ASKING QUESTIONS IN LEARNER ENGLISH:
FIRST AND SECOND LANGUAGE ACQUISITION OF MAIN AND EMBEDDED INTERROGATIVE STRUCTURES

by

LUCIA POZZAN

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of
the requirements for the degree of Doctor of Philosophy,
The City University of New York
2011

This manuscript has been read and accepted for the
Graduate Faculty in Linguistics in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

VIRGINIA VALIAN

Date

Chair of Examining Committee

GITA MARTOHARDJONO

Date

Executive Officer

MARTIN CHODOROW
EVA FERNÁNDEZ
JANET DEAN FODOR
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

ASKING QUESTIONS IN LEARNER ENGLISH:
FIRST AND SECOND LANGUAGE ACQUISITION OF MAIN AND EMBEDDED INTERROGATIVE STRUCTURES

by

Lucia Pozzan

Adviser: Professor Virginia Valian

This dissertation examines how adults and children learning English produce and judge English interrogative structures. The ultimate goal of this study is to contribute to an understanding of the extent, nature and sources of learners' persistent difficulties with some syntactic properties of the language they are acquiring.

To examine whether word order errors in the production of English interrogatives by L2 learners stem from lack of knowledge or from difficulties with automatic implementation of L2 procedures under real-time constraints, L2 learners' performance across a range of tasks (oral and written production, timed and untimed grammaticality judgments) is compared.

To examine whether errors in the production of English interrogatives by L2 learners can be imputed to transfer of L1 properties, L1 Chinese and L1 Spanish learners' production patterns are compared.

Finally, to examine whether errors in the production of L1 learners can be attributed to properties of the adult input, the results from an elicited production study with 3-5 year olds are examined in light of the frequency of different word combinations in the adult input.

Taken together, the results of the present studies indicate that difficulties with English interrogative structures (a) are a consistent phenomenon both in L1 and L2 acquisition, (b) might

be better accounted for in terms of non-target-like representations rather than difficulties with implementation of L2 procedures, and (c) do not follow in a direct way from the properties of learners' L1s or the properties of the input. Furthermore, the results of the present studies show that learners' errors are associated with specific syntactic configurations (*wh*- vs. yes/no structures) and *wh*-words (*why* and *when vs. who, what,* and *where*), suggesting that child and adult learners entertain similar grammatical hypotheses and make use of similar mechanisms for language acquisition.

Acknowledgments

First, I would like to thank my advisor, Virginia Valian, for the guidance, encouragement and intellectual stimulation she provided throughout my graduate career and for copious comments on all the projects I have embarked on since first stepping foot into the Language Acquisition Research Center in the fall of 2007. Working at LARC has been a fundamental personal and academic experience, and the lab has represented for me throughout the years an intellectual space for an open exchange of ideas and constructive criticism. Virginia's intellectual (and material) generosity has allowed me to experiment with ideas and projects that seemed unworkable at first. I deeply admire Virginia's intellectual rigor, and am grateful to her for teaching me to think about the big picture, to always consider seriously and with equanimity the opposing side in the intellectual debate, and to even rejoice in the refutation of a cherished hypothesis; this is, for me, the essence of good scientific inquiry.

I am deeply grateful to Janet Dean Fodor for her detailed comments, ideas and insights, and for investing countless hours into commenting on everything I wrote. The initial project that ultimately developed into this dissertation started in Janet's office, and would not have been carried out without her contribution and support through each phase. Above all, I am grateful to Janet for teaching me how to think and write like a psycholinguist.

I would like to thank Eva Fernández, for her advice, support, and friendship. This dissertation would not have been written without Eva generously offering her lab for testing participants and finding ways to support me through the Center for Teaching and Learning and the General Education Office at Queens College. I will greatly miss our Wednesday chats over coffee!

I wish to thank Martin Chodorow for his time, help, and patience, and for teaching me virtually everything I know about statistics; as he would say, I am now ‘dangerously knowledgeable’. Thanks to Dianne Bradley, for helping me immensely with the initial protocol for the elicited production studies, and for teaching me to find the bear before getting ready to shoot at it. I would also like to thank Carlo Cecchetto, who taught the first linguistics class I ever took – it was love at first hearing – and Luigi Rizzi, for encouraging me to pursue a Ph.D. in linguistics. Special thanks to Giuliano Bocci, partner in crime of my first linguistics investigations in Siena, whose loyal friendship and rare combination of syntactic and statistical expertise have been an essential part of my growth as a linguist and as a person, and to Daniel Altshuler, for convincing me to pursue a Ph.D. in the U.S., for continuous friendship since the 2004 EGG summer school in Cluj-Napoca, moral and practical support, and for being the most enthusiastic linguist I know.

For helpful comments and discussion at different stages of this process, I wish to thank Marcel den Dikken, Nancy Eng, Fred Eckman, Gita Martohardjono, Bill McClure, Akira Omaki, Sandeep Prasada, David Pereplyotchik, David Stringer, Lidiya Torniova, Christina Tortora, Judith Summerfield, and the audiences at GALA 2009, SLRF 2009, SLFR 2010, and BUCLD 2010.

For friendship, support, and comic relief, I thank Carolina Barrera Tobón, Chiara Bauzulli, Annalisa Benetello, Roi Ben-Itzhack, Chiara Contino, GizMo, Brett Marroquín, Jesse Mayerson, Priscilla Obeada, Alessandra Piovene, Sara Sambin, and Hélène Zago.

Thanks for friendship, support and insights to all the research assistants and interns at LARC who have helped at various stages of this process. In particular, I would like to thank Paul Feitzinger for all his help making CHILDES Excel-friendlier and for setting up the online system for collecting the online L2 corpus, Erica Knowles and Syelle Graves for running the lab smoothly and cheerfully, and supporting my research with their time and expertise. I also wish to thank the Graduate Center and the National Science Foundation for financial support.

Special thanks to the research assistants that have worked with me on the child study, Nathan Lafave and Margarita Zeitlin; this work could not have been done without their hard work and immense help running and implementing experiments, testing participants, and transcribing hours of child speech, or without Margarita's editing and checking of various drafts of this dissertation.

Thanks to the adult and child participants for their generosity with their time, and especially the child participants and their families for their kindness and the fun times! I am deeply grateful to Amit Shaked Pasman and the Temple Shaaray Tefila Nursery School for their help recruiting participants for the child study.

Thanks to my parents for their endless love and encouragement; to my mother, for always taking my side, even when I am wrong, and for transmitting to me her love for teaching and her passionate commitment to making a difference in children's lives; to my father, for not taking my side when I am wrong, for reminding me of the importance of scientific inquiry, even when any of its potential application seems – at the time – unbearably distant, and for always urging me to pursue what I love and not what I think is most useful.

In a category of its own, I would like to thank my 'New York Circle of Sisters.' Patricia Crown, for her wisdom, kindness and generosity; it still surprises me how she paid me for five years for offering me coffee and chatting with me in Italian! Erin Quirk, without whom this project would not have been born, for a memorable beginning of my graduate school career and continuing friendship, support and intellectual stimulation. Alexia Ioannidou, for being an

amazingly generous and supportive friend, for past and future adventures in the Caribbean, and for reminding me that spring break is for traveling and not for writing. Elizabeth Pratt, for discussing every project I have ever conceived of, for help with experimental items and linguistic judgments, cat-duty sharing, and, above all, for her patience and unlimited generosity; I could not have dreamed of a better roommate. Giulia Bencini, whose bigheartedness and kindness are a source of inspiration to me and the people around her, for her intellectual generosity, extensive mentorship, pivot tables, yoga challenges, and unconditional love. This dissertation is dedicated to them.

Table of Contents:

List of Tables:

List of Figures:

1. Introduction

This dissertation presents an in-depth investigation into the acquisition of main and embedded questions by first and second language learners. The studies presented here examine the extent to which learners of Standard English produce word-order errors in main and embedded interrogatives via a series of experiments. In particular, this dissertation focuses on subject-auxiliary inversion errors[1]. In main questions, these errors involve the lack of subject-auxiliary inversion in contexts where, in the native adult grammar, an auxiliary obligatorily precedes the subject, as shown in (1). Conversely, non-target productions in embedded questions take the form of over-application of subject-auxiliary inversion, as shown in (2):

(1) *Why you are saying that?

(2) *I don't know where is the restroom.

A robust finding in the acquisition literature is that learners' non-target productions usually take the form of *omission* of obligatory grammatical elements, rather than *commission* (Kolk, 2001; Snyder, 2007), especially in the domain of morpho-syntax. In the case of omission errors, it is often difficult for researchers to infer whether the speaker has omitted an element because an abstract feature (e.g., tense) is absent from the learner's grammatical representation, or whether

[1] For ease of exposition, I will refer to lack of subject-auxiliary inversion and over-application of subject-auxiliary inversion errors as "inversion errors".

the absence of such a feature in production is purely a matter of ‘spell-out’ (e.g., Prévost & White, 2000). By focusing on an error of commission, the present study does not face the same problems.

The primary goal of this study is to investigate the *extent*, *nature* and *causes* of subject-auxiliary inversion errors in first and second language learners of English, and to examine whether existing explanations adequately account for the empirical findings.

In order to investigate the nature of L2 learners’ errors in English interrogatives, L2 learners’ performance in a number of tasks (oral and written production, timed and untimed acceptability judgments) was examined. In order to investigate the role of L1 transfer as one of the possible causes for non-target productions in L2 learners, the production patterns of L2 learners’ with different native language (L1) backgrounds were examined. Finally, in order to investigate whether the native adult input could account for these errors, the production of monolingual English-speaking children was examined in light of the properties of the adult input.

1.1. Second Language Acquisition of English Questions

A central goal of second language acquisition research is that of determining the nature of errors that are persistent even at advanced proficiency levels. Very broadly speaking, the nature of these errors could be characterized either in terms of insufficient knowledge of the second language (L2, henceforth), or in terms of difficulties with the implementation of L2 procedures. Non-target productions can be attributed to insufficient L2 knowledge when an L2 property is not present in the L1 of the learners. In these cases, the L2 property is apparently extremely difficult to learn in adulthood, causing L2 learners to perform poorly across tasks, including tasks that are traditionally thought to draw more heavily on explicit knowledge and declarative memory (e.g., edited written production and untimed grammaticality judgments, see Ellis, 2004, 2005; Bowles, 2011). An example of an error stemming from insufficient knowledge is, for example, the (failed) acquisition of L2 Italian auxiliary selection (i.e., *be* vs. *have*) by speakers whose L1 lacks this distinction (Sorace, 1993).

L2 production errors can instead be said to stem from difficulties with the implementation of target procedures when the L2 requires the automatization of linguistic procedures that are not implemented consistently by learners (Antón-Méndez, 2010). An example of this type of error is the lack of implementation of third person subject-verb agreement in L2 learners of English. These errors do not stem from lack of knowledge, as shown by the fact that they only surface in tasks in which the learner is pressed for time (e.g., in oral production as opposed to written production or grammaticality judgment tasks).

Finally, implementation difficulties might take the form of excess automatization of L1 procedures; these errors happen when automatic L1 procedures get erroneously applied to the learners' L2 (Antón-Méndez, 2010). An example of this type of error could be the *occasional* production of null subjects in L2 English by speakers of a null-subject language. Errors of excess automatization do not follow from lack of knowledge and are more likely to surface in tasks where the learner is pressed for time (e.g., in oral production as opposed to written production or grammaticality judgment tasks).

Another central goal of second language acquisition research is that of determining the causes of L2 learners' non-target grammatical representations or procedures. One of the causes of non-target representations and/or procedures is the influence of L1 properties on the L2 grammar. The influence of the L1 on the L2 is an issue that is orthogonal to that of determining the nature of errors in L2 learners' performance because L1-driven errors might be due to lack of knowledge of the L2 system (e.g., learners fail to recognize the difference between the L1 and the L2), or to difficulties with the implementation of L2 procedures (e.g., learners erroneously apply L1 procedures to the L2).

In order to determine whether errors of inversion in L2 speakers are due to L1 transfer, I conducted an elicited production study in collaboration with Erin Quirk (Pozzan & Quirk, sumbmitted) with participants whose L1 was either Chinese[2] or Spanish. These L1s were chosen because they differ from each other and from English in terms of the relative order of the subject and a tensed verbal element in questions, and can thus be used to distinguish the predictions of different accounts.

In Chinese, there is no overt *wh*- or verb movement, and declarative and interrogative structures alike exhibit SVO order. In Spanish, on the other hand, there is overt *wh*-movement, while there is no subject-*auxiliary* inversion per se in questions. However, in Spanish interrogatives (main and embedded alike), the tensed lexical verb and the auxiliary (when present) precede the subject in terms of linear order, giving rise to (Aux)-Verb-Subject order in interrogative structures. Traditionally, it has been assumed in the syntactic literature that this linear order of constituents is derived by movement of the tensed verb to the head of the Complementizer Phrase (CP) via V-to-T-to-C movement. This proposal assumes that English

[2] For ease of exposition, I will refer to L1 speakers of Mandarin, Cantonese, and Shanghainese as 'Chinese' speakers. These languages are similar with respect to verb movement, the property at issue here.

and Spanish main questions are similar in that both require movement of a tensed verbal element to the left periphery (T-to-C movement), while the difference in word order between the two languages (Aux-Subject-Verb in English vs. (Aux)-Verb-Subject in Spanish) arises from the independently motivated fact (e.g., Pollock, 1989) that in English only the tensed auxiliary raises to T, but in Spanish, the tensed lexical verb raises to T. Under this view, both English and Spanish require T-to-C movement, but the specific element that moves to C is the auxiliary in English and the tensed verbal complex in Spanish.[3] Another difference between the two languages lies in the fact that VS order in Spanish questions is optional for adjunct *wh*-questions

---

[3] An ongoing debate in the literature concerns whether VS order in Spanish interrogatives (and more generally, in many Romance varieties) is caused by movement of the verb to the CP domain (V-to-C movement) or by the subject being located in a post-verbal position. While it is well beyond the scope of this dissertation to investigate the fine syntax of interrogative structures in Spanish, a good reason to think that VS order in Spanish interrogatives might be due to the presence of a subject in post-verbal position rather than V-to-C movement comes from the observation that in Spanish interrogatives, it is not just the tensed verb that precedes the subject, but the whole VP, as shown in (i):

(i) Cuando [$_{TP}$ ha [$_{VP}$ llamado [$_{DP}$ a sus padres]]] Maria?
When has called her parents Maria
When did Maria call her parents?

An analysis of (i) in terms of V-to-C movement is problematic, given that it would require a phrasal constituent (VP) to move to C, which is ruled out by the theory. In order to maintain the traditional analysis according to which VS order in Romance interrogatives is a V2 phenomenon, Rizzi (2001) assumes that VOS order in interrogative structures is due to movement of the whole VP to the left periphery (and, specifically, to a position adjoined to FinP), and then movement of the *wh*-element (when present) to Spec,FocP.
Regardless of what the best syntactic characterization for Spanish interrogative structures might be, the important fact for the present discussion is that Spanish interrogatives, analogously to English interrogatives but differently from Chinese ones, differ from declarative clauses in terms of the relative order of the subject and the tensed verb, and that this similarity could make a difference for learners.

(Torrego, 1984), and especially for *why*-questions (Baauw, 1998), while it is obligatory for argument questions.

L1-transfer accounts of second language acquisition predict that L2 learners will transfer properties of their L1 to the L2. Transfer of L1 properties can assume different forms. Here, I sketch the basic predictions for L1 Chinese and L1 Spanish learners' oral productions, while a more detailed discussion of predictions and possible experimental outcomes is presented in Section 2.1.

a. Main Questions: L1 Chinese learners of English are expected to fail to implement inversion to some degree because Chinese does not exhibit this property.

b. Main Questions: L1 Spanish learners are expected to transfer, to some extent, syntactic properties of Spanish interrogative structures to English ones. Transfer of L1 properties might consist of transfer of word order or transfer of abstract features. In the first case, we expect L1 Spanish learners to produce ungrammatical VS structures in English main questions.[4] However, if L1 transfer is to be conceived as transfer of abstract features, and if it is true that Spanish and English are similar in that both require T-to-C movement, English subject-auxiliary inversion will not create particular difficulties for L1 Spanish learners at intermediate/advanced levels of proficiency. However, given that Spanish displays an argument-adjunct asymmetry with respect to word order in questions, L1 Spanish learners might exhibit an argument-adjunct asymmetry in their L2.

[4] Previous research indicates that this error occurs only at the very beginning stages of L2 acquisition (Escutia, 2002).

c. Embedded Questions (oral production): Chinese does not display inversion in either main or embedded questions. L1 Chinese learners are thus not expected to produce inversion errors in English embedded interrogatives.

d. Embedded Questions (oral production): L1 Spanish learners are expected to transfer, to some extent, syntactic properties of Spanish interrogative structures to English ones. Transfer of L1 properties might consist of transfer of word order or transfer of abstract features. In the first case, we expect L1 Spanish learners to produce ungrammatical VS structures in embedded English questions. On the other hand, if L1 transfer is to be conceived mainly as transfer of abstract features, L1 Spanish speakers are expected to produce inversion errors in English embedded questions.

The oral production experiments presented in Section 2.3.1. were aimed at investigating whether syntactic properties of learners' L1s were responsible for L2 learners' non-target productions in English interrogatives structures.

In order to investigate the nature of these non-target productions, L2 learners' written productions and acceptability judgments of main and embedded English questions were examined. The prediction was that if inversion errors are due to genuine lack of knowledge of the L2 system, difficulties with inversion should surface across experimental tasks. On the other hand, if inversion errors stem from difficulties with the implementation of L2 procedures, non-target patterns with respect to inversion should only surface in tasks where learners are under hard real time pressures (oral production and timed acceptability judgments).

1.2. First Language Acquisition of English Questions

An additional central aim of this dissertation is that of investigating some of the possible causes of word order errors in the acquisition of English questions by child and adult learners. While the role of L1 transfer in the acquisition of L2 learners will be addressed in Chapter 2, Chapter 3 will examine whether inversion errors in production by monolingual English-speaking children could be predicted on the basis of the adult input.

If it were found that L1 child and L2 adult learners exhibit similar qualitative and quantitative patterns and that both L1 transfer and input frequency play a limited role in the acquisition of English interrogative structures should be taken to suggest that first and second language learners use the same mechanisms and procedures while acquiring a language, contrary to claims in the literature that adult L2 learners use non-domain-specific mechanisms (Bley-Vroman, 1990; 2009), or different cognitive systems subserved by separate neural circuits (declarative vs. procedural memory, Ullmann, 2004, 2005) for language acquisition.

In Chapter 3, I first investigate children's production patterns of English main and embedded questions via an elicited production study and then compare the results of the production study with data from corpora of child directed speech. While word order error patterns in L1 acquisition of main questions have been extensively studied, there is a considerable lack of consensus on some of the main empirical findings and on the explanations thereof.

In particular, while early findings showed a question-type asymmetry in main questions, with yes/no questions giving rise to higher inversion rates than *wh*-questions (e.g., Klima &

Bellugi, 1966), and an argument-adjunct asymmetry, with non-subject argument *wh*-questions being associated with higher inversion rates than adjunct *wh*-questions (e.g., Stromswold, 1990; Sarma, 1991; de Villiers, 1991), these findings were not replicated consistently in the subsequent literature (e.g., Erreich, 1984; Valian, Lasser & Mandelbaum, 1992). By using (a) a within-subjects design, (b) the same experimental protocol to elicit these different structures, and (c) two different coding schemes in order to compare the present findings with results from previous studies in the literature, the present study should provide robust results that can be used to settle some controversies in the literature.

The acquisition of embedded questions by child learners of English is still a fairly understudied topic. To the best of my knowledge, only two studies in the literature have investigated this structure (Sarma, 1991; Stromswold, 1990), with contradictory results: Sarma (1991) found that inversion errors in embedded questions are virtually non-existent in the production of English-learning children, while Stromswold found that these errors occur at a rate of approximately 10% of the time in children's spontaneous speech. By using a protocol similar to that developed for L2 learners with controlled materials, and by eliciting both yes/no and *wh*-questions, the present study should provide substantial evidence to fill this gap in the literature.

The acquisition of English questions by first language learners has seen the contraposition of two main theoretical approaches, the generativist/nativist and the empiricist/constructivist. The main difference between these two accounts is that, while constructivist accounts deny the existence of innate mechanisms for language acquisition and contend that acquisition patterns can be explained by input properties alone, the generative account assumes that innate principles guide language acquisition, above and beyond the role of the input.

In the case of the acquisition of interrogative structures in particular, the challenge for constructivist theories is that of showing that errors in child speech can be explained by the frequency with which certain word combinations occur in the input, in the absence of such a pattern in the adult speech (i.e., adult native speakers never produce main *wh*-questions without inversion). In contrast, the challenge for UG-based theories is that of showing that errors cannot be derived from the input alone and that there are a priori reasons to predict the specific error patterns observed in child speech. While the predictions of constructivist models are fairly straightforward with respect to the existence of a question-type and argument-adjunct asymmetry (see below), generative accounts of the acquisition of English interrogatives have more often been post-hoc rather than prediction-based accounts. For this reason, in this study, the predictions of generative accounts will mainly be based on the previous literature.

In the following paragraphs I will summarize some findings of my investigation of input patterns in child directed speech, and sketch the predictions of constructivist and generative accounts for inversion errors in main questions.

1.2.1. Input Analysis for English Questions

The input to children in six corpora of American English was examined. Main and Embedded yes/no and *wh*-questions were extracted. Only questions containing the *wh*-words *what*, *which*, *when* and *why* and the auxiliary *is* and *are* were included in the final analyses in order to best compare the input data with the materials in the experimental investigation. While current constructivist accounts base their predictions for inversion patterns in main questions on the frequency of particular *wh*+auxiliary combinations in main English questions in the input, it is conceivable that inversion errors could be caused by the presence of non-inverted embedded *wh*-questions in the input (e.g., Maratsos et al., 1979; Erreich, 1984; Tornyova & Valian, 2009). For this reason, I will present constructivist predictions based both on the absolute frequency of *wh*+auxiliary combinations in main questions and on the relative frequency of *wh*+auxiliary combinations (evidence for inversion) over *wh*+NP+auxiliary combinations (evidence for non-inversion) in the input. Table 1 presents the distribution of inverted and non-inverted *wh*- and main yes/no questions in the input.

Table 1: Distribution of interrogative structures in the adult input by *wh*-type and inversion

Inversion	Question				
	What	***Which***	***When***	***Why***	***Yes/No***
Inverted	6397	133	32	239	2,740
Non-inverted	608	11	8	24	1,470

Question-type

Main *wh*-questions are more frequent than main yes/no questions in the input to children (6,801 vs. 2,740). Moreover, the ratio of inverted *wh*-questions over non-inverted ones is much higher than the ratio of inverted yes/no questions over non-inverted ones (10.5 vs. 1.9), indicating that evidence is more abundant for inversion in *wh* than in yes/no contexts. Constructivist accounts should thus predict the existence of an asymmetry in errors, with yes/no questions being associated with higher non-inversion rates than *wh*-questions. In contrast, generative accounts might predict an asymmetry in the opposite direction.

Wh-type

Main *what*-questions are by far the most frequent *wh*-question in adult speech, followedat a considerable by *why*-, *which*- and *when*-questions. Constructivist accounts based on the absolute frequency of inverted structures in the input should thus predict not an argument-adjunct asymmetry but a *what*-asymmetry. On the other hand, constructivist accounts based on the relative frequency of inverted over non-inverted structures should predict *what, which* and *why* questions to exhibit similar inversion rates given that the ratio of inverted over non-inverted questions is comparable for these elements (10.5 for *what*, 12.1 for *which*, and 10 for *why*), but particularly low for *when*-questions (4). Based on previous literature, on the other hand, generative accounts predict either the existence of an argument-adjunct asymmetry (with arguments being associated with higher inversion rates, see, for example, Stromswold, 1990) or a *why*-effect (Thornton, 2008). Table 2 summarizes the contrastive predictions for argument and adjunct main questions. The symbol '√' indicates that the account predicts the asymmetry and the symbol 'X' that the account does not predict an asymmetry, while the symbol '>' indicates that the account predicts higher accuracy for one structure.

Table 2: Summary of predictions for main questions in L1 acquisition

Asymmetry	**Account**	
	Constructivist	**Generative**
Question-type	√ (*wh*- ≥ yn)	√ (yn ≥ *wh*-)
Argument-adjunct	X (*what* > everything else)	√ (argument > adjunct)

1.3. Methodology

In this dissertation, I use oral elicited production as the main technique to investigate L1 and L2 learners' knowledge of English interrogatives. While elicited production seems particularly well suited to examine structures that might be used infrequently in adult and child speech and seems not to present some of the disadvantages of other techniques (see Section 3.2.1.3), it is obviously just one of the many ways in which learners' knowledge of a language can be measured. Moreover, it is likely that, due to the fact that speech is by its very nature affected by the pressures of real time processing, oral production might be particularly prone to performance errors and this type of measure might thus underestimate learners' linguistic knowledge. However, the converse is also true, and it could be argued that untimed tasks are more likely to reflect speakers' metalinguistic knowledge.

In order to present a more complete picture of L2 acquisition of English interrogatives, written production and timed and untimed acceptability judgments were also used to investigate adult L2 learners' competence. The idea is that converging evidence from different types of performance data would constitute a solid basis from which to draw inferences about linguistic competence. In the case of L1 learners of English, on the other hand, knowledge of inversion was only investigated via elicited production. This was because the data from grammaticality judgments of interrogative structures obtained from children (Stromswold, 1990) seem far from conclusive, and because, in the L1 study, I was mainly concerned with comparing children's productions with patterns in the adult input; the possibility of investigating children's performance on interrogative structures via a variety of methodologies was beyond the scope of this study.

1.4. Coding

A surprising fact about the literature on the first language acquisition of English main questions is the lack of consistency across studies with respect to some basic empirical findings. This inconsistency is probably due to a number of factors, including differences in the coding of utterances and errors.

For example, some researchers (e.g., Thornton, 2008) code questions with an omitted auxiliary as non-inverted, while others do not include these productions or code them in a separate category. Differences in the coding of errors are particularly crucial when comparing errors in yes/no and *wh*-questions; given that lack of inversion is grammatical in English yes/no questions under specific pragmatic circumstances and that the majority of the studies in the literature examine written transcripts of spontaneous production where it is not immediately obvious if a non-inverted string is a question or a statement, researchers have often opted to exclude non-inverted questions (e.g., Rowland, 2007). In order for the different results from this study to be maximally comparable with each other and with other studies in the literature, the same coding scheme and two types of analyses were conducted throughout this dissertation. Each production was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology, and presence of target lexical items (e.g., subject and *wh*-element). Following Ambridge et al.'s (2006) coding scheme, incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors (non-inversion in main questions and inversion in embedded questions).
- Double tense/double auxiliary errors.
- Omitted auxiliary errors or errors that, due to lack of morphology, were ambiguous between non-inversion errors and omitted auxiliary errors.
- Other errors. Other errors included questions that differed in type from the target (yes/no instead of *wh*-question, and vice versa); subject *wh*-questions instead of object *wh*-questions; productions that differed from the target in the lexical items used; questions without a subject; questions with VP movement; questions where the *wh*-word differed from the target one, etc.

As mentioned above, two analyses were conducted on the data. The first analysis measured learners' *accuracy* with respect to English interrogatives; all utterances produced in a given experiment were included in this analysis. The second analysis measured learners' knowledge of *inversion;* in this analysis, only inverted and non-inverted questions were included.

1.5. Organization

This dissertation is organized as follows: in Chapter 2, I present a summary of the literature on the acquisition of main and embedded questions by L2 learners of English and investigate L2 learners' performance on English main and embedded questions via an elicited production study (Section 2.3.1.), a study of two written corpora (Section 2.3.2) and two acceptability judgment studies (untimed: Section 2.4.1; timed: Section 2.4.2).

In Chapter 3, I present a summary of the literature on the acquisition of main and embedded questions by L1 learners of English, investigate L1 learners' knowledge of English main and embedded questions via an elicited production study, and compare the results from the experiments with child directed speech data from six CHILDES corpora (MacWhinney, 2000).

Chapter 4 presents a summary of the experimental findings and suggestions for further research.

2. Second Language Acquisition of English Questions

2.1. Introduction

In this chapter, I will focus on the acquisition of English main and embedded questions by L2 learners of English. In English main questions, an auxiliary verb needs to precede the subject in main questions, while in embedded questions the relative order of the subject and the auxiliary is the same as in declarative sentences:

(3) What are you eating?

(4) Are you eating pizza?

(5) Mary doesn't know what he is eating.

(6) Mary doesn't know if he is eating pizza.

The phenomenon illustrated by sentences (3)–(4) is known in the early literature as subject-auxiliary inversion, and in the more recent generative literature as T-to-C movement. The first term refers to the fact that the relative surface order of the subject and the auxiliary is 'inverted' with respect to declarative clauses, while the second term refers to the fact that this surface order is thought to obtain from overt movement of a tense-bearing element from the head of the Tense Phrase (TP) to the head of the Complementizer Phrase, CP.

From the first days of modern psycholinguistics (e.g., Klima & Bellugi, 1966), it has been known that children learning English as their first language often use the wrong word order when producing a main question, failing to front the auxiliary, especially if a *wh*-word is present, as in (7).

(7) *What you are eating?

Difficulties with subject-auxiliary inversion in English have also been documented in adult second language learners (e.g., Spada & Lightbown, 1999).

Research on this topic is important because an understanding of why some types of syntactic errors, but not others, occur in language acquisition, and how these errors relate to the target input and the properties of individual words, should ultimately shed light on the nature of language-learning mechanisms. This is particularly true in the case of productions like (7), which are entirely absent in the input of the target language that learners are acquiring.

In this chapter, I investigate the extent to which adult second language learners' production and acceptability patterns of subject-auxiliary inversion in English interrogatives are affected by the syntax of L2 learners' native language, by the type of question they are producing (or judging), and the *wh*-word present in the interrogative sentence.

I also investigate the extent to which L2 learners' productions and acceptability judgments are consistent across experimental tasks and modalities. For example, it is conceivable that some error patterns are production-specific or only surface when task demands are high (e.g., when there are hard real-time constraints on learners' performance). In order to investigate whether inversion errors surface in learners' production both when they are under time pressure to produce speech and when they aren't, I examine the results of an elicited oral production task and a written production task. In order to examine whether second language learners' difficulties with English subject-auxiliary inversion only surface when learners are under real-time pressure, I examine adult learners' acceptability judgments of grammatical and ungrammatical main and embedded questions in two tasks: one in which learners are given a time limit to read and judge a sentence and one in which they do not have any time constraints.

Effects of grammatical properties on subject-auxiliary inversion: First-language transfer[5]

One of the main goals of the elicited production study was that of investigating the extent to which second language learners' subject-auxiliary inversion difficulties with English interrogatives can be attributed to properties of their first language (L1). While subject-auxiliary inversion errors have been documented in the L2 literature, the effect of L1 transfer has not been systematically investigated before.

