *who are you—?* more frequently than *how are you—?* or *why are you—?*. In order to investigate this possibility, the specific combinations of *wh-word* + *aux* + *subj*, that is, *who/how/why are you V-ing?* were searched for in the caregivers' utterances in CHILDES to compare the frequency of these items with the results of our production data. Crucially, searching for those specific combinations in four caregivers' speech revealed the opposite results, as shown in Table 4.15 (the other files do not have enough data).

Table 4.15 Frequency of *wh-word* + *are you—?* pattern in caregivers' speech

Subject	<i>Who are you—?</i>	<i>How are you—?</i>	<i>Why are you—?</i>	Total number of files	Word count	Age range
T-MOT	1	2	2	28	88,619	3;0
S-INV	6	13	12	47	284,107	3;2
R-FAT	1	6	13	40	237,041	4;2
A-MOT	3	6	28	42	314,744	4;0
Total	11	27	55	157	924,511	

- T-MOT: Trevor's mother (in Dematras), S-INV: Shem's investigator, R-FAT: Ross' father, Adam-MOT: Adam's mother

The number for *why are you—?* was bigger than that for *how are you—?*, and the number of *who are you—?* was the lowest. If Rowland and Pine's (2000) argument is right (i.e., that children acquire a specific combination of *wh-word* + *aux* + *subj* earlier than the other combination depending on the frequency in the input), we would expect children to do better with *why* questions than any other questions in our experiment,

whose design calls for this specific combination *why/how/who are you V-ing?*. However, the results of our experiment showed that children did better with *who* questions than with the *why* and *how* questions. In addition, the frequency difference between *how are you—?* and *why are you—?* in the input (i.e. *how* < *why* 27:55) is not reflected in the experiment. The mean difference of inversion scores in *how* questions and *why* questions (i.e. 2.94 vs. 2.82) is very similar. Therefore, it seems difficult for us to see how a lexical-based input frequency account can explain our data.

4.4 Conclusion

A possible argument/adjunct asymmetry in children's acquisition of *wh*-question and subj-aux inversion was discussed in this chapter. Two theories of language acquisition were tested: the structure-based theory and the input-frequency based theory. The results of the children's input and production data from the present study show that children do better at subj-aux inversion with argument *wh*-questions than they do with adjunct *wh*-questions, which cannot be explained by the input frequency. These results support the structure-based generative theory, rejecting Rowland and Pine's (2000) lexical-based input frequency theory. Children's poor inversion with adjunct *wh*-questions conforms to the prediction of IP adjunction account proposed by Stromswold (1990) and De Villiers (1991).

CHAPTER 5

L2 STUDY

This chapter consists of two parts as the previous chapter. The first investigates L2 learners' input for inverted *wh*-questions in English. The second part involves L2 learners' knowledge about subj-aux inversion in *wh*-questions. For the input study, *wh*-questions in various input sources for Korean learners of English were analyzed and compared to find a similar tendency of the input. For the production study, a grammaticality judgement task was conducted with Korean learners of English to investigate their knowledge about inversion in *wh*-questions. Four types of *wh*-questions were investigated in both studies: *what*, *who*, *how*, and *why* questions. The input study will show a similarity of input regarding the relative frequency of inverted *wh*-questions and the acquisition study will show a difference between argument and adjunct *wh*-questions regarding subj-aux inversion. The results from the two studies will be discussed in the aspects of two theories: the structure-based generative theory and the lexical-based input frequency theory.

5.1 L2 Input study: corpus analysis (Experiment 3)

Various English corpora were analyzed in order to find tendencies in the use of *wh*-questions in terms of frequency and inversion rates. First, English textbooks in Korea were chosen because they are one of the main sources from which Korean learners

of English get English input in Korea. Next, TV scripts of American situation comedies and movies were chosen to compare learners' input from textbooks and from ordinary English conversation. If there is any similarity among textbooks, among drama scripts, and between the two types (i.e., textbooks and drama scripts) in terms of the order of relative frequency and inversion rates for different *wh*-question types, it seems plausible to assume that learners are exposed to similar types of input regarding subj-aux inversion in *wh*-questions. In other words, learners' input from the different input sources will not be very different.

5.1.1 *Wh*-questions in English textbooks in Korea

Middle school English textbooks in Korea were analyzed in this study. The main contents of English textbooks were available on the website (www.wordnote.com) chapter-by-chapter. Most of the chapters are composed of conversation and reading texts, which means that both spoken English and written English were included in this corpus. In order to run the search program, MonoConc Pro, on the text, each chapter was downloaded and saved as a text file by the researcher. A total of 23 textbooks (seven from first-grade textbooks, eight from second-grade textbooks, and eight from third-grade textbooks) were downloaded and transformed into text format. Each textbook was published by a different publishing company and written by a different author. Some publishing companies had one or two different textbooks for each grade, and others had only one per grade. Each textbook had about 12 chapters and each

chapter had from 2,000 to 8,000 words. In total, 23 textbooks with 282 chapters were collected for analysis. Detailed information is provided in the following table:

Table 5.1 Textbook corpus

Textbooks	File name	No. of files	Word count
1	cihak (lee) M1	12	1,614
2	cwungang (kim) M1	12	3,225
3	kumsung (cang) M1	12	3,024
4	kyohak (bae) M1	12	4,305
5	titim (kim) M1	11	3,034
6	twusan (cang) M1	12	2,447
7	twusan (kim) M1	12	1,789
8	chencay (lee) M2	12	4,163
9	cihak (lee) M2	12	4,718
10	cwungang (kim) M2	11	3,967
11	kumsung (cang) M2	12	4,218
12	kyohak (bae) M2	12	4,428
13	titim (kim) M2	11	3,316
14	twusan (cang) M2	12	6,840
15	twusan (kim) M2	12	11,138
16	chencay (hwang) M3	12	6,287
17	chencay (lee) M3	13	5,981
18	cihak (lee) M3	14	5,258
19	donga (cang) M3	13	4,955
20	donga (na) M3	12	8,042
21	hanse M3	13	6,025
22	kyohaksa M3	16	7,076
23	sisu M3	12	7,275
Total		282	105,935

Middle school textbooks, rather than high school textbooks, were chosen because the construction dealt with in this study (i.e., *wh*-question and subj-aux inversion) is a relatively easy structure and is introduced and taught mainly in middle school. The

structure of high school textbooks is very similar to that of middle school textbooks (i.e., a combination of conversation and reading passages), even though the length (the size of each chapter) is different. Therefore, it was thought that the use of *wh*-questions (e.g., subj-aux inversion in *wh*-questions) would be similar in middle school textbooks and high school textbooks.

Various textbooks (seven or eight) were used for each grade to find any similarity among them. If there is a consistent similarity among the textbooks, we may assume that the learners get similar input even though they use different textbooks.

As in the L1 study, the frequency of *wh*-words in each file was first counted by searching for all the lines containing a *wh*-word (i.e. *who*, *what*, *how*, and *why*). Second, the frequency of *wh*-questions was counted by searching for only lines that contain a *wh*-question (excluding *wh*-relative clauses and *wh*-exclamations). This was done by hand, line by line. Third, each *wh*-question was analyzed and categorized into one of three groups as done for the L1 input study: inverted, uninverted and other questions. In the analysis, the “total number of *wh*-words” indicates the number of occurrences of each *wh*-word in the file. The “total number of *wh*-questions” indicates the number of *wh*-questions (which is smaller than the number of *wh*-words). “Inverted” indicates the number of questions with the word order, [*wh*-word+ Aux+Subj]. “Uninverted” indicates the number of questions with the word order, [*wh*-word + Subj + Aux]. Uninverted questions include mainly indirect questions (e.g., *I do not know what you want*). “Other questions” includes those phrases and sentences that cannot be

categorized into either of the other two (e.g., *who?* or *what about—?*). In addition, *what* and *who* subject questions (e.g., *what makes you think so?*, *who is going?* and so on) are included in this category because they do not belong to [*wh*-word+Aux+Subj] or [*wh*-word+Subj+Aux] (See Chapter 4 for detailed discussion).

In the analysis of each *wh*-question, only one-word *wh*-word phrases (e.g., *how*), not multi-word *wh*-phrases (e.g., *how many people*), were counted as the *wh*-element. The category “Subj” includes both pronominals and regular noun phrases. “Aux” includes main verbs (e.g., *who is she?* where *is* is not an auxiliary verb but a main verb) as well as auxiliary verbs (e.g., *who is he looking at?*). After the analysis, inversion rates for each *wh*-question were calculated and compared. Finally, the frequency of inverted questions of each type of *wh*-question was calculated and compared.

To begin, the frequency of *wh*-words was counted. The results are presented in the following table:

Table 5.2 Frequency of *wh*-words in textbooks

Textbook	Total number of <i>wh</i> -words					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	4	2	0	8	14	29%	14%	0%	57%	100%
2	26	2	3	41	72	36%	3%	4%	57%	100%
3	17	3	4	18	42	40%	7%	10%	43%	100%
4	16	5	7	23	51	31%	10%	14%	45%	100%
5	10	4	2	37	53	19%	8%	4%	70%	100%
6	9	0	1	15	25	36%	0%	4%	60%	100%
7	13	0	0	16	29	45%	0%	0%	55%	100%
8	12	17	6	29	64	19%	27%	9%	45%	100%
9	14	6	5	19	44	32%	14%	11%	43%	100%
10	21	2	5	37	65	32%	3%	8%	57%	100%
11	13	16	2	42	73	18%	22%	3%	58%	100%
12	35	8	2	30	75	47%	11%	3%	40%	100%
13	8	10	5	17	40	20%	25%	13%	43%	100%
14	15	6	12	43	76	20%	8%	16%	57%	100%
15	33	22	29	75	159	21%	14%	18%	47%	100%
16	15	8	7	23	53	28%	15%	13%	43%	100%
17	21	9	5	37	72	29%	13%	7%	51%	100%
18	10	5	13	29	57	18%	9%	23%	51%	100%
19	4	2	13	19	38	11%	5%	34%	50%	100%
20	37	12	18	61	128	29%	9%	14%	48%	100%
21	15	8	10	29	62	24%	13%	16%	47%	100%
22	14	14	5	36	69	20%	20%	7%	52%	100%
23	22	19	7	55	103	21%	18%	7%	53%	100%
Total	384	180	161	739	1,464	26%	12%	11%	50%	100%
Mean	16.7	7.8	7.0	32.1	63.7	26%	12%	11%	50%	100%

As shown in Table 5.2, *what* is more frequent than any other *wh*-word in all of the textbooks except text 12, where the number of *hows* is higher than that of *whats* (35 *how* vs. 30 *what*). The relative frequency of *who* and *why* varies depending on the textbook; *how* is more frequent than *who* in all textbooks except in two textbooks (18 and

19), and than *why* except in four textbooks (8, 1, 13 and 22). In sum, the frequency of *wh*-words in the textbook corpus can be summarized in the following rank order:

(1) Relative order of frequency of *wh*-words in textbooks

what > how > who, why

The relative frequency is shown in the following figure:

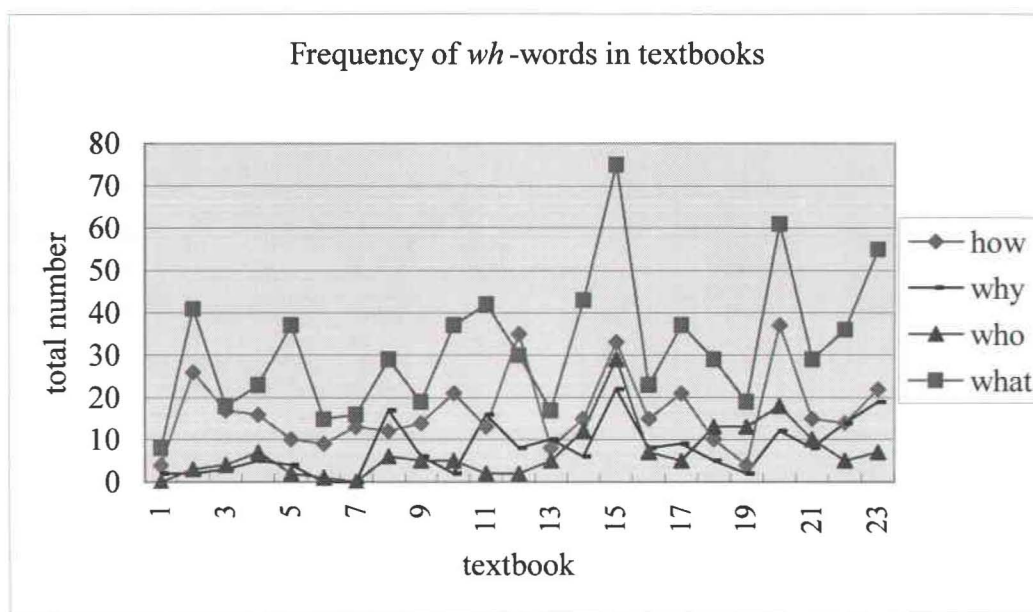


Figure 5.1 Frequency of *wh*-words in textbooks

Figure 5.1 shows that the relative frequency of *wh*-words especially among *what*, *how* and *who* (i.e., *what* > *how* > *who*) is quite consistent among the twenty-three textbooks.

Next, the frequency of *wh*-questions (excluding other lines that do not involve *wh*-questions) was searched. In this search, those questions containing a multi-word *wh*-phrase were excluded (e.g., *how many do you have?*) because the grammatical status

of certain *wh*-phrases is different in these sentences. For example, in the sentence, how many pencils do you have, the *how*-phrase is not an adjunct but an argument. The results are presented in Table 5.3.

Table 5.3 Frequency of *wh*-questions in textbooks

Textbook	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	1	2	0	7	10	10%	20%	0%	70%	100%
2	8	2	3	28	41	20%	5%	7%	68%	100%
3	5	3	4	13	25	20%	12%	16%	52%	100%
4	7	5	7	20	39	18%	13%	18%	51%	100%
5	5	4	2	31	42	12%	10%	5%	74%	100%
6	6	0	1	10	17	35%	0%	6%	59%	100%
7	4	0	0	15	19	21%	0%	0%	79%	100%
8	3	17	6	24	50	6%	34%	12%	48%	100%
9	8	6	5	14	33	24%	18%	15%	42%	100%
10	7	2	5	37	51	14%	4%	10%	73%	100%
11	6	16	2	35	59	10%	27%	3%	59%	100%
12	11	6	2	19	38	29%	16%	5%	50%	100%
13	3	10	5	14	32	9%	31%	16%	44%	100%
14	6	6	12	36	60	10%	10%	20%	60%	100%
15	18	22	29	62	131	14%	17%	22%	47%	100%
16	8	8	7	14	37	22%	22%	19%	38%	100%
17	14	7	5	24	50	28%	14%	10%	48%	100%
18	4	4	13	17	38	11%	11%	34%	45%	100%
19	2	1	5	9	17	12%	6%	29%	53%	100%
20	15	8	5	41	69	22%	12%	7%	59%	100%
21	8	7	2	20	37	22%	19%	5%	54%	100%
22	6	13	1	26	46	13%	28%	2%	57%	100%
23	10	17	5	38	70	14%	24%	7%	54%	100%
Total	165	166	126	554	1,011	16%	16%	12%	55%	100%
Mean	7.17	7.2	5.5	24.1	44.1	16%	16%	12%	55%	100%

Table 5.3 shows that *what* questions are more frequent than any other type of *wh*-question. About half of the *wh*-questions are *what* questions. *How* questions are more frequent

than *who* questions except in six textbooks (8, 13, 14, 15, 18 and 19); the frequency of *who* questions and *how* questions is the same in one textbook (4). The relative frequency of *why* questions varies from textbook to textbook. Table 5.3 shows the following rank order of frequency of *wh*-questions:

(2) Relative frequency of *wh*-questions in textbooks

i) what > how > who

ii) what > why

The relative frequency of *wh*-questions is very similar to the relative frequency of *wh*-words, and the rank order in (2) is quite consistent throughout the twenty-three textbooks.

The consistency is depicted in the following figure:

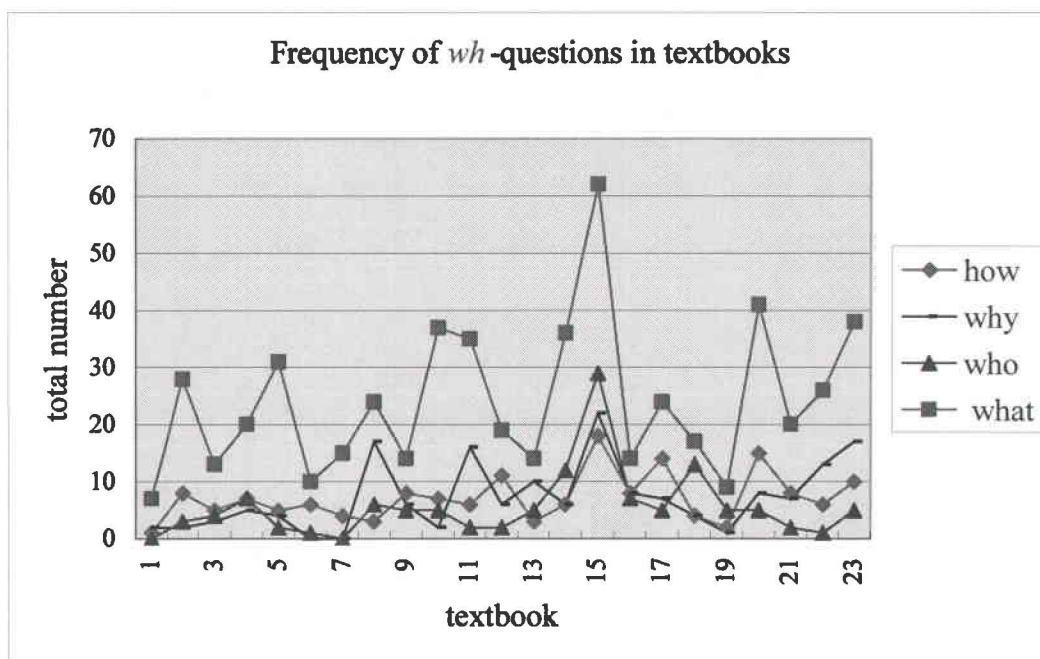


Figure 5.2 Frequency of *wh*-questions in textbooks

Third, each *wh*-question was analyzed and categorized as one of three types: inverted, uninverted and other questions. The results are shown in Table 5.4.

Table 5.4 *Wh*-questions in textbooks

T	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> +Aux+Subj)				Uninverted (<i>Wh</i> +Subj+Aux)				Other questions			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
1	1	2	0	7	1	0	0	6	0	0	0	0	0	2	0	1
2	8	2	3	28	8	0	1	20	0	0	0	0	0	2	2	8
3	5	3	4	13	5	3	4	11	0	0	0	0	0	0	0	2
4	7	5	7	20	7	2	3	19	0	0	0	0	0	3	4	1
5	5	4	2	31	5	1	1	26	0	1	0	0	0	2	1	5
6	6	0	1	10	6	0	0	9	0	0	0	0	0	0	1	1
7	4	0	0	15	4	0	0	13	0	0	0	0	0	0	0	2
8	3	17	6	24	3	13	0	17	0	1	0	0	0	3	6	7
9	8	6	5	14	7	5	0	10	0	0	0	0	1	1	5	4
10	7	2	5	37	6	0	1	20	1	0	0	0	0	2	4	17
11	6	16	2	35	6	7	0	23	0	9	0	4	0	0	2	8
12	11	6	2	19	9	3	0	18	1	0	0	0	1	3	2	1
13	3	10	5	14	3	3	0	9	0	0	0	1	0	7	5	4
14	6	6	12	36	6	4	2	28	0	0	0	1	0	2	10	7
15	18	22	29	62	18	17	1	38	0	0	0	0	0	5	28	24
16	8	8	7	14	7	2	1	10	1	2	1	3	0	4	5	1
17	14	7	5	24	12	3	2	14	1	3	0	4	1	1	3	6
18	4	4	13	17	2	1	1	9	2	1	1	2	0	2	11	6
19	2	1	5	9	2	1	1	3	0	0	0	5	0	0	4	1
20	15	8	5	41	12	2	0	19	3	2	0	12	0	4	5	10
21	8	7	2	20	3	7	1	15	4	0	0	2	1	0	1	3
22	6	13	1	26	5	9	0	18	0	0	0	4	1	4	1	4
23	10	17	5	38	6	10	1	26	3	4	0	9	1	3	4	3
T	165	166	126	554	143	93	20	381	16	23	2	47	6	50	104	126
M	7.2	7.2	5.5	24.1	6.2	4.0	0.9	16.6	0.7	1.09	0.1	2.0	0.3	2.2	4.5	5.5

In general, inverted questions were more frequent than uninverted questions in most of the types of *wh*-questions. The number of uninverted questions is smaller than (the same as at least) that of inverted questions in each type of *wh*-question in all textbooks

except in two textbooks (11 and 21). Unverted questions are more frequent than inverted ones involving *why* questions in textbook 11, and involving *how* questions in textbook 21. This means that *wh*-questions were mainly used as direct questions (e.g., *how is he doing?*), not as indirect questions (e.g., *do you know how he is doing?*). One of the reasons for infrequent indirect questions in middle school textbooks might be that the structure of indirect questions may not yet be introduced or may hardly be used in middle school textbooks. Another reason seems to be that indirect questions themselves may be infrequent compared with direct questions in everyday use of *wh*-questions. This will be shown in the analysis of American TV situation comedies and movies, too, in the next section. It is also notable that the number of “other questions” is quite large, which means that a great number of *wh*-questions (about 30 %) do not involve subj-aux inversion.

As the inverted questions are the most relevant input for learning subj-aux inversion, inversion rates for *wh*-questions were calculated. The inversion rates were calculated by finding the ratio of the number of inverted questions to the total number of *wh*-questions in each type. The results are presented in the following table:

Table 5.5 Inversion rates for *wh*-questions in textbooks

Tbk	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> +Aux+Subj)				Inversion rates			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>What</i>
1	1	2	0	7	1	0	0	6	100%	0%	N/A	86%
2	8	2	3	28	8	0	1	20	100%	0%	33%	71%
3	5	3	4	13	5	3	4	11	100%	100%	100%	85%
4	7	5	7	20	7	2	3	19	100%	40%	43%	95%
5	5	4	2	31	5	1	1	26	100%	25%	50%	84%
6	6	0	1	10	6	0	0	9	100%	N/A	0%	90%
7	4	0	0	15	4	0	0	13	100%	N/A	N/A	87%
8	3	17	6	24	3	13	0	17	100%	76%	0%	71%
9	8	6	5	14	7	5	0	10	88%	83%	0%	71%
10	7	2	5	37	6	0	1	20	86%	0%	20%	54%
11	6	16	2	35	6	7	0	23	100%	44%	0%	66%
12	11	6	2	19	9	3	0	18	82%	50%	0%	95%
13	3	10	5	14	3	3	0	9	100%	30%	0%	64%
14	6	6	12	36	6	4	2	28	100%	67%	17%	78%
15	18	22	29	62	18	17	1	38	100%	77%	3%	61%
16	8	8	7	14	7	2	1	10	88%	25%	14%	71%
17	14	7	5	24	12	3	2	14	86%	43%	40%	58%
18	4	4	13	17	2	1	1	9	50%	25%	8%	53%
19	2	1	5	9	2	1	1	3	100%	100%	20%	33%
20	15	8	5	41	12	2	0	19	80%	25%	0%	46%
21	8	7	2	20	3	7	1	15	38%	100%	50%	75%
22	6	13	1	26	5	9	0	18	83%	69%	0%	69%
23	10	17	5	38	6	10	1	26	60%	59%	20%	68%
Total	165	166	126	554	143	93	20	381	87%	55%	16%	69%
Mn	7.2	7.2	5.5	24.1	6.2	4	0.9	16.6	87%	55%	16%	69%

- Tbk: Textbook, Mn: Mean

The inversion rates vary among different types of *wh*-questions. However, there is consistency among different textbooks (i) between *who* questions and *how* questions on the one hand, and (ii) between *what* questions and *who* questions on the other. The

inversion rates for *how* questions are almost always higher than those for *who* questions, and inversion rates for *what* questions are higher than those for *who* questions except for two textbooks, 3 and 21. In textbook 3, inversion rate for *who* question is higher than that for *what* question, but the same as that for *how* question. In textbook 21, inversion rate for *who* question is higher than *how* question but lower than *what* question. However, this does not seem to affect the overall tendency because the raw number of *who* question in both textbooks is very low (i.e. four for textbook 3 and two for textbook 21). The following figure shows the consistency of the rank order of inversion rates among the four types of *wh*-questions in the different textbooks:

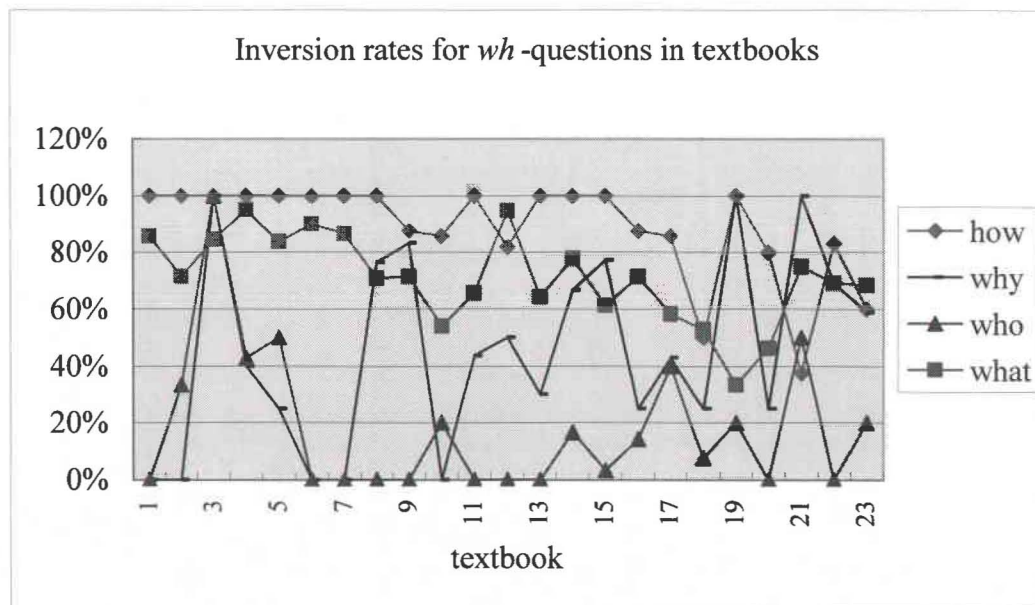


Figure 5.3 Inversion rates for *wh*-questions in textbooks

Figure 5.3 shows the rank order of inversion rates among the four types of *wh*-questions. Except for *why* questions, the relative rank order of inversion rates seems to be consistent throughout the 23 textbooks (i.e., *how* > *what* > *who*). In addition, taking into consideration the raw number of the *how*, *what* and *who* questions shown in Table 5.3, the difference between *who* questions and the other two (i.e., *what* and *how* questions) becomes larger in terms of the frequency of inverted questions. In sum, considering the number of textbooks analyzed here, it seems possible to draw a conclusion about the relative order of inversion rates for four types of *wh*-questions as follows:

(3) Relative order of inversion rates for *wh*-questions in textbooks

- (i) *what* > *who*
- (ii) *how* > *who*
- (iii) *why*: inconsistent

Differences in the rank order of frequency of *wh*-questions and inversion rates are reflected in the different frequency of inverted *wh*-questions. Comparing *what* and *who* questions, *what* questions surpass *who* questions in terms of the frequency order and inversion rates, which results in higher frequency of inverted *what* questions than inverted *who* questions. Comparing *how* and *who* questions, *how* questions surpass *who* questions in terms of frequency and inversion rates, which results in higher frequency of inverted *how* question than inverted *who* questions. The number of inverted *wh*-

questions is presented in Table 5.5 and the rank order of relative frequency is shown in (4).

(4) Relative order of frequency of inverted *wh*-questions in textbooks

what > how > (why) > who

The relative frequency in (4) is depicted in the following figure:

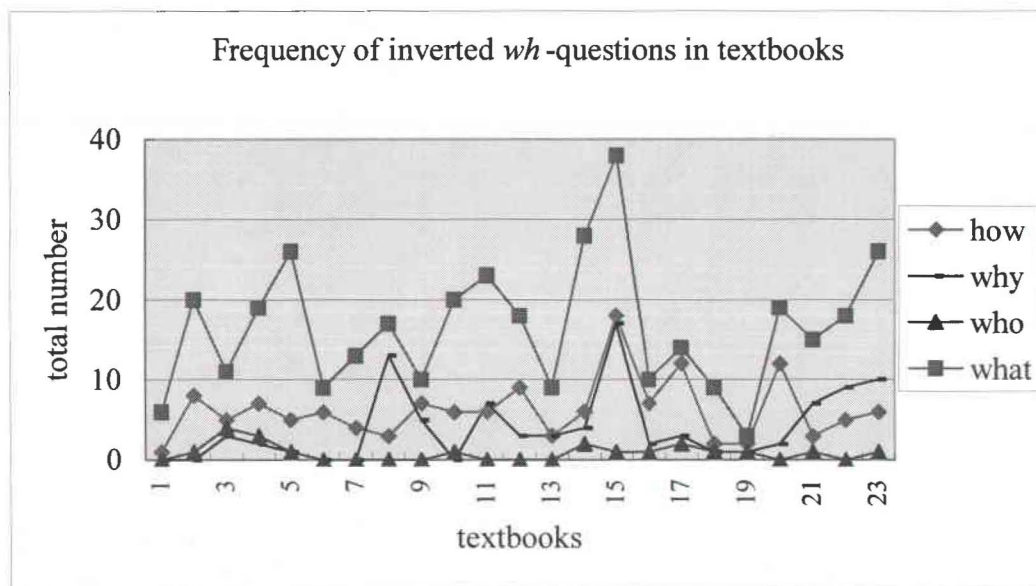


Figure 5.4 Frequency of inverted *wh*-questions in textbooks

To summarize, there is a similarity among different English textbooks in terms of (i) frequency of *wh*-words, (ii) frequency of *wh*-questions, (iii) inversion rates, and (iv) frequency of inverted *wh*-questions. The following graphs help to show the consistent rank order of relative frequency of *wh*-questions in textbook corpus:

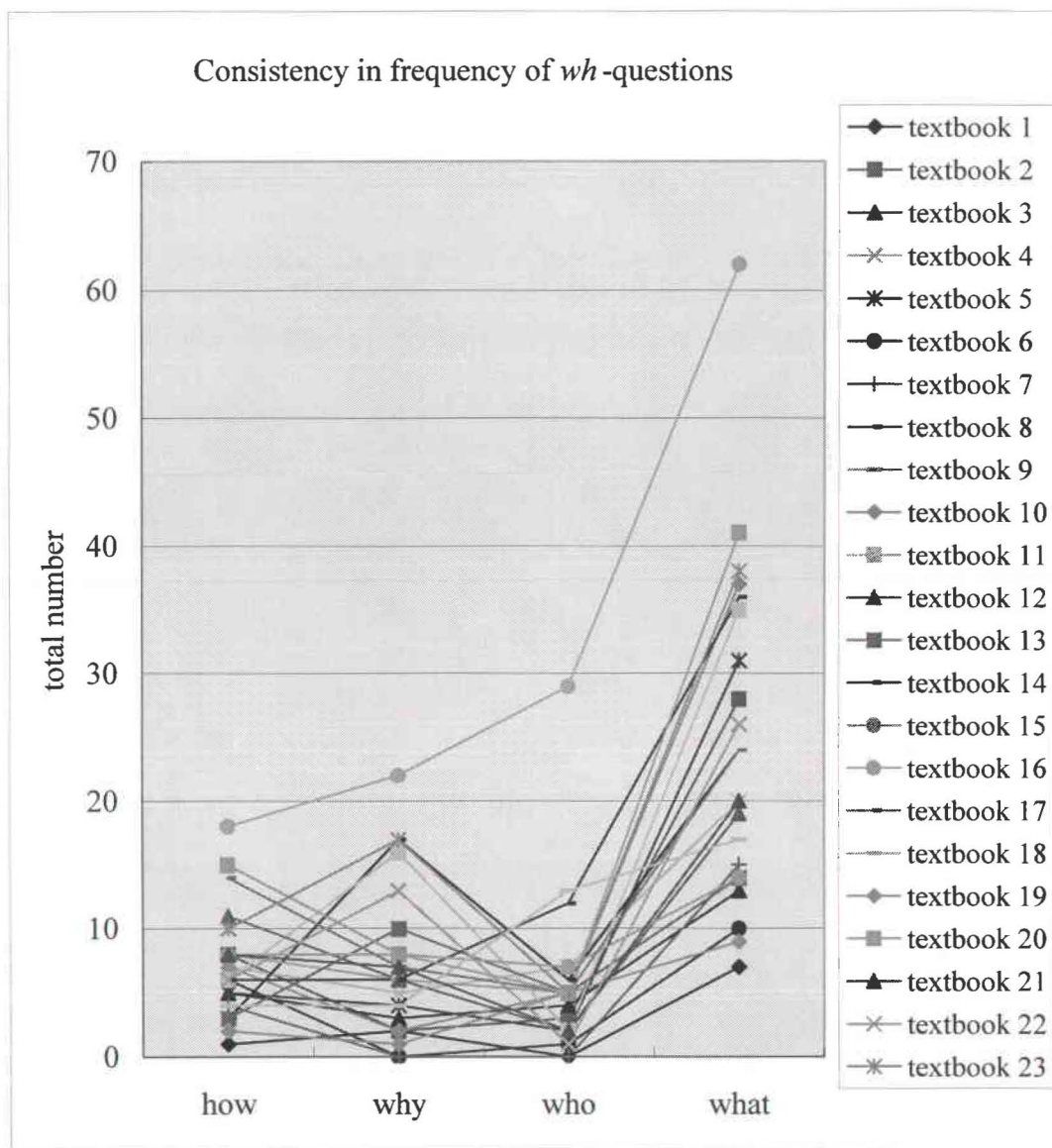


Figure 5.5 Consistency in relative frequency of *wh*-questions in textbooks

The shapes of the curves are very similar to one another (especially the order, *what* > *how* > *who*), which means that the relative frequency among the different types of *wh*-

questions is fairly consistent across all of the twenty-three textbooks.

In addition, the following graph help to show the consistency of frequency of inverted *wh*-questions in textbook corpus:

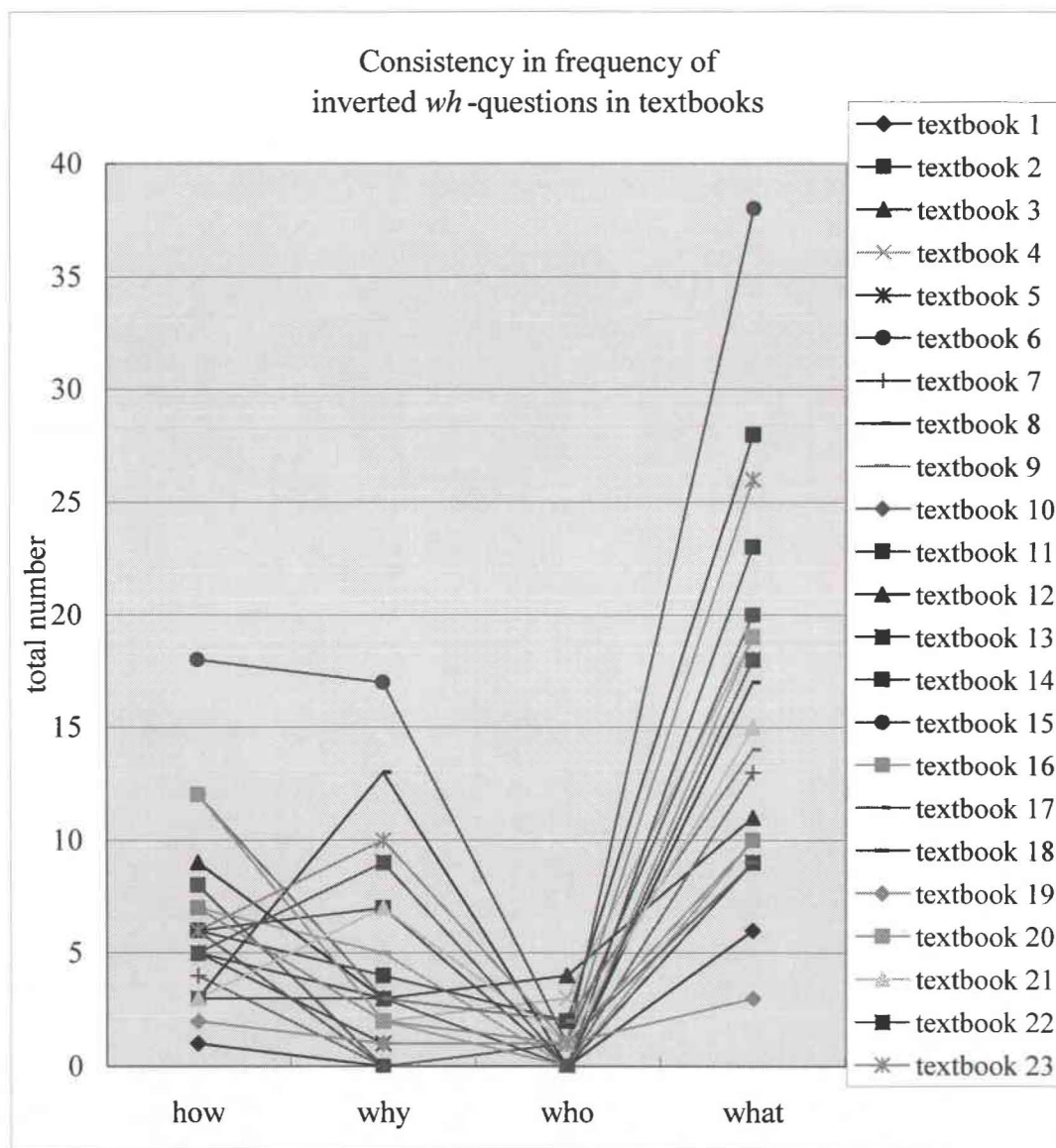


Figure 5.6 Consistency in relative frequency of inverted *wh*-questions in textbooks

The relative frequency of inverted *wh*-questions among *how*, *who* and *what* questions is consistent throughout the 23 textbooks (i.e., *what* > *how* > *who*). If this is a general tendency in textbooks, we can conclude that the learners receive a very similar type of input for subj-aux inversion and *wh*-questions no matter what textbook they are using.

5.1.2 *Wh*-questions in scripts of American television situation comedies and movies

Various American television scripts of situation comedies and movie scripts were analyzed to see if there is any consistency among different input sources for the learners.

The following scripts were analyzed in this study.

(5) Corpora for television and movie scripts

i) Television situation comedies: *Friends* (9 episodes), *Frasier* (9 episodes)

ii) Movies: *Kramer vs. Kramer*, *The Sixth Sense*, *American President*,

As Good As It Gets.

The scripts for these television situation comedies and movies were available on the internet and the researcher had these scripts in text format which had been prepared for another project by Park & Lee (2000).

5.1.2.1 Scripts of television situation comedies

To begin, nine scripts from *Friends* and nine scripts from *Frasier* were chosen to

find any similarity among the scripts from the same program, and between the two sets of scripts from different programs, too. There are advantages to compare various scripts from the episodes of the same program. As the running time of each episode is the same, the scripts are similar in length. Therefore, it is easy to compare the number of search items among the different scripts. In addition, as there are a limited number of characters in each episode, if consistency is found among scripts, it seems to show that the same group of people gives the learner (the viewer) similar input in different contexts (as each script deals with a different topic) regarding the use of *wh*-questions. If a similar tendency is found in different sit-coms (of course, with a different group of characters), it seems reasonable to conclude that the learners are exposed to similar input with respect to *wh*-questions within the specific genre (situation comedy), which is very similar to the style of everyday conversation.

The corpus of *Friends* used in this study is summarized in Table 5.6.

Table 5.6 *Friends* Corpus

Script	File name	Word count
1	Friends 2	2,604
2	Friends 3	2,622
3	Friends 4	2,739
4	Friends 5	2,757
5	Friends 6	2,725
6	Friends 8	2,323
7	Friends 9	2,623
8	Friends 10	2,926
9	Friends 11	3,050
Total		24,369
Mean		2,708

The frequency of *wh*-words searched for in the *Friends* corpus is presented in Table 5.7 and Figure 5.7.