With the exception of Full Access (without transfer) theory (Epstein, Flynn, & Martohardjono, 1996; Martohardjono & Flynn, 1995), the majority of current formal theories of second language acquisition assume initial transfer of L1 properties to L2 grammars. Theories tend to disagree, on the other hand, on the availability of Universal Grammar (UG) during second language acquisition. For example, Full Transfer/No Access theory (Bley-Vroman, 1990; Schachter, 1990) assumes no access to UG properties, while Full Transfer/Full Access (Schwartz, 1998; Schwartz & Sprouse, 1996) assumes L1 influence throughout L2 development.

[5]A modified version of the elicited production study with second language learners of English has been submitted for publication (see Pozzan & Quirk, submitted).

The L1 grammar is assumed to be the basis for the development of the L2 grammar, but the learner is thought to have access to UG when the L1 is insufficient for the learning task. According to the Minimal Trees Hypothesis (Vainikka & Young-Scholten, 1994, 1996) and its more recent development, Organic Grammar (Vainikka & Young-Scolten, 2005), the L1 grammar is available but contains no functional categories (e.g. TP). The emergence of functional categories does not depend on L1 transfer, but emerges in response to L2 input. This implies that the development of functional categories in learners with different L2s will be the same. A related proposal is that of Valueless Features (Eubank 1993, 1996), according to which the L1 grammar is the starting point of L2 development, but the strength of features, which is hypothesized to underlie phenomena such as T-to-C movement, from L1 is not available. Finally, Hawkins and Chan's (1997) Failed Functional Features hypothesis and, more recently, Tsimpli and Mastropavlou's (2007) Interpretability hypothesis maintain that features of functional categories (and specifically, uninterpretable features in Tsimpli and Mastropavlou's theory) are not accessible in adult second language acquisition. While an in-depth discussion of the different theories is well beyond the scope of the present discussion, it is important to keep in mind that while most theories recognize the role of L1 transfer at the initial stages of L2 acquisition (but see Hawkins, 2001, for the data indicating that L1 transfer is absent in the initial stages of L2 acquisition and that it only plays a role once functional categories are acquired), an area of contention is the extent and nature of L1 influence beyond initial stages of development (see Section 2.3.1.3. for further discussion of L2 theories with respect to the experimental findings).

Pozzan and Quirk (submitted) present an elicited production study focusing on learners whose L1s differed from English and from each other in terms of word order in both main and embedded questions, namely, Spanish and Chinese.

As discussed in Section 1.1., T-to-C movement and subsequent subject-verb inversion is always grammatical in Spanish main and embedded questions.[6] However, while inversion is obligatory with argument *wh*-words, it has been argued to be optional in yes/no questions and in adjunct *wh*-questions.[7] Crucially, in Spanish there is no asymmetry between main and embedded contexts. In Chinese, on the other hand, inversion is never an option, neither in main nor embedded contexts.

Spanish:

(8) Qué comió Maria?

‘What ate Maria?’

(9) *Qué Maria comió?

---

[6] Based on the fact that, in Spanish, it is not just the finite auxiliary verb that raises, but an auxiliary plus past participle, it has been proposed in the literature (e.g., Suñer, 1994) that Spanish may in fact not display T-to-C raising, but rather, fail to raise the subject to Spec, TP. For the purposes of this investigation, I will follow Torrego’s (1984) classic proposal that Spanish displays verb movement in main and embedded questions.

[7] Torrego (1984) argues that inversion is only obligatory with argument *wh*-words and not with adjunct *wh*-words. This judgment has been challenged in the literature, and a number of authors (Baauw, 1998; Rutten, 1995) have claimed that most native speakers of Peninsular Spanish consider inversion with adjuncts other than *why* to be obligatory.

‘What Maria ate?’

(10) Porqué lloró Maria?

‘Why cried Maria?’

(11) Porqué Maria lloró?

‘Why Maria cried?’

(12) Hector quiere saber qué comió Maria.

‘Hector wants to know what ate Maria.’

(13) */??Hector quiere saber qué Maria comió.

‘Hector wants to know what Maria ate.’

(14) Hector quiere saber porqué lloró Maria.

‘Hector wants to know why cried Maria.’

(15) Hector quiere saber porqué Maria lloró.

‘Hector wants to know why Maria cried.’

Chinese:

(16) Bīng xǐhuan shénme?

‘Bing like what?’

(17) *Xǐhuan Bīng shénme?

'Like Bing what?'

(18) *Shénme xǐhuan Bīng?

'What like Bing?'

(19) Jiéxī xiǎng zhidào Bīng xǐhuan shénme.

'Jessie would like know Bing like what.'

(20) *Jiéxī xiǎng zhidào xǐhuan Bīng shénme.

'Jessie would like know like Bing what.'

(21) *Jiéxī xiǎng zhidào shénme xǐhuan Bīng.

'Jessie would like know what like Bing.'

Three outcomes with respect to the performance of L1 Spanish and L1 Chinese learners on English main and embedded questions are logically possible:

(i) L1 Spanish learners invert more than L1 Chinese learners.

(ii) L1 Chinese learners invert more than L1 Spanish learners.

(iii) L1 Spanish and L1 Chinese learners invert at similar rates.

Outcome (i) is compatible with theories of SLA that assume full transfer of L1 properties throughout L2 development (e.g., Bley-Vroman, 1990; Schwartz & Sprouse, 1996), given that T-to-C movement (along with *wh*-movement) is always an option in Spanish, but never in Chinese. With respect to main questions, Outcome (i) is also compatible with views according to which similarities between L1 and L2 facilitate acquisition. Other views of transfer assume, on the contrary, that certain types of similarities can instead delay acquisition. That is, similarities between the L1 and the L2 might obscure for the learner that the L2 pattern is different from that of L1 and that the L2 pattern needs to be learned (Oller & Ziahosseiny, 1970; Ringbom, 1987). Additionally, it has been proposed that when the L1 and the L2 are extremely different in their properties, the L2 pattern might be easy to learn due to its saliency or the "surprise effect" (Kleinmann, 1977). Thus, given that the L2 input is, at a certain level of abstraction, compatible with the L1 Spanish parameter settings (i.e., T-to-C movement is always grammatical in Spanish), but not with the L1 Chinese parameter settings (i.e., T-to-C movement is never grammatical in Chinese), obligatory inversion in English main questions might be harder for L1 Spanish learners than L1 Chinese learners, giving rise to Outcome (ii). Finally, Outcome (iii) would be compatible with theories that do not assume L1 transfer to L2 in general (e.g., Epstein, Flynn, & Martohardjono, 1996) and with theories that do not assume L1 transfer of functional categories (Vainikka & Young-Scholten, 1994, 1996, 2005), or of feature strength (Eubank 1993, 1996).

For embedded questions, the same outcomes are possible, but with different implications for L1 transfer. Outcome (i), again, would be predicted by theories that assume L1 transfer in intermediate to advanced L2 learners, given that T-to-C movement is always grammatical in

Spanish embedded questions but not in Chinese. Outcome (iii) would again be compatible with views that do not assume L1 transfer to L2 in general (e.g., Epstein, Flynn, & Martohardjono, 1996) and with theories that do not assume L1 transfer of functional categories (Vainikka & Young-Scholten, 1994, 1996, 2005), or of feature strength (Eubank 1993, 1996). Specifically, Outcome (iii) could be explained by excess automatization accounts, where excess of automatization takes the form of over-application of L2 procedures from main to embedded contexts. Outcome (ii) does not seem straightforwardly compatible with any of the views described above.

Effect of grammatical properties on subject-auxiliary inversion: question-type and wh-type.

In order to investigate the extent to which L2 learners' difficulties with English interrogatives can be attributed to properties of the input language (L2) and to other subtle syntactic properties of individual words, speakers were prompted to produce and judge yes/no and *wh*-questions and, among the *wh*-questions, arguments and adjuncts. The reason for this was that previous research has suggested that these factors play a role in the L1 and L2 acquisition of English interrogatives. As Lee (2008) points out, such findings are important "because they enable us to better understand the mechanism of language acquisition. The similarities, if any, can be a piece of supporting evidence for a common mechanism operating in language acquisition that can be attributed to the properties of natural language itself" (Lee, 2008: 627). For example, L1 learners' inversion rates have been shown to be influenced by the type of question, i.e., yes/no vs. *wh*-questions (see Klima & Bellugi, 1966; Rowland 2007, among others, for the finding that inversion errors are more prevalent in *wh*-questions than yes/no questions; but see Erreich, 1984;

Valian, Lasser, & Mandelbaum, 1992 for the opposite tendency), and by the type of *wh*-word used (Rowland & Pine, 2000; Stromswold, 1990; Thornton, 2008; among many others). L2 learners' inversion rates have also been argued to be influenced by the type of question (see Eckman et al., 1989 for the finding that acquisition of yes/no questions implies acquisition of *wh*-questions; but see Pienemann, Johnston & Brindley, 1988 for the finding that yes/no questions are acquired earlier than *wh*-questions) and by the type of *wh*-word used (Lee, 2008).

With respect to the asymmetry between yes/no and *wh*-questions, it is important to notice that main yes/no questions without inversion are grammatical, albeit marked, in standard English. Specifically, according to the Cambridge Grammar of the English Language (2002: 881), non-inverted yes/no questions have an epistemic bias towards an answer "with the same propositional content as the question". That is, the expected answer to (22) is (23), while the expected answer to (24) is (25). The epistemic bias is confirmed by the impossibility of Negative Polarity Items (NPI) in positive non-inverted yes/no questions, as shown in (26) and (27):

(22) They've finished?

(23) Yes, they've finished.

(24) They haven't finished?

(25) No, they haven't finished.

(26) There's something/*anything else you need?

(27) You have (*ever) been to Paris?

It has been argued in the literature that the presence of non-inverted yes/no questions might be one of the factors causing learners to hypothesize that inversion is optional in English main clauses (see Valian, Lasser, & Mandelbaum, 1992; Tornyova & Valian, 2009). If this is the case, and learners are sensitive to this property of the input, we would expect them to produce and accept non-inverted structures in yes/no questions. On the other hand, it has been proposed in the literature (e.g., Klima & Bellugi, 1966; Gleitman, Newport and Gleitman, 1984) that non-inversion errors are due to other properties (e.g., difficulties with multiple movement operations/dependencies, reduced saliency of the auxiliary following the *wh*-word). If this latter hypothesis is on the right track, we would predict learners to produce and accept non-inverted *wh*-questions more often than non-inverted yes/no questions, given that they involve two types of dependencies (one between the *wh*-and its trace and one between the auxiliary in C^0 and its trace) and that the auxiliary is less salient than in yes/no questions, where it appears at the beginning of the clause and cannot be contracted.

With respect to the effect of question type in embedded questions, different theories make the similar prediction that production (and acceptance) of subject-auxiliary inversion should be more frequent in embedded *wh*-questions than in embedded yes/no questions. According to structural accounts, yes/no and *wh*-questions differ in that the head of the complementizer position is not filled by any overt element in embedded *wh*-questions, while in embedded yes/no questions, the complementizer position is filled by *if*. Given that the complementizer position is arguably the goal of the movement of the tensed auxiliary, this movement is blocked when C^0 is filled by *if*, making inversion in embedded yes/no questions impossible.[8] This structural

---

[8] In order for this proposal to work within more recent proposals on the syntax of the left periphery, where the CP system consists of multiple heads and projections (FocP, TopP,

difference has in turn been claimed to be responsible for the fact that, in non-standard varieties of English that allow inversion, inversion in embedded *wh*-questions is always possible, while it is only available in embedded yes/no when the complementizer *if* is absent (e.g., Labov, 1972). If applied to second language grammars, this proposal makes the straightforward prediction that subject-auxiliary inversion should only be produced and accepted in embedded *wh*-questions and in embedded yes/no questions without an overt complementizer.

Constructivist accounts of language acquisition (e.g., Tomasello, 2003), which posit that language learning is initially specific to individual words and their combination in the input, also predict a similar pattern, based on the simple co-occurrence of word combinations. The combination *wh*-word+auxiliary is frequent in the input to learners due to the fact that it is instantiated in main questions, and, hence, might result in *wh*-word+auxiliary productions/acceptance patterns in embedded contexts. The combination *if*+auxiliary, on the other hand, is never heard in the input and should not occur in learners' output.

The two theories can be distinguished in terms of the predictions they make with respect to embedded yes/no questions introduced by *whether*: while structural accounts might predict similar production and acceptability patterns for *whether* and *wh*-elements,[9] constructivist

---

Finiteness, Force, etc.), we need to assume that the complementizer *if* and the moved auxiliary still compete for the same position.

[9] *Whether* is thought to target the same position as other *wh*-elements, Spec, CP (Kayne, 1991). One reason for this is that, like other *wh*-elements and differently from *if*, it can select for infinitival clauses:

(ii) I wonder where PRO to go out tonight
(iii) I wonder whether PRO to go out tonight

accounts do not, because the combination *whether*+auxiliary is not present in the input. Another difference between structural accounts and constructivist accounts with respect to the production of embedded questions has to do with predictions for different *wh*-elements: according to constructivist accounts, it should follow that the more frequent a *wh*-word+auxiliary combination in main questions, the more likely its production/acceptance in embedded questions. On the other hand, according to structural accounts that assume that inversion in embedded questions is the result of rule overgeneralization/excess automatization, there should be a correlation between inversion rates for individual *wh*-words in main and embedded questions in learners' production rather than in the input. That is, the same *wh*-words that are associated with the production of high inversion rates in main questions should also display high inversion rates in embedded questions, above and beyond their frequency in the input.

Finally, constructivist accounts predict that subject-auxiliary inversion should be produced and accepted to a similar extent in free relatives (e.g., 'Tom likes what she is wearing', * 'Tom likes what is she wearing') and embedded *wh*-questions given the surface similarity of these structures. To the best of my knowledge, I am not aware of reports of this type of error in the literature or of studies that have specifically investigated the production of free relatives in acquisition.

(iv) *I wonder if PRO to go out tonight

The reason for this difference, in Government and Binding terms, is that the subject of infinitival clauses (PRO) needs to be ungoverned; the complementizer *if* in C^0 is a potential governor, giving rise to the ungrammaticality of (iv), while *wh*-elements in Spec,CP are not, and the sentences in (ii) and (iii) are grammatical.

Effects of task and modality on subject-auxiliary inversion

A central concern of any scientific investigation is to determine the generalizability of the research findings. In the case of the elicited production experiment, I wanted to determine the extent to which eventual findings of production errors with respect to subject-auxiliary inversion could be generalized to second language learners with different language backgrounds (L1 Chinese and L1 Spanish), different types of interrogatives (yes/no and *wh*-) and different *wh*-words.

A related question is whether these findings can be generalized to other tasks and modalities (e.g., acceptability judgments and written production). For example, a number of studies (Johnson, 1992; Haig, 1991; Murphy, 1997) have found an effect of input modality on grammaticality judgments, with accuracy on auditorily-presented materials being significantly lower (and reaction times slower) than for visually-presented materials, especially for ungrammatical sentences. Other studies have found that different performance tasks might load differently on implicit and explicit linguistic knowledge (e.g., Bialystok, 1982). For example, R. Ellis (2005) has shown that oral narration, oral imitation of sentences, and timed grammaticality judgments tap into implicit knowledge, while untimed grammaticality judgments measure explicit knowledge. In order to investigate whether difficulties with subject-auxiliary inversion characterize second language learners' production patterns beyond output modality, and whether the same grammatical properties that have been argued to affect the spoken production of subject-auxiliary inversion also affect learners' written production, I examined a corpus of L2 learners' written essays. Spoken and written productions are different because speech needs to flow at a ‘reasonable’ rate and, as such, it imposes strong real-time constraints on the planning and execution of the message. On the other hand, written production is arguably not subject to the same real-time constraints: writers can easily edit and modify (non-contiguous) elements of their production or start it all over again. The studies on L2 written production reported in Section 2.3.2. had two main objectives: to investigate whether the extent and the patterns of errors in the production of interrogative structures in L2 English are consistent across different tasks and output modalities, and to examine whether word order errors were present in speakers of a wider range of L1 language backgrounds than previously examined.

Moreover, I was interested in investigating whether the same patterns that have been argued to surface in production (i.e., effect of question type and *wh*-word) also have an effect on acceptability judgments. I wanted to investigate whether putting real-time constraints on the learner (e.g., giving them a limited amount of time to make a decision about a sentence) would affect acceptability judgment patterns. To this end, I constructed two experiments to examine L2 learners’ acceptability judgments of grammatical and ungrammatical English main and embedded questions. The acceptability judgment studies I conducted had two main objectives: to investigate whether the difficulty with subject-auxiliary inversion in L2 English is consistent across tasks that put different demands on the speakers, and to systematically investigate subject-auxiliary inversion in speakers with a wider range of first language backgrounds than previously examined.

To sum up, if learners’ difficulties with subject-auxiliary inversion were specific to production, we would expect them to surface only in oral and written production. On the other hand, if these difficulties were not specific to production, but stemmed from real-time processing constraints, we would not expect these errors to surface in writing, but to surface in oral production, and possibly in timed acceptability judgments. This finding would be compatible with proposals in the literature that time pressured tasks (e.g., speaking, timed acceptability judgments) require learners to rely on their implicit knowledge, while tasks without time constraints allow learners to rely on both implicit and explicit knowledge (R. Ellis, 2005; Bowes, 2011). Finally, if difficulties with the grammatical rules of English interrogatives are a general characteristic of learners' grammars, we expect similar patterns to emerge in learners' performance across tasks and modalities. In other words, the existence of converging performance data could be taken as a basis from which to draw inferences about linguistic competence (as suggested by Chaudron, 1983; Birdsong, 1989; Bley-Vroman & Masterson, 1989, among others), or as White put it, "linguistic competence is, of course, an abstraction […] There is no direct way to tap into competence, but various aspects of linguistic performance can give insights into competence" (White, 1989: 57–58).

This chapter is organized as follows: in Section 2.2., I review the literature on the L2 acquisition of main and embedded questions in some detail. In Section 2.3., I present the findings from the production studies and in Section 2.4., I present the findings from the two acceptability judgment studies. In Section 2.5., I discuss the findings and the implications of the production and the acceptability judgment studies for the nature of the grammatical representations of interrogative structures in L2 learners.

2.2. Previous studies of L2 Acquisition of English Questions

2.2.1. L2 Acquisition of Main questions

The acquisition of English main questions has received considerable attention in the L2 literature. In general, the literature on the L2 acquisition of English main questions has taken a qualitative, applied approach rather than a quantitative one. With few exceptions, this research has focused on establishing stages in the acquisition of questions (summarized below in Table 3) and on the methods of instruction, and providing feedback that is most effective at different stages.

Table 3: Stages in L2 acquisition of English questions (adapted from Dyson, 2008)[10]

Stage	Question type	Example	Description
1	Single words or fragments with rising intonation	Yes? What? A ball or a shoe?	Questions are formed by adding rising intonation to single words
2	Clauses with rising intonation	You like Chinese food?	Questions are formed by adding rising intonation to declarative sentences
3	*Wh*-fronting Do-fronting	What you want? Do you understand? Do the boy is beside the bus?	Questions are formed by adding a clause-initial interrogative constituent (do or *wh*-)
4	Yes-No inversion Copula inversion	Have you seen it? Where is my purse?	Subject-Auxiliary Inversion in yes/no questions is acquired, while only copula inversion is in place in *wh*-questions
5	Productive subject-auxiliary inversion.	Why did he leave? I wonder where should we go.	Subject-auxiliary inversion is fully productive and is extended to embedded questions.
6	Cancel inversion	I wonder where he is.	Subject-auxiliary inversion is confined to main questions

Little is known about effects of L1 transfer on English questions. Some initial evidence that L2 learners might transfer L1 properties to their production of English questions comes from Zobl's (1979, 1995) and Spada and Lightbown's (1999) findings with L1 French L2 English learners. These studies showed that non-inverted questions were more likely to be accepted and produced

[10] A similar pattern has been described for the acquisition of word order in German declarative clauses. German displays V2 in main clauses and SOV order in embedded clauses. A number of studies have investigated the acquisition of German word order (Klein & Dittmar, 1979; Clahsen, Meisel, & Pienemann, 1983; Clahsen & Muysken, 1986; Ellis, 1989) by speakers with different L1 backgrounds (Spanish, Italian, Portuguese, English, and Turkish). Clahsen and Muysken propose that L2 learners, irrespective of their L1, go through a number of ordered stages in their acquisition of German word order. Pienemann (1989) found that children learning German as a second language went through the same stages.

when the subject was a full NP than when it was a pronoun. Inverted questions, on the other hand, were more likely to be produced and considered grammatical when the subject was a pronoun. Given that (stylistic) inversion in French is possible only with pronouns, but not with full DPs, this result suggested that L2 learners apply properties of French questions to English. However, in the absence of a comparison group (e.g., L2 speakers whose L1 properties differed from French and English), this result cannot be unequivocally attributed to L1 transfer.

With respect to the effect of question type on inversion in main questions, the findings in the literature are inconclusive. According to some researchers (e.g., Pienemann & Johnston, 1987) the acquisition of word order in yes/no questions precedes that of *wh*-questions. This conclusion has been challenged by a study by Eckman, Moravscik, and Wirth (1989). The authors measured inversion rates in yes/no and *wh*-questions in a production study conducted with adult learners of English whose L1s were Korean, Japanese, or Turkish. They found that, in general, when inversion had been acquired in yes/no questions, it had also been acquired in *wh*-questions, but not vice versa (i.e., participants as a group inverted more in *wh*-questions than in yes/no questions). Their interpretation of these results was that inter-language grammars respect linguistic universals, given the existence of an implicational universal between inversion in yes/no questions and inversion in *wh*-questions. That is, if a language has verb-subject order in yes/no questions, it will also have it in *wh*-questions, but not vice versa.

On the other hand, other authors (e.g., White, Spada, Lightbown, & Ranta, 1990) have found no difference between yes/no and *wh*-questions. In an oral communication task that was part of a larger study with 53 L1 French 10–12-year-olds learning English, White et al. (1990) found that learners inverted in yes/no and *wh*-questions at a similar rate, reaching only 55% on average in the post-test phase. This suggests that in the acquisition of questions, target and non-target-like rules coexist, at least for some time: "cases of inversion and lack of inversion are found in the same subject" (White et al., 1991: 429).

A recent study by Youhanaee (2007) looked at Persian speakers' mastery of main and embedded questions in English. Persian does not display *wh*-movement or movement of a tensed element to C in either main or embedded questions. Eighty participants, divided evenly into four proficiency groups (elementary, lower intermediate, upper intermediate, and advanced) on the basis of a general proficiency test for English, were asked to translate 6 main yes/no questions and 15 main *wh*-questions from Persian to English. All proficiency groups were more accurate in their production of yes/no questions than *wh*-questions. The author only provides overall accuracy rates for the different groups, but from the error examples, it seems that the most frequent errors were overgeneralization of inversion to subject *wh*-questions, lack of inversion in non-subject *wh*-questions, and lack of *wh*-movement:

(28) What did cause the accident?

(29) When he played football?

(30) Did he bring flowers for whom?

This latter finding is fairly atypical; most studies have shown that when *wh*-words are included, they are always inverted (see Batmanian, Sayehli, & Valian, 2008; White et al., 1991; Kellerman, 1979; Eckman, Moravcsik, & Wirth, 1989; but see Bhatt & Hancin-Bhatt, 2002 for the opposite finding with L1 Hindi speakers).

The findings in the literature with respect to the effect of *wh*-words on inversion are very sparse. Haznedar (2003) studied the acquisition of inversion in the spontaneous production of a Turkish-speaking child learning English at age 4. She found inversion rates to be higher for *what*, *who* and *where* (173/184, 10/10, and 61/64, respectively) than for *which* (16/18), *how* (34/38) and *why* (29/32). With respect to the order of appearance of different *wh*-elements across different L1s and L2s, children learning English as a second language seem to acquire arguments *what* and *who* before adjuncts *why* and *when* (Felix, 1976; Lightbown, 1978; Park, 2000),[11] while the status of *where* and *how* is less clear.

More recently, Spada and Lightbown (1999) looked at inversion rates for different *wh*-words in a series of tasks: an oral production task, a scrambled questions task and a preference task. In the scrambled questions task, students were presented with 20 cartoons and were asked to rearrange word cards to produce appropriate main questions. Each student was presented with instances of *what* (1x), *where* (1x), *when* (3x) and *how* (1x). Overall, inversion rates were similar for *what* and *where* (40% and 37% inversion respectively at pretest and 57% at posttest for both) and lower for *when* and *how* (20–27% and 13% inversion at pretest and 27–43% and 23% inversion for post-test, respectively). In the preference task, students had to judge pairs of inverted and non-inverted main questions. Each student was presented with instances of *what* (2x), *where* (3x), *when* (2x) and *why* (3x) questions. At both pre- and post-test, participants accepted ungrammatical–non-inverted *what* and particularly *why* questions more often than

---

[11] Felix's study looked at L1-English speaking children learning German, while Lightbown (1978) investigated L1-English-speaking children learning French, and Park's (2000) study examined production in L1-Korean children learning English.

grammatical ones, while they did not seem to show a clear preference in the case of *where* and *when* questions.

To overcome some of the limitations of previous studies (e.g., few experimental items, child and adolescent participants), Lee (2008) conducted a grammaticality judgment study with 41 adult L1 Korean L2 learners of English to see whether, in line with findings for children learning English as their L1, English second language learners' inversion rates would be influenced by the argument-adjunct status of *wh*-words. The learners' proficiency was not assessed independently, but was 'assumed to be high' (Lee, 2008: 642). Each student was presented with 32 experimental items, 16 argument main questions (8 *what*, 8 *who*) and 16 adjunct main questions (8 *how*, 8 *why*), and was asked to rate them on a scale from −2 to +2. Lee found a significant effect of inversion (with inverted questions being rated higher than non-inverted questions), a marginal effect of language group (learner vs. native English speakers) and a group by inversion interaction (with native speakers' judgments being sharper than those of learners), but no main effect of *wh*-word. No argument-adjunct asymmetry was found for inverted questions, while an asymmetry for the learners was found in the non-inverted questions; specifically, non-inverted adjunct questions were considered significantly less degraded than non-inverted argument questions.

The results from the literature of the acquisition of main questions in L2 English have mainly focused on developmental, implicational stages in children and adolescents. This approach, by its own nature, focuses on the emergence of a construction (normally taking one or two productions for evidence of emergence), rather than on its mastery. Moreover, most studies of the acquisition of questions have focused on how children or adolescents learn English as a second language, and not on how adults do. While there might be similarities between child and adult second language acquisition, this cannot be taken for granted and is the topic of current debates in the field (e.g., Song & Schwartz, 2009).

Additionally, most studies in the literature did not control for the effect of the first languages of the participants in the studies, which makes it difficult to assess whether the error patterns that emerge are general characteristics of second language acquisition, or only of a subset of learners with L1 backgrounds dissimilar to English in this respect.

As a consequence, little is known at present about the extent to which adult L2 learners of English master subject-auxiliary inversion in main and embedded questions and what factors affect their production. The study reported in Section 2.3.1. is aimed at filling this gap in the literature by manipulating these factors experimentally.

2.2.2. L2 Acquisition of Embedded questions

Very little is known about the acquisition of embedded questions in second language learners of English. Being "a fairly good index of overall structural sophistication" (Johnston, 1985: 245), such structures are fairly uncommon in spontaneous production and are arguably rare in the input learners receive. Given the low frequency of these structures in the input and their difference in terms of subject-auxiliary inversion from more frequent main questions, embedded questions allow us to see to what extent learners overgeneralize subject-auxiliary inversion from main to embedded contexts or conservatively refrain from applying movement unless presented with direct positive evidence for it. As mentioned above, it has been suggested that second language learners of English with a variety of L1 backgrounds produce non-target subject-auxiliary inversion in embedded *wh*-questions. However, the data available in the literature is anecdotal in nature and no quantification of the phenomenon exists. Bley-Vroman (1997) notes that "Hebrew-speaking learners of English have been observed to apply subject-verb inversion (with *do*-support) in English embedded *wh*-interrogatives even though English doesn't do this and Hebrew, though it does invert in embedded structures, doesn't have auxiliaries" (p. 6). Escutia (2002) notes that "teachers of English as a Foreign Language (EFL) in Spain have firsthand experience with (peninsular) Spanish speaking students' tendency to produce and accept inverted indirect embedded questions" (Escutia, 2002:1). He cites two such written examples from two high-intermediate learners:

(31) She asked why hadn't he arrived yet.

(32) She wanted to know where did Helen live.

In the context of a discussion of grammatical fossilization in L2 acquisition, Finnegan (1999) cites one such example from a Mandarin-speaking learner of English:

(33) I want to see what can I buy.

A more in-depth study of embedded questions in L2 English was conducted by Johnston (1985) and also reported in Pienemann (2005). Johnston (1985) found that 4 out of 7 L1 Polish and 3 out of 8 L1 Vietnamese learners of English in his study produced both inverted and non-inverted embedded questions.[12] One Polish and one Vietnamese speaker produced only inverted embedded questions and one Vietnamese speaker produced only target embedded questions, while two Polish and two Vietnamese speakers did not produce any embedded questions that could reveal inversion or non-inversion patterns (subject questions, infinitival questions, etc.). Unfortunately, in this report, only the total number of embedded questions produced by each speaker (range 1–14) and whether or not that speaker produced inversion is reported, but inversion in the relevant productions is not quantified.

Pienemann (1998) reports a study conducted by Mackey, Pienemann and Doughty (1992) in which six L2 learners completed a series of tasks aimed at eliciting 3rd person singular morphology, interrogative and negative structures. Six L2 learners (age range 19–25) whose L1 was either Indonesian (n = 4), Korean (n = 1), or Chinese (n = 1) participated in this study. One participant did not produce any embedded question, while the other five participants produced a total of 13 embedded questions overall, with no participant producing more than 3. Inversion rates ranged from 0% (0/1) to 67% (2/3).