Table 5.7 Frequency of *wh*-words in *Friends*

Script	Total number of <i>wh</i> -words					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	16	6	6	23	51	31%	12%	12%	45%	100%
2	8	4	5	27	44	18%	9%	11%	61%	100%
3	7	3	7	36	53	13%	6%	13%	68%	100%
4	6	5	6	34	51	12%	10%	12%	67%	100%
5	6	6	6	30	48	13%	13%	13%	63%	100%
6	10	1	10	28	49	20%	2%	20%	57%	100%
7	5	10	4	21	40	13%	25%	10%	53%	100%
8	7	4	3	19	33	21%	12%	9%	58%	100%
9	11	8	10	26	55	20%	15%	18%	47%	100%
Total	76	47	57	244	424	18%	11%	13%	58%	100%
Mean	8.4	5.2	6.3	27.1	47.1	18%	11%	13%	58%	100%

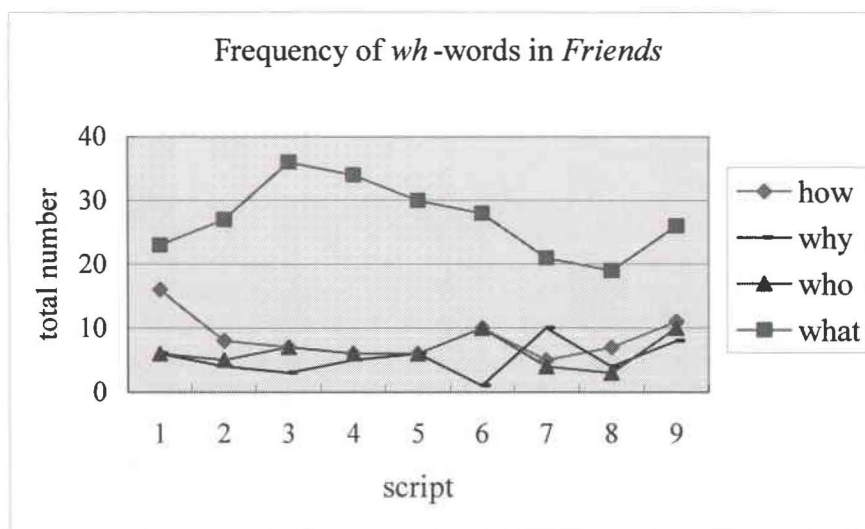


Figure 5.7 Frequency of *wh*-words in *Friends*

Figure 5.7 shows that *what* is more frequent than any other *wh*-word in every script.

Hows are more frequent than *whos* in five scripts (script 1, 2, 7, 8, and 9), but as frequent

as *whos* in the other four scripts (script 3, 4, 5, and 6). *Hows* are more frequent than *whys* except in two scripts (5 and 7). In script 5, the frequency of *how* is the same as that of *why*. In Script 7, *how* is more frequent than *why*. Finally, *whos* are more frequent than *whys* in five scripts (2, 3, 4, 6, and 9), but as frequent as *whys* in two scripts (1 and 5). In scripts 7 and 8, *whys* are more frequent than *whos*. In sum, our search gives us a relative frequency for four *wh*-words, over the nine different scripts as follows:

(6) Order of relative frequency of *wh*-words in *Friends*

i) what > how, who

ii) how > why

The next step counted only *wh*-questions, that is, *wh*-words used as relative pronouns and exclamatives were excluded, but other types of *wh*-questions were included (e.g., *what?* or *I don't know what to do?*). The results are shown in the following table:

Table 5.8 Frequency of *wh*-questions in *Friends*

S	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	7	5	0	20	32	22%	16%	0%	63%	100%
2	6	3	2	23	34	18%	9%	6%	68%	100%
3	3	3	5	33	44	7%	7%	11%	75%	100%
4	4	5	4	30	43	9%	12%	9%	70%	100%
5	6	6	4	20	36	17%	17%	11%	56%	100%
6	7	1	6	22	36	19%	3%	17%	61%	100%
7	3	9	4	18	34	9%	26%	12%	53%	100%
8	4	4	3	16	27	15%	15%	11%	59%	100%
9	8	8	4	24	44	18%	18%	9%	55%	100%
Total	48	44	32	206	330	15%	13%	10%	62%	100%
Mean	5.3	4.9	3.6	22.9	36.7	15%	13%	10%	62%	100%

Table 5.8 shows that *what* questions are more frequent than any other question. *How* questions are more frequent than *who* questions except for three scripts (3, 4 and 7). In script 3 and 7, *who* questions are more frequent than *how* questions whereas the frequency of the two types is the same in script 4. However, the differences between the two are very small (i.e., 2 and 1 respectively). Comparing *why* questions and *who* questions, *why* questions are more frequent than *who* questions except in scripts 3 and 6. The relative frequency of *wh*-questions is summarized in (7) and the rank seems to be quite consistent among different scripts.

(7) Relative order of frequency *wh*-questions in *Friends*

what > how, why > who

The following figure helps depict the rank order in (7).

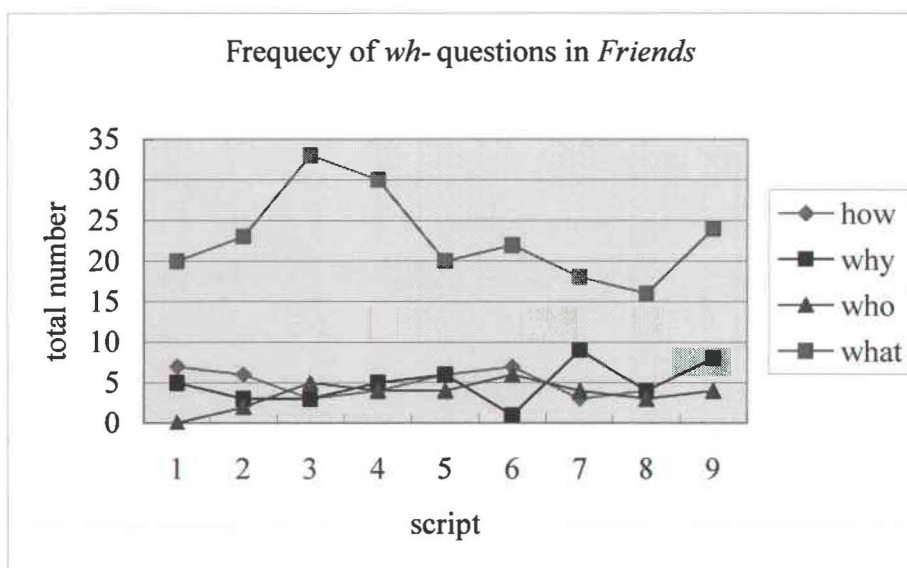


Figure 5.8 Frequency of *wh*-questions in *Friends*

The difference in frequency between *what* questions and the other types of question are so large that the relative frequency rank among the other three *wh*-questions seem to be weak.

Each *wh*-question was then analyzed and categorized into three types: inverted, uninverted and other questions. The results are presented in Table 5.9.

Table 5.9 *Wh*-questions in *Friends*

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Uninverted (<i>Wh</i> + Subj+ Aux)				Other questions			
S	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
1	7	5	0	20	5	5	0	7	0	0	0	4	2	0	0	9
2	6	3	2	23	4	3	1	6	2	0	0	7	0	0	1	10
3	3	3	5	33	3	1	4	6	0	0	1	3	0	2	0	24
4	4	5	4	30	2	3	0	6	0	0	0	2	2	2	4	22
5	6	6	4	20	2	3	3	7	1	1	0	3	3	2	1	10
6	7	1	6	22	4	0	1	5	2	0	3	4	1	1	2	13
7	3	9	4	18	2	6	0	3	1	0	0	3	0	3	4	12
8	4	4	3	16	3	2	0	3	1	0	0	2	0	2	3	11
9	8	8	4	24	4	5	1	16	3	0	1	1	1	3	2	7
Tl	48	44	32	206	29	28	10	59	10	1	5	29	9	15	17	118
Mn	5.3	4.9	3.6	22.9	3.2	3.1	1.1	6.5	1.1	0.1	0.6	3.2	1	1.7	1.9	13.1

Overall, only *what* questions are used as inverted and uninverted questions in all scripts, but the other types of *wh*-questions are mainly used as inverted questions. Many of *who* questions (17 out of 32) and *what* questions (118 out of 206) belong to “other questions” compared to *why* and *how* questions. One reason for the large number of “other questions” in the case of *who* questions is related to *who* subject questions. Since *who* subject questions do not permit subj-aux inversion (e.g., *who is leaving?*), all the *who* subject questions are categorized as “other questions”. In fact, there were twelve *who* subject questions among seventeen “other questions” involving *who*. On the other hand, in the case of *what*, there are not only *what* subject questions (e.g., *what happened?*) but also many questions composed of only one word (e.g., *what?*).

In order to find any consistency in the inversion rate of the different scripts,

inversion rates for each *wh*-question type in each script were calculated. The results are presented in Table 5.10 and Figure 5.9.

Table 5.10 Inversion rates for *wh*-questions in *Friends*

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Inversion rates			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
S												
1	7	5	0	20	5	5	0	7	71%	100%	N/A	35%
2	6	3	2	23	4	3	1	6	67%	100%	50%	26%
3	3	3	5	33	3	1	4	6	100%	33%	80%	18%
4	4	5	4	30	2	3	0	6	50%	60%	0%	20%
5	6	6	4	20	2	3	3	7	33%	50%	75%	35%
6	7	1	6	22	4	0	1	5	57%	0%	17%	23%
7	3	9	4	18	2	6	0	3	67%	67%	0%	17%
8	4	4	3	16	3	2	0	3	75%	50%	0%	19%
9	8	8	4	24	4	5	1	16	50%	63%	25%	67%
Tl	48	44	32	206	29	28	10	59	60%	64%	31%	29%
Mn	5.3	4.9	3.6	22.9	3.2	3.1	1.1	6.5	60%	64%	31%	29%

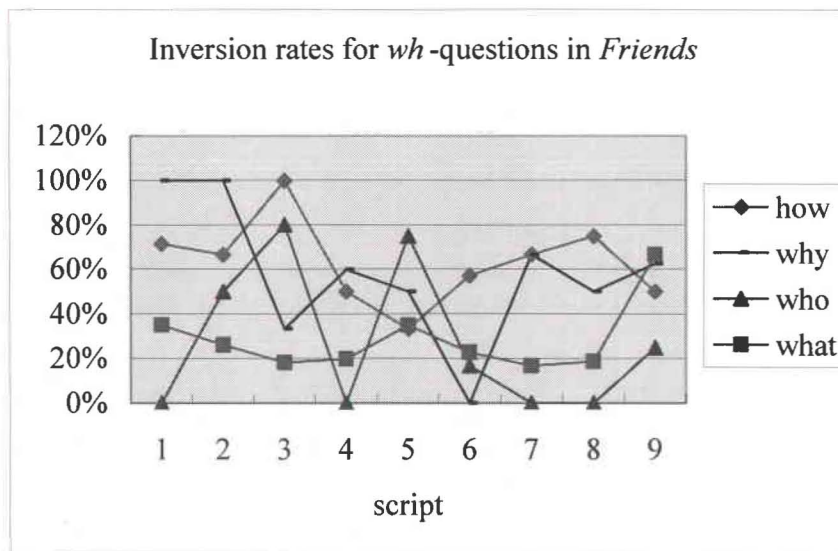


Figure 5.9 Inversion rates for *wh*-questions in *Friends*

In general, inversion rates for *how* and *why* questions are higher than those for *what* and *who* questions. Inversion rates for *how* and *why* questions are above 50% in almost all scripts except for three cases (i.e., 33% in script 3 and 0% in script 6 for *why* questions; 33% in script 3 for *how* question). By contrast, inversion rates for *who* and *what* questions are lower than 50% except four cases (i.e., 50% in script 2, 80% in script 3 and 75% in script 5 for *who* questions; 67% in script for *what* questions). In sum, there seems to be the following rank order for inversion rate for *wh*-questions among different scripts in *Friends*:

(8) Relative order of inversion rates for *wh*-questions in *Friends*

why, how > what, who

There seems to be consistency among the different scripts of *Friends* in terms of (i) the frequency of *wh*-words, (ii) the frequency of *wh*-questions and (iii) the inversion rate for *wh*-questions as shown in (6), (7) and (8).

Here, again, differences in the frequency of *wh*-questions and inversion rates among different *wh*-questions is reflected in a different frequency of inverted *wh*-questions. First, even though the inversion rates for *what* and *who* questions are similar, inverted *what* questions are more frequent than inverted *who* questions because of the higher frequency of *what* questions than *who* questions. Second, comparing *how* and

who questions, the frequency and inversion rate is higher for *how* questions than for *who* questions, inverted *how* questions are more frequent than inverted *who* questions. The frequency number is shown in Table 5.10 and the relative order among the three types of questions is shown in (9).

(9) Relative order of frequency of inverted *wh*-questions in *Friends*

i) what > how > who

ii) what > why

The order in (9) is depicted in the following figure:

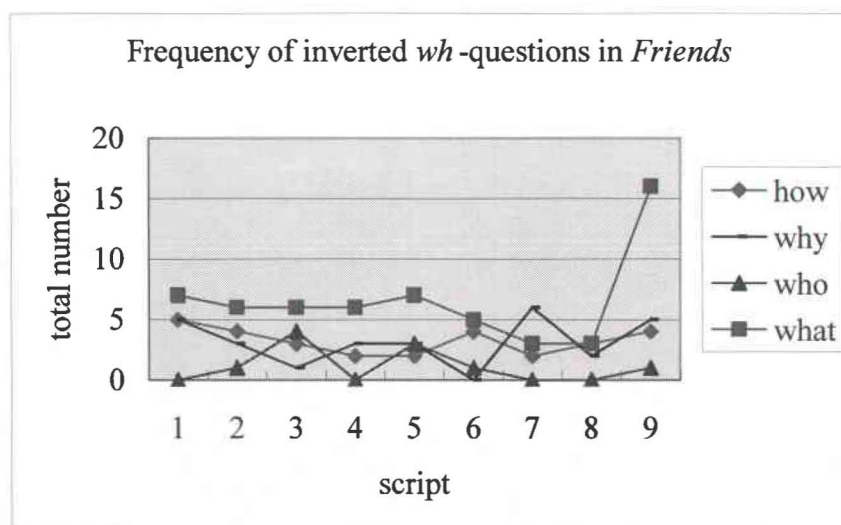


Figure 5.10 Frequency of inverted *wh*-questions in *Friends*

The relative frequency of inverted *wh*-questions is quite consistent across the

scripts except for script 3 and 5, where the number of inverted *how* question is larger than the number of inverted *who* question by one. Even though the frequency of *what* questions was much higher than any other question, the number of inverted *what* questions is relatively small because of the low inversion rate. The number of inverted *who* questions is very small because of their low frequency and inversion rate.

In the next step, the data from another situation comedy, *Frasier* were analyzed and compared with the data from *Friends*, to find any consistency between the two different situation comedies. Nine scripts from *Frasier* were collected and analyzed in this study. A summary of the *Frasier* corpus is presented in Table 5.11.

Table 5.11 *Frasier* corpus

Script	File name	Word count
1	Frasier 1	4,484
2	Frasier 2	4,045
3	Frasier 3	4,175
4	Frasier 4	4,354
5	Frasier 5	4,218
6	Frasier 6	4,376
7	Frasier 7	4,328
8	Frasier 8	4,356
9	Frasier 9	4,412
Total		38,748
Mean		4,305

The results of a frequency search of *wh*-words are summarized in Table 5.12 and Figure 5.11.

Table 5.12 Frequency of *wh*-words in *Frasier*

Script	Total number of <i>wh</i> -words					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	6	4	5	29	44	14%	9%	11%	66%	100%
2	15	4	9	28	56	27%	7%	16%	50%	100%
3	6	8	12	20	46	13%	17%	26%	43%	100%
4	12	3	10	29	54	22%	6%	19%	54%	100%
5	9	3	10	32	54	17%	6%	19%	59%	100%
6	9	5	9	22	45	20%	11%	20%	49%	100%
7	10	1	6	28	45	22%	2%	13%	62%	100%
8	9	8	3	31	51	18%	16%	6%	61%	100%
9	15	5	4	37	61	25%	8%	7%	61%	100%
Total	91	41	68	256	456	20%	9%	15%	56%	100%
Mean	10.1	4.6	7.6	28.4	50.7	20%	9%	15%	56%	100%

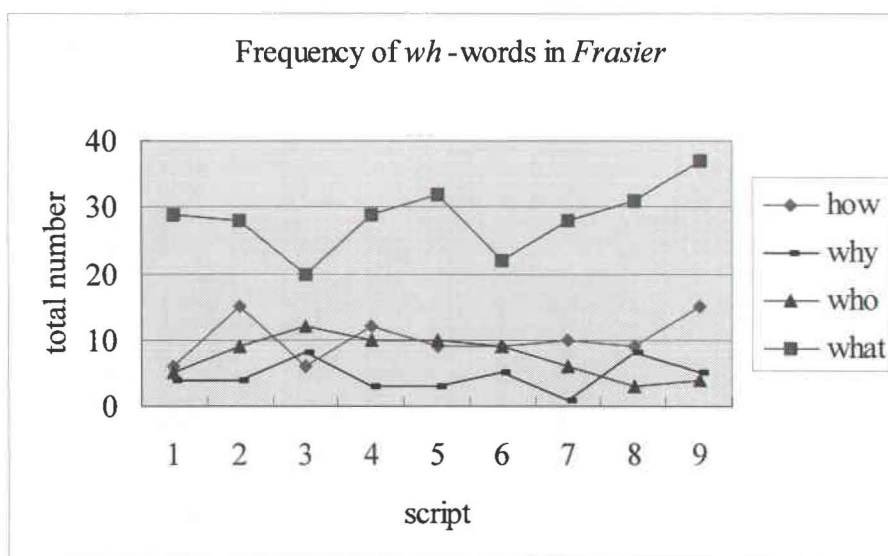


Figure 5.11 Frequency of *wh*-words in *Frasier*

The relative frequency of *wh*-words found in *Friends* is similar to that in *Frasier*. *Hows* are more frequent than *whos* except in scripts, 3, 5 and 6. In scripts, 3 and 5, *whos* are

more frequent than *hows* by six and one respectively. In general, *whos* are more frequent than *whys* except for two scripts (8 and 9), but less frequent than *hows* except for one script (3). In script 6, the frequencies of *hows* and *whos* are the same. This is summarized in (10).

(10) Relative order of frequency of *wh*-words in *Frasier*

what > how > who > why

Next, only *wh*-questions were counted. The results are presented in the following table:

Table 5.13 Frequency of *wh*-questions in *Frasier*

S	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	4	3	2	26	35	11%	9%	6%	74%	100%
2	7	2	4	19	32	22%	6%	13%	59%	100%
3	2	5	4	18	29	7%	17%	14%	62%	100%
4	8	3	5	18	34	24%	9%	15%	53%	100%
5	3	3	9	25	40	8%	8%	23%	63%	100%
6	7	4	6	15	32	22%	13%	19%	47%	100%
7	6	1	2	18	27	22%	4%	7%	67%	100%
8	8	8	1	21	38	21%	21%	3%	55%	100%
9	6	5	3	27	41	15%	12%	7%	66%	100%
Tl	51	34	36	187	308	17%	11%	12%	61%	100%
Mn	5.7	3.8	4.0	20.8	34.0	17%	11%	12%	61%	100%

The relative frequency of *wh*-questions was similar to that of *wh*-words. *What* questions are more frequent than any other type of *wh*-question. *How* questions are more frequent than *who* questions except in two scripts (3 and 5). There seems to be a pattern in frequency order for *wh*-questions in *Frasier* as follows:

(11) Relative order of frequency of *wh*-questions in *Frasier*

i) what > how > who

ii) what > why

The rank order in (11) seems to be consistent among different scripts as depicted in Figure 5.12.

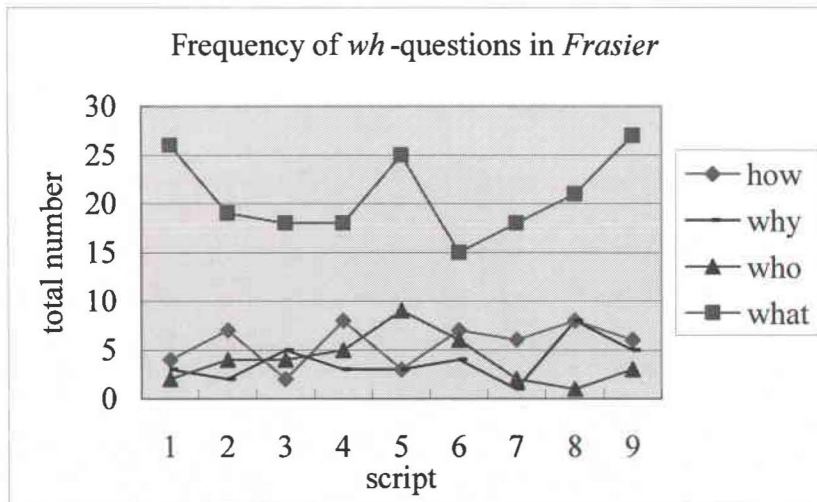


Figure 5.12 Frequency of *wh*-questions in *Frasier*

As in scripts of *Friends*, the difference between *what* questions and the other three types

of questions is so great that the difference among the other three seems to be trivial.

However, a similar tendency emerges in different scripts in terms of the relative frequency of *how* and *who* questions (i.e., *how* > *who*).

All the *wh*-questions were analyzed again and categorized into three groups: inverted, uninverted and other questions. The results are shown in Table 5.14.

Table 5.14 *Wh*-questions in *Fraiser*

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Uninverted (<i>Wh</i> + Subj+ Aux)				Other questions			
S	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
1	4	3	2	26	3	3	0	6	1	0	0	6	0	0	2	14
2	7	2	4	19	2	1	1	6	4	1	0	7	1	0	3	6
3	2	5	4	18	2	3	2	6	0	1	0	4	0	1	2	8
4	8	3	5	18	6	3	1	6	2	0	0	3	0	0	4	9
5	3	3	9	25	3	3	4	6	0	0	0	6	0	0	5	13
6	7	4	6	15	3	2	1	3	4	1	2	4	0	1	3	8
7	6	1	2	18	6	1	1	7	0	0	0	3	0	0	1	8
8	8	8	1	21	5	6	0	5	3	1	0	6	0	1	1	10
9	6	5	3	27	3	4	2	12	2	0	0	3	1	1	1	12
Tl	51	34	36	187	33	26	12	57	16	4	2	42	2	4	22	88
Mn	5.7	3.8	4.0	20.8	3.7	2.9	1.3	6.3	1.8	0.4	0.2	4.7	0.2	0.4	2.4	9.8

Overall, the relative frequency of the three types is very similar to that in *Friends*. Only *what* question is used as inverted, uninverted and other questions in all scripts. On average, inverted *wh*-questions are more frequent than uninverted *wh*-questions (*how*: 33 vs. 16, *why*: 26 vs. 4, *who*: 12 vs. 2 and *what*: 57 vs. 42 respectively). Uninverted *who* questions are found only in one script (6). *Who* questions are used as either inverted or other questions; more than half of the *who* questions (i.e., 22 out of 36) belong to “other

questions”, which is due to the large number of subject questions (7 out of 22).

In order to identify any consistency among the scripts, the inversion rates of each *wh*-question were calculated. The results are presented in Table 5.15 and Figure 5.13.

Table 5.15 Inversion rates for *wh*-questions in *Frasier*

	Total number of <i>wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Inversion rate			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
S												
1	4	3	2	26	3	3	0	6	75%	100%	0%	23%
2	7	2	4	19	2	1	1	6	29%	50%	25%	32%
3	2	5	4	18	2	3	2	6	100%	60%	50%	33%
4	8	3	5	18	6	3	1	6	75%	100%	20%	33%
5	3	3	9	25	3	3	4	6	100%	100%	44%	24%
6	7	4	6	15	3	2	1	3	43%	50%	17%	20%
7	6	1	2	18	6	1	1	7	100%	100%	50%	39%
8	8	8	1	21	5	6	0	5	63%	75%	0%	24%
9	6	5	3	27	3	4	2	12	50%	80%	67%	44%
Tl	51	34	36	187	33	26	12	57	65%	76%	33%	30%
Mn	5.7	3.8	4.0	20.8	3.7	2.9	1.3	6.3	65%	76%	33%	30%

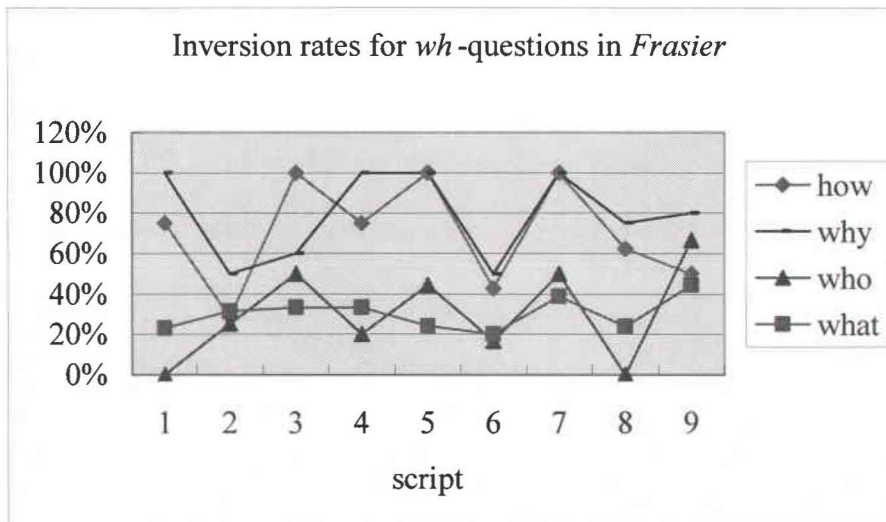


Figure 5.13 Inversion rates for *wh*-questions in *Frasier*

In general, inversion rates for *how* and *why* questions are higher than those for *who* and *what* questions (average 65% and 76% vs. 33% and 30% respectively). The inversion rates for *how* is above 50% except for two scripts (29% in script 2 and 43% in script 6). The relative order of inversion rates among the four types of *wh*-questions is consistent in the scripts of *Frasier*, as it was in *Friends* corpus. In sum, we find the following rank order of inversion rates for *wh*-questions in *Frasier* corpus.

(12) Relative order of inversion rates for *wh*-questions in *Frasier*

why , how > what, who

Here again, the different frequency and inversion rates among different types of *wh*-questions is reflected in a different frequency of inverted questions. First, even though the inversion rates for both *wh*-questions are similar, inverted *what* questions are more frequent than inverted *who* questions (*what* > *who*) because of the higher frequency of *what* questions than *who* questions. Second, because the frequency and inversion rates are higher in *how* questions than in *who* questions, inverted *how* questions are more frequent than inverted *who* questions. In other words, because *who* questions are lower in frequency and inversion rates than any other type of question, the frequency of inverted *who* questions is very low. The total number of inverted *wh*-questions is shown in Table 5.15, and the rank order of relative frequency is shown in (13).

(13) Relative order of frequency of inverted *wh*-questions in *Frasier*

what > how, why > who

The order in (13) is depicted in Figure 5.14:

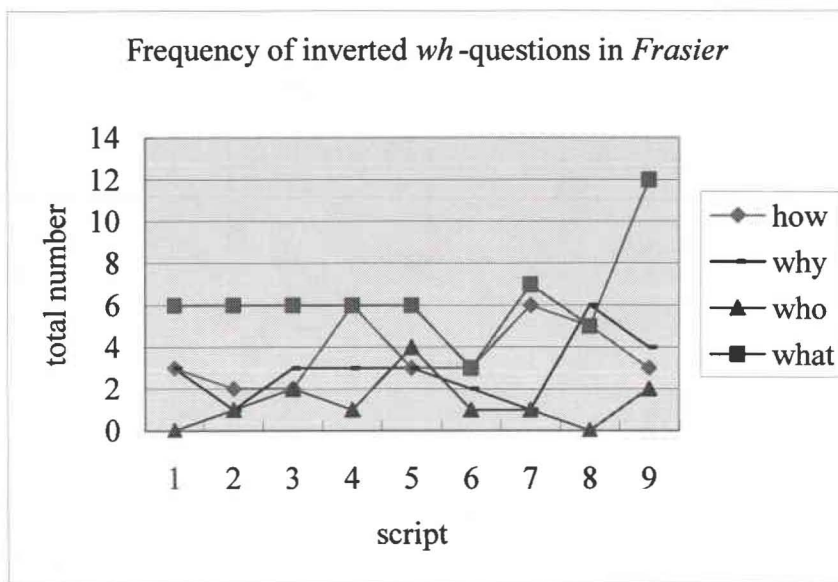


Figure 5.14 Frequency of inverted *wh*-questions in *Frasier*

In sum, there seems to be consistency among the different scripts from different situation comedies in terms of (i) the relative frequency of *wh*-words, (ii) the relative frequency of *wh*-questions (iii) the relative inversion rate for *wh*-questions and (iv) the relative frequency of inverted *wh*-questions. The following figures help to summarize the consistency among the different scripts in *Friends* and *Frasier*.

First, compare the frequency of *wh*-words in *Friends* (Figure 5.15) and *Frasier* (Figure 5.16).

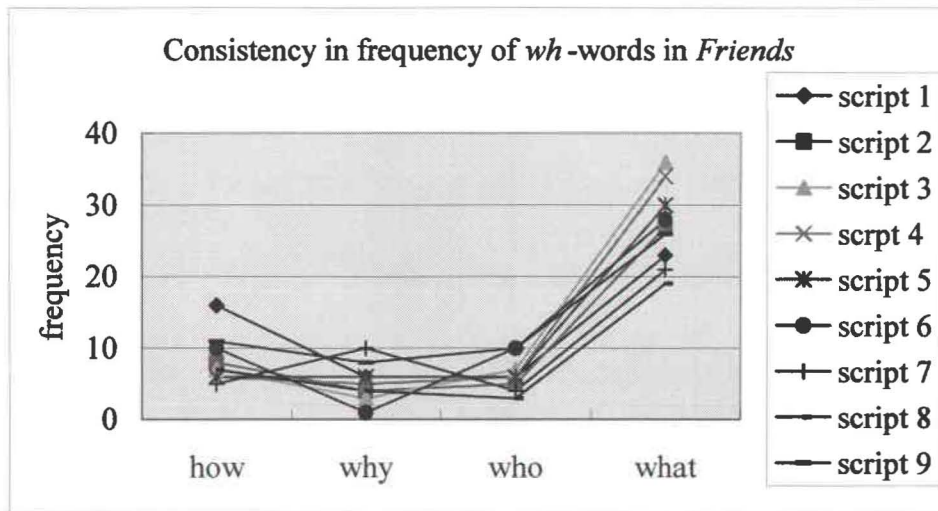


Figure 5.15 Consistency in relative frequency of *wh*-words in *Friends*

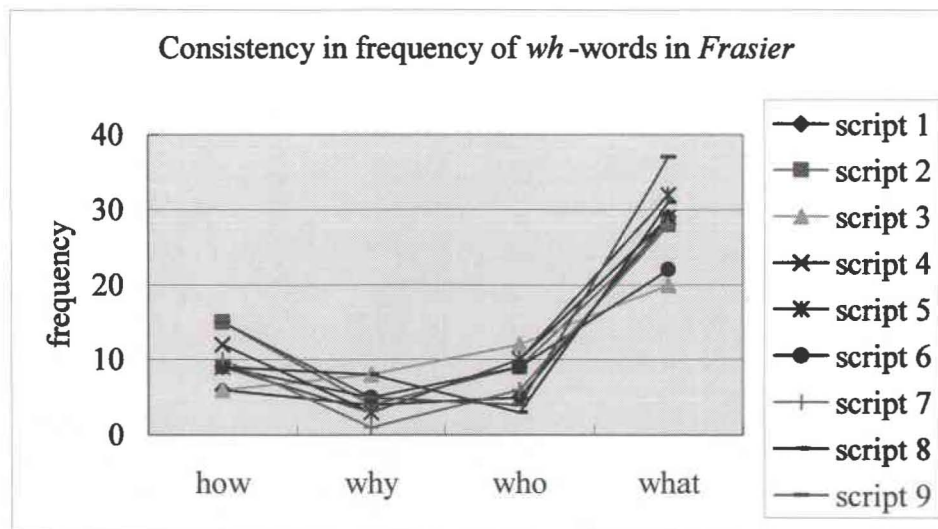


Figure 5.16 Consistency in relative frequency of *wh*-words in *Frasier*

The shape of the curves in the two scripts is very similar, which means that the two different sitcom scripts are very similar to each other in terms of the relative frequency of *wh*-words.

Second, compare the frequency of *wh*-questions in *Friends* (Figure 5.17) and *Frasier* (Figure 5.18).

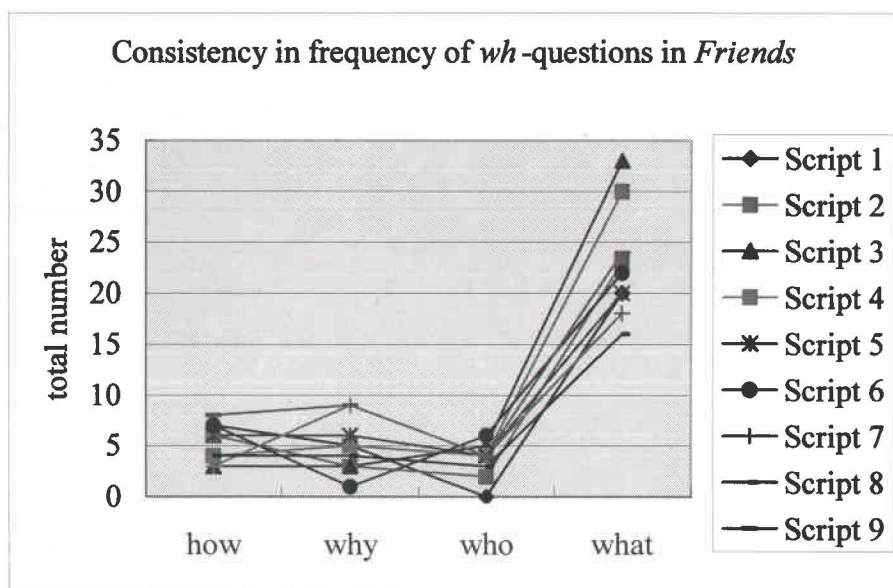


Figure 5.17 Consistency in relative frequency of *wh*-questions in *Friends*

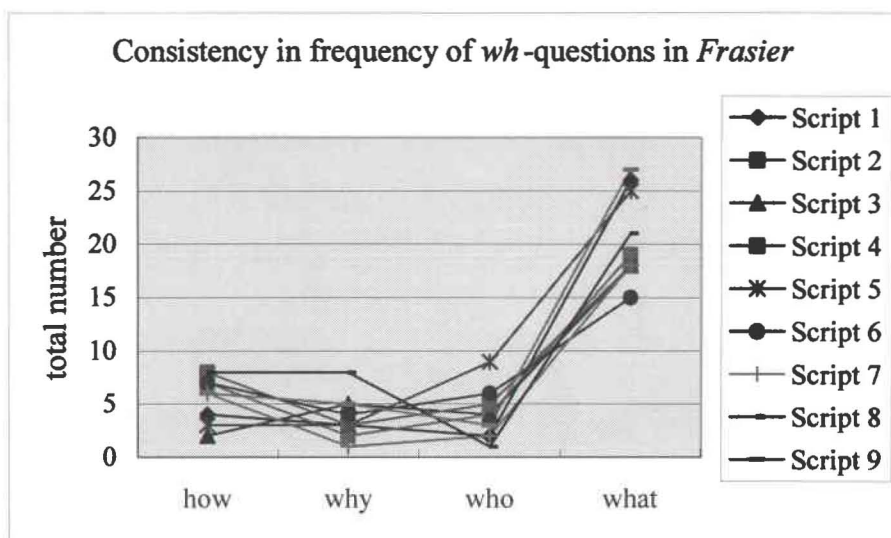


Figure 5.18 Consistency in relative frequency of *wh*-questions in *Frasier*

Again, the shape of the curves in two scripts is very similar in the two sets of scripts, which means that they are quite similar to each other in terms of the relative frequency of *wh*-questions.

Third, compare the inversion rates for *wh*-questions in *Friends* (Figure 5.19) and *Frasier* (Figure 5.20).

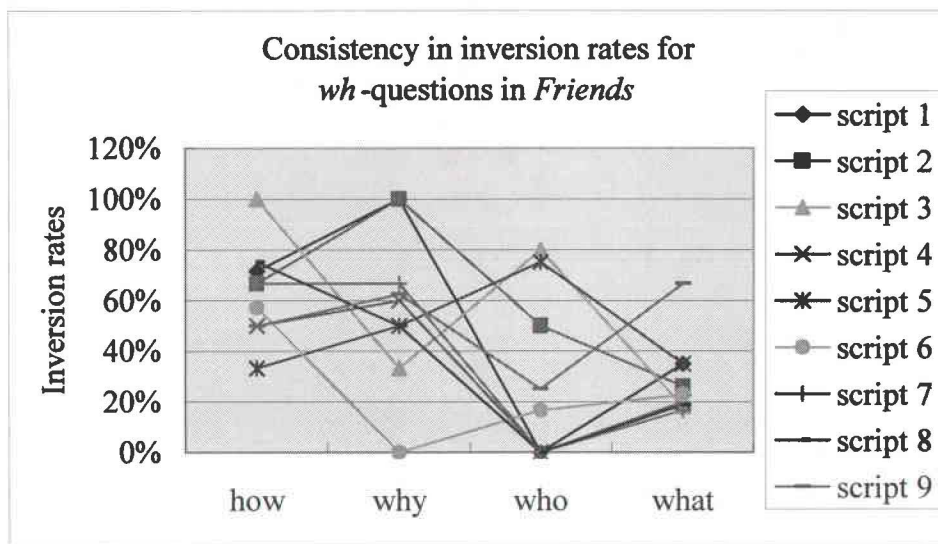


Figure 5.19 Consistency in inversion rates for *wh*-questions in *Friends*

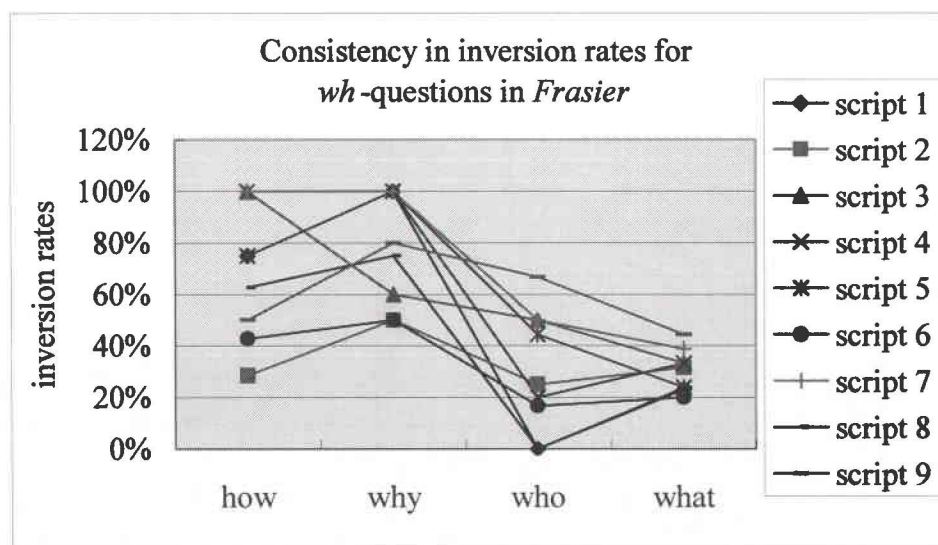


Figure 5.20 Consistency in inversion rates for *wh*-questions in *Frasier*

The shapes of curves are similar to each other in the two sets of scripts from different sitcoms except for a few cases. This tells us that the two scripts are very similar in terms of the inversion rates for *wh*-questions.

Fourth, compare the relative frequency of inverted *wh*-questions in *Friends* (Figure 5.21) and *Frasier* (Figure 5.22).