Youhanaee (2007) presented 80 L1 Persian L2 English speakers with 10 embedded yes/no and 6 embedded *wh*-questions to translate from Persian into English. The elementary and lower intermediate groups did significantly worse than the upper intermediate and advanced groups (see Figure 1 below). The author does not report the rates of embedded inversion errors

---

[12] In the brief section on embedded questions, the author does not distinguish between yes/no and *wh*-questions and no relevant examples are provided. As a result, we do not know whether the corpus contained both types and whether inversion occurred in both.

produced by the participants, but inversion errors and co-presence of a *wh*-element after a complementizer (or both) are the only error types reported by the author:

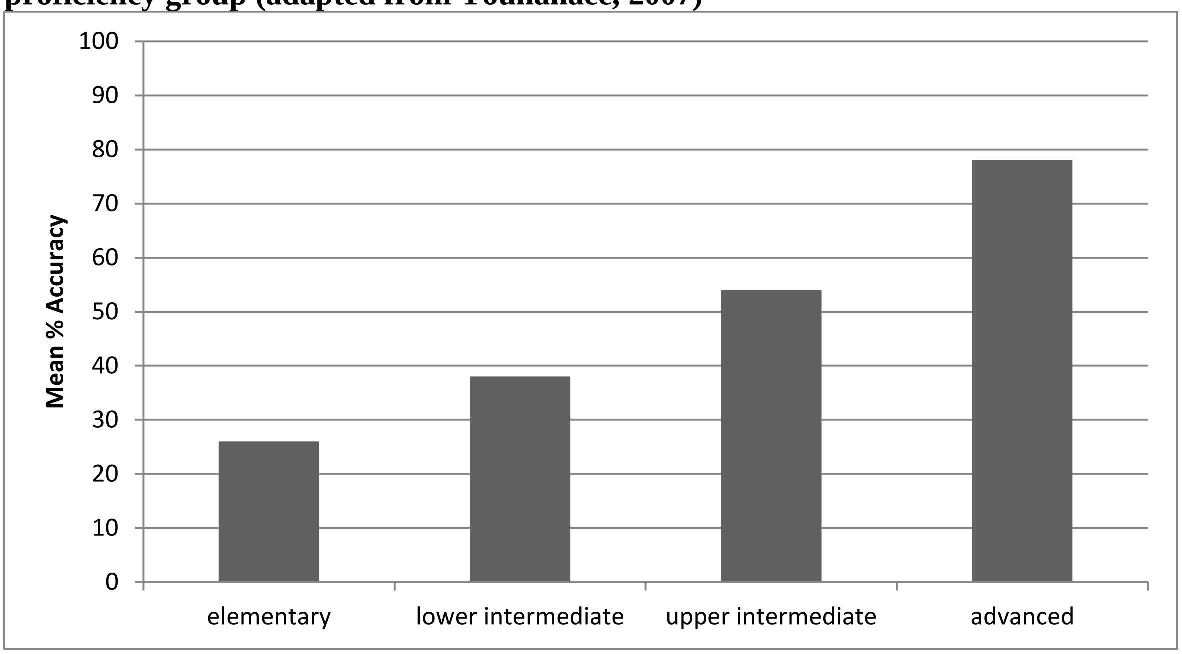
(34) I don't know that whether he finished his work.

(35) I'm not sure when will the film start.

(36) I doubt how old is he.

(37) I don't remember that whose book did he give to Ali.

Figure 1: L1 Persian learners' average accuracy in embedded English questions by proficiency group (adapted from Youhanaee, 2007)



2.3. Present Study: L2 Production of English Questions

2.3.1. Elicited Oral production

2.3.1.1. Experiment 1: Main questions

The goal of this experiment was to investigate the production of subject-auxiliary inversion in English main questions by adult second language learners. Specifically, we examined whether intermediate/advanced L2 learners differed from native speakers in their mastery of subject-auxiliary inversion, whether there was an effect of L1 properties on subject-auxiliary inversion, and whether L2 learners' inversion rates were influenced by linguistic factors that have been argued to influence L1 learners' inversion rates, such as question type (*wh*- vs. yes/no) and *wh*-word type (e.g., argument vs. adjunct).

2.3.1.1.1. Method

2.3.1.1.1.1. Participants

One of the aims of the present study was to examine the role of L1 in the acquisition of English interrogatives. As such, participants included native speakers of English and L2 learners of English whose L1 was either Chinese or Spanish—languages that differ from one another and from English with respect to the relative word order of the subject and the tensed verb in main questions.

Participants were recruited through online and paper advertising, personal contacts, and the Introduction to Psychology subject pool at Queens College, which allowed for language background screening. Participants either received course credit or were compensated $10 for their participation.[13]

A total of 88 participants were tested. Six participants were excluded because of data loss and two were excluded because they arrived in the U.S. around age six and considered themselves dominant in English. Data from a total of 80 participants (16 English native speakers, 32 L1 Chinese speakers, and 32 L1 Spanish speakers) underwent further analyses.

L2 participants were judged to be intermediate/advanced with respect to their English proficiency. Proficiency was assessed through a portion of the Michigan Test of English Language Proficiency (MTELP) designed to assess listening proficiency. Table 2 provides a summary of participants' proficiency scores, age, age at arrival, and years of stay in an English-speaking country.

Table 4: Demographics of L2 Participants – Oral Elicited Production Experiment

Demographics	**L1**			
	Spanish (N = 32)		**Chinese (N = 32)**	
MTELP (SD)	39.5/45 (5.2)		37.2/45 (4)	
Age (SD)	28.6	(9.7)	25.7	(5.9)
Years of stay in US (SD)	7.4	(8.1)	3.1	(3.9)
Age of arrival in US (SD)	21.2	(8)	22.6	(6.8)

Spanish and Chinese participants did not differ in terms of age and age of arrival (*t* (62) = 1.4, n.s., and *t* (62) = .76, n.s., respectively), but they differed in terms of years of stay in the US (*t* (62) = 2.7, p = .008) and English proficiency (*t* (62) = 1.9, p = .05), as measured by the MTELP.

[13] This research was supported by a CUNY Doctoral Research Grant to Lucia Pozzan and Erin Quirk.

On average, Spanish speakers had spent more years in the U.S. and scored higher on the MTELP.

2.3.1.1.1.2. Materials

Experiment 1 consisted of a computerized elicitation task, administered through the E-prime software 1.0 (Psychology Software Tools, Inc.). Participants were told that they needed to help a shy student ask his teacher some questions. In items intended to elicit yes/no questions, a prompt such as “Maybe Gloria called Jim. Ask Miss Brainy” appeared written in a speech bubble on the screen. Participants were instructed to advance to the next screen, which showed a teacher in front of a blackboard, and to produce a question aloud. They were not able to return to the previous screen to revisit the prompt. In the final screen, the teacher provided a written answer to the participant’s question. Items eliciting *wh*-questions followed the same general format, differing only in the prompt (e.g., “Gloria called Jim. Ask Miss Brainy why”.). Screenshot examples of slides aimed at eliciting yes/no and *wh*-questions are provided in Appendix A.

Four experimental lists were constructed, each containing six practice items (three yes/no and three *wh*-), and 32 experimental items.[14] Each participant was assigned to one of the four experimental lists. The two fully *within* factors were question type (yes/no vs. *wh*-) and presence of auxiliary in the prompt (auxiliary vs. lexical verbs in the prompt). The two auxiliaries used in the prompt were *be* and *have*, while prompts with lexical verbs required the insertion of do-

[14] We decided not to insert any fillers due to time constraints. In order to minimize subject attrition noticed during pilot testing, we decided to administer Experiment 1, Experiment 2, the MTELP, and a language questionnaire in one single experimental session, which lasted, on average, a little over one hour.

support by the participant. The type of *wh*-word in *wh*-questions was also manipulated. There were four types of *wh*-words: arguments *what* and *who* and adjuncts *why* and *where*.[15] Type of *wh*-word was a within-subjects factor but a between-items factor. Each participant was presented with a total of 32 main questions as follows:

- 16 yes/no:
    - 8 with lexical verbs
    - 8 with auxiliary verbs
        - 4 with *be*
        - 4 with *have*

- 16 *wh*-questions
    - 8 with lexical verbs
        - 2 *who*, 2 *what*, 2 *where*, 2 *why*
    - 8 with auxiliary verbs
        - 4 with *be*
            - 1 *who*, 1 *what*, 1 *where*, 1 *why*
        - 4 with *have*
            - 1 *who*, 1 *what*, 1 *where*, 1 *why*

[15] While *where* can function either as an argument (e.g., Where did you put the book?) or an adjunct (Where did you have dinner?) depending on the verb, it functioned as an adjunct in our experimental materials.

Experimental materials for the four experimental lists are given in Appendix B.

Each sentence frame (item) occurred in the four conditions across experimental lists (in a prompt aimed at eliciting a yes/no question with a lexical verb, a prompt aimed at eliciting a yes/no question with an auxiliary, a prompt aimed at eliciting a *wh*-question with a lexical verb, or a prompt aimed at eliciting a *wh*-question with an auxiliary). Half of the verbs used in the experimental items appeared in past tense, while the other half were in present tense. The experimental items were pseudo-randomized so that no more than two consecutive experimental sentences shared any of the features relevant to the investigation (i.e., yes/no, *wh*-type). Finally, to control for order effects, four additional lists in which the order of the experimental items was reversed were created.

2.3.1.1.1.3. Procedure

After reading and signing the consent forms, participants were seated in front of a computer. The experimenter guided the participant through the six practice items and answered questions, but did not provide corrective feedback. After the practice session, the experimenter turned on a digital recorder and left the room. Experiment 1 took approximately twenty minutes.

2.3.1.1.1.4. Transcription and coding

Participants' responses were transcribed by the authors. A subset (30%) of the responses was transcribed independently by both transcribers, and no major disagreements were found. Coding was performed independently. Inter-coder agreement was high (> 96%) and disagreements were resolved by discussion.

Each production was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology, and presence of target lexical items (e.g., subject and *wh*-element).[16] Following Ambridge et al.'s (2006) coding scheme, incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors, as in (38), or raising errors as in (39).
- Double tense/double auxiliary errors, as in (40)–(41).
- Omitted auxiliary errors, as in (42), or errors that, due to lack of morphology, were ambiguous between non-inversion and omitted auxiliary errors, as in (43).
- Other errors. Other errors included questions that differed in type from the target (yes/no instead of *wh*-question, and vice versa); subject *wh*-questions instead of object *wh*-questions, as in (44); productions that differed from the target in the lexical items used, questions without a subject, questions with VP movement, as in (45); questions where the *wh*-word differed from the target one, as in (46); and questions where the *wh*-element was not clause-initial, as in (47), or where it was doubled by a dummy element in the sentence, as in (48).

(38) Why your brother has fired Mark?

(39) Why your brother fired Mark?

(40) Who did the boss complimented?

(41) Is Laura is visiting Bill in London?

(42) Why your husband walking to work?

(43) Why you call Jim?

(44) Who is seeing Mark for brunch? (Target: Who is Mark seeing for brunch?)

(45) Where is going Julia on vacation?

(46) Where is the cat hiding? (Target: Why is the cat hiding?)

---

[16] Lack of inversion in main yes/no questions was coded as incorrect for ease of comparison with *wh*-questions.

(47) The babysitter is driving who to school?

(48) What does Mary carry something in her bag?

Only on a handful of occasions did learners produce more than one response for a trial. On these occasions, the second utterance was judged as an attempt to correct the first response; only the second utterance was scored.

2.3.1.1.2. Results and interim discussion

The main goal of Experiment 1 was to examine the relative contribution of L1, question type, and *wh*-type to the production of subject-auxiliary inversion errors in main English questions by adult L2 learners of English. Two sets of analyses were performed: the first set of analyses used arcsine transformed mean percent *correct* productions as the dependent variable. This was calculated by dividing the number of correct responses by the *total* number of productions. The second set of analyses used arcsine transformed mean percent *inversion* as the dependent variable, and this was calculated by dividing the number of correct responses by inverted and non-inverted responses only; productions that provided no evidence one way or another with respect to inversion or that contained errors unrelated to subject-auxiliary inversion (i.e., auxiliary omission, different structure, morphological errors, etc.) were thus excluded from this second set of analyses.

2.3.1.1.2.1. First language

Table 5 reports the raw number (and percentage) of productions in each coding category for L1 Chinese and L1 Spanish L2 learners of English. English native speakers produced 100% target inverted main questions and were thus excluded from further analyses.

Table 5: L2 learners' oral production of main questions by coding category and L1

Coding	**L1**	
	Chinese	**Spanish**
Correct	868 (84.8%)	737 (71.9%)
Non-inverted	29 (2.8%)	78 (7.6%)
Double aux/tense	16 (1.6%)	43 (4.2%)
No auxiliary	10 (1%)	20 (2%)
Other errors	101 (9.9%)	146 (14.3%)

L1 Chinese speakers produced a significantly higher number of correct responses than L1 Spanish speakers in their production of main questions (t_1 (62) = 2, p = .047; t_2 (31) = 5.5, p < .0001). Moreover, L1 Chinese speakers produced significantly fewer inversion errors than Spanish speakers (t_1 (62) = 2.4, p = .02; t_2 (31) = 4.3, p < .0001). The item analysis also showed a significant difference between L1 Chinese and L1 Spanish in terms of auxiliary omission errors, (t_1 (62) = 1.5, p = .14; t_2 (31) = 2.5, p = .02), double tense errors (t_1 (62) = 1.9, p = .06; t_2 (31) = 3.1, p = .004), and other errors (t_1 (62) =1.3, p = .2, t_2 (31) = 2.7, p = .009).

Overall, L1 Chinese speakers produced fewer errors than L1 Spanish speakers. Once productions that contained errors unrelated to subject-auxiliary inversion (i.e., auxiliary omission, double tense, and other errors) were removed from the analyses, average inversion for L1 Chinese speakers was 96.8% and 90.4% for L1 Spanish speakers. This difference was statistically significant (t_1 (62) = 2.6, p = .01; t_2 (31) = 4.5, p < .0001).

Overall, L1 Chinese speakers produced fewer errors and inverted more often than L1 Spanish speakers. This difference is not likely to be due to differences in proficiency, age of acquisition (AoA), or length of exposure to English, given that L1 Spanish speakers' proficiency scores (as measured by the MTELP) and length of stay in the US were higher than those of L1 Chinese speakers, and the two groups did not differ in terms of AoA.

2.3.1.1.2.2. Question-type

Table 6 reports the the raw number (and percentage) of productions in each coding category by question type, while Table 7 and Table 8 the raw number (and percentage) of productions in each category by question type for the L1 Spanish and L1 Chinese groups, respectively. Figure 2 presents a summary of productions by coding type for the two groups of learners in *wh-* and yes/no questions.

Table 6: L2 learners' oral production of main questions by coding category and question-type

Coding	**Question Type**	
	Yes/No	***Wh-***
Correct	872 (85.2%)	733 (71.6%)
Non-inverted	14 (1.4%)	93 (9.4%)
Double aux/tense	39 (3.8%)	20 (2%)
No auxiliary	5 (0.6%)	25 (2.4%)
Other errors	94 (9.1%)	153 (14.9%)

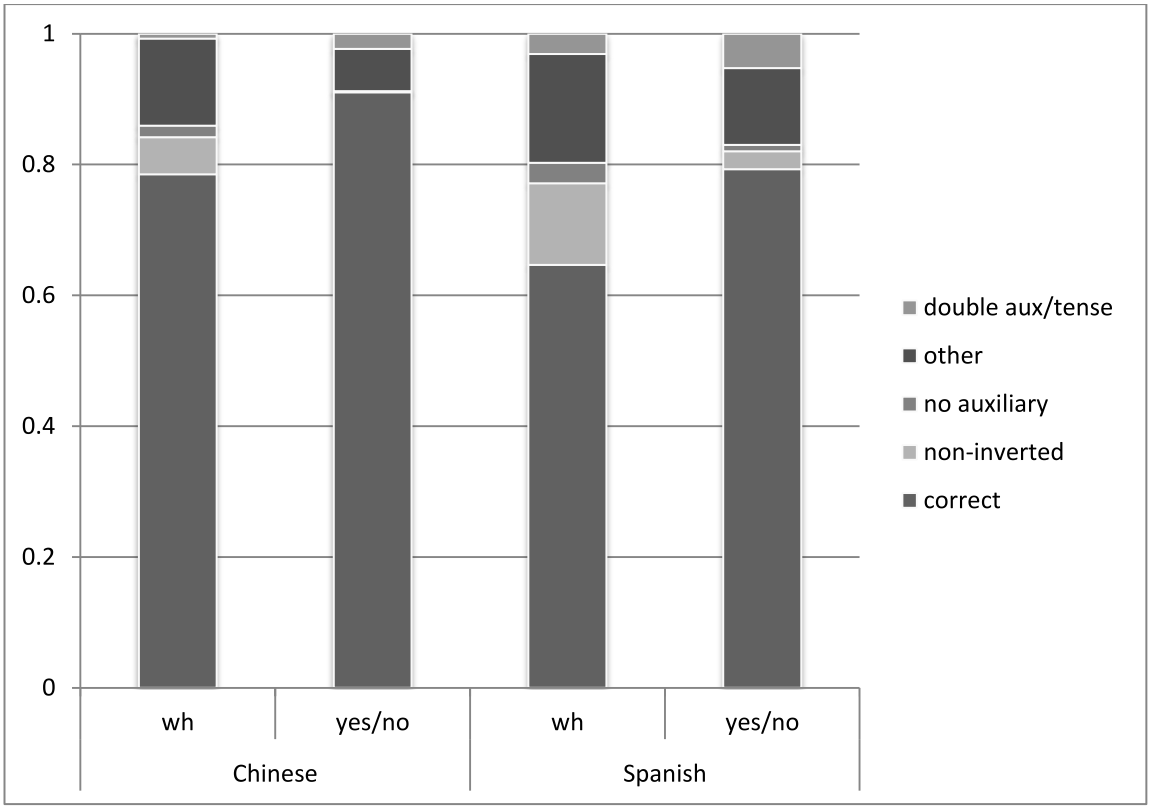
Table 7: L1 Spanish learners' oral production of main questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh*-**
Correct	406 (79.3%)	331 (64.6%)
Non-inverted	14 (2.7%)	64 (12.5%)
Double aux/tense	27 (5.3%)	16 (3.1%)
No auxiliary	4 (1%)	16 (3.1%)
Other errors	61 (11.7%)	85 (16.6%)

Table 8: L1 Chinese learners' oral production of main questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh*-**
Correct	466 (91%)	402 (78.5%)
Non-inverted	0	29 (5.7%)
Double aux/tense	12 (2.3%)	4 (0.8%)
No auxiliary	1 (0.2%)	9 (1.8%)
Other errors	33 (6.4%)	68 (13.3%)

Figure 2: L2 learners' oral production of main questions by coding category, question-type and L1



A 2 (question type) x 2 (L1) mixed design ANOVA using arcsine transformed percent *correct* as the dependent variable showed a significant effect of question type (F_1 (1,62) = 39.6, p < .0001; F_2 (1,31) = 20.4, p < .0001) and L1 (F_1 (1,62) = 5.9, p = .02; F_2 (1,31) = 30.3, p < .0001). There was no interaction between question type and L1 (all Fs <1).

A second 2 (question type) x 2 (L1) mixed design ANOVA using arcsine transformed percent *inversion* as the dependent variable showed a significant effect of question type F_1 (1,62) = 36.2, p < .0001; F_2 (1,31) = 32.5, p < .0001) and L1 (F_1 (1,62) = 4.6, p = .04; F_2 (1,31) = 25.2, p < .0001) and no interaction (F_1 < 1; F_2 (1,31) = 2.2, p = .15).

On average, yes/no questions were associated with higher rates of correct and inverted productions than *wh*-questions, and L1 Chinese speakers produced, on average, higher rates of correct and inverted questions than L1 Spanish speakers.

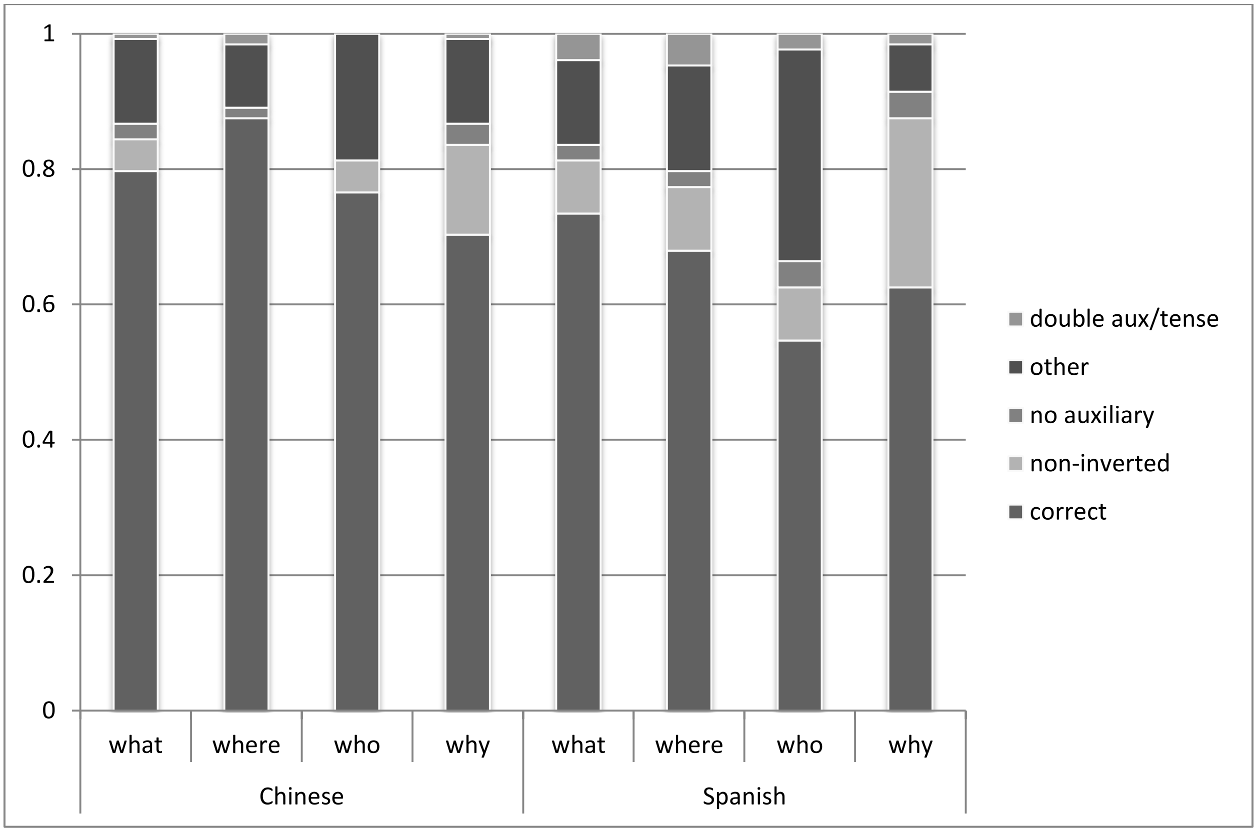
2.3.1.1.2.3. Wh-type

Table 9 reports the raw number of inverted responses along with percentage of correct and inverted responses by *wh*-word by L1, and Figure 3 presents a summary of responses by coding type, language group and *wh*-type.

Table 9: L2 learners' oral production of target main questions by *wh*-word and L1

***Wh*-Type**	**L1**					
	Chinese			**Spanish**		
	# correct	**% correct**	**% inverted**	**# correct**	**% correct**	**% inverted**
what	102	79.7%	94.4%	94	73.4%	90.4%
where	112	87.5%	100%	87	68%	87.8%
who	98	76.6%	94.2%	70	54.7%	87.5%
why	90	70.3%	84.1%	80	62.5%	71.4%

Figure 3: L2 learners' oral production of main questions by coding category, *wh*-type and L1



A 4 (*wh*-type) x 2 (L1) mixed design ANOVA with arcsine transformed percent correct as the dependent variable was performed. In the omnibus ANOVA, the effect of *wh*-type was significant (F_1 (3,186) = 4.1, p = .008; F_2 (3,28) = 3.8, p = .02). The effect of L1 was marginally significant in the subject analysis, but significant in the item analysis (F_1 (1,62) = 3.6, p = .06; F_2 (1,28) = 22, p < .0001). The interaction between *wh*-type and L1 was marginally significant in the subject analysis and not significant in the item analysis (F_1 (3,186) = 2.5, p = .06; F_2 (3,28) = 1.7, p = .2).

A second 4 (*wh*-type) x 2 (L1) mixed design ANOVA with arcsine transformed percent inversion as the dependent variable was performed. In the omnibus ANOVA, the effect of *wh*-type was significant (F_1 (3,168) = 12.2, p < .0001; F_2 (3,28) = 14.6, p = .001). The effect of L1 was also significant (F_1 (1,62) = 5.6, p = .02; F_2 (1,28) = 9.6, p < .0001). No interaction with L1 was found (all $F_1 < 1$). Pairwise comparisons were carried out to explore the effect of individual *wh*-elements on correct and inverted responses. Bonferroni correction was applied to prevent Type I error inflation. The first set of analyses (performed on percent correct) revealed a marginally significant difference between *what* and *who* (p = .07) and *where* and *who* (p = .06) both in the subject and item analyses, with *who* being associated with fewer correct responses.

The second set of analyses (performed on percent inversion), showed a significant difference between *why* and all other *wh*-elements (subject analysis: *what* vs. *why* (p < .0001), *where* vs. *why* (p < .0001), *who* vs. *why* (p = .04); item analysis: *what* vs. *why* (p = .004), *where* vs. *why*, (p < .0001), *who* vs. *why* (p = .001)), with *why* being associated with lower inversion rates. This is the only case in which the correct and inverted analysis revealed a different pattern of responses. The reason for this difference is likely to be due to the high number of other responses with *who*[17] (18.8% for Chinese and 31.3% for Spanish) and the relative low number of

[17] Most "other error" responses with *who* were subject questions in place of non-subject questions (e.g., *Who saw Sarah for brunch?* instead of the target *Who did Sarah see for brunch?).*This pattern was also noticed in the pilot elicited production study with English-speaking children, and, for this reason, non-subject *who* prompts were removed and *which* questions were elicited in the latter experiment. It is possible that this is due to the fact that *who* is the only *wh*-element that is exclusively used for human referents and that learners have a preference to assign an agent/subject role to the [+animate +human] sentence-initial *wh*-word. This pattern would follow the canonical sentence bias found in L1 acquisition of English (Bever,

other responses with *why* (12.5% for Chinese and 7% for Spanish). Once other responses were removed, the difference between *who* and other *wh*-words disappeared while the difference between *why* and the other *wh*-elements became significant.

Experiment 1 was aimed at investigating the effects of L1, question type, and *wh*-type on the production of main questions in L2 learners of English.

As for the effect of L1 (Chinese vs. Spanish), we found that L1 Chinese speakers produced overall higher rates of correct and inverted main questions than L1 Spanish speakers, and both groups were significantly less accurate than native-speaker controls. As noted before, this result cannot be explained by differences in English proficiency, given that, on average, L1 Spanish participants scored higher than L1 Chinese participants on the MTELP and had lived in an English-speaking country for a longer period of time. This difference is also unlikely to be due to dialectal differences among Spanish speakers for two reasons. First, only four L1 Spanish speakers were native speakers of Caribbean Spanish varieties in which subject-verb inversion in questions is disfavored (Toribio, 2000; Suñer, 1994), and three of them never produced inversion errors. Second, although inversion is dispreferred in these dialects, it is nonetheless grammatical, while in Chinese, it is never grammatical.

The direction of the L1 effect is unexpected under a simple transfer account of the acquisition of subject-auxiliary inversion in L2 English given that Chinese lacks inversion in questions altogether, while Spanish allows subject-verb inversion in all questions. One way of

---

1970), according to which children have the tendency to interpret sentences with surface order NVN as agent-action-object sequences.

accounting for this pattern, as mentioned in the introduction, is to hypothesize that properties that are present in learners' L1s but require restructuring (e.g., English inversion patterns for L1 Spanish) are more difficult to acquire than those that are entirely absent in the L1 (e.g., English inversion patterns for L1 Chinese).

As for the effect of question type (yes/no vs. *wh*-questions), we found that learners produced, on average, higher rates of correct inverted structures in yes/no than in *wh*-questions. This result is in line with the findings from the literature on acquisition stages (Spada & Lightbown, 1999; Pienemann & Johnston, 1987). On the other hand, this result contrasts with the findings reported by Eckman et al. (1989), and might be surprising considering that non-inverted yes/no questions (e.g., *She likes fish?)* are present in the native Standard English input, while non-inverted *wh*-questions aren't (e.g.,**What she likes?*). Given that native English controls never produced non-inverted yes/no questions in this experiment, this result suggests that learners might know the conditions under which non-inverted yes/no questions are appropriate in English. Eckman et al. (1989) proposed that second language acquisition follows linguistic universals. The relevant universal states that inversion in yes/no questions is less common than inversion in *wh*-questions, so if a language displays inversion in yes/no questions, it will also display it in *wh*-questions. In acquisition terms, the prediction is that if a learner has acquired inversion in yes/no questions, she will also have acquired inversion in *wh*-questions.

Eckman and colleagues used two criteria for acquisition: 90% or 80% inversion rates. To best compare the present results with the ones from Eckman et al.'s study, Appendix C reports average *inversion* rates for each L2 learner and indicates whether the participant has acquired inversion in yes/no and *wh*-questions by those criteria (by means of a "+" sign). Whenever a participant has acquired inversion in *wh*-questions, he or she has also acquired it in yes/no questions. There are no exceptions to this generalization when using the 80% criterion, while there is one exception if we use the 90% accuracy threshold (participant 37). Conversely, there are 13 and 19 participants that have not acquired *wh*-questions but have acquired yes/no questions according to the 80% and 90% inversion criteria, respectively. The implicational relationship in these data exhibits the opposite pattern than the one reported in Eckman et al. (1989).

With respect to the effect of *wh*-type, we did not find an argument-adjunct asymmetry in production. The adjunct *where* was the element with the highest rates of inversion, while the adjunct *why* was the element with the lowest rates of inversion. The present result follows a general tendency: low inversion rates with *why* have been consistently documented in L1 acquisition of English (Labov & Labov, 1978; Berk, 2003; de Villiers, 1991; Rowland & Pine, 2000; Thornton, 2008), while findings of argument-adjunct asymmetries have failed to be replicated by a number of studies in the acquisition literature (e.g., Kuczaj & Brannick, 1979; Ambridge et al. 2006; Valian, Lasser, & Mandelbaum, 1992). This result, on the other hand, contrasts with the results of a grammaticality judgment task conducted by Lee (2008) with a group of L1 Korean L2 learners of English. In this study, Lee found an argument-adjunct asymmetry in ungrammatical non-inverted main questions; ungrammatical non-inverted adjunct questions were judged as more acceptable than non-inverted argument questions. This difference might be due to the different tasks, or to the fact that the two studies used different adjunct *wh*-words: Lee used *why* and *how*, while the present experiment used *why* and *where*.

2.3.1.2. Experiment 2: Embedded questions

In this experiment, we investigated the mastery of English embedded questions by adult second language learners. Specifically, we examined whether advanced L2 learners produced subject-auxiliary inversion in embedded questions. The main issue we were interested in investigating was whether subject-auxiliary inversion errors in embedded questions could be attributed to transfer of L1 properties onto subject-auxiliary inversion or to overgeneralization of inversion patterns from main to embedded contexts.

We hypothesized that if subject-auxiliary inversion in embedded contexts was due to L1 transfer, only L1 Spanish speakers would produce the errors. On the other hand, if these errors were due to overgeneralization of inversion from main to embedded contexts, both groups would produce them to a certain extent. We also hypothesized that, if subject-auxiliary inversion errors in embedded contexts were due to overgeneralization of inversion from main to embedded structures, L2 learners' inversion rates would also be influenced by *wh*-word type, being high for elements that invert consistently in main questions (i.e., *where*, *what*) and low for elements that do not invert consistently in main questions (i.e., *why*). A similar protocol and materials were used in Experiments 1 and 2 in order to examine whether inversion rates in main questions correlated with inversion rates in embedded questions.