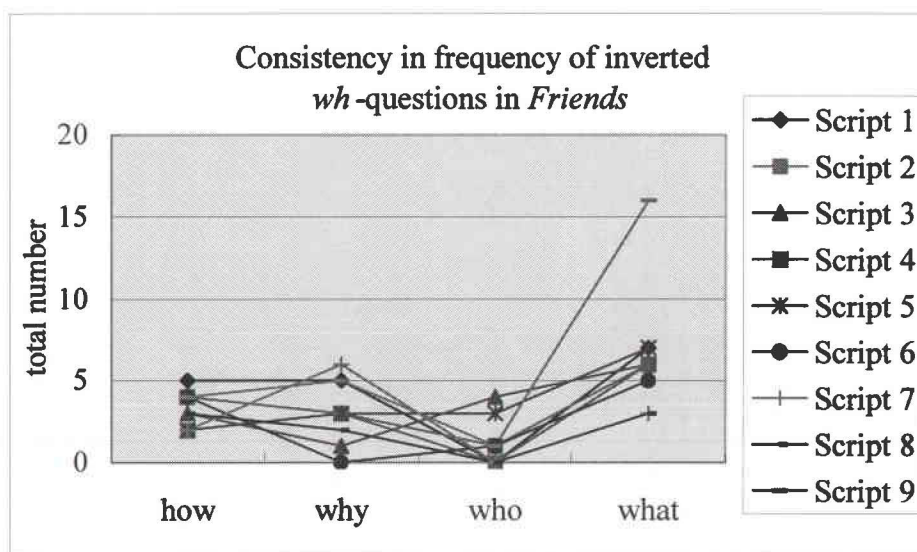


Figure 5.21 Consistency in relative frequency of inverted *wh*-questions in *Friends*

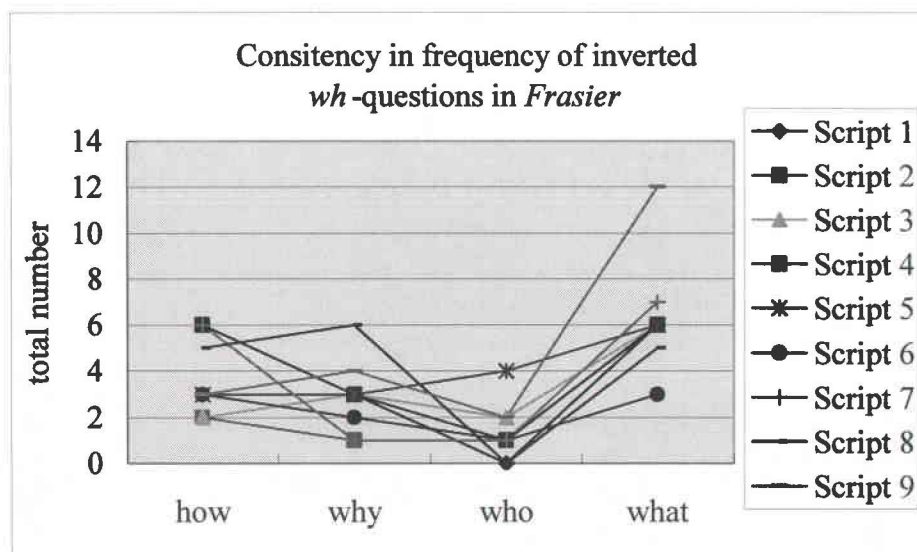


Figure 5.22 Consistency in relative frequency of inverted *wh*-questions in *Frasier*

Again, the shape of curves is very similar to each other in the two different sets of scripts.

This tells us that the two sets of scripts are very similar in terms of the relative frequency of inverted *wh*-questions.

In sum, the shape of curves for the two different sets of television sitcom scripts is similar in the aspects of (i) relative frequency of *wh*-words, (ii) relative frequency of *wh*-questions, (iii) relative inversion rate for *wh*-questions and (iv) relative frequency of inverted *wh*-questions. The similarity among the scripts in each sitcom and the similarity between the two sets of scripts from different sitcoms lead us to the conclusion that learners almost certainly receive similar input from the different input sources regarding these four measures. The relative frequency orders are summarized as follows.

(14) Relative frequency order in TV sitcom scripts

- i) Frequency of *wh*-words: *what* > *how* > *why*
- ii) Frequency of *wh*-questions: *what* > *how* > *who*
- iii) Inversion rate: *how, why* > *who, what*
- iv) Frequency of inverted *wh*-questions
 - a) *what* > *how*
 - b) *how* > *who*

5.1.2.2 Movie scripts

Four movie scripts were also analyzed in this input study. The movie corpus is summarized in the following table:

Table 5.16 Movie corpus

Scripts	File name	Word count
1	<i>Kramer vs. Kramer</i>	13,315
2	<i>The Sixth Sense</i>	7,432
3	<i>American President</i>	25,930
4	<i>As Good As It Gets</i>	15,185
Total		61,862
Mean		15,466

The same methods used for the television sitcom scripts were employed here.

To begin, the frequency of *wh*-words was calculated. The results are shown in Table 5.17.

Table 5.17 Frequency of *wh*-words in movies

Script	Total number of <i>wh</i> -words					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
1	41	22	10	107	180	23%	12%	6%	59%	100%
2	28	13	18	83	142	20%	9%	13%	58%	100%
3	44	28	50	141	263	17%	11%	19%	54%	100%
4	57	28	36	115	236	24%	12%	15%	49%	100%
Total	170	91	114	446	821	21%	11%	14%	54%	100%
Mean	42.5	22.8	28.5	111.5	205.3	21%	11%	14%	54%	100%

As shown in Table 5.17, the length of each script varies; however, the relative frequency of *wh*-words seems to be the same as in the previous television sitcom scripts.

(15) Relative order of frequency of *wh*-words in movies

what > how > who, why

This order is depicted in Figure 5.23.

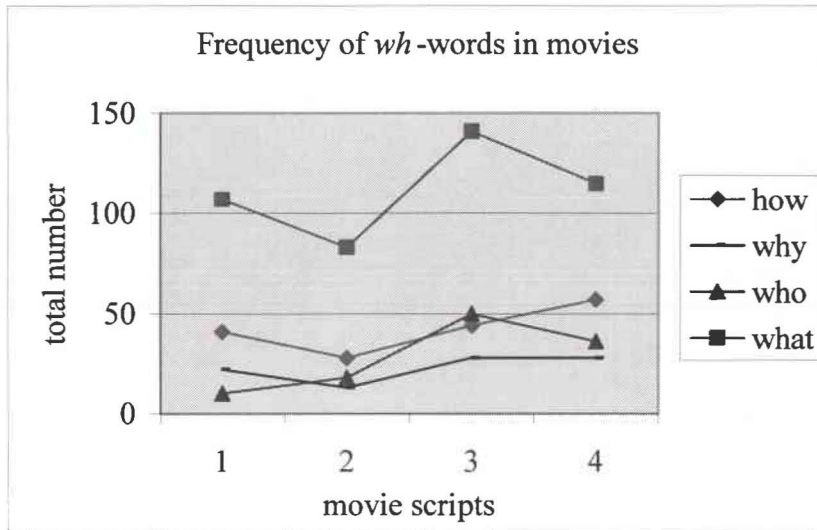


Figure 5.23 Frequency of *wh*-words in movies

Next, the frequency of *wh*-questions was calculated by hand. The results are presented in the following table.

Table 5.18 Frequency of *wh*-questions in movies

	Total number of <i>wh</i> -questions					Percentage				
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Total
S										
1	15	16	7	75	113	13%	14%	6%	66%	100%
2	12	13	8	66	99	12%	13%	8%	67%	100%
3	22	28	16	71	137	16%	20%	12%	52%	100%
4	30	24	4	75	133	23%	18%	3%	56%	100%
Tl	79	81	35	287	482	16%	17%	7%	60%	100%
Mn	19.8	20.3	8.8	71.8	121	16%	17%	7%	60%	100%

Table 5.18 shows that the relative frequency among the four types of *wh*-questions is quite well defined in the different movies. *What* questions are more frequent than any other type of question, and the least frequent are *who* questions. The relative frequency is summarized in (16).

(16) Order of relative frequency of *wh*-questions in movies

what > how, why > who

The consistency in the relative frequency of *wh*-questions is depicted in Figure 5.24.

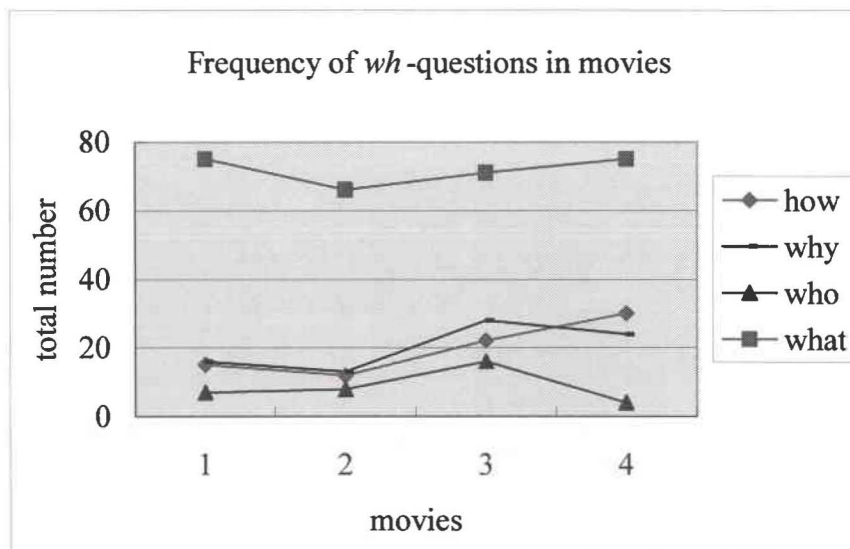


Figure 5.24 Frequency of *wh*-questions in movies

Even though the lengths of the scripts are very different (from 7,432 words to 25,930 words), and even though the contents are also very different from one another, the relative frequency of the four types of *wh*-questions is very similar across the different scripts.

In the third step, each *wh*-question was analyzed and categorized into one of three groups: inverted, uninverted and other questions. The results are presented in Table 5.19.

Table 5.19 *Wh*-questions in movies

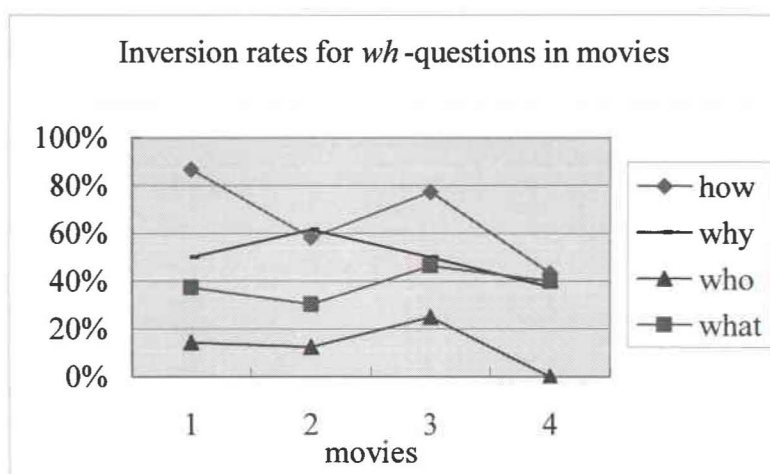
	Total number of <i>Wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Uninverted (<i>Wh</i> + Subj+ Aux)				Other questions			
S	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
1	15	16	7	75	13	8	1	28	2	4	2	13	0	4	4	34
2	12	13	8	66	7	8	1	20	5	3	0	15	0	2	7	31
3	22	28	16	71	17	14	4	33	5	10	1	12	0	4	11	26
4	30	24	4	75	13	9	0	30	13	5	1	14	4	10	3	31
Tl	79	81	35	287	50	39	6	111	25	22	4	54	4	20	25	122
Mn	19.8	20.3	8.8	71.8	13.0	9.6	1.5	28.0	6.3	5.5	1.0	13.5	1.0	5.0	6.3	31.0

As in the previous television scripts, on average, inverted *wh*-questions are used more frequently than uninverted *wh*-questions, that is, indirect *wh*-questions (50 vs. 25 for *how*; 39 vs. 22 for *why*; 6 vs. 4 for *who*; and 111 vs. 54 for *what*). However, as in the other television sitcom scripts, more than half of the *who* questions belong to other questions (25 out of 35) due to the *who* subject questions (18 out of 25).

In order to identify any consistency in inversion rates for *wh*-questions among the different movie scripts, inversion rates for each type of question in each script were calculated. The results are presented in Table 5.20 and Figure 5.25.

Table 5.20 Inversion rates for *wh*-questions in movies

	Total number of <i>Wh</i> -questions				Inverted (<i>Wh</i> + Aux + Subj)				Inversion rate			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>What</i>
S												
1	15	16	7	75	13	8	1	28	87%	50%	14%	37%
2	12	13	8	66	7	8	1	20	58%	62%	13%	30%
3	22	28	16	71	17	14	4	33	77%	50%	25%	46%
4	30	24	4	75	13	9	0	30	43%	38%	0%	40%
Tl	79	81	35	287	50	39	6	111	63%	48%	17%	39%
Mn	19.8	20.3	8.8	71.8	12.5	9.8	1.5	27.8	63%	48%	17%	39%

Figure 5.25 Inversion rates for *wh*-questions in movies

Overall, inversion rates for *how* and *why* questions are higher than those for *who* and *what* questions, which is similar to *Friends* and *Frasier*. However, comparing *what* questions and *who* questions, the inversion rate for *what* questions is higher than that for *who* questions in all scripts, which is different from the two sitcom scripts. Overall, the order of inversion rates in movie scripts seems to be as follows:

(17) Relative order of inversion rate for *wh*-questions in movies

how, why > what > who

Finally, the frequency of inverted *wh*-questions is compared for different *wh*-questions. Inverted *what* questions are more frequent than inverted *who* questions because of the higher frequency and inversion rates of *what* question than those of *who* questions (*what* > *who*). Comparing *how* questions and *who* questions, inverted *how* questions are more frequent than inverted *who* questions because of the higher frequency and inversion rates of *how* questions than those of *who* questions. The frequency of inverted *wh*-questions for each type is shown in Table 5.20, and the rank order of relative frequency of inverted *wh*-question is shown in (18).

(18) Relative order of frequency of inverted *wh*-questions

what > how > why > who

The order in (18) is depicted in the following figure.

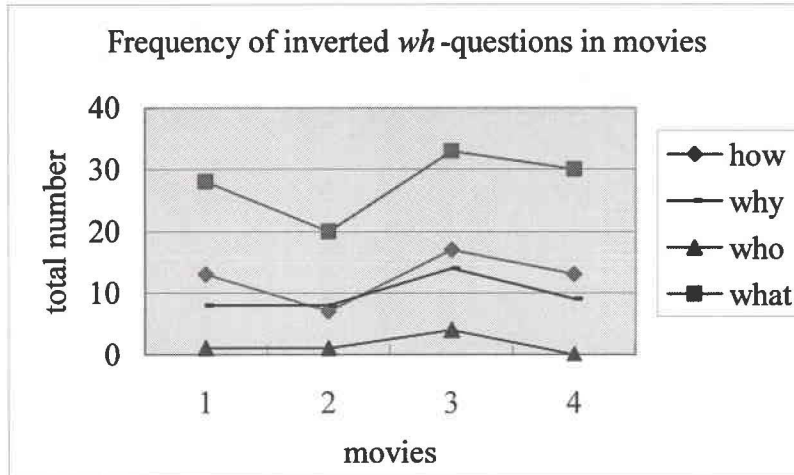


Figure 5.26 Frequency of inverted *wh*-questions in movies

In sum, there seems to be a consistency among different movie scripts in terms of (i) the relative frequency of *wh*-words, (ii) relative frequency of *wh*-questions, (iii) relative inversion rates for *wh*-questions, and (iv) relative frequency of inverted *wh*-questions.

Figure 5.27 shows the relative frequency of *wh*-words across different scripts.

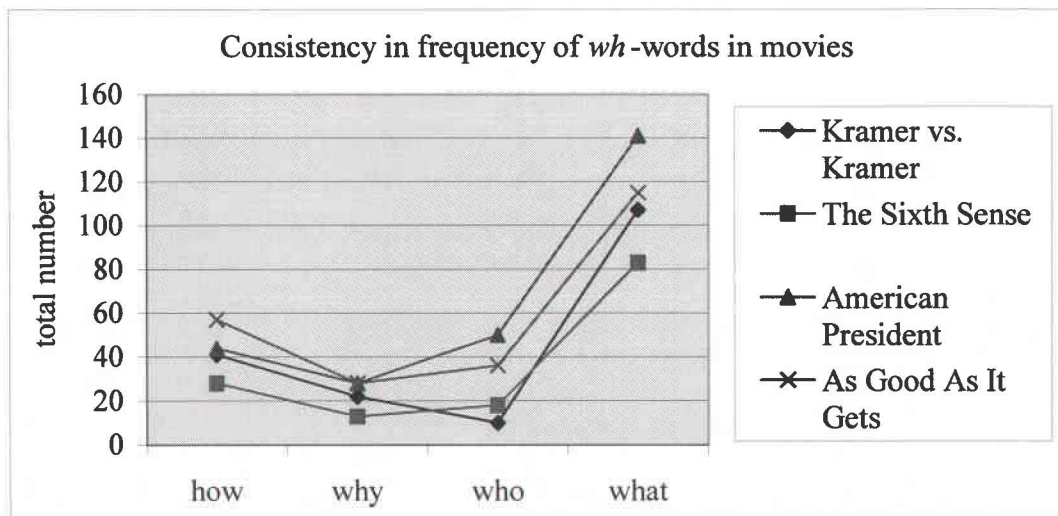


Figure 5.27 Consistency in relative frequency of *wh*-words in movie scripts

The shape of curves is very similar. The similar patterns indicate that the scripts are very similar to each other in terms of the relative frequency of *wh*-words (i.e. *what* > *how* > *who*, *why*).

Next, consider the following figure concerning the frequency of *wh*-questions.

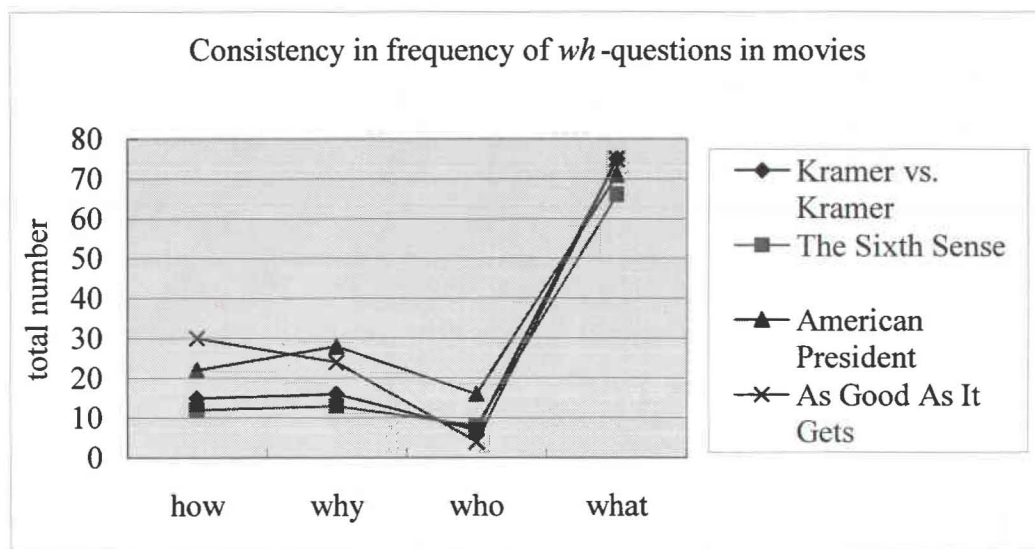


Figure 5.28 Consistency in frequency of *wh*-questions in movies

Again, the shapes of the curves are very similar to each other. The similar patterns indicate that the scripts are very similar to each other in terms of the relative frequency of *wh*-questions (i.e., *what* > *how*, *why* > *who*).

Finally, consider the following figure, which shows the relative frequency of inverted *wh*-questions.