2.3.1.2.1. Method

2.3.1.2.1.1. Participants

The same participants from Experiment 1 participated in Experiment 2. The two experiments were administered during the same session. Experiment 2 was always administered after Experiment 1, given that main questions were used to elicit embedded questions.

2.3.1.2.1.2. Materials

Experiment 2 consisted of a computerized elicitation task that was similar to Experiment 1 but was modified for eliciting embedded questions rather than main questions. Participants were instructed that a new character, Sarah, would be part of Experiment 2 and that this character was curious about the questions that Phil asked. The first screen showed Phil asking his teacher a question written in a speech bubble. Participants were instructed to read the question aloud. The second screen showed Sarah asking either “What did Phil not know?” or “What did Phil want to know?” in a speech bubble. Participants’ task was that of producing an embedded question in response to Sarah’s question (e.g., “Phil wanted to know where John had gone”). Participants were instructed that they could not return to the previous screen once they had advanced to the next. They were also instructed to produce “complete sentences”, i.e., to always start their answers with either “Phil wanted to know” or “Phil didn’t know”.

Screenshot examples of slides aimed at eliciting embedded yes/no and *wh*-questions are provided in Appendix D. Four experimental lists were constructed, each containing six practice items and thirty-two experimental items. The same main questions that participants were prompted to produce in Experiment 1 were now used to elicit embedded questions. Participants were assigned to lists so that, in Experiment 2, they would be presented with the most different version of the item they were prompted to produce in Experiment 1 (e.g., if a participant had been prompted to produce item 1 as a *wh*-question with a lexical verb in Experiment 1, they would see item 1 as a yes/no question with an auxiliary verb in Experiment 2). Each participant was randomly assigned to one of the four experimental lists. The two fully *within* factors were question type (yes/no vs. *wh*-) and presence of auxiliary in the prompt (auxiliary vs. lexical verbs in the prompt). The two auxiliaries used in the prompt were *be* and *have*. The type of *wh*-word in *wh*-questions was also manipulated. There were four types of *wh*-words: arguments *what* and *who* and adjuncts *why* and *where*. Type of *wh*-word was a within-subjects factor but a between-items factor. Each participant was presented with a total of 32 main questions as follows:

- 16 yes/no:
    - 8 with lexical verbs
    - 8 with auxiliary verbs
        - 4 with *be*
        - 4 with *have*

- 16 *wh*-questions
    - 8 with lexical verbs
        - 2 *who*, 2 *what*, 2 *where*, 2 *why*
    - 8 with auxiliary verbs
        - 4 with *be*

- 1 *who*, 1 *what*, 1 *where*, 1 *why*

- 4 with *have*
  - 1 *who*, 1 *what*, 1 *where*, 1 *why*

In a quarter of the prompts, the subject was second person (e.g., your brother). The presence of the second person was included to ensure that participants were producing authentic embedded questions instead of quotative questions (e.g., *He wanted to know:* "*What are* ***you*** *doing?*"). As a result, productions were only considered correct if the second person possessive pronoun in the prompt was transformed into a third person possessive pronoun. So, for example, if the question in the prompt was: "What is **your** husband cooking?," the target response would be "Phil wanted to know what **her**/**Ms. Brainy's** husband was cooking", while a response like "Phil wanted to know what **your** husband was cooking" would be considered non-target. This was done to ensure that if participants were producing embedded questions with inversion, this was not due to their directly quoting the prompt. The experimental items were pseudo-randomized so that no more than two consecutive experimental sentences shared any of the features relevant to the investigation (e.g., question-type, *wh*-type). Finally, to control for order effects, four additional lists were created in which the order of the experimental items was reversed. The four experimental lists are given in Appendix E.

2.3.1.2.1.3. Procedure

Experiment 2 took place in the same room as Experiment 1. Participants were given the option of taking a break between the two experiments. As in Experiment 1, the experimenter guided the participant through the six practice items and provided clarification about the procedure. At the end of this combined experimental session, a questionnaire on biographical data, language background, and language use was administered. Non-native participants were also administered a standard language assessment test (MTELP). Experiment 2 took approximately 25 minutes.

2.3.1.2.1.4. Transcription and coding

The same coding and checking procedures used in Experiment 1 were used in Experiment 2. Each production was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology, and presence of target lexical items (e.g., subject and *wh*-element). Incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors, as in (49).
- Double tense/double auxiliary errors, as in (50)–(51).
- Omitted auxiliary/omitted morphology errors, as in (52)–(53).
- Other errors. Other errors included questions that differed in type from the target (e.g., yes/no instead of *wh*-question and vice versa, as in (54)); subject *wh*-questions instead of object *wh*-questions; productions that differed from the target in the lexical items used; questions without a subject; questions with VP movement, as in (55); questions where the *wh*-word differed from the target one; questions where the *wh*-element was not clause-initial, as in (56), or where it was doubled by a dummy element in the sentence; skipped sentences; and sentences where a second person pronoun in the prompt failed to be substituted with a third person pronoun as in (57)–(58).

(49) Phil wanted to know what is Ms. Brainy's brother drinking for dinner.

(50) Phil want to know who did her husband hired.

(51) He wanna know was her husband is cooking now.

(52) Phil want to know who the math teacher helping now.

(53) Phil want to know if John smoke a lot.

(54) Phil did not know if the president had traveled. (Target: Phil didn't know where the president had traveled.)

(55) Phil wanted to know what drinks Ms. Brainy's brother with dinner.

(56) Philip want to know the babysitter drive who to school.

(57) Phil didn't know where your brother forgot his keys.

(58) Philip wanted to know where does your brother wash his clothes.

Only on a handful of occasions did learners produce more than one response for a given trial. On these occasions, the second utterance was judged as an attempt to correct the first response. Only the second utterance was scored.

2.3.1.2.2. Results and interim discussion

The main goal of this experiment was to quantify subject-auxiliary inversion errors in embedded English questions and to examine the relative contribution of L1, question type, and *wh*-type to these productions in adult L2 learners of English. We report the analyses for the three factors separately.

Two sets of analyses were performed: in the first set of analyses, arcsine transformed mean percent *correct* production was the dependent variable. This was calculated by dividing the number of correct responses by the *total* number of productions. The second set of analyses used arcsine transformed mean percent *inversion* as the dependent variable, and this was calculated by dividing the number of correct responses by inverted and non-inverted responses; productions that provided no evidence one way or another with respect to inversion (i.e., auxiliary omission, different structure, morphological errors, etc.) were thus excluded from this second set of analyses.

2.3.1.2.2.1. First language

Table 10 reports reports the raw number (and percentage) of productions in each coding category for L1 Chinese and L1 Spanish L2 learners of English. English native speakers produced only 1 inverted structure and 14 other responses[18] over a total of 1378 embedded questions and were excluded from further analyses.

Table 10: L2 learners' oral production of embedded questions by coding category and L1

Coding	**L1**	
	Chinese	**Spanish**
Correct	570 (55.7%)	617 (60.3%)
Inverted	127 (12.4%)	142 (13.9%)
Double aux/tense	0	3 (0.3%)
No auxiliary	168 (16.4%)	95 (9.3%)
Other errors	159 (15.5%)	167 (16.3%)

Overall, L1 Chinese and L1 Spanish speakers did not differ in terms of inverted responses (t_1 (62) = .9; t_2 (31) = 1, p = .3), double auxiliary errors (t_1 (62) = 1.8, p = .08; t_2 (31) = 1.8, p = .08) or other errors (all *t*s < |1|). L1 Chinese and L1 Spanish speakers differed in terms of auxiliary omission responses (t_1 (62) = 2.5, p = .02; t_2 (31) = 4.9, p < .0001). Moreover, the item analysis indicated that L1 Chinese and L1 Spanish speakers differed in terms of their correct responses (t_1 (62) = .8; t_2 (31) = 2.5, p = .02). L1 Spanish produced more correct responses and fewer auxiliary omission responses than L1 Chinese speakers.

[18] The other errors were all change of structure (e.g., a yes/no question produced in place of a *wh*-question). The single inversion error and three other errors were all produced by the same speaker.

This result is fairly surprising when compared with the results of main questions, where Chinese speakers were producing significantly more accurate responses and fewer errors than L1 Spanish speakers. The higher rates of auxiliary/morphology omission errors in L1 Chinese learners might be explained by differences in L1 properties, given that Chinese lacks inflectional morphology altogether, while Spanish has a rich morphological system. More specifically, it could be that Chinese speakers encounter more difficulties with English morphology as sentences become longer and more complex. However, once productions that were mute with respect to inversion (i.e., auxiliary omission, double tense, and other errors) were removed from the analyses, average non-inversion for L1 Chinese speakers was 81.8%, while the average for L1 Spanish speakers was 81.3%. This difference was not significant (all ts $< |1|$).).

2.3.1.2.2.2. Question type

Table 11 reports the raw number (and percentage) of productions in each coding category by question type, while Table 12 and Table 13 report the raw number (and percentage) of productions in each category by question type for the L1 Spanish and L1 Chinese group, respectively. Figure 4 presents a summary of the productions by coding, L1 group and question-type.

Table 11: L2 learners' oral production of embedded questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh-***
Correct	712 (69.5%)	475 (46.4%)
Inverted	28 (2.7%)	241 (23.5%)
Double aux/ tense	1 (0.1%)	2 (0.2%)
No auxiliary	155 (15.1%)	108 (10.5%)
Other errors	128 (12.5%)	198 (19.3%)

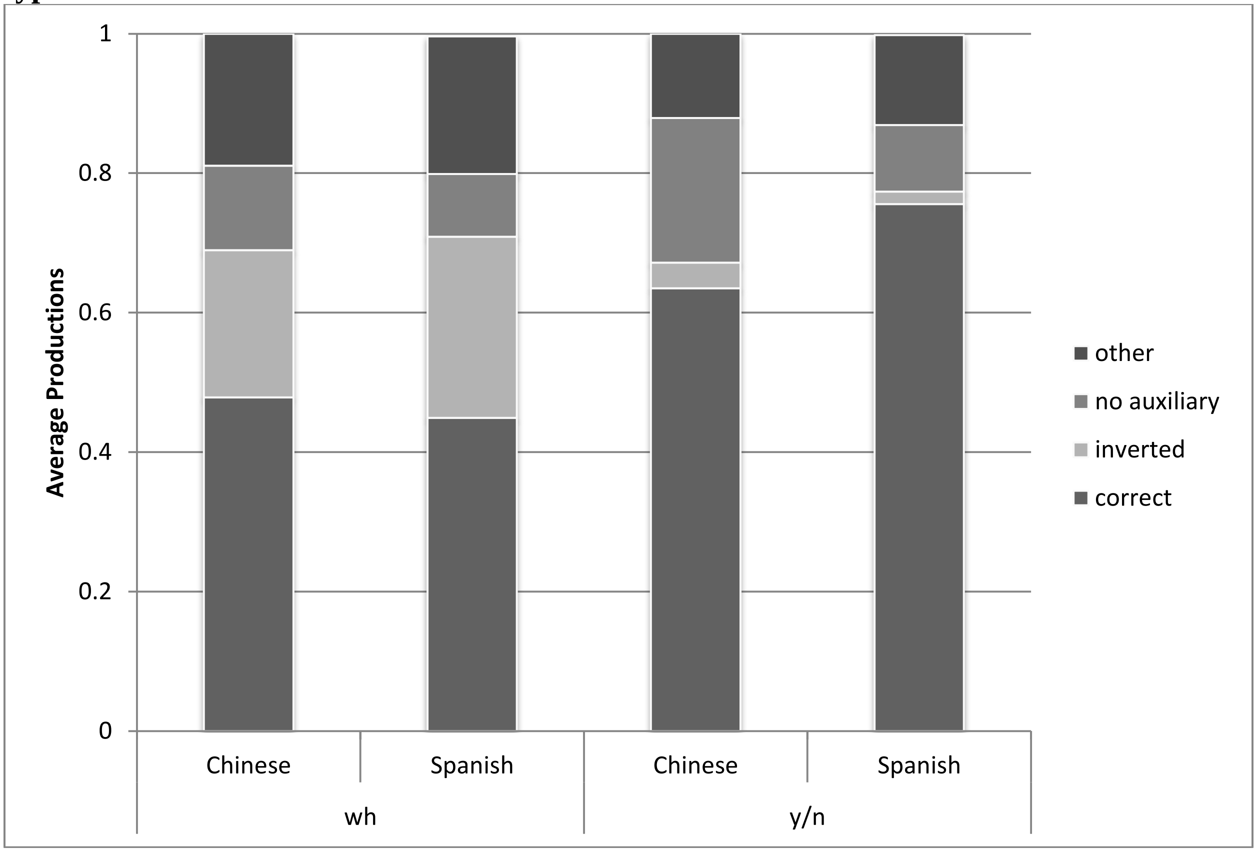
Table 12: L1 Spanish learners' oral production of embedded questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh-***
Correct	387 (75.6%)	230 (44.9%)
Inverted	9 (1.8%)	133 (26%)
Double tense	1 (0.2%)	2 (0.4%)
No auxiliary	49 (9.6%)	45 (9%)
Other errors	66 (12.9%)	101 (19.7%)

Table 13: L1 Chinese learners' productions of embedded questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh-***
Correct	325 (63.5%)	245 (47.9%)
Inverted	19 (3.7%)	108 (21.1%)
Double tense	0	0
No auxiliary	106 (20.7%)	62 (12.1%)
Other errors	62 (12.1%)	97 (18.9%)

Figure 4: L2 learners'oral production of embedded questions by coding category, question-type and L1



A 2 (question type) x 2 (L1) mixed design ANOVA using arcsine transformed percent correct as the dependent variable showed a significant effect of question type (F_1 (1,62) = 66.7, p < .0001; F_2 (1,31) = 105.6, p < .0001) and no effect of L1 (*F1* < 1; *F2* (1,31) = 2.9, p = .09). There was a significant interaction between question type and L1 (F_1 (1,62) = 10.6, p = .002; F_2 (1,31) = 9.9, p = .004), reflecting the fact that L1 Spanish speakers produced significantly more correct responses than L1 Chinese speakers in the yes/no condition (t_1 (62) = 2.5, p = .01; t_2 (31) = 3.6, p = .001), but not in the *wh*-condition (t_1 (62) = .86, n.s.; t_2 (31) = 1.5, n.s.). In particular, Chinese speakers produced more auxiliary and morphology omission errors than L1 Spanish speakers in embedded yes/no questions (t_1 (62) = 3.3, p = .002; t_2 (31) = 5.7, p < .0001), but not in embedded *wh*-questions (t_1 (62) = 1.1, p = .3; t_2 (31) = 1.7, p = .09).

A second 2 (question type) x 2 (L1) mixed design ANOVA using arcsine transformed percent inversion as the dependent variable showed a significant effect of question type (F_1 (1,62) = 133.7, p < .0001; F_2 (1,31) = 139.6, p < .0001), no effect of L1 (all *F*s < 1), and a significant interaction between L1 and question type (F_1 (1,62) = 3.9, p = .052; F_2 (1,31) = 7.7, p = .009), reflecting the fact that L1 Spanish speakers produced more non-inverted responses than L1 Chinese speakers in the yes/no condition (t_1 < 1; t_2 (31) = 2.3, p = .03), but not in the *wh*-condition (t_1 (31) = 1.3, p = .2; t_2 (31) = 1.8, p = .08).

To sum up, yes/no questions were, on average, associated with higher rates of correct non-inverted productions than *wh*-questions. Spanish speakers and Chinese speakers did not differ in terms of their overall accuracy and non-inversion rates. However, on average, L1 Spanish speakers produced higher rates of correct non-inverted responses than L1 Chinese speakers in the yes/no condition under both analyses. On the other hand, L1 Spanish speakers did not differ from L1 Chinese speakers in terms of their overall correct responses in the *wh*-condition in both analyses.

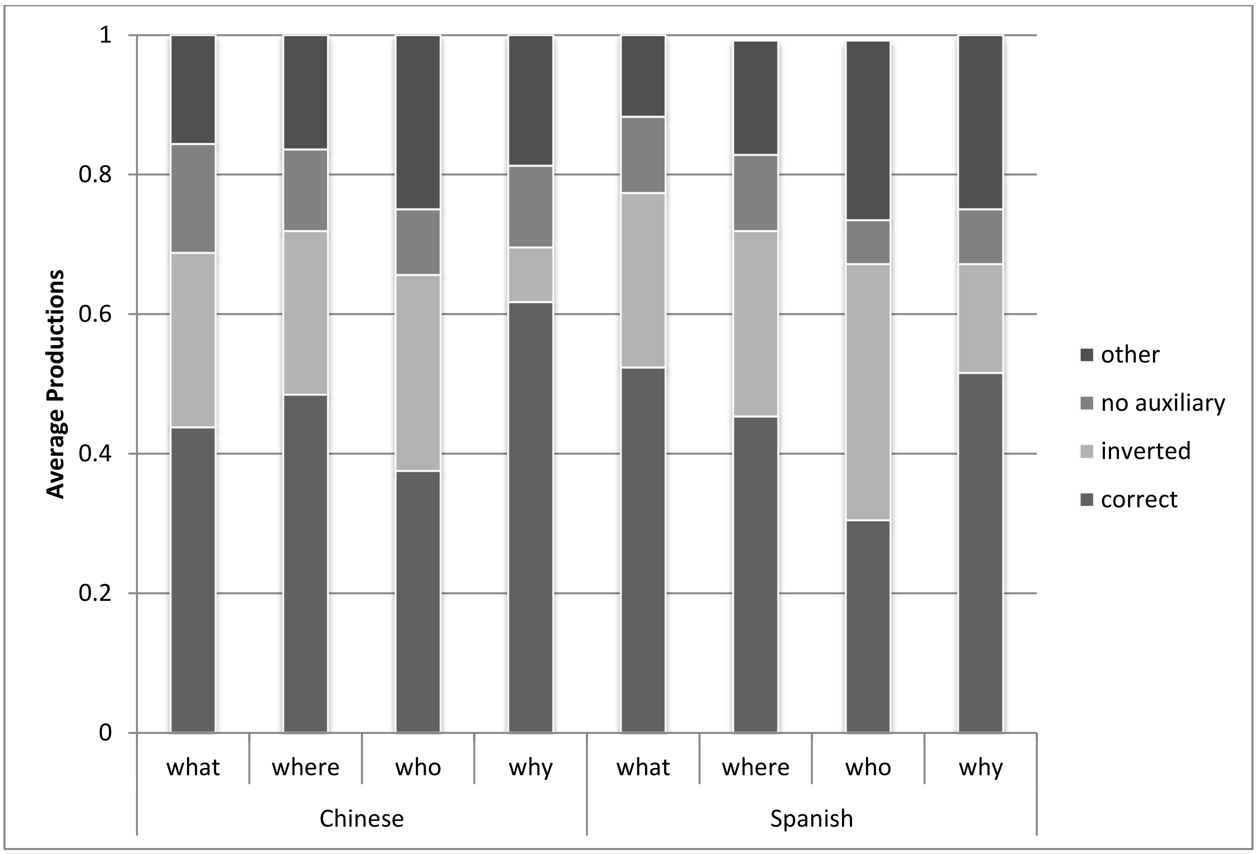
2.3.1.2.2.3. Wh-type

Table 14 reports the raw number of correct responses along with percene correct and percent non-inversion by *wh*-word and by L1, and Figure 5 presents a summary of the productions by coding, L1 group and *wh*-type.

Table 14: L2 learners' oral production of target embedded questions by *wh*-word and L1

***Wh*-Type**	**L1**					
	Chinese			**Spanish**		
	# correct	**% correct**	**% non-inverted**	**# correct**	**% correct**	**% non-inverted**
what	56	43.8%	63.6%	67	52.3%	67.7%
where	62	48.4%	67.4%	58	45.3%	63%
who	48	37.5%	57.1%	39	30.5%	45.3%
why	79	61.7%	88.8%	66	51.6%	76.7%

Figure 5: L2 learners' oral production of embedded questions by coding category, *wh*-type and L1



A 4 (*wh*-type) x 2 (L1) mixed design ANOVA with arcsine transformed percent correct as the dependent variable was performed. The overall ANOVA was significant, indicating a significant effect of *wh*-type (F_1 (3,186) = 11.9, p < .0001; F_2 (3,28) = 7.7, p = .001). The effect of L1 was not significant (F_1 < 1; F_2 (1, 28) = 2.5, p = .13). The interaction between *wh*-type and L1 was not significant in (F_1 (3,186) = 1.9, p = .13; F_2 (3,28) = 1.8, p = .16).

A second 4 (*wh*-type) x 2 (L1) mixed design ANOVA with arcsine transformed percent inversion as the dependent variable was performed. The overall ANOVA was significant, indicating a significant effect of *wh*-type (F_1 (3,168) = 11.7, p < .0001; F_2 (3,28) = 6.7, p = .001). The effect of L1 was not significant (F_1 (1,62) = 1.2, p = .3; F_2 (1,28) = 3.5, p = .07). No interaction with L1 was found (F_1 (3,186) = 1.6, p = .2; F_2 (3,28) = 1.4, p = .3).

Pairwise comparisons were carried out to explore the effect of individual *wh*-elements on correct and inverted responses. Bonferroni correction was applied to prevent Type I error inflation. The first set of analyses (performed on overall percernt correct) revealed a significant difference between *who* and all other *wh*-elements in the subject analysis: (*what* vs. *who* (p < .0001), *where* vs. *who* (p = .005) and *why* vs. *who* (p < .0001)) and a significant difference between *who* and *what* (p = .04) and *who* and *why* (p < .0001) in the item analysis. In all cases, *who* was associated with fewer correct responses.

The second set of analyses (performed on percernt correct non-inversion rates) showed a significant difference between *why* and all other *wh*-elements (subject analysis: *what* vs. *why* (p = .001), *where* vs. *why* (p = .002), *who* vs. *why* (p = .001); item analysis: *why* vs. *what* (p = .01), *why* vs. *where*, (p = .02), *why* vs. *who* (p = .002)), with *why* being consistently associated with lower inversion rates.

Experiment 2 was aimed at investigating the effects of L1, question type and *wh*-type on the production of embedded questions in L2 learners of English.

While both groups produced more non-target inverted sentences than native-speaker controls, the learner groups showed similar non-inversion rates in embedded questions. This result is unexpected under an L1 transfer account, given that Chinese lacks inversion in questions altogether, while Spanish allows subject-verb inversion in embedded questions.

As for the effect of question type (yes/no vs. *wh*-questions), we found that learners produced on average higher rates of non-target inverted responses in *wh*-questions than in yes/no questions. This result suggests that learners are sensitive to structural differences between embedded yes/no and *wh*-questions. Of 1024 tokens of embedded yes/no questions produced by the experimental groups, there were only 28 instances of inversion in embedded yes/no questions, 13 of which lacked an overt complementizer.

As for the effect of *wh*-type, we did not find an argument-adjunct asymmetry in inversion patterns in the production of embedded questions. Rather, our findings in Experiment 2 mirror those in Experiment 1 in that *who* questions are associated with a high number of other responses and *why* questions are associated with lower inversion rates than all other *wh*-elements. Thus, a *why* effect on inversion rates is found in both main and embedded contexts. These findings are consistent with the idiosyncratic behavior of *why* in English first language acquisition and cross-linguistically.

In Appendix F, we report average inversion rates by L2 learner in embedded *wh*- and yes/no questions respectively and indicate whether the participant has acquired the target inversion rule in main and embedded questions according to two criteria: 90% and 80% accuracy. According to both criteria, if a participant has acquired the pattern of inversion in main questions, he/she has also acquired it in embedded questions . There were five exceptions to this rule in the case of *wh*-interrogatives when using the 80% inversion criterion (participants 24, 36, 47, 48, and 56) and only 2 when using the 90% criterion (participants 24 and 29). In the case of yes/no interrogatives, there were two exceptions to the rule when using the 80% criterion (participants 50 and 58) and three exceptions (participants 16, 50, and 58) when using the 90% inversion criterion. This is compatible with the claim that acquisition of more complex structures (embedded questions) follows the acquisition of simpler structures (main questions), as predicted by Spada and Lightbown's stages of acquisition of questions. Specifically, if we use the more conservative 90% accuracy criteria for *wh*-questions, 19 participants had acquired neither pattern (Stages 1–3); 29 participants had acquired the main question pattern but not the embedded question pattern (Stages 4–5); two had not acquired the main question pattern but had acquired the embedded question pattern (not predicted); and fourteen had acquired both the main and embedded question patterns (Stage 6). With respect to yes/no questions, on the other hand, one had acquired neither pattern (Stages 1–3); ten participants had acquired the main question pattern but not the embedded question pattern (Stages 4–5); three had not acquired the main question pattern but had acquired the embedded question pattern (not predicted); and 50 had acquired both the main and embedded question patterns (Stage 6). McNemar's test shows a significant association between acquisition of inversion patterns embedded *wh*-questions and main *wh*-questions (χ^2 (1) = 21.81, p <.0001), while the association was not significant in the case of yes/no questions (χ^2 (1) = 2.77, p = .096).

2.3.1.3. General discussion

One aim of the present study was to quantify the production of main and embedded questions by adult L2 English speakers, filling a gap in the L2 literature. The results show that accuracy and inversion patterns in intermediate/advanced L2 learners of English differ significantly from those of native speakers. Non-target inversion in main and embedded *wh*-questions occur at a non-trivial rate, approximately 10% and 20% overall, respectively.

Another aim of this study was to investigate the role of L1 in non-target L2 productions in English questions. This was done by selecting L2 speakers from L1s that differ from English and one another with respect to inversion in questions. Our results show that L1 Spanish speakers were less accurate than L1 Chinese speakers in their production of English main questions. This cannot be explained by assuming a simple transfer of properties from the L1 to the L2. According to this hypothesis, L1 Chinese speakers should encounter more difficuties with subject-auxiliary inversion than Spanish speakers, given that verb movement is not instantiated in Chinese. On the other hand, this pattern might be explained by a version of the subset principle: due to the fact that Spanish allows, but does not always require T-to-C movement in main questions, Spanish can be thought of as a superset of English with respect to verb movement in main questions. It is thus possible that L1 Chinese learners performed better than L1 Spanish learners because the English input is plainly incompatible with the Chinese parameter settings, making the English property salient for L1 Chinese speakers. This explanation is not viable for embedded questions, given that in embedded questions the English input is incompatible with the L1 Spanish setting but compatible with L1 Chinese in terms of lack of verb movement. With respect to embedded *wh*-questions, the two groups did not differ from each other and both made a consistent number of inversion errors. This lack of L1 transfer observed in embedded questions is compatible with any version of the Full Access Hypothesis (Full Transfer/Full Access, Schwartz & Sprouse 1994, 1996; Full Access without transfer, Flynn & Martohardjono, 1994; Flynn, 1996; Epstein et al. 1996, 1998) and with the proposal that, as learners move to later stages of acquisition, errors that can be imputed to L1 transfer diminish, while errors of overgeneralization increase (Taylor, 1975a, b). Future work focusing on speakers at low/intermediate proficiency levels should investigate this possibility.

Another aim of the present study was to assess the contribution of linguistic factors (e.g., question type and *wh*-type) on inversion. In main questions, we found that learners were more accurate in inversion with yes/no questions than *wh*-questions. This does not reflect the relationship found cross-linguistically in which inversion in yes/no questions implies inversion in *wh*-questions (Eckman, 1989). Moreover, this pattern does not reflect properties of the native English input in that non-inverted yes/no questions are grammatical in casual speech (Crowley & Rigsby, 1987; van Herk, 2000) when they presuppose an answer with the same propositional content, differently from non-inverted *wh*-questions. The fact that yes/no questions are associated with higher rates of inversion than *wh*-questions is consistent with findings from first language acquisition research. Proposals to account for this pattern have been put forth in the first language acquisition literature; for example, Gleitman, Newport, and Gleitman (1984) hypothesize that this pattern can be accounted for in terms of the added saliency of the auxiliary in first position in yes/no questions. Klima and Bellugi (1966), on the other hand, account for the same pattern in terms of the number of operations required in target *wh*-questions (*wh*-fronting and subject-auxiliary inversion) in contrast to yes/no questions (subject-auxiliary inversion only).

These accounts have some plausibility, but they are difficult to distinguish on empirical grounds. The present data do not resolve this issue.

With respect to the effect of question type in embedded questions, we found that learners were also more accurate with yes/no questions than with *wh*-questions. There are several ways to account for this pattern. For example, according to an input-based account, this general pattern could be predicted by the fact that *wh*-elements can be followed immediately by an auxiliary (e.g., in main *wh*-questions) while *if* is never followed immediately by an auxiliary.

Conversely, it could be argued that learners, to some extent, overgeneralize the inversion pattern of main questions to less frequent embedded structures. Clearly, overgeneralization does not occur uniformly in all embedded questions, and learners' sensitivity to the structural difference between embedded *wh*-questions and yes/no questions (e.g., the presence of an overt complementizer) needs to be accounted for. We suggest that learners are sensitive to structural differences between embedded yes/no and *wh*-questions. This finding mirrors a pattern found in non-standard varieties of English (e.g., AAVE, Scottish English, Hiberno English, and Appalachian English, among others), in which inversion is licit in embedded *wh*-questions and embedded yes/no questions that lack an overt complementizer (Labov, 1972; Henry, 1995; Filppula, 1999; Green, 2002). As has been hypothesized for these varieties of English, we propose that embedded inversion in L2 grammars cannot take place in embedded yes/no questions due the presence of an overt complementizer *if*[19] in the position targeted by inverted auxiliaries (C^0).

Under the hypothesis that embedded inversion stems from syntactic overgeneralization, we predicted a correlation between inversion rates of individual *wh*-elements in main and embedded questions (see Stromswold, 1990). In our data, there was no correlation between accuracy in main *wh* and embedded *wh* questions across subjects and across items (r (62) = .2, p = .11 and r (30) = −.12, p = .5). The correlation between inversion in main and non-inversion in embedded was also not significant across subjects (r (62) = −.09, p = .5). Crucially, however, there was a significant negative correlation between inversion rates in main and non-inversion rates in embedded across items, r (30) −.66, p < .0001. Given that inversion scores for subjects collapsed across different *wh*-words, while inversion scores for items did not (*wh*-word was a between item factor), it is possible that the correlation between inversion rates in main and embedded questions that depends on *wh*-words. In other words, the more a *wh*-word inverts in main questions, the more it will invert in embedded questions, and vice versa.