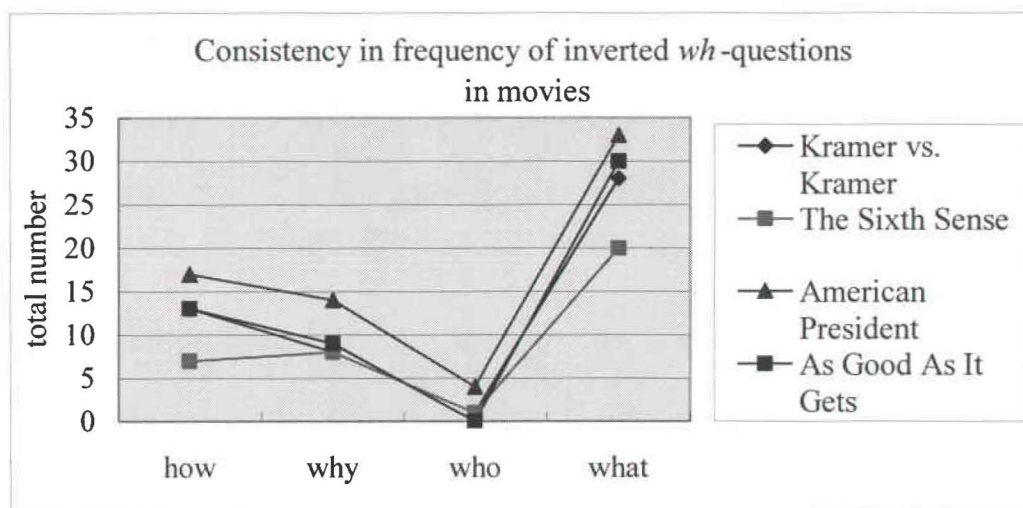


Figure 5.29 Consistency in frequency of inverted *wh*-questions in movie scripts

Figure 5.29 also shows a similarity in the shape of curves among different scripts. This indicates that the scripts are very similar to each other in terms of the order of the relative frequency of inverted *wh*-questions (i.e., *what* > *how*, *why* > *who*).

Finally, it is worth noting that the shape of curves is very similar to the ones found in the analysis of two sets of television scripts (*Friends* and *Frasier*) in the previous section.

5.1.3 Discussion

The similarity among different corpora indicates that L2 learners are exposed to very similar type of input involving subj-aux inversion in *wh*-questions even though they use different materials to learn English (i.e., textbooks, television sitcoms and movies, which are one of the main sources from which learners get English input). The scripts for serial television sitcoms and movies reflect everyday conversations in terms of

the use of *wh*-questions. Therefore, the results drawn from the materials analyzed in this study are presumably not far from what the learners get from their exposure to English input in everyday life either in America or in Korea. This is also supported by our findings from the textbook corpus. It was found that the results from textbooks are very similar to those from movies and TV scripts. Generally speaking, it seems plausible to propose that the learners are exposed to very similar types of input in the aspect of inversion in *wh*-questions, regardless of the learning materials.

Before looking at the frequency data of our input study, it might be thought that each learner gets a different type of input. However, considering the consistency found in this study, the learners' input seems to be similar in terms of the relative frequency of *wh*-words and *wh*-questions, subj-aux inversion rates in *wh*-questions, and the relative frequency of inverted *wh*-questions. If this is right, a frequency-based theory would presumably predict that learners will show a parallel developmental order when they learn *wh*-questions, especially with respect to the acquisition of auxiliary inversion in *wh*-questions.

Let's consider this possibility more closely. The results of our input study of four types of corpora are summarized in the following table:

Table 5.21 Summary of L2 input study: corpus analysis

Corpus	Relative frequency of <i>wh</i> -words	Relative frequency of <i>wh</i> -questions
Textbooks	<i>what</i> > <i>how</i> > <i>who</i> , <i>why</i>	<i>what</i> > <i>how</i> > <i>who</i> <i>what</i> > <i>why</i>
Friends	<i>what</i> > <i>how</i> , <i>who</i> <i>how</i> > <i>why</i>	<i>what</i> > <i>how</i> , <i>why</i> > <i>who</i>
Frasier	<i>what</i> > <i>how</i> > <i>who</i> > <i>why</i>	<i>what</i> > <i>how</i> > <i>who</i> <i>what</i> > <i>why</i>
Movies	<i>What</i> > <i>how</i> > <i>who</i> , <i>why</i>	<i>what</i> > <i>how</i> , <i>why</i> > <i>who</i>
Corpus	Inversion rates for <i>wh</i> -questions	Relative frequency of Inverted <i>wh</i> -questions
Textbooks	<i>what</i> > <i>who</i> <i>how</i> > <i>who</i>	<i>what</i> > <i>how</i> > (<i>why</i>) > <i>who</i>
Friends	<i>why</i> , <i>how</i> > <i>what</i> , <i>who</i>	<i>what</i> > <i>how</i> > <i>who</i> <i>what</i> > <i>why</i>
Frasier	<i>why</i> , <i>how</i> > <i>what</i> , <i>who</i>	<i>what</i> > <i>how</i> , <i>why</i> > <i>who</i>
Movies	<i>why</i> , <i>how</i> > <i>what</i> > <i>who</i>	<i>what</i> > <i>how</i> > <i>why</i> > <i>who</i>

More importantly, we find the same order of frequency of inverted *wh*-questions, namely:

(19) Relative order of frequency of inverted *wh*-questions

what > *how* > *who*

Because of the similar rank order of relative frequency of *wh*-questions and inversion rates, it is possible that the same order in (19) is found among different scripts.

This rank order allows us to propose an expected order of acquisition from the lexical-based input frequency account as follows:

(20) Expected order of acquisition of subj-aux inversion with *wh*-questions

- i) what > how > who
- ii) No argument/adjunct asymmetry

In other words, we may make the following predictions.

- i) Korean learners of English will do best at inversion in *what* questions, next best in *how* questions, and least well in *who* questions.
- ii) Korean learners of English will not show any argument/adjunct asymmetry in inversion because input frequency does not show any asymmetrical order between argument questions and adjunct questions.

5.2 L2 Acquisition study: Grammatical judgment task (Experiment 4)

The results of L2 input study enables us to test the two theories: the lexical-based input frequency theory and the structure-based generative theory because the two theories make different predictions on the acquisition order of subj-aux inversion in different *wh*-questions as follows:

(21) Lexical-based input frequency theory

- i) *what* > *how* > *who*
- ii) No argument adjunct asymmetry

(22) Structure-based generative theory

- i) *who* > *how*, *why*
- ii) *what* > *how*, *why*
- iii) Argument adjunct asymmetry

The input theory predicts that learners will do better at subj-aux inversion in the order of *what* > *how* > *who* so that there will be no argument/adjunct asymmetry. It does not make any prediction on the acquisition order for *why* questions because the frequency of *why* varies depending on the text.

On the other hand, the structure-based generative theory predicts a different order of acquisition as shown in (22) (i.e., *what* > *how*, *why*, and *who* > *how*, *why*), so that the learners will show argument/adjunct asymmetry in acquisition of *wh*-questions in terms of subj-aux inversion. However, we do not know yet if there will be any difference between *what* and *who*. Even though they are both argument questions, the linguistic properties of the two items are different. For example, *what* usually refers to an inanimate entity whereas *who* refers to an animate entity. The generative theory does not provide any information about the influence of animacy on the acquisition of subj-aux inversion (e.g., *Who did you see?* vs. *What did you see?*). The structural theory

does not make any prediction between *how* and *why*, either. Even though both are adjunct questions, the linguistic properties of the two items are very different. For example, *how* refers to the manner of action (e.g., *how did he fix it?*) and *why* refers to the reason (e.g., *why is he laughing?*).

A grammaticality judgment task is one of the well-known methods in SLA research to test L2 learners' knowledge of certain grammatical phenomena. In this study, a grammaticality judgement task was used as a method of measuring L2 learners' acquisition of inversion in *wh*-questions. The results will allow us to see any difference in L2 learners' judgment of inversion in different types of *wh*-questions.

If the lexical-based input frequency approach is right, the learners' judgments will have the accuracy order, *what* > *how* > *who*. If the structure-based generative approach is right, the learners' judgement will be better on *what* and *who* questions than on *why* and *how* questions. Since the learners might use two possible structures for adjunct *wh*-questions (i.e., IP adjunction for uninverted ones and CP for inverted ones) and only one structure for argument *wh*-questions, it is more likely that the learners will have difficulty with the adjunct *wh*-questions. Therefore, the learners' judgment on uninverted questions will be more accurate in argument *wh*-questions than in adjunct *wh*-questions.

However, there is an overlap among the predictions made by the two theories. Both theories predict that learners' judgment will be more accurate at *what* questions than at *how* questions regarding subj-aux inversion as shown in (21) and (22). As a result,

the only way to contrast the two theories is to compare *how* questions and *who* questions because the two theories predict an opposite acquisition order for these two. This brings us to the following research questions:

(23) Research questions

1. If the lexical-based input frequency theory is right, grammaticality judgment on inversion by Korean learners of English will be better on *how* questions than on *who* questions (*how* > *who*). If the structure-based generative theory is right, the opposite order (*who* > *how*) will be found.
2. If the structure-based generative theory is right, the following order will also be found: (i) *who* > *how*, *why* (ii) *what* > *how*, *why*.

5.2.1 Materials

In this study, a listening version of a grammaticality judgment task was prepared to test L2 learners' knowledge of inversion in *wh*-questions (it was thought that the target items might be too easy to notice if the task were done as a paper-based test). The same types of *wh*-questions investigated in our input study were chosen for this study for direct comparison (i.e., *what* and *who* for argument questions, *how* and *why* for adjunct questions). Four pairs of inverted and uninverted questions were prepared for each type of *wh*-question. Half of the pairs had a pronoun subject, *you* and *he*; the other two pairs

had noun phrase subjects, *the children* and *the old man*. The same kinds of subjects (*you, he, the old man* and *the children*) were used in all four types of questions to control variability. The same kinds of verbs and adverbs were used both with pronoun subject and with noun subjects. The target sentences are shown in (24).

(24) Target sentences

a. Pronouns subj. Be-aux.

1. Who (are you/*you are) meeting in the cafeteria?
2. What (are you /*you are)reading in the library?
3. Why (are you/*you are) jumping on the bed?
4. How (are you/*you are) going to the campground?

b. Pronoun subj. Do-aux

5. Who (did he visit/*he visited) last night?
6. What (did he find/*he found) last night?
7. Why (did he cry/*he cried) last night?
8. How (did he sleep/*he slept) last night?

c. Noun subj. Be-aux.

9. Who (are the children/*the children are) meeting in the cafeteria?
10. What (are the children/*the children are) reading in the library?
11. Why (are the children/*the children are) jumping on the bed?
12. How (are the children/*the children are) going to the campground?

d. Noun subj. Do-aux

13. Who (did the old man vist/*the old man visited) last night?
14. What (did the old man find/ *the old man found) last night?
15. Why (did the old man cry/*the old man cried) last night?
16. How (did the old man sleep/*the old man slept) last night?

A total of thirty-two test items (four pairs of inverted and uninverted items for each *wh*-word) and fifty-seven distractors were prepared for the test. The test items and distractors (a total of eighty-nine sentences) were randomized and recorded on a cassette tape by a male native speaker of English. There was an eight-second pause between each sentence to give the students time to think and check the answer. The subjects were asked to indicate whether each sentence was possible in English on a Likert scale, the range of which were from -2 (not possible), to +2 (perfectly O.K.) (See Appendix I).

5.2.2 Subjects

A total of 41 Korean learners of English and 9 native speakers of English participated in this study. The learners were students at Hanguk University of Foreign Languages in Seoul, Korea. Two classes of students majoring in English were asked to volunteer for the test. The native speakers of English were graduate students from the Linguistics and Second Language Studies departments at the University of Hawai'i at Manoa.

5.2.3 Methods

The task was performed in the classroom during class time. The students sat in their seats and a tape recorder was placed on the teacher's desk at the front. The subjects were told that there were no vocabulary errors in the sentences and also that *whom* could be replaced by *who* in the object questions. In addition, they were asked not to go back to the previous sentences to correct them. Before the test, the subjects were requested to fill out a background information form for further analysis. It took about 20 minutes for the listening test session and about 40 minutes for the entire session. The native speakers took the same test individually in a classroom in the presence of the researcher.

5.2.4 Results

The results are presented in the following table and figures.

Table 5.22 Mean of grammaticality judgment task

Type	Subject	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	Mean
Inverted (+)	Learner (N=41)	0.74	1.27	0.81	1.20	1.01
	N.speakers (N=9)	1.97	2.00	2.00	2.00	1.99
Uninverted (-)	Learner (N=41)	-0.09	0.00	-0.48	-0.54	-0.28
	N.speakers (N=9)	-1.44	-1.42	-1.56	-1.61	-1.51

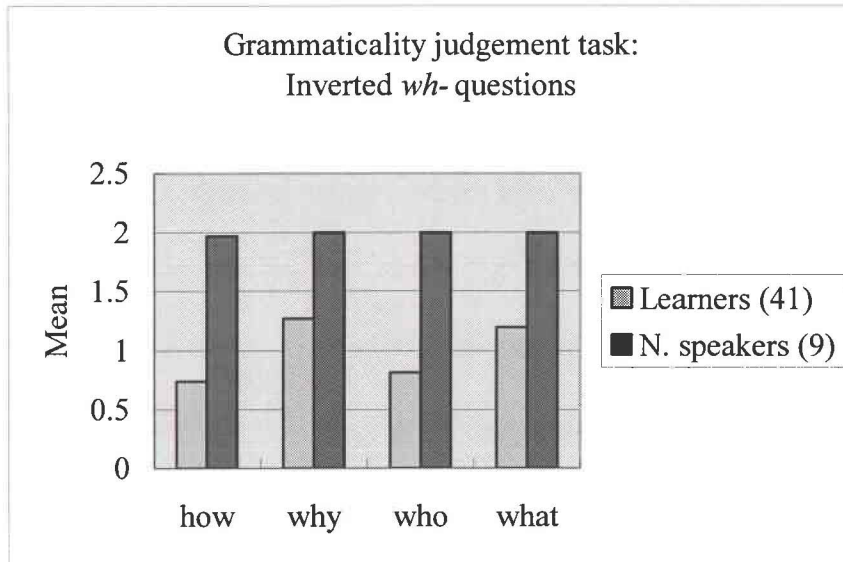


Figure 5.30 Grammaticality judgement task: Inverted *wh*-questions

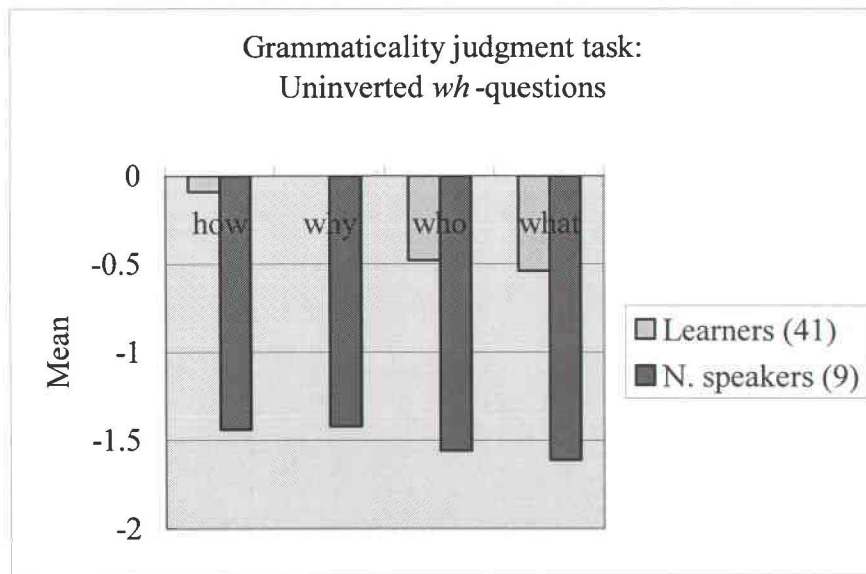


Figure 5.31 Grammaticality judgement task: Uninverted *wh*-questions

Overall, both native speakers and L2 learners give the right judgment (+ for inverted and – for uninverted). However, learners' judgments are quite weak compared to native

speakers' (1.01 vs. 1.99 for inverted ones and -0.28 vs. -1.51 for uninverted ones).

Native speakers' judgment on inverted ones is very consistent through the four types of *wh*-word. In contrast, learners' judgments look different depending on the type (0.74 for *how*, 1.27 for *why*, 0.81 for *who* and 1.2 for *what*). A one-way repeated measures ANOVA was conducted and it was found that the effect of *wh*-word (inverted) was significant ($F(3, 120) = 7.68^*, p < 0.0001$). Subsequent Tukey's tests were conducted. The results are shown in the following table:

Table 5.23 Mean differences of grammaticality judgment task (inverted *wh*-questions)

	what (1.20)	who (0.81)	how (0.74)
why (1.27)	0.07	0.46*	0.53*
what (1.20)		0.39*	0.67*
who (0.81)			0.07

* Minimum significant difference = 0.3556, 2-tailed, $\alpha = 0.05$

Table 5.23 shows that the mean difference between *what* and *why* and that between *how* and *who* is not significant, but that the mean differences between (i) *who* and *why*, (ii) *how* and *why*, (iii) *what* and *who*, and (iv) *what* and *how* are significant.

The results seem to indicate that the distinction between argument and adjunct is not shown in L2 learners' judgment on inverted *wh*-questions. However, the results does not seem to reveal the input frequency effect, either, because our prediction based on input study suggested the order, *how* > *who*, as stated in Research Question 1 in (23).

Native speakers' judgments on uninverted questions are quite consistent throughout the four types of *wh*-questions. However, learners' judgments on ungrammatical ones are quite different depending on the type of *wh*-question. A one-way repeated measures ANOVA was conducted and it was found that the effect of *wh*-word (uninverted) was significant ($F(3, 120) = 6.24^*$, $p = 0.0006$). In order to test our Research Questions 1 and 2, subsequent Tukey's tests were conducted. The results are shown in the following table.

Table 5.24 Mean differences of grammaticality judgment task (uninverted *wh*-questions)

	how (-0.09)	who (-0.48)	what (-0.54)
why (0.00)	0.09	0.48*	0.54*
how (-0.09)		0.39	0.45*
who (-0.48)			0.06

* Minimum significant difference = 0.4025, 2-tailed, $\alpha = 0.05$

Table 5.24 shows that the mean difference between *who* and *what* (argument) questions is not significant, and that the mean difference between *why* and *how* (adjunct) questions is not significant, either. By contrast, the mean difference between *who* and *how* questions is significant, and the mean difference between *who* and *why* questions is also significant. In addition, the mean differences between *what* and *why* questions and between *what* and *how* are also significant.

The results do not conform to the prediction of the lexical-based input frequency theory stated in our research question 1 in (23). Instead, they conforms to the prediction

of the structure-based generative theory. That is, the learners' judgment on ungrammaticality of uninverted *wh*-questions is better for *who* questions than for *how* questions. In addition, the results also conforms to the prediction of structure-based generative theory stated in our research question 2 in (23). That is, there is an argument/adjunct distinction in L2 learners' judgments on uninverted *wh*-questions. The learners' judgment on the ungrammaticality of uninverted *wh*-questions is better in argument *wh*-questions than in adjunct *wh*-questions.

In sum, the results from uninverted *wh*-questions show that the learners' judgements on uninverted *wh*-questions are better in argument *wh*-questions (*what* and *who*) than in adjunct *wh*-questions (*how* and *why*). On the other hand, the learners' judgments on inverted *wh*-questions did not show any argument/adjunct distinction. However, the judgment scores for inverted *wh*-questions were above 0.7 throughout the four types, which means that the learners' overall judgment on grammaticality of inverted *wh*-questions is correct in general. In contrast, the learners' judgment on uninverted adjunct *wh*-questions (i.e., *how* and *why*) was not correct because the mean is about 0.00 (i.e., -0.09 for *how* and 0.00 for *why*), but the learners' judgment on uninverted argument *wh*-questions were much better (i.e., -0.48 for *who* and -0.54 for *what*). This means that the learners have problems judging uninverted adjunct *wh*-questions.