While we failed to find an argument-adjunct asymmetry in the production of main and embedded questions, we showed that *why* is associated with the lowest rates of inversion in both main and embedded questions. Whether the difference in inversion rates between *why* and the other *wh*-elements in our data is derived from L1 transfer or Universal Grammar cannot be

[19] Participants overwhelmingly produced the complementizer *if* instead of *whether* when prompted to produce a yes/no embedded question. Out of the 1024 embedded yes/no productions in this experiment, 104 productions contained the complementizer *whether*. 79/104 were coded as correct, 19 as omitted auxiliary responses, and 6 as other responses.

determined based on the present data given that *why* behaves idiosyncratically in both Chinese and Spanish. For example, Lin (1992) and Ko (2005) claim that a higher base-generation site of *weishenme*, the translation equivalent of *why* in Chinese, is responsible for the fact that only elements which can be A′-moved over the specifier of CP can precede it. With respect to Spanish, Goodall (1991) similarly hypothesizes that Spanish *por qué* 'why' is base-generated in a position higher than other *wh*-elements, and, as such, does not trigger obligatory subject-verb inversion.

Further investigation with speakers of a language that does not evidence a *why*-distinction is thus needed to address this question. There are, however, two pieces of evidence that argue against an L1 transfer explanation: the lack of clear L1 transfer effects elsewhere in our experiment, and the fact that *why* has been shown to associate with low inversion rates both in L1 acquisition of English and cross-linguistically (de Villiers, 1991; Berk, 2003; Thornton, 2008, among others). It seems natural to explain low inversion rates in L1 and L2 English in a similar fashion. For example, Thornton (2008) has proposed that in child English *why* is base-generated in Spec, IntP, which is already endowed with an interrogative feature, making the movement of a [+interrogative] auxiliary to the left periphery optional. Regardless of what explanation turns out to best account for the behavior of *why* in the acquisition of English, it is important to bear in mind that early explanations attributing difficulties with *why* to its conceptual complexity (e.g., Kay, 1980) cannot account for its behavior in adult L2 acquisition cross-linguistic data from adult native speakers and should hence be rejected if we ultimately seek a unitary explanation for the *why*-asymmetry.

An important fact, which has been overlooked by studies that only consider one or two productive utterances to establish whether a structure has been mastered by learners but emerges from the present quantitative study, is that target-like and non-target-like rules coexist in adult intermediate and advanced second language grammars. A variational model of second language acquisition, such as the one proposed for first language acquisition by Yang (2002, 2004) and further developed by Legate and Yang (2007), seems to be well equipped to account for the accuracy profiles that emerged in this study. In this model, different hypotheses compete to best parse the adult native input by first being accessed probabilistically and then punished or rewarded in terms of probability dependent upon their success. Alternatively, the Acquisition by Processing Theory model (Truscott & Sharwood Smith, 2004), in which there are no dedicated language acquisition mechanisms and in which language development is seen a consequence of processing procedures, predicts the intermittent appearance of non-target productions in L2 speakers. According to this model, the production system is shared between the two (or more) languages of a speaker, and L1-transfer is a result of competition between L1 and L2 procedures. The appearance of non-target productions is thus an effect of the L1 procedure having won the competition, possibly due to its ease for the production system. The theory thus explains why non-target productions might still appear in the speech of advanced speakers.

While the model was initially proposed as an alternative to the traditional view that L1-transfer is the result of erroneous parameter setting in L2, given its reliance on UG, it might be possible to extend it to explain the appearance of intermittent non-target productions that cannot be imputed to the speaker's L1, but to UG. Further investigation into the properties of the input that second language learners receive with respect to inversion is needed to explore this possibility.

2.3.2. Written production

The findings from the elicited production experiments described in the previous sections indicate that inversion errors in L2 learners' productions do not derive in a simple way from the properties of their L1 grammars: L1 Chinese learners produced overall more native-like inverted responses in main questions than L1 Spanish learners despite the fact that subject-verb inversion is instantiated in Spanish and not in Chinese. Moreover, L1 Chinese and L1 Spanish learners did not differ in terms of inversion rates in English embedded *wh*-questions, despite the fact that Spanish always allows inversion in embedded questions, while Chinese never does.

A question that arises is whether the reported high rate of errors is due to the particular task or whether the pattern uncovered in the previous section would also surface in another output modality and in tasks where demands on speakers' processing systems are different.

In this section, I examine a corpus of L2 learners' written productions to investigate whether the error patterns seen in elicited spoken production also characterize language production in the written modality.

Written examples of lack of inversion in main questions, and especially examples of embedded inversion, abound on the web (scientific papers, newspaper articles, emails, Facebook statuses, etc.). Here are some examples on the latter type of error that I have collected over the years:

(59) Have you ever wondered what would Carrie Bradshaw do if she were you or you were here (or…Samantha, Miranda, Charlotte)? [Facebook group]

(60) Sabrina is wondering where are her glasses. [Facebook status update; L1 Italian speaker]

(61) But again I'm not sure how is that going to affect the Writing Fellows. [Email exchange; L1 Russian speaker]

(62) Just let me know what it is about and how does it work. [Email; L1 Spanish speaker]

(63) Caterina is wondering where did Bjørn disappear yesterday [Facebook status; L1 Italian speaker]

(64) Please let me know how was your stay and if there is anything I should know about, good or bad. [Email; L1 Hebrew speaker]

(65) In other words, the relevant question should not be whether children are sensitive to pragmatics or not but rather when, and in what circumstances, do pragmatic factors affect children's (and adults') comprehension. [Journal article; L1 French speaker]

(66) Let's not nitpick or wrangle over to what extent is reading in decline. (New York Times, 11.9.2007 Arts, col.8, http://www.nytimes.com/2007/11/19/arts/19nea.html?_r=1&oref=slogin)

(67) I am a Romanian and I'm just curious what can some American people do here. [Online discussion group; L1 Romanian speaker]

(68) I am a English learner and I just could not understand what does that phrase mean, i would appreciate it if someone could give me a detail explanation. [Yahoo India Answers, http://in.answers.yahoo.com/question/index?qid=20071127183923AA8wbar]

(69) This is a continuation of an earlier post on what is an Auditory Processing Disorder (APD) and the behavioural symptoms of APD. [http://healthbitesonline.blogspot.com/2011/01/symptoms-of-auditory-processing.html]

(70) Can you tell what is his level of severity?

[http://healthbitesonline.blogspot.com/2011/01/symptoms-of-auditory-processing.html]

(71) STEP 1: Enter what type of move is this [http://www.mychangeaddress.com/]

The aim of this corpus study was to investigate how pervasive and consistent inversion errors are in L2 written production and what factors affect them. To do so, I examined the relevant portions of the International Corpus of Learner English (Granger et al., 2009), and constructed a new written corpus of essays where participants were asked to write about a job interview.

2.3.2.1. International Corpus of Learner English (ICLE)

In order to determine the extent of inversion errors in second language learners' written productions, I examined the International Corpus of Learner English (Granger et al., 2009). The ICLE consists of 6,085 texts, with a total of 3,753,030 words. Each student in the corpus contributed 1 text; a total of 6,085 learners and 16 different L1s were represented. The learners were all university undergraduate students (usually in their third or fourth year) in their twenties (average age: 22.3), who had learned English in a non-English-speaking country (45% reported no stay in an English-speaking country; 19% reported a stay of less than 3 months; and 25% a stay of 3 or more months). The majority of the learners were female (76%). Proficiency was not measured independently, but a random sample of 20 essays per L1 corpus was rated for writing proficiency. Overall, 60% of the sample essays were rated as advanced. There were clear differences among the sub-corpora: while 20/20 of the Swedish and 19/20 of the Dutch essays were rated as advanced, only 1/20 of the Chinese and 2/20 of the Tswana and Japanese essays were. Information about number of years of formal education in English (school and university) and months spent in an English-speaking country is provided for each learner in a separate database.

The majority of the essays (91%) were argumentative, untimed (62%), and not part of an exam (61%). About half of the essays were written with the support of reference tools (48%). The composition of the ICLE corpus is summarized in Table 15 below.

Table 15: ICLE composition summary

First Language	Word Count	Text Count	Average length words/essay	Text Type
Bulgarian	200,194	302	663	302 Argumentative
Chinese	490,617	982	500	982 Argumentative
Czech	201,687	243	830	197 Argumentative 46 Literary
Dutch	234,732	263	893	252 Argumentative 11 Literary
Finnish	274,628	390	704	357 Argumentative 33 Literary
French	226,922	347	654	295 Argumentative 52 Literary
German	229,698	437	526	422 Argumentative 15 Literary
Italian	224,222	392	572	133 Argumentative 61 Literary 198 Other
Japanese	198,241	366	542	366 Argumentative
Norwegian	211,725	317	668	312 Argumentative 4 Literary 1 Other
Polish	233,920	365	641	361 Argumentative 3 Literary 1 Other
Russian	229,584	276	832	275 Argumentative 1 Other
Spanish	198,131	251	789	199 Argumentative 52 Literary
Swedish	200,033	355	564	302 Argumentative 53 Literary
Tswana	199,173	519	384	519 Argumentative
Turkish	199,532	280	713	280 Argumentative
TOTAL	*3,753,030*	*6,085*	*617*	*5554 Argumentative* *331 Literary* *200 Other*

The ICLE corpus is tagged with the CLAWS7 tagset from WMatrix (Rayson, 2008).

2.3.2.1.1. Main questions

Due to restrictions in the ICLE search functions, main questions were extracted by searching for all sequences containing a question mark and were manually coded for question type, *wh*-type, auxiliary type and errors. A total of 9,147 utterances were extracted and coded. Given that the primary focus of this investigation was to determine the rate of inversion errors in the written production of interrogative structures of second language learners of English, only main questions where inversion could have taken place were included in the final corpus.

I thus excluded subject *wh*-questions as in (72); globally ungrammatical sentences, as in (73) and (74); reduced questions lacking both a subject and an auxiliary, as in (75) and (76); echo questions, as in (77); sentences where the question mark marked the transcriber's uncertainty with respect to some words in the text, as in (78); sentences that were ambiguous between a direct quotation and an embedded question with inversion, as in (79); and non-English utterances, as in (80).

(72) What is happy then?

(73) At what point to male first mark?

(74) Now what Japanese children happen?

(75) So why study at university?

(76) Strange?

(77) Useful for what?

(78) In order to <?>

(79) A little child in conversation with his mother asks her what sound does the mouse produce?[20]

(80) However, nihil nove sub sole?

The final corpus consisted of a total of 6,609[21] sentences produced by 2,558 speakers. There were 3023 *wh*-questions (produced by 1,715 speakers) and 3,596 yes/no questions (produced by 1,758 speakers). Across the corpus, speakers produced an average of .5 *wh*-questions and .6 yes/no questions each. Speakers that produced both *wh*- and/or yes/no questions produced an average of 1.7 main *wh*-questions and 2 main yes/no questions. The composition of the main question sub-corpus is summarized in Table 16 below.

[20] There were 37 such productions with a *wh*-element, 14 of which were inverted (38%). There were also three productions where the auxiliary followed the verb of the main clause directly, without a complementizer. There were 12 productions introduced by if or whether ending with a question mark. None of them were inverted.

[21] There were an additional 10 *how come* questions in the corpus, all correctly non-inverted. They were excluded from the final corpus because their behavior with respect to inversion is opposite to that of all other *wh*-elements.

Table 16: ICLE writers' production of main questions by question-type and L1

First Language	*Wh*-questions and learners that produced them (N)	Yes/No questions and learners that produced them (N)
Bulgarian	181 (96)	210 (107)
Chinese	129 (101)	146 (97)
Czech	256 (119)	290 (118)
Dutch	159 (95)	239 (108)
Finnish	289 (152)	369 (158)
French	222 (135)	310 (148)
German	217 (131)	301 (152)
Italian	167 (107)	199 (130)
Japanese	159 (103)	172 (96)
Norwegian	272 (135)	416 (164)
Polish	108 (71)	158 (88)
Russian	250 (115)	245 (107)
Spanish	121 (66)	114 (58)
Swedish	239 (147)	238 (127)
Tswana	62 (38)	29 (24)
Turkish	192 (104)	160 (76)
TOTAL	*3023 (1715)*	*3596 (1758)*

2.3.2.1.1.1. Coding

Each main question was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology, and presence of target lexical items (e.g., subject and *wh*-element).[22] Following Ambridge et al.'s (2006) coding scheme, incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors, as in (81)–(82), or raising errors as in (83).
- Double tense/double auxiliary errors, as in (84)–(85).

[22] Lack of inversion in main yes/no questions was coded as incorrect for ease of comparison with *wh*-questions.

- Omitted auxiliary errors, as in (86), or errors that, due to the lack of morphology, were ambiguous between non-inversion and omitted auxiliary errors, as in (87).
- Other errors. Other errors included questions without a subject, as in (88); questions without a main verb, as in (89); questions where movement of the whole VP had been applied, as in (90); questions with incorrect auxiliary/morphology, as in (91)–(92); and questions in which the order of negation (or another adverb) was incorrect as in (93)–(94).

(81) What our society will be then?

(82) TV is an addiction or work wonder?

(83) Why then it happens that today we lead a much busier and tenser life than our grandparents did?

(84) Does society owes something to man?

(85) Why do the people will agree to kill the cadre?

(86) Why many students using and owning it?

(87) Why on Earth prospective Bulgarian teachers need a profound knowledge of Old Greek?

(88) How is possible that people had access to information such as the way to make drugs or bombs?

(89) Should smoking totally banned in all restaurants?

(90) What would be the world like if Columbus did not have any dreams of new lands?

(91) But what is it mean?

(92) What do he do?

(93) Is not there a place for dreams and imagination?

(94) Is really television the opium of the masses?

2.3.2.1.1.2. Results

Overall, accuracy with respect to the relevant parameters was very high. The most common error was lack of inversion, but this error accounted for only about 4% of the productions in the corpus. Table 17 shows the raw number and percernt of productions in each coding category.

Table 17: ICLE writers' production of main questions by coding category

Coding	**Productions**
Correct	6,145 (93%)
Non-inverted	249 (3.8%)
Double aux/tense	18 (0.3%)
No auxiliary/No morphology	62 (0.9%)
Other errors	135 (2%)

2.3.2.1.1.2.1. First language

In order to investigate whether high accuracy rates in the production of main questions were homogeneous across the corpus and whether some language groups exhibited higher rates of non-target productions than others, I examined accuracy and inversion rates for the different first language groups. As discussed in the section on oral elicited production, speakers' productions of English interrogatives were measured in two ways. The first measure was overall percernt correct, i.e., the number of correct (inverted) responses over the total number of responses. The second measure was percent inversion, which was calculated by dividing the number of correct responses by responses that were either correct or contained a non-inverted auxiliary and no other errors. As shown in Table 18, accuracy rates were around or below 90% in some language groups, while inversion rates are consistently above 90% for all groups; accuracy and inversion rates that are below or close to 90% are in bold.

Table 18: ICLE writers' production of target main questions by L1

L1	# correct	% correct	% inverted
Bulgarian	364	93.3%	97.1%
Chinese	227	**82.5%**	93.0%
Czech	492	**90.1%**	94.1%
Dutch	378	95.5%	96.4%
Finnish	641	97.4%	98.6%
French	521	97.9%	99.4%
German	503	97.5%	98.1%
Italian	318	**86.9%**	93.3%
Japanese	295	**89.4%**	93.4%
Norwegian	642	93.3%	96.3%
Polish	255	95.9%	98.8%
Russian	438	**88.5%**	**90.7%**
Spanish	210	**89.4%**	94.6%
Swedish	453	95.8%	97.4%
Tswana	78	**85.7%**	91.8%
Turkish	330	93.8%	98.2%
Total	*6145*	*93%*	*96.1%*

Non-inversion rates in main questions were 6% or higher in speakers whose first language was Chinese, Czech, Italian, Japanese, Russian, Spanish, or Tswana, while for other language groups, lack of inversion was virtually non-existent (French, German, Polish, Turkish).

Given that ICLE speakers' proficiency levels were not assessed via any independent measure, it is hard to say whether the fact that inversion errors are more common in some languages than in others is due to proficiency. It should be noted, however, that inversion errors were low in language groups whose essays were rated as advanced (e.g., Swedish and Dutch) and higher in language groups whose essays, as a whole, were not rated as advanced (e.g., Chinese, Tswana and Japanese).

In order to investigate whether L1 was a significant predictor of overall accuracy and inversion rates, a logistic regression analysis was conducted. In the 'overall correct' analysis, each L1 group was compared to L1 Chinese because this group had the lowest rates of correct responses. In the Analysis of Inverted Responses, each language group was compared to Russian, because this group had the lowest inversion rates.

Overall, L1 was a significant predictor of both accuracy and inversion responses. All L1s but Italian and Tswana (highlighted in bold) differed significantly from L1 Chinese in terms of correct responses. All L1s but Chinese, Italian, Japanese, Spanish and Tswana (highlighted in bold) differed significantly from L1 Russian in terms of inverted responses.

Analysis of Correct Responses:

								95% C.I. for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	L1			152.534	15	.000			
	Bulgarian	1.088	.258	17.818	1	.000	2.968	1.791	4.920
	Czech	.656	.214	9.391	1	.002	1.927	1.267	2.930
	Dutch	1.496	.289	26.828	1	.000	4.464	2.534	7.863
	Finnish	2.076	.293	50.338	1	.000	7.973	4.493	14.148
	French	2.304	.344	44.965	1	.000	10.015	5.107	19.640
	German	2.106	.323	42.583	1	.000	8.214	4.364	15.462
	Italian	.337	.222	2.309	1	**.129**	1.401	.907	2.164
	Japanese	.581	.239	5.908	1	.015	1.788	1.119	2.858
	Norwegian	1.082	.220	24.130	1	.000	2.951	1.916	4.545
	Polish	1.590	.347	21.045	1	.000	4.902	2.485	9.667
	Russian	.485	.212	5.229	1	.022	1.625	1.072	2.463
	Spanish	.574	.265	4.715	1	.030	1.776	1.058	2.983
	Swedish	1.575	.278	32.047	1	.000	4.832	2.801	8.336
	Tswana	.238	.339	.493	1	**.483**	1.269	.653	2.466
	Turkish	1.154	.272	18.073	1	.000	3.172	1.863	5.400
	Constant	*1.554*	*.159*	*95.652*	*1*	*.000*	*4.729*		

Note R_L^2 = .054 (Hosmer & Lemeshow), .026 (Cox & Snell), .065 (Nagelkerke). Model $\chi^2(15) = 174.5$, $p < .0001$.

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B) Lower	Upper
Step 1[a]	L1			92.256	15	.000			
	Bulgarian	1.226	.344	12.731	1	.000	3.409	1.738	6.687
	Chinese	.316	.296	1.139	1	**.286**	1.372	.768	2.452
	Czech	.489	.242	4.066	1	.044	1.631	1.014	2.623
	Dutch	1.026	.314	10.671	1	.001	2.789	1.507	5.160
	Finnish	1.990	.370	28.876	1	.000	7.317	3.541	15.122
	French	2.882	.600	23.081	1	.000	17.842	5.507	57.810
	German	1.646	.356	21.431	1	.000	5.188	2.584	10.417
	Italian	.351	.267	1.732	1	**.188**	1.420	.842	2.396
	Japanese	.370	.275	1.816	1	**.178**	1.448	.845	2.481
	Norwegian	.970	.257	14.247	1	.000	2.638	1.594	4.366
	Polish	2.167	.601	12.982	1	.000	8.733	2.687	28.387
	Spanish	.587	.336	3.056	1	**.080**	1.798	.931	3.471
	Swedish	1.364	.332	16.915	1	.000	3.913	2.042	7.496
	Tswana	.135	.424	.102	1	**.750**	1.145	.498	2.631
	Turkish	1.732	.441	15.443	1	.000	5.651	2.382	13.403
	Constant	*2.276*	*.157*	*211.307*	*1*	*.000*	*9.733*		

Note R_L^2 = .053 (Hosmer & Lemeshow), .018(Cox & Snell), .063 (Nagelkerke). Model $\chi^2(15) = 113.9$, $p < .0001$.

2.3.2.1.1.2.2. Question-type

The second question under investigation was whether inversion rates were affected by the question type produced (*wh*- vs. yes/no). First, I examined the distribution of productions across the coding categories. The distribution is extremely similar for the two question types, as shown in Table 19 below.

Table 19: ICLE writers' production by coding category and question-type

Coding	**Question Type**	
	Wh-	**Yes/No**
Correct	2805 (93.1%)	3340 (92.9%)
Non-inverted	111 (3.7%)	138 (3.8%)
Double aux/tense	7 (0.2%)	11 (0.3%)
Non-auxiliary	41 (1.4%)	21 (0.6%)
Other	49 (1.6%)	86 (2.4%)
Total	*3013*	*3596*

Overall, question type was not a significant predictor of either correct (Model χ^2 (1) = .14, p = .7) or inverted responses (Model χ^2 (1) = .13, p = .7).

Next, I investigated whether the distribution of correct and inverted productions in *wh*- and yes/no questions was consistent across L1 groups. In Table 20, I report the distribution of overall correct and inverted responses by question type and speakers' L1s; accuracy and inversion rates that are below or close to 90% are in bold.

Table 20: ICLE writers' productions of target main questions by question-type and L1

L1	*Wh*-questions			Yes/No questions		
	# correct	% correct	% inverted	# correct	% correct	% inverted
Bulgarian	172	95.6%	98.3%	192	**91.4%**	96%
Chinese	101	**78.3%**	**89.4%**	126	**86.3%**	96.2%
Czech	226	**88.3%**	**90.4%**	266	**91.7%**	97.4%
Dutch	155	98.7%	99.4%	223	93.3%	94.5%
Finnish	283	97.9%	99.3%	358	97%	98.1%
French	219	98.6%	99.5%	302	97.4%	99.3%
German	215	100%	100.0%	288	95.7%	96.6%
Italian	140	**83.8%**	**91.5%**	178	**89.4%**	94.7%
Japanese	138	**87.3%**	**92.0%**	157	**91.3%**	94.6%
Norwegian	260	95.6%	99.6%	382	**91.8%**	94.1%
Polish	104	96.3%	97.2%	151	95.6%	100%
Russian	225	**90%**	92.6%	213	**86.9%**	**88.8%**
Spanish	106	**87.6%**	**91.4%**	104	**91.2%**	98.1%
Swedish	233	99.1%	100%	220	92.4%	94.8%
Tswana	52	83.9%	89.7%	26	89.7%	96.3%
Turkish	176	**91.7%**	97.2%	154	96.3%	99.4%
Total	*2805*	*93.1%*	*96.2%*	*3340*	*91.4%*	*96.0%*

Inversion errors were produced by a minority of speakers in the corpus (around 6% of the speakers in each language group for *wh*- and yes/no questions alike). Lack of inversion occurred between 0 and 3 times per speaker in *wh*-questions and between 0 and 6 times per speaker in yes/no questions. For the learners that produced non-inversion errors in *wh*-questions, lack of inversion ranged from 20% to 100% of their productions of main *wh*-questions (from 1/5 to 2/2). For the learners that produced inversion errors in yes/no questions, lack of inversion ranged from 8.3% to 100% of their production of main yes/no questions (from 1/12 to 2/2).

In the case of *wh*-questions, half of the speakers that failed to produce inversion did so consistently (53/100 of the speakers always produced non-inverted main *wh*-questions), while in the case of yes/no questions, lack of inversion was less consistent (39/112 of the speakers always produced non-inverted yes/no questions). Unfortunately, given that more than half of the speakers that produced a main question only produced one instance of it (997/1715 in the case of *wh*-questions and 914/1758 in the case of yes/no questions), it is difficult to infer anything definitive about the obligatoriness of inversion in these speakers' idiolects.

Overall accuracy and inversion rates were very high. Lack of inversion in *wh*-questions occurred most often in speakers whose first language was Chinese, Czech, Italian, Spanish or Tswana, while it never occurred in L1 German and L1 Swedish and virtually never in L1 Dutch, L1 Finnish, L1 French, or L1 Norwegian speakers. Lack of inversion in yes/no questions seems to be a more homogeneous phenomenon: it was present in all but one language group (i.e., Polish) but only in one case did it exceed 6% (i.e., Russian, where lack of inversion was 11.2%). It occurred most often in speakers whose L1 was either Dutch, Italian, Japanese, Norwegian, Russian or Swedish, never in first language speakers of Polish, and virtually never in first language speakers of French or Turkish. While there was a significant correlation between accuracy rates in *wh*- and yes/no questions for the L1 groups in this corpus ($r = .75$, $p = .001$), there was no correlation between inversion rates in *wh*- and yes/no questions ($r = .17$, $p = .52$). This might suggest that subject-auxiliary inversion is regulated by different factors in the two question types.

To investigate whether lack of inversion in English main questions might be due to L1 transfer, each language group was categorized with respect to whether it requires (++), allows (+), or disallows (−) subject-auxiliary and subject-verb order (indicated by $V_{\mathbf{AUX}}$ - $DP_{\mathbf{SUBJECT}}$ -

V_{LEX} and (V_{AUX}) - V_{LEX} - $DP_{SUBJECT}$[23], respectively) in main yes/no and *wh*-questions. This is shown in Table 21.

Table 21: Availability of subject-verb and subject-auxiliary inversion in main questions by question-type and L1

L1	**Main *wh*-Q:** (V_{AUX}-)V_{LEX} - $DP_{SUBJECT}$	**Main *wh*-Q:** V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}	**Main yes/no Q:** (V_{AUX}-)V_{LEX} - $DP_{SUBJECT}$	**Main yes/no Q:** V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}
Bulgarian	++	–	+	–
Chinese	–	–[24]	–	–
Czech	+	+	+	+
Dutch	++	++	+	+
Finnish	–	–	+	+
French	+	+[25]	+	+
German	++	++	+	+
Italian	++[26]	–	+	–
Japanese	–	–	–	–
Norwegian	++	++	++	++
Polish	+	+	–	+
Russian	+	+	+	+
Spanish	+	–	+	–
Swedish	++	++	++	++
Turkish	–	–	–	–
Tswana	–	–	–	–

[23] In languages where the lexical verb precedes the subject, if the auxiliary is present, it precedes the subject as well.
[24] For sake of simplicity, languages where there are no auxiliaries are categorized with those that do not allow subject-auxiliary inversion.
[25] French allows subject-auxiliary inversion with subject clitics, but not with full NPs.
[26] Subject-verb inversion is obligatory with all *wh*-elements except *why*.

Next, languages were grouped with respect to whether they require, allow or disallow subject-verb and subject-auxiliary inversion in main yes/no and *wh*-questions. Overall inversion rates for these groups are provided in Table 22 and Table 23.

Table 22: ICLE writers' production of target main *wh*-questions by L1 typology

***Wh*-Q:** V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}	**% inverted**	***Wh*-Q:** (V_{AUX}-)V_{LEX} - $DP_{SUBJECT}$	**% inverted**
Obligatory	98.5%	Obligatory	99.8%
Possible	94%	Possible	94.4%
Impossible	95.3%	Impossible	94.9%

Table 23: ICLE writers' productions of target main yes/no questions by L1 typology

Yes/No Q: V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}	**% inverted**	**Yes/No Q:** (V_{AUX}-)V_{LEX} - $DP_{SUBJECT}$	**% inverted**
Obligatory	94.4%	Obligatory	94.4%
Possible	96.1%	Possible	96.5%
Impossible	97.5%	Impossible	96.3%

With respect to *wh*-questions, speakers' from L1s with obligatory subject-auxiliary/verb inversion show the highest rates of inversion, as one might expect if simple transfer of L1 properties to English were at play. With respect to yes/no questions, inversion rates are higher for languages that do not allow subject-auxiliary or subject-verb inversion and are the lowest for languages where subject-auxiliary/verb inversion is obligatory.

The effect of availability of subject-auxiliary and subject-verb inversion in yes/no and *wh*-questions was first analyzed separately for each question type. A forward stepwise logistic regression was used, so that predictors were entered in the model only if they made a significant improvement to the model fit. The levels used for comparison were impossibility of inversion.

Availability of subject-auxiliary inversion in *wh*-questions in speakers' L1s was a significant predictor of inverted responses in *wh*-questions: languages in which subject-auxiliary inversion is obligatory were associated with a significant increase in the odds of an inverted response compared with languages where subject-auxiliary inversion was impossible.

Variables in the equation:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	SAI in *wh*-Q			20.134	2	.000			
	SAI in *wh*-Q oblig.	3.147	.720	19.127	1	.000	23.274	5.680	95.376
	SAI in *wh*-Q poss.	−.097	.199	.237	1	.627	.908	.614	1.342
	Constant	2.920	.129	509.642	1	.000	18.540		

Note R_L^2 = .067 (Hosmer & Lemeshow), .022 (Cox & Snell), .078 (Nagelkerke). Model χ^2 (2) = 63.4, p < .0001

Availability of subject-verb inversion in yes/no questions in speakers' L1s was a significant predictor of inverted responses: languages in which subject-verb inversion in yes/no questions was obligatory were associated with a decrease in the odds of an inverted response compared with L1s where subject-verb inversion was impossible, against what could have been expected based on the previous results. This asymmetry between yes/no and *wh*-questions is quite puzzling and I do not have an explanation for this result.

Variables in the equation:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	SVI in yn-Q			7.887	2	.019			
	SVI in yn-Q oblig.	−.831	.306	7.374	1	.007	.436	.239	.794
	SVI in yn-Q poss.	−.441	.276	2.548	1	.110	.644	.375	1.106
	Constant	3.647	.253	207.451	1	.000	38.375		

Note R_L^2 = .007 (Hosmer & Lemeshow), .002 (Cox & Snell), .008 (Nagelkerke). Model χ^2 (2) = 8, p = .02

In order to explore the effect of L1 syntax on inversion, question-type, availability of subject-verb inversion and subject-auxiliary inversion in *wh*- and yes/no questions and the interaction between availability of inversion and question type were entered as factors to predict subject-auxiliary inversion responses. A forward stepwise model was used, so that predictors were entered in the model only if they made a significant improvement to the model fit. The levels used for comparison were *wh*-questions and impossibility of inversion. The details of the regression analyses are presented below, but I will only discuss the final model, where all the significant predictors and their interactions were added.

The final model shows that availability of subject-auxiliary inversion in *wh*- and yes/no questions and the interaction between question type and availability of subject-auxiliary inversion in *wh*-questions were significant predictors. For *wh*-questions, there was a significant *increase* in the odds of having an inverted response when subject-auxiliary inversion in *wh*-questions was obligatory in the speakers' first languages as opposed to when subject-auxiliary inversion in *wh*-questions was impossible in the L1. On the other hand, this change in odds was *negative* when subject-auxiliary inversion in *wh*-questions was possible in the speakers' first language. Conversely, in *wh*-questions, there was a significant *positive* change in the odds of having an inverted response overall when subject-auxiliary inversion in yes/no questions was either obligatory or possible in the learners' first languages as opposed to when subject-auxiliary inversion in yes/no questions was impossible in the learners' L1s.

For yes/no questions there was a significant decrease in the odds of having an inverted response when subject-auxiliary inversion in *wh*-questions was obligatory in learners' first languages. The effect of question-type is somewhat difficult to interpret, in light of the interaction with availability of subject-auxiliary inversion in *wh*-questions: for languages where subject-auxiliary inversion was impossible in *wh*-questions, yes/no questions were associated with a significant increase in the number of inverted responses, compared with *wh*-questions; this basically indicates that only for L1s in which inversion was impossible in *wh*-questions were yes/no questions associated with higher inversion than *wh*-questions.