5.2.5 Discussion

The results of the grammaticality judgment task show us that the learners

judgments on uninverted *wh*-questions are better on *what* and *who* questions than on *how* and *why* questions. In other words, it is more likely that the learners judge uninverted questions as being correct in *why* and *how* questions than in *what* and *who* questions. That is, the learners' performance is better in *what* and *who* questions than in *why* and *how* questions. The similarities between *what* and *who* questions and between *why* and *how* questions in the learners' judgment on the uninverted *wh*-questions clearly show the distinction between argument and adjunct *wh*-questions.

On the other hand, learners' judgment on inverted *wh*-questions are relatively correct for all types of *wh*-questions (i.e., higher than 0.74) even though the scores vary significantly depending on the type of *wh*-word. This means that learners have a problem catching the ungrammaticality of the uninverted *wh*-questions. The difficulty lies in adjunct questions more than in argument questions as shown in Table 5.22.

5.3 Comparison of input and acquisition

From input frequency point of view, our input study suggested the following order of acquisition of inversion with *wh*-questions:

- i) what > how > who
- ii) No argument/adjunct asymmetry

However, our acquisition data from a grammaticality judgement task found the following

order:

- i) what, who > how, why
- ii) Argument/adjunct asymmetry.

Consequently, the acquisition data do not seem to be accounted for by the input frequency account.

However, there is another aspect to account for the phenomena from the input frequency approach. One may suggest that we need to consider the possibility of item-by-item learning, strictly following Rowland and Pine (2000). For example, the target structure of our study involves a specific combination of [*Wh*+ Aux + Subj], that is, *wh* + *are you/NP V-ing?* and *what did he/NP V?*. So the results of the study might be influenced by the frequency of these specific combinations of lexical items in the learners' input. This needs to be explored. In order to investigate this possibility, inverted and uninverted questions involving specific combinations (i.e., *wh* + *are you/NP V-ing?*) were sought in the same corpora.

In this search, eighteen scripts from *Friends* and one script from *Frasier* were added to the original corpus in order to increase the frequency of the search items. However, the scripts are not separated in the analysis, which is different from our input study in the previous section, because each script is too short to search for very specific structures.

First, inverted questions were searched by using the command, *wh+ are* (i.e., *what are, who are, why are* and *how are*) and categorized by hand into [*wh + are you*] and [*wh + are NP*] types (As most questions had *you* subjects, as shown in Table 5.25, it was quite easy to go over all the items.) As for the *how* questions, only adverbial *how* was considered (not *how + adjectives* such as *how tall are you*).

The results of the search for inverted questions are as follows:

Table 5.25 Frequency of [*wh+ are you/NP---?*] (Inversion) in L2 corpus

Corpus	Total number of script	Word count	<i>Wh + are you—?</i>				<i>Wh + are NP—?</i>			
			<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
Friends	27	74,202	10	4	3	44	2	1	3	6
Fraiser	10	42,767	2	2	1	11	0	0	0	2
Movies	5	80,107	5	7	0	18	1	0	1	3
Textbooks M1	7	19,438	12	0	0	10	0	1	0	3
Textbooks M2	8	42,788	5	13	0	17	2	0	1	9
Textbooks M3	8	50,899	3	3	0	12	0	0	0	7
Total	65	310,201	37	29	4	112	5	2	5	30
Mean		51,700	6.2	4.8	0.7	18.7	0.8	0.3	0.8	5.0

- M1 means middle school 1st grade.

- NP means “not *you*” so it includes noun phrase subjects and other pronoun subjects.

- [*Wh + are NP—?*] does not include subject questions or relative clauses.

In the [*wh+ are you—?*] pattern, *what are you—?* is the most frequent of all the *wh*-questions (112). The next most frequent is *how are you—?* (37) and *why are you—?* (29). *Who are you —?* is very rare (4). In the [*wh+ are NP—?*] pattern, *what are NP—?* is more frequent than any other question, which is similar to [*wh+ are you —?*] pattern. On the other hand, comparing *who* and *how*, *who are NP—?* is as frequent as *how are NP—?*.

It seems possible that the large number of *how are you—?* questions is due to the greeting *how are you?* or *how are you doing?* and that the large number of *what are you—?* questions is due to the *what are you doing?* pattern. If the learners take this type of input as a formulaic expression without analyzing it as [*wh+ are+ you —?*], the frequency of these items might not be relevant to our discussion. Therefore, another frequency search was done excluding the formulaic expressions, *how are you?*, *how are you doing?* and *what are you doing?*. The results are presented in Table 5.26.

Table 5.26 Frequency of [*wh+ are you/NP—?*] (Inversion) excluding formulaic expressions in L2 corpus

Corpus	Total number of script	Word count	<i>Wh + are you—?</i>			
			<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
Friends	27	74,202	5	4	3	30
Fraiser	10	42,767	2	2	1	6
Movies	5	80,107	3	7	0	14
Textbooks M1	7	19,438	1	0	0	5
Textbooks M2	8	42,788	1	13	0	11
Textbooks M3	8	50,899	0	3	0	7
Total	65	310,201	12	29	4	73
Mean		51,700	2.0	4.8	0.7	12.2

Even after excluding the formulaic expressions, the frequency order is the same among *what*, *how* and *who* questions (i.e., *what* > *how* > *who*). And the *why are you—?* pattern is more frequent than *how are you—?* and *who are you —?* question patterns. In addition, the rank order of relative frequency among these three types is consistent throughout the different scripts. Consequently, we can draw a conclusion about inverted [*wh* + *are you/NP —?*] question as follows:

(25) Frequency order of [*wh*+ *are you/NP—?*]

what > how, why > who

In the next step, uninverted questions were searched for the same structure. The results are presented in the following table:

Table 5.27 Frequency of [*wh*+ *you/NP are—?*] (Noninversion) in L2 corpus

Corpus	Total number of script	Word count	<i>wh</i> + <i>you are</i>				<i>wh</i> + <i>NP are</i>			
			<i>How</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
Friends	27	74,202	0	1	1	0	0	1	1	3
Frasier	10	42,767	0	0	0	0	0	0	0	0
Movies	5	80,107	1	1	0	0	0	0	1	1
Textbooks M1	7	19,438	0	0	0	0	1	0	0	0
Textbooks M2	8	42,788	0	0	0	0	0	3	0	1
Textbooks M3	8	50,899	0	1	0	1	0	0	0	4
Total	65	310,201	1	3	1	1	1	4	2	9
Mean		51,700	0.2	0.5	0.2	0.2	0.2	0.7	0.3	1.5

The numbers in this search are so low that it seems impossible to make any comparison. Evidently, indirect questions are rarely used in conversation and textbooks. Therefore, uninverted questions involving this specific combination are presumably not very influential for learners' acquisition of subj-aux inversion in *wh*-questions.

In sum, the frequency of the particular patterns we have been examining does not conform to the results from the grammaticality judgment task (acquisition order), so that the possibility of item-by-item learning cannot explain the argument and adjunct *wh*-questions found in the grammaticality judgment task.

5.4 Conclusion

L2 learners input and acquisition of *wh*-questions were investigated in this chapter. The input study based on a corpus analysis of various input sources for Korean learners of English found the rank order of relative frequency of inverted *wh*-questions, *what* > *how* > *who*. However, the acquisition study based on a grammaticality judgment task with Korean learners of English found argument/adjunct asymmetry; that is, learners' judgment was better at subj-aux inversion in argument *wh*-questions than in adjunct *wh*-questions. Therefore, the structure-based generative theory, not the lexical-based input frequency theory, is supported to explain L2 learners' acquisition of inversion in *wh*-questions. Even though frequency has often been discussed as one of the main issues in second language acquisition (Gass and Mackey 2002, Bley-Vroman

2002, Larsen-Freeman 2002), it does not seem to affect acquisition of inversion in *wh*-questions.

CHAPTER 6

COMPARISON BETWEEN L1 AND L2

In this chapter, our input and acquisition studies on L1 and L2 will be compared in terms of their differences and similarities. Based on the findings of these studies, I will discuss the following two considerations, in particular: (i) similarity of input in L1 and L2 (ii) similarity of acquisition in L1 and L2.

6.1. Comparison between L1 and L2 input

In our input studies on L1 and L2, we found very similar results between children's input and L2 learners' input. First, the relative frequency of *wh*-words is very similar among *what*, *how* and *who* (i.e., *what* > *how* > *who*).

(1) Order of relative frequency of *wh*-words

L1: what > why > how > who

L2: what > how > who, why

Remember that the L1 input study is based on caregivers' speech to their children while the L2 input study is based on the adult-to-adult speech (TV sitcom scripts and movie scripts) and textbooks. The similarity found in (1) means that caregivers' speech to the children is not different from adult-to-adult speech in terms of the relative frequency of

wh-words. In general, children and second language learners hear the three *wh*-words in the same relative order of frequency (i.e., *what* > *how* > *who*).

Second, the relative frequency of *wh*-questions is also similar among *what*, *why* and *who* (i.e., *what* > *why* > *who*) questions.

(2) Order of relative frequency of *wh*-questions

L1: *what* > *why* > *how*, *who*

L2: *what* > *how* (*why*) > *who*

A similarity is found among *what*, *why* and *who* questions (i.e., *what* > *why* > *who*). In other words, children and second language learners hear *what* questions more frequently than any other type of *wh*-question. And they hear *why* questions more frequently than *who* questions.

Third, compare the relative inversion rates for *wh*-questions (i.e., *how*, *why* > *who*, and *how* > *what*).

(3) Order of inversion rates for *wh*-questions

L1: *how* > *what* > *why* > *who*

L2: *why*, *how* > *what*, *who*

There is also a similarity between L1 and L2 among *how*, *why* and *who* questions and among *how* and *what* questions. In both L1 and L2, inversion rate is higher for *how* and *why* questions than for *who* questions; inversion rate is higher for *how* questions than for *what* questions.

Finally, compare the relative frequency of inverted *wh*-questions (i.e., *what* > *how*, *why* > *who*).

(4) Order of relative frequency of inverted *wh*-questions.

L1: *what* > *why* > *how* > *who*

L2: *what* > *why*, *how* > *who*

The orders in (4) show a similarity between L1 and L2 among *what*, *how*, *why* and *who* questions (i.e., *what* > *how*, *why* > *who*). Both L1 and L2 learners receive input for inverted *wh*-questions in the frequency order, *what* > *how*, *why* > *who*.

The results of the L1 and L2 input studies seem to be very plausible conceptually. Comparing *what* and *who* questions, a greater proportion of *what* questions are object questions and a greater proportion of *who* questions are subject questions (which does not involve subj-aux inversion) in our L1 and L2 input studies. The difference between subject questions and object questions seems to relate to the animacy of *what* and *who* as briefly mentioned in Chapter 2. *What* involves an inanimate entity whereas *who* involves an animate entity. As the subject of a sentence is usually an agent in English, it

is likely that *who* is used for subject questions, and *what* for object questions. This also seems relevant to the rank order between *how* and *who*. In direct questions, *how* must involve subj-aux inversion whereas *who* does not (i.e., in a subject question). The only way *how* can appear without inversion is in the cases of indirect questions (e.g., *I do not know how you did it*) and exclamatives (e.g., *how beautiful you are!*). However, indirect questions and exclamatives are much less frequent than direct questions as we found in our input studies. Since *who* questions are likely to be used as subject questions, and *how* questions must involve subj-aux inversion, it is likely that children and L2 learners hear inverted *how* questions more than inverted *who* questions.

In sum, with regard to subj-aux inversion in *wh*-questions, very similar results were found between children's input and L2 learners' input. It is likely that children and L2 learners receive very similar input, that is (i.e., *what* > *how* > *who*), in terms of the frequency of inverted *wh*-questions, which is the most relevant input for subj-aux inversion in *wh*-questions.

6.2 Comparison between L1 and L2 acquisition

In our acquisition studies on L1 and L2, very similar results were found in children's production and L2 learners' grammaticality judgments. In the L1 production study, it was found that the children produced *who* questions more correctly than *how* and *why* questions in terms of subj-aux inversion. In the L2 grammaticality judgement task, the L2 learners judged *what* and *who* questions more correctly than *how* and *why*

questions. The L2 learners also differentiated argument questions from adjunct questions.

Even though we cannot compare the results from different types of experiments (elicited production vs. grammaticality judgement task), the implications of our findings can nonetheless be identified. First, both L1 learners and L2 learners seem to undergo a very similar developmental process: acquisition of subj-aux inversion in argument *wh*-questions before adjunct *wh*-questions. The similarity of language development has been noticed in many aspects of syntax in the L1 and L2 acquisition literature (e.g., relative clauses, interrogatives, negations etc.). If the similarity between our L1 and L2 studies (i.e., argument and adjunct distinction) is not caused by chance, it seems that we may add another substage to the developmental stages of English interrogatives: inversion with argument *wh*-question > inversion with adjunct *wh*-question. This additional developmental stage raises a question about what kind of learning mechanism operates in the acquisition of subj-aux inversion in *wh*-questions: what causes the learners to master (or do better) on subj-aux inversion earlier for argument *wh*-questions than for adjunct *wh*-questions?

6.3 Conclusion

We found similarities in L1 and L2 learners' input and their acquisition of subj-aux inversion in *wh*-questions. Even though our input study showed that both L1 and L2 learners' input for subj-aux inversion reflects the order *what* > *how* > *who* (with no

argument/adjunct distinction), our acquisition study showed that both L1 and L2 learners did better at inversion in argument *wh*-questions than in adjunct *wh*-questions. The results of the L2 input and acquisition studies support the structure-based generative approach to the acquisition of *wh*-questions.

CHAPTER 7

CONCLUSION

This dissertation has focused on the argument/adjunct asymmetry in the acquisition of subj-aux inversion in *wh*-questions by English-speaking children and ESL learners. The lexical-based input frequency approach and the structure-based generative approach have been tested by comparing L1 and L2 learners' input and their acquisition of inversion. This chapter will provide a brief summary of these studies, and discuss the implications of this research for the frequency theory. Other alternative accounts for the results of our studies will also be discussed from the perspective of input frequency. It will be argued that these possibilities are not very convincing, which leads us to conclude that the input frequency theory cannot explain the argument/adjunct asymmetry in the acquisition of *wh*-questions. Finally, a possible implicational relationship involving this asymmetry is suggested as an area for further research.

7.1 Summary

Our L1 input study found that caregivers' input is similar across different caregivers for *what*, *who*, *why* and *how* questions in terms of (i) the relative frequency of *wh*-words, (ii) the relative frequency of *wh*-questions, (iii) inversion rates for *wh*-questions, and (iv) the frequency of inverted *wh*-questions. It was found that children

receive input for inverted *wh*-questions in the relative order of frequency, *what* > *why* > *how* > *who*. In contrast, our L1 production study found that children did better on inversion in *who* questions than in *how* and *why* questions (i.e., *who* > *why*, *how*).

Our L2 input study found that L2 learners' input is similar across different input sources (i.e., textbooks, television situation comedy scripts, and movie scripts) in terms of (i) the relative frequency of *wh*-words, (ii) the relative frequency of *wh*-questions, (iii) inversion rates for *wh*-questions and (iv) the frequency of inverted *wh*-questions. It was found that L2 learners receive input for inverted *wh*-questions in the relative order of frequency, *what* > *why*, *how* > *who*.

In contrast, our L2 acquisition study found that L2 learners judged the ungrammaticality of uninverted *wh*-question better in *what* and *who* questions than in *why* and *how* questions. Moreover, L2 learners' judgment of uninverted questions was very similar for *what* and *who* questions on the one hand, and for *why* and *how* questions on the other. Meanwhile, L2 learners' judgment of inverted questions was better for *what* and *why* than for *who* and *how*.

This asymmetry did not reflect either the frequency order (i.e., *what* > *why*, *how* > *who*) or the argument/adjunct distinction (i.e., *what*, *who* > *why*, *how*). However, the judgement scores for inverted *wh*-questions were above 0.7 throughout the four types, whereas the judgement scores for uninverted *wh*-questions were about 0.00 for adjunct questions and -0.5 for argument questions. This means that L2 learners have difficulty judging the ungrammaticality of uninverted *wh*-questions, especially uninverted adjunct

wh-questions, compared to the grammaticality of inverted *wh*-questions.

Comparison of input data and acquisition data from our L1 and L2 studies showed that the frequency order of inverted *wh*-questions was not reflected in the acquisition data. Both L1 and L2 learners showed an argument/adjunct asymmetry in their use of inversion in *wh*-questions which was different from the frequency order, resulting in support for the structure-based generative approach.

7.2 Implications for input frequency theory

If the findings of our input and acquisition studies are correct, what are the implications of these findings for the input frequency theory?

7.2.1 Lexical-based frequency

We found that the lexical-based input frequency theory cannot explain the asymmetrical development of *wh*-questions. In other word, the input frequency theory whose counting is based on the frequency of individual lexical items cannot predict the acquisition order of *wh*-questions. Remember the design of our input study. We counted the combination, [*wh*-word + Aux + Subj] for each type of *wh*-questions to get the frequency of different types of inverted *wh*-questions. In other words, the relative frequency order of our units of counting is [*what* + Aux + Subj] > [*why* + Aux + Subj] > [*how* + Aux + Subj] > [*who* + Aux + Subj] in L1 input, for example. And we found this order was not manifested in the acquisition order.

Then, strictly following Rowland and Pine's study, we considered another possibility, which was counting a particular *wh*-word + Aux + Subj. The relative frequency order was [*why* + *are* + *you* —?] > [*how* + *are* + *you* —?] > [*who* + *are* + *you* —?]. We also found that this order was not reflected in the acquisition order, either. It was therefore concluded that this type of lexical-based input frequency theory cannot explain the asymmetrical development of inversion in *wh*-questions.

7.2.2 Frequency of structures

Another possibility within frequency theories can be raised—should frequency of structures, instead of the frequency of individual lexical items, be what is calculated? A possible unit of counting in this approach can be “argument vs. adjunct”. That is to say, we can compare the frequency of the combinations, [*Wh*-word (Argument) + Aux + Subj] vs. [*Wh*-word (Adjunct) + Aux + Subj]. In this case, the frequency order of our input study parallels our acquisition order. Mainly because of the enormous number of inverted *what* questions in the input, the frequency of argument *wh*-questions (i.e., *what* and *who*) surpasses that of adjunct *wh*-questions (*why* and *how*). This seems to suggest that an input frequency theory whose unit of counting is structures (argument vs. adjunct), not individual lexical items, can explain the acquisition order.

However, the important question is whether the acquisition order is determined by frequency of structures or by some inherent property of structures. To answer this question, I'd like to approach this problem from the learners' point of view. To make

the input frequency theory (i.e., frequency of structures) plausible, we have to assume that the learners intake *who* questions and *what* questions in the same manner in terms of inversion, but differently from adjunct *wh*-questions such as *why* and *how* questions. That is to say, we have to assume that the learners are able to separate the relevant linguistic input (i.e., *wh*-elements → argument *wh*-elements, adjunct *wh*-elements), and don't just treat all *wh*-elements alike. Without presupposing this ability, an input frequency theory based on the frequency of structures cannot claim that the earlier acquisition of inversion in argument *wh*-questions is due to the frequency of argument *wh*-questions.

Another way of proposing the effect of structural input frequency on learners' acquisition of inversion relates to the direction of generalization of the inversion rule. From the frequency point of view, inversion is more likely to be generalized from *what* questions to *who* questions (i.e., another argument question) rather than to *why* and *how* questions (i.e., adjunct questions) because of the structural similarity of *what* and *who* questions. This would explain earlier inversion with argument *wh*-questions than with adjunct *wh*-questions.

If this is true, we expect the learners' behavior to differ on *what* and *who* questions in terms of inversion (i.e., *what* > *who*) because of the great difference in the frequency of inverted questions between the two (i.e., *what* > *who*). However, this does not seem to be true. Unfortunately, our L1 acquisition study does not provide any information on this because we did not include *what* questions in the experiment.

However, our L2 acquisition study shows that the learners' judgement of uninverted *wh*-questions was not different for *what* and *who* questions. The acquisition order might have followed the order *what* > *who* > *why*, *how* because of the enormous number of inverted *what* questions compared to the inverted *who* questions in the input. This difference is not shown in L1 acquisition study, either. The relative acquisition order for *why* and *how* should have been *why* > *how*, considering the difference of frequency in the input between the two (i.e., *what* > *why* > *how* > *who*). However, this was not reflected in the results of L1 production study. The inversion scores of *how* and *why* questions were quite similar to each other.