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	9.497	2	.009
	Block	9.497	2	.009
	Model	9.497	2	.009
Step 2	Step	34.714	2	.000
	Block	44.211	4	.000
	Model	44.211	4	.000
Step 3	Step	18.969	2	.000
	Block	63.180	6	.000
	Model	63.180	6	.000
Step 4	Step	23.370	2	.000
	Block	86.550	8	.000
	Model	86.550	8	.000
Step 5[a]	Step	-2.021	2	.364
	Block	84.529	6	.000
	Model	84.529	6	.000
Step 6	Step	5.029	1	.025
	Block	89.558	7	.000
	Model	89.558	7	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2095.018[a]	.001	.005
2	2060.304[b]	.007	.025
3	2041.335[b]	.010	.035
4	2017.965[b]	.013	.048
5	2019.986[b]	.013	.047
6	2014.957[b]	.014	.050

Variables in the Equation:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1	SAI in yn-Q			9.805	2	.007			
	SAI in yn-Q oblig.	.455	.197	5.327	1	.021	1.576	1.071	2.318
	SAI in yn-Q poss.	.402	.141	8.124	1	.004	1.494	1.134	1.970
	Constant	*2.933*	*.104*	*792.251*	*1*	*.000*	*18.784*		
Step 2	SAI in yn-Q			15.653	2	.000			

	SAI in yn-Q oblig.	3.268	1.006	10.541	1	.001	26.246	3.651	188.686
	SAI in yn-Q poss.	.439	.179	6.016	1	.014	1.550	1.092	2.201
	Q-type*SAI in yn-Q			11.216	2	.004			
	Yn-Q*SAI in yn-Q oblig.	−3.384	1.016	11.101	1	.001	.034	.005	.248
	Yn-Q*SAI in yn-Q poss.	−.065	.192	.115	1	.735	.937	.643	1.365
	Constant	*2.933*	*.104*	*792.251*	*1*	*.000*	*18.784*		
Step 3	SAI in *wh*-Q			15.767	2	.000			
	SAI in *wh*-Q oblig.	−.660	.394	2.804	1	.094	.517	.239	1.119
	SAI in *wh*-Q poss.	-1.232	.354	12.106	1	.001	.292	.146	.584
	SAI in yn-Q			20.843	2	.000			
	SAI in yn-Q oblig.	3.928	1.081	13.205	1	.000	50.798	6.106	422.600
	SAI in yn-Q poss.	1.383	.369	14.058	1	.000	3.988	1.935	8.218
	Q-type * SAI in yn-Q			11.310	2	.003			
	Yn-Q*SAI in yn-Q oblig.	−3.384	1.016	11.101	1	.001	.034	.005	.248
	Yn-Q*SAI in yn-Q poss.	−.088	.193	.210	1	.647	.916	.628	1.335
	Constant	*2.933*	*.104*	*792.251*	*1*	*.000*	*18.784*		
Step 4	SAI in *wh*-Q			17.379	2	.000			
	SAI in *wh*-Q oblig.	.961	1.227	.613	1	.434	2.615	.236	28.981
	SAI in *wh*-Q poss.	−2.129	.726	8.611	1	.003	.119	.029	.493
	SAI in yn-Q			7.968	2	.019			
	SAI in yn-Q oblig.	2.306	1.587	2.112	1	.146	10.037	.447	225.232
	SAI in yn-Q poss.	2.019	.717	7.927	1	.005	7.533	1.847	30.724
	Q-type*SAI in *wh*-Q			11.966	2	.003			
	Yes/No*SAI in *wh*-Q oblig.	−1.795	1.303	1.898	1	.168	.166	.013	2.135
	Yes/No*SAI in *wh*-Q poss.	1.449	.837	2.993	1	.084	4.257	.825	21.969
	Q-type*SAI in yn-Q			1.752	2	.416			
	Yn-Q*SAI in yn-Q oblig.	−1.589	1.652	.925	1	.336	.204	.008	5.201

	Yn-Q*SAI in yn-Q poss.	−1.018	.806	1.595	1	.207	.361	.075	1.753
	Constant	*2.933*	*.104*	*792.251*	*1*	*.000*	*18.784*		
Step 5	SAI in *wh*-Q			33.506	2	.000			
	SAI in *wh*-Q oblig.	1.961	.802	5.980	1	.014	7.104	1.476	34.195
	SAI in *wh*-Q poss.	−1.443	.368	15.342	1	.000	.236	.115	.486
	SAI in yn-Q			15.348	2	.000			
	SAI in yn-Q oblig.	1.069	.442	5.839	1	.016	2.912	1.224	6.928
	SAI in yn-Q poss.	1.333	.351	14.381	1	.000	3.792	1.904	7.551
	Q-type*SAI in *wh*-Q			22.544	2	.000			
	Yn-Q*SAI in *wh*-Q oblig.	−3.138	.720	18.966	1	.000	.043	.011	.178
	Yn-Q*SAI in *wh*-Q poss.	.431	.228	3.579	1	.059	1.539	.985	2.404
	Constant	*2.933*	*.104*	*792.251*	*1*	*.000*	*18.784*		
Step 6[e]	SAI in *wh*-Q			28.978	2	.000			
	SAI in *wh*-Q oblig.	2.189	.807	7.356	1	.007	8.925	1.835	43.409
	SAI in *wh*-Q poss.	−1.215	.380	10.232	1	.001	.297	.141	.625
	SAI in yn-Q			14.841	2	.001			
	SAI in yn-Q oblig.	1.046	.443	5.588	1	.018	2.846	1.196	6.776
	SAI in yn-Q poss.	1.310	.352	13.873	1	.000	3.707	1.860	7.385
	Q-type (yn)	.450	.203	4.937	1	.026	1.568	1.054	2.333
	Q-type*SAI in *wh*-Q			23.640	2	.000			
	Yn-Q*SAI in *wh*-Q oblig.	−3.588	.748	22.980	1	.000	.028	.006	.120
	Yn-Q*SAI in *wh*-Q poss.	−.019	.305	.004	1	.950	.981	.540	1.783
	Constant	*2.727*	*.133*	*420.831*	*1*	*.000*	*15.294*		

2.3.2.1.1.2.3. Additional linguistic factors: verb type and *wh*-type

Three additional linguistic factors, among the many that could influence overall accuracy and inversion rates, were considered: verb type (auxiliary vs. lexical), verb sub-type (type of auxiliary or lexical verb, e.g., *have, be, can*, etc.) and *wh*-type. Table 24 reports the distribution of productions by verb type, sub-type and question type. Raw numbers of correct productions, percernt correct and prercent inversion are reported for each combination.

Table 24: ICLE writers' production of main questions by verb-type, verb sub-type and question-type

Verb-type	Verb sub-type	*Wh*-Questions			Yes/No Questions		
		# correct	% correct	% inverted	# correct	% correct	% inverted
Auxiliary	BE	124	92.5%	96.9%	91	91.9%	96.8%
	can	357	94.2%	95.5%	242	98.4%	98.8%
	could	120	93.8%	95.2%	74	89.2%	91.4%
	HAVE	50	90.9%	94.3%	153	90%	95.6%
	may	4	66.7%	**66.7**%	8	100%	100%
	might	3	75%	**75**%	4	80%	**80**%
	must	5	83.3%	83.3%	5	71.4%	71.4%
	ought	2	66.7%	**66.7**%	0	0%.	0%
	shall	25	100%	100%	23	100%	100%
	should	226	94.2%	97%	228	95.4%	97%
	will	125	92.6%	95.4%	149	88.7%	89.2%
	would	169	94.9%	96%	169	96%	97.1%
Lexical	BE	840	96%	96.9%	1395	92.3%	95.6%
	HAVE	11	68.8%	**68.8**%	9	69.2%	**75**%
	other	744	89.7%	96.9%	790	93.4%	97.8%

Verb type (auxiliary vs. lexical) did not have a significant effect on correct (Model χ (1) = 2.1, n.s.) and inverted responses (Model χ (1) = 2.5, n.s.). In order to examine the effect of verb sub-type on inversion, I ran a binary logistic regression on correct and inverted main questions. However, as can be seen in the two tables above, some auxiliaries were used very infrequently. In order to limit the degrees of freedom in the regression analysis, auxiliary verbs were grouped into eight categories (be, can/could, have, may/might, must/ought, shall/should, will/would).

The verb used for comparison was must/ought because this category was associated with the lowest rates of correct and inverted responses. Verb sub-type was a significant predictor of correct and inverted responses: there was a significant *positive* change in the odds of having a correct, inverted response when the verb was lexical or the auxiliary used was either be, can/could, shall/should, will/would or have[27]. The details of the regression analyses are provided below:

Analysis of Correct Responses:

								95% C.I. for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Verb sub-type			34.511	7	.000			
	BE	1.570	.583	7.260	1	.007	4.808	1.534	15.066
	CAN/COULD	1.819	.598	9.242	1	.002	6.163	1.908	19.905
	HAVE	.875	.608	2.067	1	.151	2.398	.728	7.900
	Lexical	1.292	.584	4.895	1	.027	3.641	1.159	11.437
	MAY/MIGHT	.460	.797	.332	1	.564	1.583	.332	7.558
	SHALL/SHOULD	1.901	.613	9.630	1	.002	6.693	2.014	22.239
	WILL/WOULD	1.489	.597	6.220	1	.013	4.435	1.376	14.297
	Constant	*1.099*	*.577*	*3.621*	*1*	*.057*	*3.000*		

Note R_L^2 = .009 (Hosmer & Lemeshow), .005 (Cox & Snell), .012 (Nagelkerke). Model $\chi^2(7) = 31.9$ $p < .0001$

---

[27] *Have* only has a significant effect on inverted responses.

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	Verb sub-type			42.436	7	.000			
	BE	2.121	.586	13.081	1	.000	8.340	2.642	26.325
	CAN/COULD	2.083	.604	11.893	1	.001	8.030	2.458	26.237
	HAVE	1.418	.627	5.113	1	.024	4.130	1.208	14.118
	Lexical	2.503	.598	17.514	1	.000	12.222	3.784	39.473
	MAY/MIGHT	.460	.797	.332	1	.564	1.583	.332	7.558
	SHALL/SHOULD	2.481	.638	15.132	1	.000	11.952	3.424	41.718
	WILL/WOULD	1.735	.602	8.295	1	.004	5.667	1.740	18.450
	Constant	*1.099*	*.577*	*3.621*	*1*	*.057*	*3.000*		

Note R_L^2 = .017 (Hosmer & Lemeshow), .006 (Cox & Snell), .02 (Nagelkerke). Model $\chi^2(7) = 35.4$ p = .000

Table 25 reports the distribution of productions by *wh*-word. The raw numbers of correct productions, percernt correct and percent inversion are reported for each *wh*-word.

Table 25: ICLE writers' production of target main questions by *wh*-word

***Wh*-type**	**# correct**	**% correct**	**% inverted**
how	665	95.1%	96.5%
what	1210	95.7%	97.7%
when	31	**88.6**%	96.9%
where	138	96.5%	96.5%
which	44	95.7%	97.8%
who/whom	50	94.3%	96.2%
whose[28]	7	100%	100%
why	660	**86.2**%	**93.1**%

In order to examine the effect of *wh*-type on inversion, I ran a binary logistic regression on correct and inverted *wh*-questions. The level used for comparison was *why*. *Wh*-type was a significant predictor of correct and inverted responses: there was a significant *positive* change in the odds of having a correct response when the *wh*-word was either *how*, *what*, or *where*, and a significant positive change in the odds of having an inverted response when the *wh*-word was either *how* or *what*. The details of the regression analyses are given below:

[28] Due to the fact that logistic regression is based on difference in the odds and that odds of a correct response are calculated by dividing the number of correct responses over the number of incorrect ones, odds are indeterminate for cases in which there are no incorrect responses (i.e. denominator = 0). Whenever this situation arose, I randomly added an error in the relevant condition, so that the function could be computed.

Analysis of Correct Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	*Wh*-type			72.312	7	.000			
	how	1.145	.205	31.295	1	.000	3.141	2.103	4.691
	what	1.281	.174	54.133	1	.000	3.599	2.559	5.062
	when	.219	.541	.163	1	.686	1.245	.431	3.597
	where	1.489	.467	10.161	1	.001	4.433	1.774	11.073
	which	1.262	.731	2.985	1	.084	3.533	.844	14.792
	who	.985	.604	2.661	1	.103	2.677	.820	8.737
	whose	−.037	1.085	.001	1	.973	.964	.115	8.084
	Constant	*1.829*	*.105*	*305.460*	*1*	*.000*	*6.226*		

Note R_L^2 = .05 (Hosmer & Lemeshow), .02 (Cox & Snell), .06 (Nagelkerke). Model χ^2 (7) = 71 p < .0001

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	*Wh*-type			26.074	7	.000			
	how	.721	.255	7.993	1	.005	2.057	1.248	3.392
	what	1.131	.239	22.337	1	.000	3.098	1.938	4.951
	when	.834	1.027	.659	1	.417	2.302	.308	17.218
	where	.717	.479	2.246	1	.134	2.049	.802	5.237
	which	1.184	1.022	1.341	1	.247	3.267	.441	24.216
	who	.618	.736	.706	1	.401	1.856	.439	7.856
	whose	−.809	1.090	.550	1	.458	.445	.053	3.774
	Constant	*2.600*	*.148*	*308.447*	*1*	*.000*	*13.469*		

Note R_L^2 = .03 (Hosmer & Lemeshow), .009 (Cox & Snell), .031 (Nagelkerke). Model χ^2 (7) = 25.4 p = .001

Finally, all the factors considered above (speakers' L1, question-type, subject-auxiliary/verb inversion availability in speakers' L1, verb type, verb sub-type and *wh*-type) and their two-way interactions were entered as predictors of correct and inverted responses in a stepwise forward binary logistic regression.

Analysis of Correct Responses:

Seven factors were significant in predicting correct responses: speakers' L1; *wh*-type; verb sub-type; the interaction between availability of subject-auxiliary inversion in *wh*-questions and question type; the interaction between availability of subject-verb inversion in *wh*-questions and verb sub-type; the interaction between subject-auxiliary inversion in yes/no questions and verb type; and the interaction between L1 and *wh*-type, in that order. For reasons of space, the output of the regression model is not provided here, but t all the steps are discussed below, after the summary of the model:

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	174.087	15	.000
	Block	174.087	15	.000
	Model	174.087	15	.000
Step 2	Step	64.085	8	.000
	Block	238.172	23	.000
	Model	238.172	23	.000
Step 3	Step	38.355	7	.000
	Block	276.527	30	.000
	Model	276.527	30	.000
Step 4	Step	32.216	2	.000
	Block	308.743	32	.000
	Model	308.743	32	.000
Step 5	Step	36.624	14	.001
	Block	345.367	46	.000
	Model	345.367	46	.000
Step 6	Step	10.758	2	.005
	Block	356.124	48	.000
	Model	356.124	48	.000
Step 7	Step	114.484	98	.122
	Block	470.609	146	.000
	Model	470.609	146	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3185.596	.026	.065
2	3121.511	.035	.089
3	3083.156	.041	.103
4	3050.940	.046	.115
5	3014.316	.051	.128
6	3003.558	.052	.132
7	2889.074	.069	.172

In the first step, L1 was entered as a predictor: every L1 other than Italian and Tswana was associated with higher rates of correct responses than L1 Chinese. In the second step, *wh*-type was added as a predictor: *when*, *which*, *who*, and *whose* did not differ from *why* in terms of correct responses but *where*, *what* and *how* did. In the third step, verb sub-type was added as a predictor: every verb differed from MUST/OUGHT in terms of accuracy rates except HAVE, MAY/MIGHT and WILL/WOULD. In the fourth step, the interaction between question type and availability in the L1 of subject-auxiliary inversion in *wh*-questions was added to the model: compared to impossibility in the L1 of subject-auxiliary inversion in *wh*-questions, obligatoriness of subject-auxiliary inversion in *wh*-questions was associated with a significant *positive* change in odds of a correct response in *wh*-questions, while it was associated with a negative change in the odds of a correct response in yes/no questions. In the fifth step, the interaction between availability of subject-verb inversion in *wh*-questions interacted with *wh*-type: compared to impossibility in the L1 of subject-auxiliary inversion in *wh*-questions availability of subject-auxiliary inversion in *wh*-questions was associated with a positive change in the odds of a correct response in *why*-questions, but with a negative change in odds in *what* and *how* questions. In the sixth step, the interaction between verb sub-type and availability in speakers' L1s of subject-auxiliary inversion in yes/no questions was added to the model, but none of the individual contrasts reached the significance threshold. Finally, in the last step, the interaction between L1 and question-type was added: while in L1 Chinese there was a positive change in odds of a correct response when the question was yes/no as opposed to *wh*-, the change in the odds was negative in L1 Bulgarian and L1 Norwegian.

Analysis of Inverted Responses:

Six factors were significant in predicting inverted responses in the corpus: speakers' L1; verb sub-type; the interaction between L1 and question type; *wh*-type; the interaction between subject-verb inversion in *wh*-questions and verb sub-type; and the interaction between subject-verb inversion in yes/no questions and verb-type.

In the first step, L1 was added to the model. Every L1 except L1 Chinese, Italian, Japanese and Tswana was associated with a positive change in the odds of an inverted response compared with L1 Russian. In the second step, verb sub-type was added to the model: every verb type other than *have* and *may* was associated with a positive change in the odds of an inverted correct response compared with *must/ought*. In the third step, the interaction between L1 and question type was added: while in Russian there was a negative change in the odds of an inverted response in yes/no compared to *wh*-questions, in L1 Czech and L1 Spanish the odds of producing an inverted response increased significantly in yes/no questions. The opposite happened when the L1 was Dutch and Norwegian, where the odds of producing an inverted response in yes/no questions compared to *wh*-questions decreased significantly. In the fourth step, *wh*-type was added to the model: the odds of having a correct response increased significantly when the *wh*-word was *how* or *what* as opposed to *why*. In the fifth step the interaction between verb sub-type and availability of subject-verb inversion in *wh*-questions was added to the model, but none of the individual contrasts was significant. In the sixth step, the interaction between verb type and availability of subject-auxiliary inversion in yes/no questions was added to the model: there was a significant positive change in the odds of having an inverted response when the L1 allowed subject-verb inversion in yes/no question compared to when it didn't, but only if the verb was an auxiliary.

The picture that emerges so far from the written production of main questions is fairly complicated and somewhat different from what we might have expected from the results of the oral elicited production study. First of all, overall accuracy in the written the corpus was a little lower than in the oral production experiment (93% vs. 95%), while inversion was a little higher (96% vs. 94%). In the present corpus, moreover, it seems that transfer of L1 properties might have had a positive effect on inversion rates: there was a positive relationship between obligatoriness of subject-auxiliary inversion in *wh*-questions and inversion rates. However, this effect was reversed when subject-auxiliary inversion was possible as opposed to obligatory in *wh*-questions, suggesting, in line with previous results from oral production, that parameter resetting might be harder if the L1 is in a superset/subset relation with the L2. In yes/no questions, the availability of subject-auxiliary inversion in yes/no questions was not a significant predictor of inversion rates, and obligatoriness of subject-verb inversion had a negative effect on subject-auxiliary inversion rates. Overall, the effect of question-type was not significant in this corpus, differently from what was observed in the elicited production study. This might be due to the fact that in this corpus, pragmatic conditions for lack of inversion in yes/no questions were met. Finally, the effect of *wh*-type was replicated in this corpus: accuracy rates were lower for *why* compared to *how*, *where* and *what,* and inversion rates were lower for *why* compared to *what* and *how*, respectively. Inversion rates were, on the other hand, similar for adjunct *when* and *why*. This finding replicates the existence of a *why*-asymmetry and the lack of an argument-adjunct asymmetry, as already noted in the section on elicited production. Finally, in the present corpus, a wide range of auxiliary verbs was used, showing that accuracy and inversion rates might also be modulated by the type of auxiliary produced.

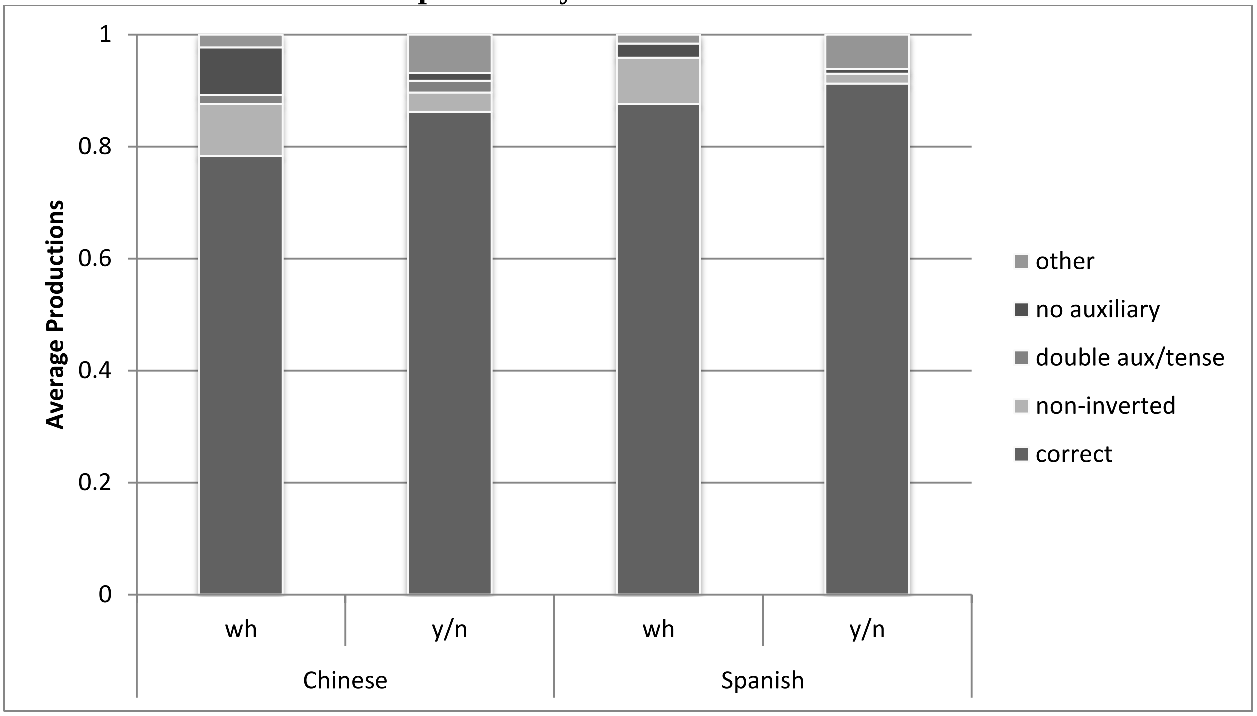
2.3.2.1.1.2.4. Main Questions: L1 Chinese and L1 Spanish

In order to best compare the present results with the findings from the oral production study, I decided to look at the effect of question-type and *wh*-type in L1 Chinese and L1 Spanish L2 learners. In order to investigate whether L1, question type and their interaction were significant predictors of overall accurate and inverted responses, a forward stepwise model was used, so that predictors were entered in the model only if they made a significant improvement to the model fit. Table 26 and Figure 6 present a summary of productions by coding, L1 and question-type for these two sub-corpora.

Table 26: ICLE writers' production of main target questions by question-type and L1 – L1 Chinese and L1 Spanish only

Question-type	L1					
	Chinese			Spanish		
	# correct	**% correct**	**% inverted**	**# correct**	**% correct**	**% inverted**
wh-	129	78.3%	89.4%	121	87.6%	91.4%
yes/no	145	86.2%	96.2%	114	91.2%	98.1%
Total	*274*	*82.5%*	*93.0%*	*235*	*89.4%*	*94.6%*

Figure 6: ICLE writers' production of main questions by coding category, question-type and L1 – L1 Chinese and L1 Spanish only



Overall, L1 was a significant predictor of accurate responses, while the effect of question-type was marginally significant (p = .055), indicating that, overall, Spanish speakers were more accurate than Chinese speakers at their production of English main questions, and that, on average, yes/no questions were associated with slightly higher inversion rates. The model summary is presented below:

Variables in the Equation:

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Spanish	.579	.265	4.786	1	.029	1.784
	Constant	*1.549*	*.159*	*95.036*	*1*	*.000*	*4.708*

Note R_L^2 = .01(Hosmer & Lemeshow), .010 (Cox & Snell), .017 (Nagelkerke). Model χ^2 (1) = 4.96, p = .026

On the other hand, only question-type was a significant predictor of inverted responses, indicating that yes/no questions were associated with higher inversion rates than *wh*-questions. The model summary is presented below:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	yes/no	1.246	.444	7.862	1	.005	3.477
	Constant	2.242	.224	99.932	1	.000	9.409

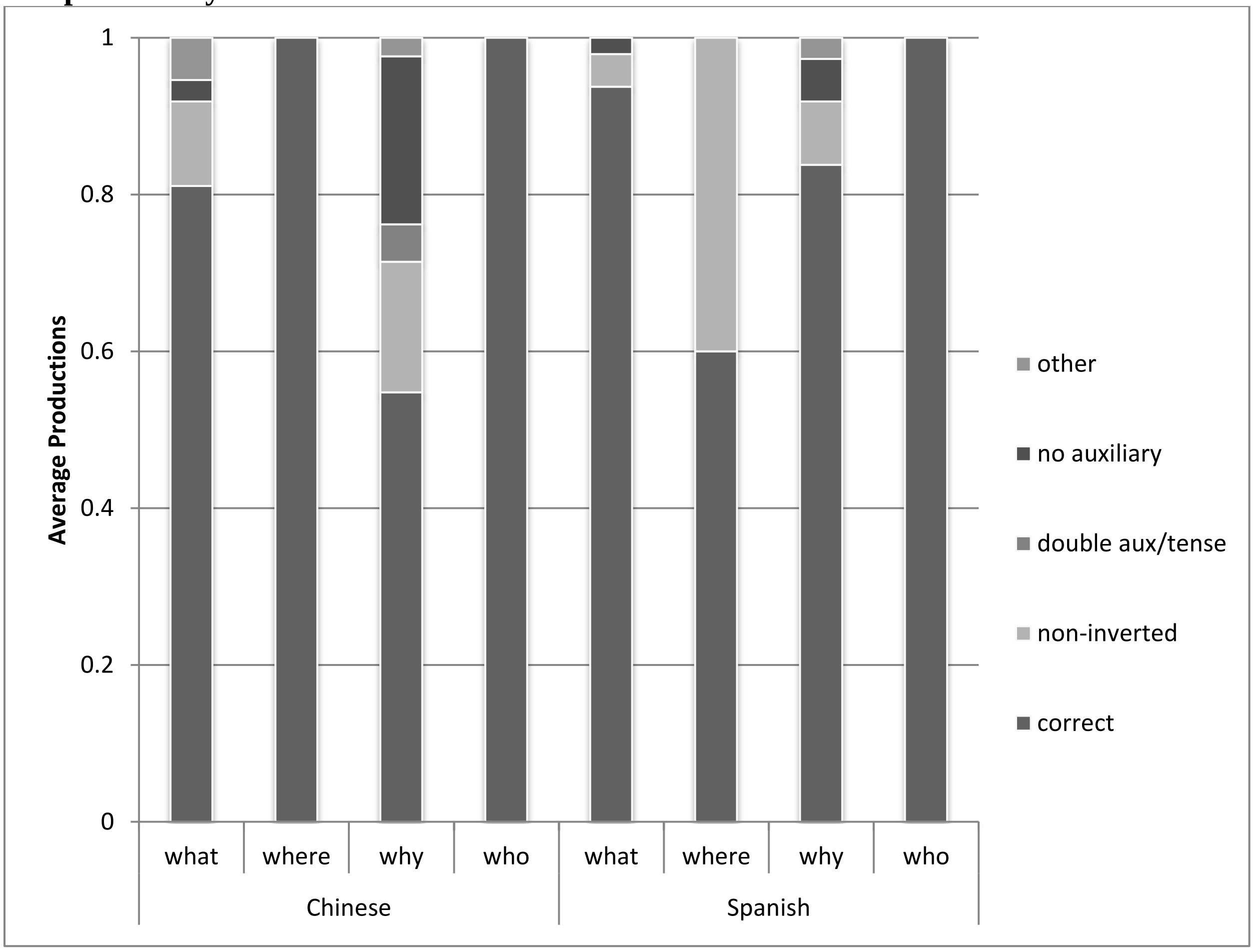
Note R_L^2 = .04 (Hosmer & Lemeshow), .02 (Cox & Snell), .052 (Nagelkerke). Model χ^2 (1) = 9.2, p = .002

Finally, in order to examine the effect of *wh*-type on inversion, I ran a forward stepwise binary logistic regression on correct and inverted *wh*-questions. In order for the comparison between the oral and the written production to be maximally informative, only performance with the *wh*-elements used in the elicited production experiment (what, where, who, why) was examined. Table 27 and Figure 7 present a summary of productions by coding, L1 and *wh*-type.

Table 27: ICLE L1 Spanish and L1 Chinese participants' productions of main question by *wh*-type and L1

***Wh*-Type**	**L1**					
	Chinese			**Spanish**		
	# correct	**% correct**	**% inverted**	**# correct**	**% correct**	**% inverted**
what	37	81.1%	88.2%	48	93.8%	95.7%
where	4	100%	100%	5	60.0%	60.0%
who	1	100%	100%	3	100%	100%
why	42	54.8%	76.7%	37	83.8%	91.2%
Total	*84*	*69.0%*	*84.1%*	*93*	*88.2%*	*92.1%*

Figure 7: ICLE writers' productions by coding category, *wh*-type and L1 – L1 Chinese and L1 Spanish only



Visual inspection suggested that the two groups did not behave similarly with respect to different *wh*-words: while L1 Chinese speakers exhibited high non-inversion rates with *why*, Spanish speakers exhibited particularly high non-inversion rates for *where*. There were only four *who*-questions in this sub-corpus, and they were all correctly inverted. Logistic regression is based on difference in odds; given that the odds of a correct response are calculated by dividing the number of correct responses by the number of incorrect ones, odds are indeterminate for casesin which there are no incorrect responses. For this reason, *who*-questions were eliminated from this analysis and the level used for comparison was *what*-questions.

Both L1 and *wh*-type were significant predictors of correct responses: there was a significant *positive* change in the odds of having a correct response when the L1 was Spanish compared to Chinese and when the *wh*-word was either *why* or *where* compared to *what*.

None of the factors on the other hand, was a significant predictor of inverted responses (Model χ^2 (5) = 10.3, p = .67); the effect of *wh*-type approached significance (p = .064), indicating a trend for *where*-questions to be associated with slightly lower inversion rates (p = .07). The details of the regression analyses for correct responses are given below:

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	9.551	1	.002
	Block	9.551	1	.002
	Model	9.551	1	.002
Step 2	Step	8.637	2	.013
	Block	18.188	3	.000
	Model	18.188	3	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	170.038[a]	.054	.083
2	161.401[a]	.100	.155

Variables in the Equation:

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Spanish	1.187	.399	8.824	1	.003	3.276
	Constant	*.785*	*.237*	*11.002*	*1*	*.001*	*2.192*
Step 2[b]	Spanish	1.140	.409	7.768	1	.005	3.125
	Wh-type			7.970	2	.019	
	where	-.785	.893	.774	1	.379	.456
	why	-1.196	.424	7.970	1	.005	.302
	Constant	*1.494*	*.371*	*16.205*	*1*	*.000*	*4.455*

2.3.2.1.2. Embedded Questions

Embedded *wh*-questions were extracted semi-automatically by searching for sequences that represent the surface order V_{LEX} - *wh*-word - NP (non-inverted structures) and V_{LEX} - *wh*-word - auxiliary - NP (inverted structures), while embedded yes/no questions were extracted by searching for sequences that represent the surface order V_{LEX}-(if/whether)-NP (non-inverted structures) and V_{LEX}-(if/whether)-aux-NP (inverted structures). Only *wh*-questions selected by lexical verbs were included; embedded questions selected by adjectives or nouns, as in (95) and (96), were not included in the search.

(95) He was unsure (of) what major to choose.

(96) The question of what Americans want fiscal policy to be has been asked.

This first search output a superset of structures containing embedded questions: the V_{LEX}-*wh*-word search output embedded questions (e.g., She wonders what Bill has built) and free relative clauses (e.g., She bought what Bill built), while the search V_{LEX}-if/whether output embedded questions (e.g., She wonders if Bill will come to the party) and conditional clauses (e.g., Bill will leave if John does too). All sentences were then manually coded as either 'embedded question', 'free relative' or 'conditional' clauses and only the first underwent further analyses.

Given that the primary focus of this investigation was to determine the rate of inversion errors in the written production of embedded questions of second language learners of English, only embedded questions where inversion could have taken place were included in the final corpus. Free relatives, conditional clauses, subject *wh*-questions, infinitival questions, ungrammatical/mis-tagged sentences, and embedded *wh*-questions whose embedded predicate was a copula and whose arguments were both DPs were excluded (see below). Distinguishing between free relative and embedded questions is aften problematic.[29] Out of context, many sentences are ambiguous between the two meanings. For example, the string in (97) is ambiguous between a free relative and an embedded question interpretation:

(97) Mary knows what Bill knows.

---

[29] Distinguishing between free relative and embedded *wh*-questions is often problematic. As an example, see this discussion on an online grammar discussion forum: http://www.englishforums.com/English/RelativeClausesIndirect-Questions/vjbxv/post.htm [Accessed on 1-15-2011].

In its free relative interpretation, (97) means ‘if Bill knows x, then Mary knows x’, while in its embedded question interpretation, it means ‘if Bills knows x, then Mary knows that he knows x.’ In the free relative clause structure, the verb of the main clause takes an NP as its complement, while in the embedded question structure, it takes a clause. Free relatives and embedded *wh*-questions can be distinguished by using a number of tests (see Baker, 1970, 1995). Below, I present eight tests that can be used to distinguish between embedded questions and free relatives along with unambiguous examples of free relative clauses (as in (98)) and embedded *wh*-questions (as in (99)):

(98) She wondered what Albert bought (unambiguous embedded *wh*-question)

(99) She bought what Albert liked (unambiguous free relative clause)

Test	Embedded Question	Free Relative
1) Compatibility with *whether*	YES She wondered whether Albert liked something	NO *She bought whether Albert liked something
2) Compatibility with cleft	YES She wondered what it was that Albert liked	NO *She bought what it was that Albert liked
3) Compatibility with *else*	YES She wondered what else Albert liked	NO *She bought what else Albert liked
4) Compatibility with *ever*	NO *She wondered whatever Albert liked	YES She bought whatever Albert liked
5) Compatibility with multiple *wh*-elements	YES She wondered who liked what	NO *She bought who liked what
6) Paraphrase with 'answer to the question'	YES She didn't know[30] the answer to the questions: "what did Albert buy?"	NO *She bought the answer to the question: "what did Albert buy?"
7) Obligatorily finite	NO She wondered what to buy	YES *She bought what to like
8) Pronoun substitution	NO She didn't know[31] where Alex lived → *She didn't know *there*	YES She moved where Alex lived → She moved *there*

Free relative and conditional clauses were then excluded from further analyses, while ambiguous cases such as (100) or (101) were included:

(100) She knows what he knows

(101) She understands if you tell her the truth

[30] *Didn't know* was used in this case because *wonder* cannot be used with NP complements.
[31] See previous note.

Subject *wh*-questions and infinitival questions (as in (102) and (103), respectively) were excluded because inversion cannot take place in these structures.

(102) We can not say which colour is the most beautiful one - green, red or blue.

(103) I feel like if people are standing on the edge and now they must decide which way to go.

Ungrammatical sentences that could not be analyzed (as (104)–(105)) and sentences that were mistagged by CLAWS and thus did not contain the relevant sequence of grammatical categories (as in (106) and (107)) were also excluded from further analyses.

(104) Upon this social phenomenon, we would like to know what reasons become they like to possess a credit card.

(105) This means what students knowledge on the Japanese language itself.

(106) showing that **harm which** feminism has caused [*harm* mistagged as a verb]

(107) This prostitutes [sic] can get Hiv from this **guys who** they slept with them and they can become pregnant. [*guys* mistagged as a verb]

Finally, I decided to exclude all embedded *wh*-questions in which the predicate of the embedded clause was a copula and in which both the subject and the predicate of the copular clause were DsP. Given the numeric relevance of such examples and the fact that this seems an issue of theoretical relevance, I will discuss in detail the reason why these sentences were excluded. In the corpus there were many examples like (108):

(108) $^{??/*}$Do you know who is her father?

Determining whether inversion has taken place in sentences like (108) or whether we are dealing with subject *wh*-questions is not straightforward.[32] For example, notice the following example from a linguistics article on the use of grammaticality judgments[33]:

(109) The absence of clear criteria to determine **what is the exact nature of grammaticality** has raised a few concerns on the part of some researchers.

While native speakers of English tend to reject sentences like the one in (108) as ungrammatical, they often do not find them as deviant as cases of embedded inversion in non-copular constructions like (110):

(110) *Mary doesn't know what is he going to eat.

[32] For a discussion on the web on whether the inverted or non-inverted version of this sentence is grammatical, see http://forum.wordreference.com/showthread.php?t=438569
[33] A more entertaining example can be found here: http://icanhascheezburger.files.wordpress.com/2010/11/d0a02a60-e422-4884-a67b-0047303a33a3.jpg

Notice, moreover, that the intermediate status of a sentence like (109) cannot just be imputed to the presence of a heavy DP; the presence of a heavy DP, in fact, does not make embedded inversion in non-copular constructions any more acceptable:

(111) *Mary doesn't know what is her young cousin from Australia going to eat.

I would like to suggest that sentences like the one in (108) are indeed grammatical, but are often judged as unacceptable due to reasons related to processing and information structure.

In the syntactic literature, it is customary to distinguish between predicational and specificational copular clauses, as in (112) and (113), respectively:

(112) Susan is the winner of the race. [Predicational Clause]

(113) The winner of the race is Susan. [Specificational Clause]

(112) and (113) are truth-conditionally identical, but differ in terms of information structure. Predicational structures have a free topic-focus information structure (either DP can be topic and either DP can be focus), while specificational sentences have a fixed information structure: the subject is always the topic and the predicate DP is always focus (Mikkelsen, 2005). Let's now consider a sentence like (114):

(114) I don't know who the winner of the race is.

In (114) the verb in the embedded *wh*-question is a copula and the embedded clause is specificational. The relevant part of the derivation of this sentence is fairly straightforward: the *wh*-element, which acts as the predicate of the copular clause, starts out in the VP[34], while the DP 'the winner of the race' is in the specifier position of the predicative phrase. The *wh*-element then moves to Spec, CP position, giving rise to the order *wh*-DP-copula. Let's now consider a sentence like (115):

(115) I don't know who is the winner of the race.

Speakers who find (115) ungrammatical analyze it as if it had the same derivation as (114). Under this analysis, (115) is an embedded question, and the embedded clause is specificational. The *wh*-element, which starts out in VP moves to Spec, CP. Given the relative order of the copula and the subject, (115), however, also involves raising the copula past the subject DP; (115) is then judged ungrammatical because it involves subject-auxiliary inversion in an embedded context.

There is however an alternative derivation for (115): if we assume that the embedded clause has a predicational structure, the *wh*-element would start out in the specifier of the copular

[34] There is an ongoing debate in the syntactic literature on whether the order of the DPs in specificational clauses is derived from that of predicational clauses. For example, Moro (1997) and Mikkelsen (2005) assume that specificational clauses are inverted, in that the predicate DP raises to subject position. Heycock and Kroch (1999), on the other hand, show that this analysis is problematic and adopt an analysis without movement. For the present purposes, I remain agnostic with respect to verb movement in specificational clauses and only focus on *wh*-movement of the referring DP.

clause and the DP 'the winner of the race' would be in the complement of the copular head. In this alternative derivation, the *wh*-element would ultimately move from Spec, TP to Spec, CP, maintaining the original order *wh*-copula-DP. This derivation should result in a grammatical sentence, because it only involves *wh*-movement from a subject position to Spec, CP. Nothing, in fact, prevents *wh*-movement from subject position to Spec, CP, as shown by (116):

(116) I don't know who won the race.

The question is then why native English speakers find (115) deviant. I would like to suggest that (115) is not ungrammatical, but hard to process due to its information structure: speakers tend to assign to it a specificational structure by default. One possibility for this, suggested to me by Marcel den Dikken (p.c.) is that non d-linked *wh*-expressions like *who* are preferentially assigned the *least referential function* in the sentence, which, in this case, would be the function of the predicate in a specificational structure (as opposed to the subject of the predicate in a predicational structure). That is, if a predicate parse is available for *who*, that's the parse that must be assigned to it, resulting in (114) being judged grammatical and (115) being judged ungrammatical. In (116), however, there is no alternative to a subject parse for the *wh*-element and the sentence is judged as grammatical. Notice also that it is not impossible for bare non d-linked *wh*-expressions like *who* to be construed as D-linked: while "what did who do?" is considered a superiority violation, one can provide contexts in which *who* is d-linked; under such circumstances, superiority effects go away (see Pesetsky (1987) for discussion and example).

One would imagine, if the line of reasoning above is on the right track, that in a context in which bare *who* is explicitly d-linked, the status of (115) should improve, probably to the point of full grammaticality, as shown in (117)-(118):

(117) I don't know which of the athletes is the winner of the race.

(118) I don't know who of them is the winner of the race.

This proves once again that there is nothing in the grammar preventing the extraction of a *wh*-element from the subject position of a predicational clause. More importantly, it shows that d-linked *wh*-elements are easier to reconstruct in subject position. This is expected if, as argued above, the problem with (115) is due to the parser preferentially assigning bare *wh*-elements a predicate function. D-linked elements, being referential, shouldn't qualify as predicates, making (117) and (118) structurally unambiguous: the d-linked *wh*-phrase cannot arguably be construed as the predicate of the copular sentence, so it is unambiguously construed as the subject and the sentence becomes acceptable. One problem with this analysis is that it predicts (119) to be ungrammatical:

(119) I don't know which of the athletes the winner of the race is.

Contrary to predictions, English speakers seem to find (119) grammatical, even if they might not find (119) as natural as (117). A possible explanation for this relies on the interaction between the information structure of *wh*-questions and that of copular clauses. As noticed above, in a specificational clause, the post-copular DP is in focus, while the DP in subject position is a topic. That is, a specificational clause is used when the predicate is given information, while the new information is the identity of the winner. This means that there is maximal featural compatibility between the post-copular DP in a specificational structure and its *wh*-status. This is exactly what we have in (114), where the *wh*-element moves from its post-copular focal position to a focus position in the left periphery. On the other hand, subjects are not typically foci. This means that there isn't maximal featural compatibility between the subject of the predicational structure in (115) and its *wh*-status. The idea, then, is that the perfect compatibility between the post-copular DP in a specificational clause and its *wh*-status leads the hearer to assign a specificational 'deep structure' to a sentence like (115) and reconstruct the *wh*-element in post-copular position. This causes the sentence to sound ungrammatical, because under this interpretation, the copula is inverted. Finally, there is maximal featural compatibility between D-linked elements and subject position, in that they share topic features (Rizzi, 2005). This explains why (117) is judged as more natural than (119), but why both are considered grammatical: there is maximal feature compatibility between the subject position in a copular sentence and the D-linked status of the *wh*-element; hence, the subject parse for the D-linked *wh*-element is preferred.

There were 201 copular sentences in the corpus, 49 of which displayed the order *wh*-copula-DP (24.4%). As I hope to have shown, it is often hard to judge whether sentences like (115) are incorrectly inverted or not. These utterances were thus excluded from further analyses.

The final corpus of embedded questions contained 3148 utterances[35]. 2168 of these embedded questions were *wh*- and 980 were embedded yes/no questions (616 were introduced by *whether* and 396 were introduced by *if*). Across the corpus, speakers produced an average of .36 embedded *wh*-questions and .16 embedded yes/no questions each. For the speakers that actually produced embedded questions, they produced, on average, 1.4 embedded *wh*-questions and 1.2 embedded yes/no questions each. The composition of the embedded sub-corpus is summarized in Table 28 below.

[35] Embedded questions that ended with a question mark (as in (v) and (vi)) were coded separately, given that these could be also be considered (inverted or non-inverted) direct quotations:

(v) I have to admit that sometimes during my early university years I thought what on earth do I need this course for?

(vi) Then, some condemnation may ask when the fetus can be considered as humanhood?

There were 30 embedded *wh*-questions ending with a question mark, 7 of which were inverted (23.3%). On the other hand, there were only 6 instances of embedded yes/no questions ending with a question mark (2 introduced by *whether* and 4 by *if*), none of which were inverted.

Table 28: ICLE writers' production of embedded questions by question-type and L1

L1	*Wh*-questions and learners that produced them (N)	Yes/No questions and learners that produced them (N)
Bulgarian	119 (87)	45 (42)
Chinese	117 (97)	142 (134)
Czech	124 (84)	35 (33)
Dutch	137 (95)	61 (44)
Finnish	198 (138)	70 (59)
French	128 (99)	63 (57)
German	151 (114)	83 (65)
Italian	128 (98)	55 (54)
Japanese	153 (104)	58 (40
Norwegian	199 (135)	74 (62)
Polish	68 (59)	57 (50)
Russian	111 (86)	51 (44)
Spanish	149 (87)	41 (33)
Swedish	158 (113)	59 (57)
Tswana	133 (101)	41 (38)
Turkish	95 (65)	45 (38)
Total	*2168 (1562)*	*980 (850)*

2.3.2.1.2.1. Coding

Each embedded question was coded as either correct (native-like) or incorrect (non-native-like) with respect to word order, verbal morphology and presence of target lexical items (e.g., subject and *wh*-element). Incorrect questions were further coded into three additional categories:[36]

- Subject-auxiliary inversion errors, as in (120)–(121) .
- Omitted auxiliary/morphology errors, as in (122).

[36] There is no double tense/double aux coding category because no such errors were produced in this corpus.

- Other errors. Other errors included questions without a main verb, as in (123); questions where movement of the whole VP had applied, as in (124); and questions with incorrect auxiliary/verbal morphology, as in (125)–(126).

(120) I wonder how can one really claim that science is in clash with imagination.

(121) This would give us the opportunity to know more about what are we going to do in future.

(122) In case of homicides, of course we can understand how the family of the killed feel.

(123) Some imagine what it would like to be someone else, somewhere else.

(124) Think about how would be your house without the last century's inventions.

(125) Immediately we went and asked asked [sic] why was women beaten so badly by men.

(126) One can go no further than one generation to see how time is changed.

2.3.2.1.2.2. Results

Overall, accuracy in the corpus was very high. The most common error was subject-auxiliary inversion, but this error accounted only for 2.6% of the productions in the corpus. Table 29 shows the number (and percentage) of productions in each coding category.

Table 29: ICLE writers' production of embedded questions by coding category

Coding	**Productions**
Correct	2963 (94.1%)
Inverted	83 (2.6%)
No aux/no morph.	34 (1.1%)
Other	68 (2.2%)
Total	*3154*

2.3.2.1.2.2.1. First Language

In order to investigate whether accuracy in the production of embedded questions was homogeneous across the corpus and whether some language groups had more difficulties than others, I examined accuracy and non-inversion rates for the different L1 groups.

Speakers' productions of English interrogatives were measured in two ways, as before: the first measure was overall percent correct, i.e., the number of correct (non-inverted) responses over the total number of responses. The second measure was percent non-inversion, and it was calculated by dividing the number of correct responses by responses that were either correct or contained an inverted auxiliary and no other errors.

As shown in Table 30, accuracy and non-inversion rates were above 90% for most language groups. Accuracy and inversion rates close to 90% are bolded.

Table 30: ICLE writers' production of target embedded questions by L1

L1	**# Correct**	**% Correct**	**% Non-inverted**
Bulgarian	152	92.8%	93.9%
Chinese	227	**87.3%**	96.2%
Czech	145	**91.2%**	**93.5%**
Dutch	197	99.5%	100.0%
Finnish	261	97.4%	98.5%
French	189	99%	99.5%
German	233	99.6%	100%
Italian	174	95.1%	97.2%
Japanese	191	**90.5%**	96.5%
Norwegian	266	97.4%	100%
Polish	124	99.2%	99.2%
Russian	154	95.1%	96.9%
Spanish	179	94.2%	98.4%
Swedish	213	98.2%	99.5%
Tswana	129	**74.1%**	**83.8%**
Turkish	129	92.1%	98.5%
Total	*2963*	*94.1%*	*95.4%*

Overall, accuracy and non inversion in embedded questions was high in the corpus. However, accuracy and non-inversion rates were fairly low (around or lower than 90%) for L1 speakers of Chinese, Czech, Japanese and Tswana. On the other hand, accuracy was at ceiling for L1 speakers of Dutch, French, German and Polish, and non-inversion was at ceiling for Dutch, French, German, Norwegian, Polish and Swedish. Given that ICLE speakers' proficiency levels were not assessed via any independent measure, it is hard to say whether the fact that inversion errors are more common in some L1 groups than in others is due to proficiency or to syntactic characteristics of the different L1s. It should be noted that inversion errors did not occur in language groups whose essays were rated as advanced (e.g., Swedish and Dutch) and were generally high in language groups whose essays were not rated as such (e.g., Chinese, and Tswana).

In order to investigate whether L1 was a significant predictor of accurate and non-inverted responses, a logistic regression analysis was conducted. Each L1 group was compared to L1 Tswana because this group had the lowest accuracy and the highest inversion rates. Overall, L1 was a significant predictor of both accuracy and inversion rates. All L1s exhibited significantly higher accuracy and non-inversion rates than L1 Tswana. A second analysis was run, in which each language was compared to L1 Chinese, which exhibited the second lowest accuracy and highest non-target inversion rates. All L1s but Czech and Japanese differed significantly from Chinese in terms of correct responses, while all languages except Bulgarian, Czech, Italian and Japanese differed significantly from L1 Chinese in terms of non-inverted responses.

Analysis of Correct Responses (L1 Tswana used as comparison group):

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	L1			136.468	15	.000			
	Bulgarian	1.499	.346	18.754	1	.000	4.477	2.272	8.822
	Chinese	.875	.254	11.845	1	.001	2.400	1.458	3.950
	Czech	1.285	.329	15.236	1	.000	3.613	1.896	6.886
	Dutch	4.235	1.017	17.329	1	.000	69.070	9.404	507.303
	Finnish	2.565	.420	37.255	1	.000	13.007	5.707	29.644
	French	3.495	.732	22.826	1	.000	32.965	7.858	138.296
	German	4.398	1.017	18.701	1	.000	81.279	11.074	596.542
	Italian	1.914	.383	24.966	1	.000	6.783	3.201	14.373
	Japanese	1.203	.292	16.995	1	.000	3.331	1.880	5.903
	Norwegian	2.584	.420	37.824	1	.000	13.256	5.817	30.207
	Polish	3.775	1.019	13.730	1	.000	43.605	5.920	321.178
	Russian	1.904	.402	22.460	1	.000	6.715	3.055	14.760
	Spanish	1.736	.356	23.839	1	.000	5.677	2.827	11.397
	Swedish	2.922	.534	29.990	1	.000	18.576	6.528	52.856
	Turkish	1.409	.359	15.428	1	.000	4.091	2.025	8.262
	Constant	1.053	.173	37.003	1	.000	2.867		

Note R_L^2 = .013 (Hosmer & Lemeshow), .055 (Cox & Snell), .152 (Nagelkerke). Model $\chi 2$ (15) = 177.9, $p < .0001$

Analysis of Inverted Responses (L1 Tswana used as comparison group):

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	L1			87.069	15	.000			
	Bulgarian	1.093	.393	7.751	1	.005	2.984	1.382	6.444
	Chinese	1.587	.404	15.422	1	.000	4.888	2.214	10.791
	Czech	1.033	.393	6.903	1	.009	2.810	1.300	6.074
	Dutch	3.642	1.026	12.600	1	.000	38.178	5.110	285.241
	Finnish	2.537	.549	21.347	1	.000	12.645	4.310	37.101
	French	3.601	1.026	12.313	1	.000	36.628	4.901	273.714
	German	3.806	1.026	13.767	1	.000	44.961	6.022	335.673
	Italian	1.914	.503	14.459	1	.000	6.783	2.529	18.195
	Japanese	1.665	.443	14.162	1	.000	5.288	2.221	12.589
	Norwegian	3.939	1.025	14.754	1	.000	51.357	6.882	383.223
	Polish	3.187	1.027	9.623	1	.002	24.225	3.233	181.496
	Russian	1.787	.504	12.554	1	.000	5.969	2.222	16.036
	Spanish	2.448	.622	15.496	1	.000	11.563	3.418	39.119
	Swedish	3.720	1.026	13.151	1	.000	41.279	5.527	308.296
	Turkish	2.526	.745	11.484	1	.001	12.500	2.901	53.868
	Constant	1.641	.219	56.389	1	.000	5.160		

Note $R_L^2 = .015$ (Hosmer & Lemeshow), .037 (Cox & Snell), .164 (Nagelkerke). Model $\chi^2(15) = 113.7$, $p < .0001$

2.3.2.1.2.2.2. Question-type

The second issue under investigation was whether inversion rates were affected by the type of question produced (*wh*- vs. yes/no). First, I examined the distribution of productions across the coding categories. The distribution of errors other than inversion is similar for the two question types, as shown in Table 31 below. However, inversion was virtually nonexistent in embedded yes/no questions, while it was about 4% in *wh*-questions. There was just one inversion error in yes/no questions, produced by one Tswana speaker, reported in (127):

(127) Think **whether are you** doing the right thing and think of the consiquences [sic] that will follow

Table 31: ICLE writers' production of embedded questions by question-type and coding category

Coding	Question-Type	
	Wh	**Yes/No**
Correct	2014 (92.9%)	949 (96.8%)
Inverted	82 (3.8%)	1 (0.1%)
No auxiliary/No morphology	22 (1%)	12 (1.2%)
Other	50 (2.3%)	18 (1.8%)

Question type was a significant predictor of accuracy and inversion rates: embedded *wh*-questions were associated with a negative change in the odds of having a correct, non-inverted response:

Analysis of Correct Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	Wh	−.824	.198	17.282	1	.000	.439	.297	.647
	Constant	3.395	.180	356.847	1	.000	29.812		

Note R_L^2 = .01 (Hosmer & Lemeshow), .006 (Cox & Snell), .018 (Nagelkerke). Model $\chi^2(1) = 20.2$, $p < .0001$

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B	
								Lower	Upper
Step 1[a]	Wh	−3.660	1.007	13.210	1	.000	.026	.004	.185
	Constant	6.861	1.001	47.019	1	.000	954.000		

Note R_L^2 = .07 (Hosmer & Lemeshow), .018 (Cox & Snell), .079 (Nagelkerke). Model $\chi^2(1) = 54$, $p < .0001$

Next, I wanted to investigate whether the distribution of correct and non-inverted productions in *wh*- and yes/no questions was consistent across L1 groups. In Table 32, I report the distribution of overall correct and non-inverted responses by question type and speakers' L1s; accuracy and non-inversion rates below or close to 90% are bolded.

Table 32: ICLE writers' production of target embedded questions by question-type and L1

L1	*Wh*-questions			Yes/No questions		
	# correct	**% correct**	**% non-inverted**	**# correct**	**% correct**	**%non-inverted**
Bulgarian	107	**89.9**%	91.5%	47	100%	100%
Chinese	98	**83.8**%	91.6%	129	**90.8**%	100%
Czech	111	**89.5**%	91.7%	34	97.1%	100%
Dutch	137	100%	100%	61	98.4%	100%
Finnish	191	96.5%	97.9%	70	100%	100%
French	127	99.2%	99.2%	62	98.4%	100%
German	150	99.3%	100%	83	100%	100%
Italian	120	93.8%	96%	55	98.2%	100%
Japanese	136	**88.9**%	95.1%	55	94.8%	100%
Norwegian	194	97.5%	100%	72	97.3%	100%
Polish	67	98.5%	98.5%	58	100%	100%
Russian	104	93.7%	95.4%	50	98%	100%
Spanish	138	92.6%	97.9%	41	100%	100%
Swedish	154	97.5%	99.4%	59	100%	100%
Tswana	93	**69.9**%	**80.2**%	37	**90.2**%	97.4%
Turkish	88	92.6%	97.8%	41	91.1%	100%

As already noted, accuracy and non-inversion rates were higher for yes/no than for *wh*-questions: accuracy rates were consistently above 90% in embedded yes/no questions across writers' L1. Inversion errors in embedded *wh*-questions occurred most often in speakers whose first language was Bulgarian, Chinese, Japanese or Tswana, while they never occurred in L1 Dutch, L1 German and L1 Norwegian speakers.

It is also worth noticing that inversion errors were produced only by a minority of speakers: inversion errors occurred from a minimum of 0 to a maximum of 4 times per speaker in *wh*-questions. For the learners that did produce inversion errors in embedded *wh*-questions, these ranged from 25% to 100% of their relevant productions (from 1/4 to 1/1). The majority of the speakers that produced inversion errors did so consistently (53/78 of the speakers that produced an inversion error did so whenever they produced an embedded *wh*-question). However, given that the majority of the speakers that produced an embedded *wh*-question only produced one instance of it (1130/1562) it's hard to infer something definitive about the obligatoriness of inversion in these speakers' grammars.

To investigate whether inversion errors in English embedded questions are at least in part due to L1 transfer, each language group was categorized with respect to whether it requires (++), allows (+) or disallows (−)inversion in yes/no and *wh*-questions.

Table 33: Availability of subject-verb and subject-auxiliary inversion in embedded questions by question-type and L1

L1	**Embedded *wh*:** V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}	**Embedded *wh*:** $(V_{AUX}\text{-})V_{LEX}$ - $DP_{SUBJECT}$	**Embedded yes/no:** V_{AUX} - $DP_{SUBJECT}$ - V_{LEX}	**Embedded yes/no:** $(V_{AUX}\text{-})V_{LEX}$ - $DP_{SUBJECT}$
Bulgarian	–	++	–	+
Chinese	–	–	–	–
Czech	+	+	+	–
Dutch	–	–	–	–
Finnish	–	–	–	–
French	–	+	–	–
German	–	–	–	–
Italian	–	+	–	+
Japanese	–	–	–	–
Norwegian	–	–	–	–
Polish	+	+	+	+
Russian	–	+	–	–
Spanish	–	+	–	+
Swedish	–	–	–	–
Turkish	–	–	–	–
Tswana	–	–	–	–

From Table 33, it emerges that while a number of L1s allow subject-verb inversion in embedded *wh*-questions, subject-auxiliary inversion in yes/no questions is extremely rare (it is only instantiated in Czech and Polish). Moreover, subject-verb inversion is less common in embedded yes/no questions than in embedded *wh*-questions: of the languages represented in this corpus, only Bulgarian, Italian and Spanish allow it.

Next, I grouped languages with respect to whether they allow[37] or disallow subject-verb and subject-auxiliary inversion in embedded *wh*-questions and provide overall inversion rates for

---

[37] Only one language (i.e., Bulgarian) displays obligatory subject-verb inversion in *wh*-questions, so it was grouped with the languages that allow it.

these groups. This was not done for embedded yes/no questions, given that there was just one inversion error in this condition. Languages where subject-verb/auxiliary inversion is possible are associated with slightly higher inversion rates, as shown in Table 34.

Table 34: ICLE writers' production of target embedded questions by question-type and language typology

Embedded *Wh*-: (Aux-)Lex-Sub	**% non-inverted**	**Embedded *Wh*-: Aux-Subj-Lex**	**% non-inverted**
Possible	95.7%	Possible[38]	94.2%
Impossible	96.4%	Impossible	96.3%

In order to explore the effect of language structure on inversion in embedded *wh*-questions, availability of subject-verb/auxiliary inversion in learners' L1s was entered as a predictor of correct and inverted responses. A forward stepwise regression analysis was used so that predictors were entered in the model only if they made a significant improvement to the model fit. Availability of subject-verb inversion and subject-auxiliary inversion were not significant predictors of correct or non-inverted responses (p = .5, p = .6 for the effect of subject-verb and subject auxiliary inversion on correct responses, respectively, and p = .4 and p = .07 for the effect of subject-verb and subject-auxiliary inversion on inverted responses.[39])

[38] Only Polish and Czech seem to allow subject-auxiliary inversion in embedded questions.

[39] This latter marginal effect was in the direction predicted by simple L1-transfer: L1s that disallow subject-auxiliary inversion were associated with a marginally significant positive change in the odds of having a non-inverted response.

2.3.2.1.2.2.3. Additional linguistic factors: verb type and *wh*-type

Three additional linguistic factors (among the many that could influence overall accuracy and inversion rates) were considered: verb type (auxiliary vs. lexical), verb sub-type and *wh*-type. Table 35 reports the distribution of productions by verb sub-type and questions type. Raw numbers of correct productions, percent correct and percent inversion (only for *wh*-questions) are reporteded for each verb and question type.

Table 35: ICLE writers' production of target embedded questions by verb type, verb sub-type and question-type

		Question-type				
		***Wh*-Questions**			**Yes/No Questions**	
Verb-type	**Verb sub-type**	**# correct**	**% correct**	**% non-inverted**	**# correct**	**% correct**
Auxiliary	BE	226	90.9%	95.8%	33	94.3%[40]
	can	68	85%	88.3%	62	98.4%
	could	31	93.9%	93.9%	14	100%
	HAVE	115	96.6%	99.1%	37	97.4%
	may	10	100%	100%	2	100%
	might	13	100%	100%	4	100%
	must	9	81.8%	81.8%	1	100%
	shall	2	100%	100%	2	100%
	should	67	94.4%	94.4%	77	95.1%
	will	78	88.6%	92.9%	46	100%
	would	66	91.7%	94.3%	32	100%
Lexical	BE	458	84.2%	85.8%	397	97.3%
	HAVE	49	100%	100%	38	95%
	other	822	92.6%	96.7%	208	96.3%

[40] This is the only verb with which there was an inversion error. Percent inversion for this category was 97.1%.