7.2.3 Negative effect of uninverted questions

There is another aspect of input frequency theory. In order to test frequency theory, we have focused on the frequency of inverted *wh*-questions (e.g., mainly the frequency of inverted *wh*-questions and inversion rates) because it is the most relevant input for the acquisition of inversion in *wh*-questions. However, it is also possible to think that inversion in a particular type of *wh*-question might be delayed (or retarded) if the frequency of uninverted questions (and non-inversion rates, that is, the ratio of uninverted questions to total questions) is high for that *wh*-word in the input. In order to investigate this possibility briefly, non-inversion rates were calculated using the total number of *wh*-questions and uninverted *wh*-questions, which were available from our input study. The results are shown in the following table (for L1)

Table 7.1 Non-inversion rates for *wh*-questions in caregivers' utterances

	Total number of <i>wh</i> -questions				Uninverted (<i>Wh</i> + Subj +Aux)				Percentage			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-MOT	35	67	44	324	4	2	1	36	11%	3%	2%	11%
P-PAT	52	225	170	1,267	7	2	4	61	13%	1%	2%	5%
P-LOI	52	176	175	1,285	9	5	0	100	17%	3%	0%	8%
R-FAT	132	241	117	865	5	8	2	20	4%	3%	2%	2%
S-INV	208	376	160	2,460	43	12	11	200	21%	3%	7%	8%
H-MOT	15	20	40	959	6	0	0	116	40%	0%	0%	12%
Total	494	1,105	706	7,160	74	29	18	533	18%	2%	2%	8%
Mean	82.3	184.2	117.7	1,193.3	12.3	4.8	3.0	88.8	18%	2%	2%	8%

Note, first, that uninverted questions are more frequent in *what* questions than in any other *wh*-question (i.e., 533 vs. 18 – 74). Second, the rate of non-inversion is higher in *how* questions than in any other *wh*-question (18% vs. 2 – 8%). Considering only the frequency of uninverted questions, inversion should be delayed for *what* questions because of the high frequency of uninverted *what* questions. Considering the non-inversion rates, inversion should be retarded more for *how* questions than for any other question because of the high uninversion rate for *how* questions.

However, previous studies (De Villier 1991, Van Valin 2002) report that children's *why* questions are uninverted more frequently than any other question (including *how* questions). Our L1 production study also found more uninverted *why* questions than uninverted *how* questions. The difference is shown in the following figure, which is a re-presentation of Figure 4.11 of Chapter 4 for convenience.

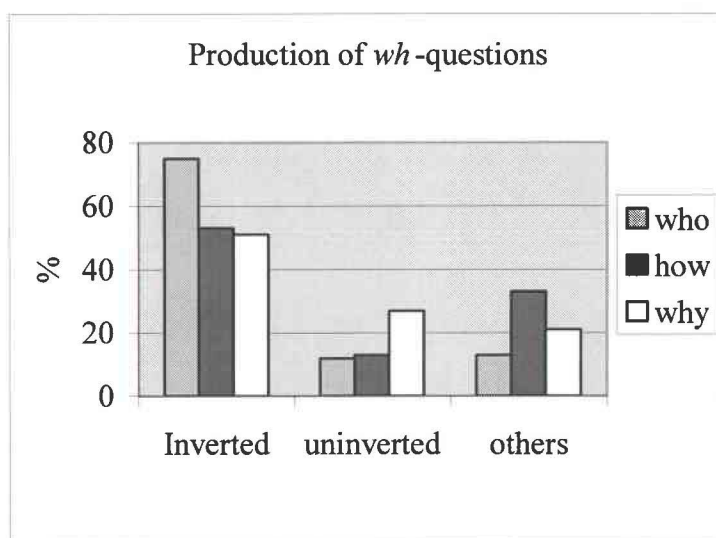


Figure 7.1 Children's production of *wh*-questions

Another possible negative effect of uninverted questions on the acquisition of inversion in *wh*-questions might relate to the ratio of inverted questions to uninverted questions. From the frequency point of view, the higher the ratio of inverted questions to uninverted questions is, the less likely children will get confused about inversion. With this rationale, the ratio of the number of inverted questions to the number of uninverted questions was calculated. The results are shown in the following table:

Table 7.2 Inversion (non-formulaic) vs. non-inversion

	Inverted (<i>Wh</i> + Aux +Subj)				Uninverted (<i>Wh</i> + Subj +Aux)				Ratio (Inverted/uninverted)			
	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>	<i>how</i>	<i>why</i>	<i>who</i>	<i>what</i>
P-MOT	15	22	4	114	4	2	1	36	3.8	11.0	4.0	3.2
P-PAT	21	40	17	364	7	2	4	61	3.0	20.0	4.3	6.0
P-LOI	22	28	22	481	9	5	0	100	2.4	5.6	N/A	4.8
R-FAT	82	97	18	545	5	8	2	20	16.4	12.1	9.0	27.3
S-INV	107	257	48	1,049	43	12	11	200	2.5	21.4	4.4	5.2
H-MOT	5	1	5	614	6	0	0	116	0.8	N/A	N/A	5.3
Total	252	445	114	3,167	74	29	18	533	4.8	14.0	5.4	8.6
Mean	42.0	74.2	19.0	527.8	12.3	4.8	3.0	88.8	4.8	14.0	5.4	8.6

The ratio was highest for *why* questions for all the caregivers except Ross's father. That is to say, children should be less likely to get confused about inversion with *why* questions than with any other *wh*-question, if the idea of negative effect for uninverted questions is right. In fact, however, uninverted questions are found in *why* questions more frequently than in any other *wh*-question in children's speech, as already mentioned above.

These results therefore do not seem to support the idea that inversion in a particular type of *wh*-question might be delayed if the frequency of uninverted questions (and non-inversion rates, that is, the ratio of uninverted questions to total questions) is high for that *wh*-word in the input.

In conclusion, it seems very difficult to accept the input frequency theory to account for argument/adjunct asymmetry in the acquisition of *wh*-questions, even when

we consider various alternative accounts involving lexical-based input frequency, structure-based input frequency, and the effect of the frequency of uninverted questions. It seems that the acquisition order is decided by the inherent property of the structures, rather than their frequency or the frequency of component lexical items.

7.3 Further research: Implicational relationship

We found from our input and acquisition studies that the structure-based generative approach can provide a better explanation for the argument/adjunct asymmetry in the acquisition of *wh*-questions. This approach attributes the argument and adjunct asymmetry to the difference in their syntactic properties, as described in GB theory.

If this is true, we predict that a child who uses the inversion rule in adjunct *wh*-questions must necessarily also use it in argument *wh*-questions, but not vice versa. In addition, it seems plausible to suggest that there should be languages in which inversion is required with only argument *wh*-questions, but not with adjunct *wh*-questions.

This seems to be true. The results from our experimental study of English-speaking children's acquisition of *wh*-question (in Chapter 4) showed a possible implicational relationship between argument and adjunct *wh*-questions. Consider the results represented here.

Table 7.3 Total number of inverted questions from an elicited production task

Subject	<i>who</i>	<i>how</i>	<i>why</i>	STAGE
21	5	5	5	All inversion
29	5	5	5	
19	5	5	4	Who Inversion
20	5	5	4	
23	5	5	3	
26	5	5	2	
30	5	4	4	
17	5	4	2	
25	5	2	4	
33	5	0	3	
27	5	1	1	
18*	4	3	5	Inversion trial
24	4	4	4	
31	3	2	2	
14	0	0	0	Non-inversion
16	0	0	0	
28	0	0	0	
Total (17)	66	50	48	
Mean	3.88	2.94	2.82	

- Maximum number: 5

Table 7.3 shows that inversion scores in *who* questions are always higher than (or at least the same as) those in *how* and *why* questions (except for child 18). Each child who produced *how* and *why* questions with inversion 100% of the time also produced *who* questions with inversion 100 % of the time (subjects 21, 29), but not vice versa (subjects 30, 17, 25, 33, and 27). In other words, there are no children who invert *why* and *how* questions 100% of the time without also inverting *who* questions 100% of the time, but not vice versa. Therefore, generally speaking, acquisition of subj-aux inversion with adjunct *wh*-questions implies acquisition of subj-aux inversion with argument *wh*-

questions. The results seem to indicate a possible implicational relationship between argument and adjunct *wh*-questions.

Second, another piece of evidence for the implicational relationship between argument and adjunct *wh*-questions is found in Baković's (1998) discussion of inversion in Spanish. Baković (1998) describes variation in inversion in *wh*-questions among dialects of Spanish. He argues that 'whether or not a *wh*-question will require inversion in a given dialect depends on two factors: whether it is a matrix or a subordinate *wh*-question, and *how argumental the moved wh-phrase is.*' (emphasis added p.35). Acknowledging Torrego (1984) as the first to discover the second factor, he says that in one dialect of Spanish, argument *wh*-phrases (*quién* 'who', *qué* 'what') require inversion while adjunct *wh*-phrase (*dónde* 'where', *cuándo* 'when', *cómo* 'how' and *por qué* 'why') do not. Integrating the data from Goodall (1991ab), Suñer (1994), and his own study (1995), Baković (1998) proposes the following implicational relationship among *wh*-phrase types (in matrix clause only).

(1) Implicational relationship among *wh*-phrase types.

A *wh*-question requires inversion if the *wh*-phrase is

<i>quién</i> , <i>qué</i>	>	<i>dónde</i> , <i>cuándo</i>	>	<i>cómo</i>	>	<i>por qué</i>
'who' what'		'where' 'when'		'how'		'why'

ARGUMENT >	LOCATION >	MANNER >	REASON
Pattern 1	Pattern 2	Pattern 3	Pattern 4

The implicational relationship in (1) shows that inversion is either not possible, or required in the order of implicational hierarchy in (1). For example, if inversion is required in a dialect of Spanish, it is most likely to be with an argument *wh*-phrase (i.e., Pattern 1). If inversion is required with a locative *wh*-phrase in a dialect, it is also required with argument *wh*-phrases, but not vice versa. (The placement of time and location *wh*-phrases between the argument and the other adjunct *wh*-phrases is plausible, because they can function as either an argument or an adjunct depending on the verb. For example, in the sentence, *where shall I put this?*, *where* is an argument of the locative verb *put*. However, in *where did you learn French?*, *where* is an adjunct.)

If these findings are right, we might expect to find this implicational relationship in the area of second language acquisition, paralleling the familiar effects of the implicational hierarchy in the case of the acquisition of relative clauses by ESL learners (Hyltenstam 1977, 1984, Gass 1978, Doughty 1991). Leaving these possibilities open for future research, I believe that this line of research will provide us with more information about L1 and L2 acquisition of *wh*-questions, leading us to better understanding of this phenomenon.

7.4 Conclusion

I have summarized the findings of our L1 and L2 studies, arguing that lexical-based input frequency theory cannot explain the argument/adjunct asymmetry in the acquisition of *wh*-questions. Possible alternative accounts were also discussed from the

input frequency perspective, but it was concluded that these alternative accounts are not supported from our data. A possible implicational relationship between argument and adjunct *wh*-questions was suggested for further research.

In conclusion, the research questions of this dissertation posed in Chapter 1 can be answered. First, the L1 and L2 acquisition studies found that there is an argument/adjunct asymmetry in the acquisition of inversion in wh-questions by L1 and L2 learners of English; both L1 and L2 learners of English did better at inversion with argument questions than with adjunct questions. Second, the comparison of L1 and L2 studies revealed that there is a similarity between L1 and L2 regarding input and the acquisition of inversion in wh-questions. Finally, based on these findings, it was argued that the structure-based generative approach provides a better explanation for the acquisition of wh-questions than the lexical-based input frequency theory.

APPENDIX I: TARGET ITEMS IN EXPERIMENT 2

CHILDREN'S ACQUISITION OF *WH*-QUESTIONS: AN ELICITED PRODUCTION TASK

1. *Who are you pushing?*
2. *Who are you chasing?*
3. *Who are you biting?*
4. *Who are you stinging?*
5. *Who are you scratching?*
6. *How are you flying?* (with a picture of witch flying on a broom)
7. *How are you going very fast?*
8. *How are you moving?*
9. *How are you flying?* (with a picture of mole flying on a magic carpet)
10. *How are you eating?*
11. *Why are you smiling?*
12. *Why are you laughing?*
13. *Why are you sweating?*
14. *Why are you running?*
15. *Why are you shivering?*

APPENDIX II: TARGET ITEMS IN EXPERIMENT 4

ESL LEARNERS' ACQUISITION OF *WH*-QUESTIONS: A GRMMATICALITY JUDGMENT TASK

(Sound grammatical or not?)

	←-----→			
	Not possible		Perfectly O.K.	
Pronouns subj. Be aux.	-2	-1	+1	+2
1. Who are you meeting in the cafeteria?				
2. What are you reading in the library?				
3. Why are you jumping on the bed?				
4. How are you going to the campground?				

Pronoun subj. other aux's (did)

5. Who did he visit last night?
6. What did he find last night?
7. Why did he cry last night?
8. How did he sleep last night?

Noun subj. Be aux.

9. Who are the children meeting in the cafeteria?
10. What are the children reading in the library?
11. Why are the children jumping on the bed?
12. How are the children going to the campground?

Noun subj. other aux's

13. Who did the old man visit last night?
14. What did the old man find last night?
15. Why did the old man cry last night?
16. How did the old man sleep last night?

Pronouns subj. Be aux.

17. *Who you are meeting in the cafeteria?
18. *What you are reading in the library?
19. *Why you are jumping on the bed?
20. *How you are going to the campground?

Pronoun subj. other aux's (did)

21. *Who he visited last night?
22. *What he found last night?

- 23. *Why he cried last night?
- 24. *How he slept last night?

Noun subj. Be aux.

- 25. *Who the children are meeting in the cafeteria?
- 26. *What the children are reading in the library?
- 27. *Why the children are jumping on the bed?
- 28. *How the children are going to the campground?

Noun subj. other aux's

- 29. *Who the old man visited last night?
- 30. *What the old man found last night?
- 31. *Why the old man cried last night?
- 32. *How the old man slept last night?

REFERENCES

- Baković, E. 1998. Optimality and inversion in Spanish. In (eds.) Barbosa, D. Fax, P. Hagstrom, M. McGinnis and D. Pesetsky. *Is the Best good enough: Optimality and competition in syntax*. Cambridge, MIT Press. 35-58.
- Bley-Vroman, R. 2002. Frequency in production, comprehension, and acquisition. *Studies on Second Language Acquisition* 24: 209-213.
- Chomsky, N. 1986. *Barriers*. Cambridge: MIT Press.
- De Villiers, J. 1991. *Why* questions? In T.L. Maxfiel & B. Plunkett (eds.), *Papers in the acquisition of wh: Proceedings of the Umass Roundtable, May 1990*. (P.155-173). Amherst, MA: University of Massachusetts Occasional Papers.
- Doughty, C. 1991. Second language instruction does make a difference: Evidence from an empirical study of ESL relativization. *Studies in Second Language Acquisition* 13: 431-469.
- Eckman, F., Moravcsik, E & J. Wirth. 1989. Implicational universals and interrogative structure in the interlanguage of ESL learners. *Language Learning* 39 (2): 73-205.
- Erreich, A. 1984. Learning how to ask: patterns of inversion in *yes-no* and *wh*-questions. *Journal of Child Language* 11. 579-592.

- Gass, S. 1979. Language Transfer and universal grammatical relations, *Language Learning*, 29:327-344.
- _____ and A. Mackey. 2002. Frequency effects and second language acquisition: A Complex Picture?. *Studies on Second Language Acquisition* 24: 249-260.
- Goodall, G. 1991a. On the status of SPEC of IP. In D. Bates (eds.). *The proceedings of the tenth West Coast Conference on Formal Linguistics*. Stanford CA. CLSI, 175-182.
- _____. 1991b. Spec of IP and CP in Spanish *wh*-questions. Paper presented at the twenty-first annual Linguistics Symposium on Romance Languages, University of California, Santa Barbara. CA.
- Greenberg, J. H. 1966. *Language universals with special reference to feature hierarchies*. Janua Linguarum, Series minor 59. The Hague: Mouton.
- Haznedar, B. 2003. The status of functional categories in child second language acquisition: evidence from the acquisition of CP. *Second Language Research* 19,1: 1-41.
- Hyltenstam, K. 1977. Implicational patterns in interlanguage syntax variation. *Language Learning* 27: 383-411.
- Hyltenstam, K. 1984. The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In R. Andersen (eds.) *Second languages: A cross-linguistic perspective*. Rowley,

MA: Newbury House.

- Ingram, D. and D. Tyack. 1979. Inversion of subject NP and aux in children's questions. *Journal of psycholinguistic Research* vol.8, no.4: 333-341.
- Keenan, E. L. and Comrie, B. 1977. 'Noun phrase accessibility and Universal grammar. *Linguistic Inquiry*, 8.: 63-99.
- Kim, S-C. 1995. *The acquisition of wh-questions in English and Korean*. Ph.D dissertation, University of Hawaii at Manoa.
- Klima, E. S. and Bellugi, U. 1966. Syntactic regularities in the speech of children. In (eds.) J. Lyons and R.J. Wales, *Psycholinguistic papers*, 183-208. Edinburgh: Edinburgh University Press.
- Larsen-Freeman, D. and M. Long. 1991. *An introduction to second language acquisition theory and research*. London. Longman.
- _____. 2002. Making sense of frequency. *Studies on Second Language Acquisition* 24. 275-285.
- Labov, W. and T. Labov. 1978. Learning the syntax of questions. In (eds.) R.Campbell and P. Smith *In Recent advances in the psychology of language: Formal approaches*. (P.1-44). New York: Plenum Press.
- MacWhinney, B. 1995. *The CHILDES Project: Computational Tools for Analyzing Talk*. Lawrence Erlbaum.
- MacWhinney, B. 1998 (April). The CHILDES Database.
- Pienemann, M. 1985. Learnability and syllabus construction. In K. Hytlenstam &

- M. Pienemann (eds.) *Modeling and assessing second language acquisition* (P. 23-75). Clevedon, Avon: Multilingual Matters.
- Pienemann, M., Johnston, M. and G. Brindley. 1988. Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition* 10 (2): 217-243.
- Ravem, R. 1974. The development of *Wh*-questions in first and second language learners. In Richards, J.C. (eds.). *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman, 134-55.
- Rowland, C. F. and J. Pine. 2000. Subject-auxiliary inversion errors and *wh*-question acquisition: 'what children do know?' *Journal of Child Language* 27:157-81.
- _____. 2003. The development of inversion in *wh*- questions: a reply to Van Valin. *Journal of Child Language* 30: 197-212.
- Spada, N. and P. M. Lightbown. 1999. Instruction, first Language influence, and developmental readiness in second language acquisition. *The Modern Language Journal* 83,i: 1-22.
- Stromswold, Karin. 1988. Linguistic representations of children's *wh*-questions. *Papers and reports on Child Language Development* 20:107-14.
- _____. 1990. *Learnability and the acquisition of auxiliaries*. Ph.D. dissertation, MIT
- _____. 1995. Acquisition of subject and object *wh*-questions. *Language*

Acquisition 4: 5-48.

Suñer, M. 1994. V-movement and the licencing of argumental *wh*-phrases in Spanish.

Natural Language and Linguistics Theory. 335-372.

Torrego, E. (1984) On inversion in Spanish and some of its effects. *Linguistics*

Inquiry 15: 103-129.

Van Valin, R.D. 2002. The development of subject-auxiliary inversion in English *wh*-questions: an alternative analysis. *Journal of Child Language* 29: 161-75.

Yoshinaga, N. 1996. *Wh-questions: A comparative study of their form and acquisition in English and Japanese*. Ph.D dissertation, University of Hawaii at Manoa.

White, L., N. Spada, P. M. Lightbown and L. Ranta. 1991. Input enhancement and

L2 question formation. *Applied Linguistics* 12: 416-432.

Wilhelm, A. and K. Hanna. 1992. On the acquisition of *wh*-questions. *Calgary*

Working Papers in Linguistics 15:89-98.

Wode, H. 1978. The L1 vs. L2 acquisition of English interrogation. *Working Papers*

in Bilingualism 15: 37-57.

Wordnote:<http://www.wordnote.com>.