In order to examine the effect of verb type and verb sub-type on inversion, I ran a binary logistic regression on correct and inverted embedded questions. However, as can be seen in the table above, some auxiliaries were used infrequently. In order to limit the degrees of freedom and in line with the grouping used in main questions, verbs were grouped in seven categories (be, can/could, have, may/might/must/, shall/should, will/would).[41] The verb used for comparison for the correct and the inverted analysis was BE, given that it was associated with the lowest correct and highest non-target inversion rates.

Verb-type was not a significant predictor of correct (Model χ^2 (1) = 1.6, n.s.) and inverted responses (Model χ^2 (1) = 2.5, n.s.). Verb sub-type, on the other hand, was a significant predictor of correct and responses: *have*, lexical verbs[42] and, *shall/should* were associated with a positive change in the odds of having a correct, non-inverted response.

[41] *Ought* was not present in this corpus, so the 9 productions with *must* were grouped with *may* and *might*.

[42] Non-inversion rates for lexical verbs were calculated by dividing non-inverted responses without an auxiliary over responses with a lexical verb only (non-inverted responses) and responses with a lexical verb and *do*-support (inverted responses).

Analysis of Correct Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Verb sub-type			19.530	6	.003	
	CAN/COULD	.249	.285	.759	1	.384	1.282
	HAVE	1.322	.395	11.201	1	.001	3.753
	LEXICAL	.453	.155	8.549	1	.003	1.572
	MAY/MIGHT	.357	.607	.346	1	.556	1.429
	SHALL/SHOULD	.710	.375	3.576	1	.059	2.033
	WILL/WOULD	.422	.276	2.340	1	.126	1.525
	Constant	2.208	.095	535.929	1	.000	9.098

Note R_L^2 = .001 (Hosmer & Lemeshow), .007 (Cox & Snell), .017 (Nagelkerke). Model χ^2 (6) = 22.5, p = .001

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	Verb sub-type			32.724	6	.000			
	CAN/COULD	.221	.330	.447	1	.504	1.247	.653	2.382
	HAVE	2.930	1.008	8.446	1	.004	18.732	2.596	135.152
	LEXICAL	1.059	.222	22.846	1	.000	2.883	1.868	4.451
	MAY/MIGHT	.424	.734	.334	1	.563	1.528	.363	6.436
	SHALL/SHOULD	1.065	.519	4.212	1	.040	2.900	1.049	8.017
	WILL/WOULD	.554	.342	2.624	1	.105	1.740	.890	3.401
	Constant	2.546	.111	523.041	1	.000	12.759		

Note R_L^2 = .04 (Hosmer & Lemeshow), .015 (Cox & Snell), .047 (Nagelkerke). Model χ^2 (6) = 45.4, p < .0001

Finally, the effect of *wh*-type was examined.

Table 36 reports the distribution of correct and non-inverted productions by *wh*-word. The raw numbers of correct productions, percent correct and percent non-inverted responses are reported for each *wh*-word. Accuracy and non-inversion rates close to 90% are bolded.

Table 36: ICLE writers' production of target embedded questions by *wh*-type

***Wh*-type**	**# correct**	**% correct**	**% non-inverted**
how	480	89.9%	94.9%
what	1064	90.3%	96.9%
when	37	**88.1**%	92.5%
where	59	**84.3**%	**88.1**%
which	44	95.7%	100%
who/whom	71	**85.5**%	100%
whose	1	100%	100%
why	258	94.2%	95.9%

In order to examine the effect of *wh*-type on inversion, I ran a binary logistic regression on correct and non-inverted responses. The single correct production with *whose* was excluded.

The level used for comparison was *where* because it was associated with the lowest correct and non-inverted rates. *Wh*-type was a significant predictor of correct responses: there was a significant *positive* change in the odds of having a correct response when the *wh*-word was either *what, who, which* or *why* compared to when the *wh*-word was *where*. *Wh*-type was also a significant predictor of non-inverted responses: there was a significant *positive* change in the odds of having a correct response when the *wh*-word was either *how*, *what*, *who* or *why* compared to *where*.

Analysis of Correct Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	*Wh*-Type			22.050	6	.001			
	how	.505	.358	1.987	1	.159	1.657	.821	3.346
	what	1.131	.353	10.286	1	.001	3.100	1.553	6.188
	when	.322	.579	.309	1	.578	1.380	.444	4.289
	which	2.105	1.063	3.918	1	.048	8.203	1.021	65.928
	who	1.484	.675	4.840	1	.028	4.412	1.176	16.558
	why	1.101	.417	6.954	1	.008	3.006	1.327	6.813
	Constant	1.680	.328	26.157	1	.000	5.364		

Note R_L^2 = .02 (Hosmer & Lemeshow), .01 (Cox & Snell), .025 (Nagelkerke). Model $\chi^2(6) = 21.6$ p = .001

Analysis of Inverted Responses:

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	*Wh*-Type			15.995	6	.014			
	how	.918	.427	4.614	1	.032	2.503	1.084	5.783
	what	1.445	.415	12.124	1	.000	4.243	1.881	9.573
	when	.514	.709	.526	1	.468	1.672	.417	6.708
	which	1.763	1.079	2.668	1	.102	5.831	.703	48.365
	who	2.250	1.075	4.380	1	.036	9.492	1.154	78.094
	why	1.157	.487	5.654	1	.017	3.180	1.225	8.253
	Constant	*1.998*	*.377*	*28.125*	*1*	*.000*	*7.375*		

Note R_L^2 = .007 (Hosmer & Lemeshow), .007 (Cox & Snell), .024 (Nagelkerke). Model $\chi^2(6) = 14.5$, p =.025

Finally, all the factors considered above (speakers' first language, question-type, subject-auxiliary/verb inversion availability in speakers' first languages, verb type, verb sub-type and *wh*-type) and their two-way interactions were entered as predictors in a final logistic regression.

Analysis of Correct Responses:

In this analysis, embedded *wh*- and yes/no questions were analyzed together. The factors that made a significant contribution to the model fit were speakers' L1s, question-type, verb sub-type, the interaction between question-type and verb-type.

In the first step, L1 was added as a predictor. All languages were different from Tswana in that they were associated with a significant increase in the odds of having a correct response. A second analysis was run comparing all languages to Chinese. All L1s but Czech and Japanese were different from L1 Chinese.

In the second step, question-type was entered in the model: yes/no questions were associated with a significant positive change in the odds of having a correct response compared to *wh*-questions. In the third step, verb sub-type was entered in the model: *have, shall/should* and lexical verbs were associated with a significant positive change in the odds of having a correct response compared to BE. In the fourth step, the interaction between question-type and verb sub-type was added to the model: the positive change in the odds of having a correct response associated with HAVE and lexical verbs as opposed to BE was only true for *wh* questions and not for yes/no questions. In the fifth step, verb type was added to the model: auxiliary verbs were associated with a positive change in the odds of having a correct response compared with lexical verbs. The summary of the model is provided below:

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	199.993	15	.000
	Block	199.993	15	.000
	Model	199.993	15	.000
Step 2	Step	57.380	1	.000
	Block	257.374	16	.000
	Model	257.374	16	.000
Step 3	Step	24.734	6	.000
	Block	282.108	22	.000
	Model	282.108	22	.000
Step 4	Step	16.889	6	.010
	Block	298.996	28	.000
	Model	298.996	28	.000
Step 5	Step	4.935	1	.026
	Block	303.931	29	.000
	Model	303.931	29	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1515.684	.060	.146
2	1458.304	.077	.186
3	1433.570	.084	.203
4	1416.681	.089	.215
5	1411.746	.090	.218

Analysis of Inverted Responses (only for embedded *wh*-questions):

In this analysis, only *wh*-questions were examined. The significant predictors in this analysis were speakers' L1, verb sub-type, *wh*-type, and the interaction between availability of subject-verb inversion in *wh*-questions and verb type. All languages differed from Tswana in that they were associated with a significant increase in the odds of a non-inverted response. An additional analysis was run in which all L1s were compared to Chinese instead of Tswana. Half of the L1s differed significantly from L1 Chinese: L1 Dutch, Finnish, French, German, Norwegian, Spanish and Turkish were associated with a significant positive change in the odds of a non-inverted response, while L1 Tswana was associated with a negative change.

In the second step, verb sub-type was entered in the model: have, shall/should and lexical verbs were associated with a positive change in the odds of having a non-inverted response when compared to BE.

In the third step, *wh*-type was added to the model: questions containing *how*, *what* and *why* were associated with a positive change in the odds of having a non-inverted response compared to *where*-questions. In the fourth and last step, the interaction between availability of subject-verb inversion in *wh*-questions and verb type was added to the model: for auxiliary verbs compared to lexical verbs, there was a significant positive change in the odds of having a non-inverted response when the L1 did not allow subject-verb inversion in embedded *wh*-questions compared to languages that did. The summary of the model is provided below:

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	97.685	15	.000
	Block	97.685	15	.000
	Model	97.685	15	.000
Step 2	Step	26.020	7	.000
	Block	123.706	22	.000
	Model	123.706	22	.000
Step 3	Step	17.992	6	.006
	Block	141.697	28	.000
	Model	141.697	28	.000
Step 4	Step	4.106	1	.043
	Block	145.804	29	.000
	Model	145.804	29	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	613.610	.046	.158
2	587.590	.057	.199
3	569.598	.065	.227
4	565.492	.067	.234

The picture that emerges so far from the written production of embedded questions is fairly different from the one that emerged from the oral elicited production study. Most strikingly, overall accuracy in the written the corpus is much higher than in the oral production experiment (92% vs. 58%), and so is target non-inversion (95% vs. 82%).

Transfer of L1 properties did not have a clear effect on accuracy and non-inversion rates: availability of subject-auxiliary or subject-verb inversion in *wh*-questions was not a significant predictor of non-inversion rates in this model.

Overall, the effect of question-type was significant in this corpus: inversion errors in embedded yes/no questions were virtually non-existent. This confirms what has been found in the elicited production study and what has been reported for non-standard varieties of English, where inversion in embedded questions is only grammatical when there is no overt complementizer.

Finally, an effect of *wh*-type on accuracy and inversion rates was found in this corpus: accuracy rates were lower for *where* compared to *what*, *who*, *which* and *why* and non-inversion rates were lower for *where* compared to *how*, *what*, *who* and *why*. This finding does not directly replicate the existence of a *why*-asymmetry, given that inversion rates for *why* were similar to those for *what*. A wider range of auxiliary verbs was used in the present corpus, suggesting that accuracy and non-inversion rates might also be modulated by the type of auxiliary used.

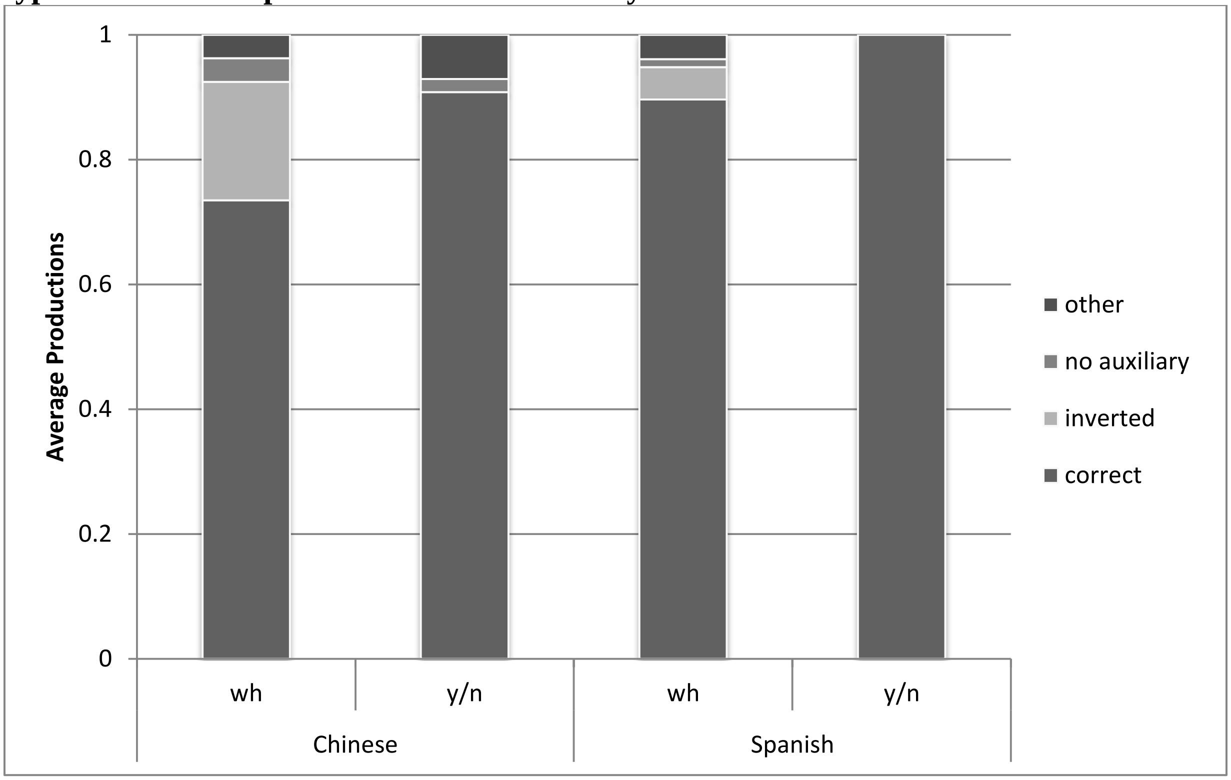
2.3.2.1.2.2.4. Embedded Questions: L1 Chinese and L1 Spanish

In order to best compare the present results with the findings from the oral production study, I decided to look at the effect of question-type and *wh*-type in L1 Chinese and L1 Spanish L2 learners. In order to investigate whether L1, question type and their interaction were significant predictors of overall accurate and non-inverted responses, a forward stepwise model was used, so that predictors were entered in the model only if they made a significant improvement to the model fit. Table 37 and Figure 8 present a summary of productions by coding, L1 and question-type.

Table 37: ICLE writers' production of target embedded questions by question-type and L1 – L1 Spanish and L1 Chinese only

Question-Type	**L1**					
	Chinese			**Spanish**		
	# correct	**% correct**	**% non-inverted**	**# correct**	**% correct**	**%non-inverted**
wh-	132	73.5%	79.5%	154	89.6%	94.5%
yes/no	141	90.8%	100%	41	100%	100%
Total	*273*	*82.4%*	*90%*	*195*	*91.8%*	*95.7%*

Figure 8: ICLE writers' production of embedded questions by coding category, question-type and L1 – L1 Spanish and L1 Chinese only



Overall, both question type and L1 were significant predictors of correct responses, indicating that, on average, Spanish speakers were more accurate than Chinese speakers at their production of English embedded questions, and that yes/no questions were associated with higher accuracy rates than *wh*-questions. The model summary is summarized below:

Omnibus Tests of Model Coefficients:

		Chi-square	df	Sig.
Step 1	Step	11.646	1	.001
	Block	11.646	1	.001
	Model	11.646	1	.001
Step 2	Step	17.305	1	.000
	Block	28.951	2	.000
	Model	28.951	2	.000

Model Summary:

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	361.840[a]	.025	.045
2	344.535[a]	.060	.109

Variables in the Equation:

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1[a]	yes/no	1.037	.327	10.082	1	.001	2.821
	Constant	*1.528*	*.154*	*97.810*	*1*	*.000*	*4.608*
Step 2[b]	Spanish	1.252	.319	15.437	1	.000	3.497
	yes/no	1.401	.340	17.030	1	.000	4.061
	Constant	*.979*	*.192*	*25.943*	*1*	*.000*	*2.661*

Given that embedded yes/no questions were all correctly non-inverted, only L1 and the interaction between L1 and question-type were entered as predictors of non-inverted responses. Only L1 was a significant predictor (Model χ^2 (1) = 5.3, p = .021; R_L^2 = .02 (Hosmer & Lemeshow), indicating that there was a significant *positive* change in the odds of having a non-inverted response when the L1 was Spanish compared to Chinese.

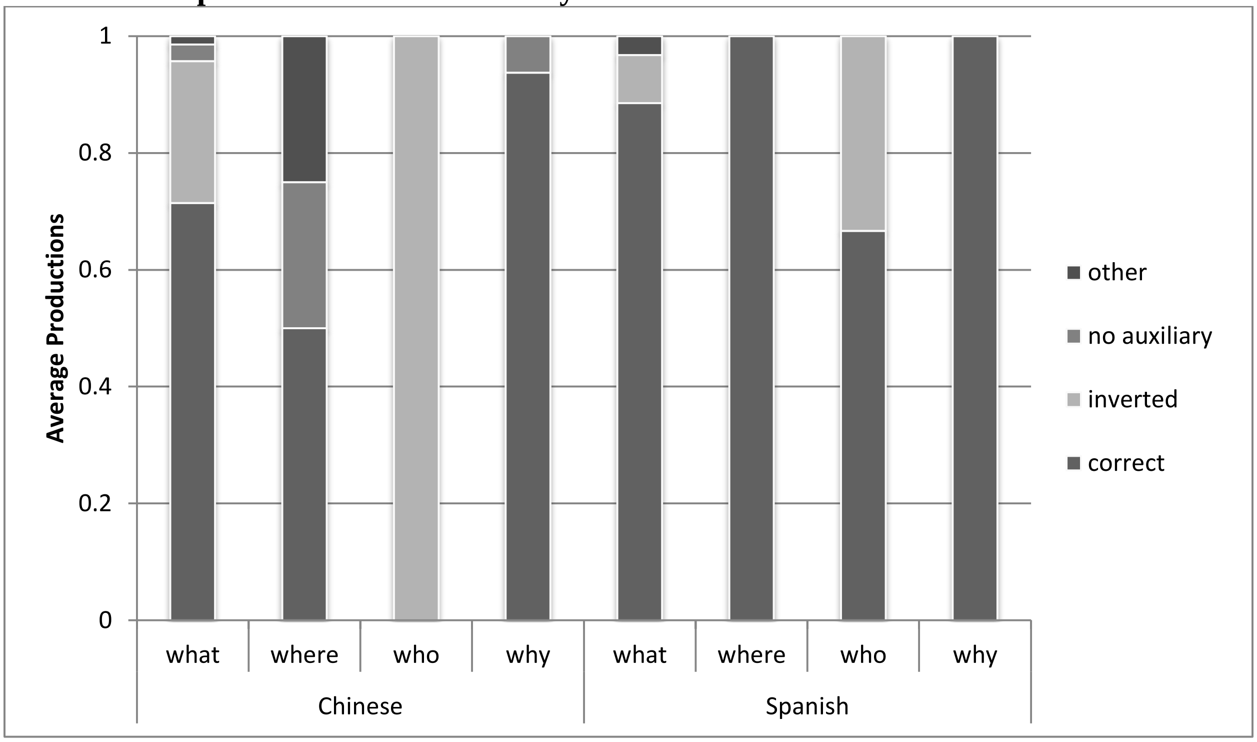
With respect to the effect of *wh*-type, in order for the comparison between the oral and the written production studies to be maximally informative, only performance with the *wh*-elements used in the elicited production experiment (*what*, *where*, *who*, *why*) was examined.

However, no statistics were computed on this sub-corpus, because L1 Spanish speakers exhibited 100% accuracy rates for *where*- and *why*-questions and Chinese speakers exhibited 0% accuracy rates for *who*-questions. Table 38 and Figure 9 present a summary of productions by coding, L1 and *wh*-type.

Table 38: ICLE writers' production of target embedded questions by *wh*-type and L1 – L1 Spanish and L1 Chinese only

	L1					
	Chinese			Spanish		
Question-type	**# correct**	**% correct**	**% non-inverted**	**# correct**	**% correct**	**% non-inverted**
what	70	71.4%	74.6%	61	88.5%	91.5%
where	4	50%	100%	1	100%	100%
who	2	0%	0%	6	66.7%	66.7%
why	16	93.8%	100%	7	100%	100%
Total	*92*	*72.8%*	*77.9%*	*75*	*88%*	*90.4%*

Figure 9: ICLE writers' production of embedded questions by coding category, *wh*-type and L1 – L1 Spanish and L1 Chinese only



A qualitative analysis of the data suggests that, overall, Chinese speakers were less accurate than Spanish speakers. More crucially, there were no inversion errors for adjunct *where* and *why* in neither group, while both groups produced non-target inverted argument questions. This suggests that the high non-target inversion rates for *where*-questions reported in the previous section were due to the influence of other L1s.

2.3.2.2. English Questions: CUNY Learner Corpus

In order to further investigate the acquisition of English subject-auxiliary inversion in second language learners' written production, I collected a small corpus of essays specifically aimed at eliciting main and embedded questions.

2.3.2.2.1. Method

2.3.2.2.1.1. Participants

Sixty eight second language learners participated in this study. The learners were all university undergraduate students (usually in their first year), in their twenties or thirties (average age: 24.9), who had moved to the US after puberty (average age of arrival to the US was 19.5). Average length of stay in the US was 5.2 years. The majority of the writers in the corpus were female (41 vs. 17).

Each speaker contributed one text to the corpus for a total of 68 essays. Eighteen different first languages were represented. The most represented L1 in the corpus was Chinese (29 texts). Proficiency was addressed independently through a shortened version of the Michigan Test of English Proficiency (MTELP). The average MTELP score for this group of wirters was 33/45. The composition of the corpus is summarized in Table 39:

Table 39: CUNY Corpus Composition Summary

L1	Text Count	Average MTELP	Average length (words/essay)
Albanian	1	44	342
Bengali	1	36	62
Burmese	2	40.5	149
Chinese	29	33	159.9
French	3	33	171.2
Georgian	1	44	56
German	1	45	353
Hebrew	2	43	257.5
Italian	1	38	61
Japanese	1	43	142
Korean	6	38.8	173.8
Nepali	1	43	200
Polish	1	43	263
Portuguese	2	39[43]	169.5
Punjabi	1	41	246
Romanian	1	34	157
Russian	5	40.2[44]	146.5
Spanish	6	40.2	182.2
Tibetan	2	41	285.5
Ukrainian	1	44	276
TOTAL	*68*	*37.12*	*179.3*

2.3.2.2.1.2. Procedure

Participants sat down in front of a computer and were asked to write a story about a job interview. The instructions were given in written form to participants on the computer screen and are reproduced below:

[43] One of the MTELP scores was lost due to program malfunctioning.
[44] See previous note.

In this part of the study, you are asked to write for 15 minutes about the following situation: Mark has recently graduated from Queens College. His major was psychology. He is interested in working in finance. He is bright and motivated, but does not have a lot of experience in this field. Today, Mark is interviewing for a job at J.P. Morgan. I would like you to write for 15 minutes about this job interview. Part of your essay should consist of a dialogue between Mark and his interviewer and should describe the things each person asks and answers. For example, Mark will probably have to answer questions regarding his qualifications, academic preparation, career interests and experience, and some general questions about his personality.

Please make sure to use verbs like 'ask',' say',' wonder', 'answer', 'want to know'.

The next screen will be on for 15 minutes.

The essays were typed on the computer using an online system developed by Paul Feitzinger for this purpose (http://www.cunylarc.org/experiments/timed-response).

2.3.2.2.2. Main Questions: Coding and Results

Main questions were manually extracted and coded for question type, *wh*-type, auxiliary type and inversion. Given that the primary focus of this investigation was to determine the rate of inversion errors in the written production by second language learners of English, only main questions where inversion could have taken place were included in the final corpus.

Unambiguous subject *wh*-questions, as in (128), globally ungrammatical sentences, as in (129), sentences that were ambiguous between a direct quotation and an embedded question with inversion[45], as in (130), and long distance questions, as in (131), were excluded from further analyses.

(128) What makes you the best candidate for this job?

(129) So can you tell me about you that can make this company hired you?

(130) And i want to know how do you position yourself among all candidates?

(131) Which one do you think is better?

The final corpus contained 182 main questions, 88 of which were yes/no and 94 of which were *wh*-questions. Across the corpus, speakers produced an average of 1.3 main yes/no questions and 1.4 main *wh*-questions, but some speakers did not produce any main questions. If we compare this average with the average of main questions in the ICLE (.5 and .6 for *wh*- and yes/no, respectively), the task seems to have been fairly successful at eliciting main questions. Overall, only 42 speakers produced at least one *wh*-question and 41 produced at least one yes/no question. The speakers that produced main questions produced on average 2.1 yes/no and 2.2 *wh*-questions each. The composition of the main question corpus is summarized in Table 40 below.

---

[45] There were eight embedded *wh*-questions ending with a question mark, five of which were inverted and four instances of embedded yes/no question ending with a question mark, one of which was inverted.

Table 40: CUNY corpus writers' production of main questions by question-type and L1

L1	Main *wh*-questions and learners that produced them (N)	Main yes/no questions and learners that produced them (N)
Albanian	2 (1)	1 (1)
Bengali	3 (1)	2 (1)
Burmese	3 (2)	2 (1)
Chinese	43 (19)	39 (20)
French	3 (1)	5 (1)
German	1 (1)	1 (1)
Hebrew	6 (2)	3 (2)
Italian	0 (0)	1 (1)
Japanese	1 (1)	3 (1)
Korean	4 (3)	3 (2)
Nepali	0 (0)	0 (0)
Polish	0 (0)	3 (1)
Portuguese	3 (1)	12 (2)
Punjabi	0 (0)	0 (0)
Romanian	4 (1)	2 (1)
Russian	7 (4)	6 (3)
Spanish	8 (3)	2 (2)
Tibetan	4 (1)	3 (1)
Ukrainian	2 (1)	0 (0)
TOTAL	*94 (42)*	*88 (41)*

2.3.2.2.2.1. Coding

Each main question was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology and presence of relevant grammatical categories (subject, *wh*-word, lexical verb).The coding scheme was the same as the one used for coding of the ICLE corpus and the elicited production experiment. Incorrect questions were further coded in three categories:

- Subject-auxiliary inversion or raising errors, as in (132).
- Omitted auxiliary errors or errors that, due to the lack of morphology, were ambiguous between non-inversion and omitted auxiliary errors, as in (133).
- Other errors. Other errors included questions without a subject, as in (134); questions without a main verb, as in (135); and questions with incorrect auxiliary/morphology, as in (136).

(132) What food you can cook?

(133) Why you want to pick up finance as your job?

(134) How long have been in this country?

(135) How much salary would you happy with?

(136) Do u afraid to be leader in your group?

2.3.2.2.2.2. Results

The distribution of yes/no and *wh*-questions across the four coding categories by question type is shown in Table 41 below.

Table 41: CUNY corpus writers'production of main questions by coding category and question-type

Coding	**Question-type**	
	***Wh*-questions**	**Yes/No questions**
Correct	78 (82.9%)	83 (94.3%)
Non-inverted	5 (5.3%)	1 (1.1%)
No auxiliary/No morphology	2 (2.1%)	1 (1.1%)
Other	9 (9.6%)	3 (3.4%)

As can be seen from Table 41, there were very few non-inverted utterances in the corpus. Speakers were overall more accurate in their production of yes/no questions than in their production of *wh*-questions (χ^2 (1) = 7.09, p = 0.008). Utterances that lacked inversion or an auxiliary/morphology are listed below. Information about the speakers that produced them and their MTELP score is provided in parentheses:

(137) There is any other experience you have in the financial world? (L1 Spanish; MTELP = 43)

(138) OK that's great, so you have some experience in this kinda job? (L1 Tibetan; MTELP = 40)

(139) Why you want to pick up finance as your job? (L1 Chinese; MTELP = 37)

(140) Why you want to change job? (L1 Chinese; MTELP = 38)

(141) So, why you still want to do this job? (L1 Chinese; MTELP = 29)

(142) What food you can cook? (L1 Chinese; MTELP = 23)

(143) What kinds of work I have to do? (Chinese; MTELP = 23)

(144) Wait a min, if your major was Psychology, why you are here to apply for a Finance? (L1 Burmese; MTELP = 39)

(145) Can you please tell me something about you. A short description, what you've done? (Romanian; MTELP = 34)

Very little can be said about inversion errors in this corpus due to the low number of productions and errors, but it is interesting to notice that 4/7 errors in *wh*-questions were produced with *why*[46], 2/7 were with d-linked *wh*-elements and that 5/9 inversion errors were produced by L1 Chinese speakers. No further analyses were conducted on this corpus.

[46] In the corpus, there were 20 *why* main questions.

2.3.2.2.3. Embedded Questions: Coding and Results

Embedded questions were manually extracted and coded for question type, *wh*-type, auxiliary type and inversion. Only *wh*-questions selected by verbs were included; embedded questions selected by nouns or adjectives, as in (146), were not included:

(146) He mentioned that the reason why he landed in pursuing a finance [sic] was that he saw a brighter an steadier future.

Given that the primary focus of this investigation was to determine the rate of inversion errors in the written production of second language learners of English, only embedded questions where inversion could have taken place were included in the final corpus. I thus did not include subject *wh*-questions, as in (147); embedded questions where the main verb was in its infinitival form, as in (148); embedded questions preceded by a period, as in (149)–(150); and embedded *wh*-questions in which the embedded predicate was a copula and both arguments were DPs (see section 2.3.2.1.2.1.)

(147) John asked who was sending Mark to the interview.

(148) The boss told Mark where to go to get his I.D.

(149) He asked him if he had any experience in finance. Or if he had ever volunteered in any finance company.

(150) He might be asked upon getting the job, how soon can he start working. Or would he be available for extra hours in chase [sic] the work is not done.

The final corpus of embedded questions contained a total of 68 sentences, 57 embedded *wh*-questions and 11 embedded yes/no questions. Across the corpus, speakers produced on average .85 embedded *wh*- and .15 embedded yes/no questions each. Compared to the ICLE, a corpus that was not designed to elicit interrogative clauses, the assignment was moderately successful at eliciting embedded *wh*-questions, while it was not successful at eliciting embedded yes/no questions (the average number of embedded *wh*- and yes/no questions produced by the writers in the ICLE was .4 and .2). Only 30/68 speakers produced embedded questions and 10 produced embedded yes/no questions. For the speakers that did produce them, 1.9 embedded *wh*-questions and 1.1 embedded yes/no questions each were produced. The composition of the embedded sub-corpus is summarized in Table 42 below:

Table 42: CUNY corpus writers' production of embedded questions by question-type and L1

L1	Embedded *wh*-questions and learners that produced them (N)	Embedded yes/no questions and learners that produced them (N)
Albanian	0 (0)	0 (0)
Bengali	0 (0)	0 (0)
Burmese	0 (0)	1 (1)
Chinese	27 (13)	5 (5)
French	5 (2)	1 (1)
Georgian	0 (0)	0 (0)
German	2 (1)	0 (0)
Hebrew	4 (1)	0 (0)
Italian	1 (1)	0 (0)
Japanese	0 (0)	0 (0)
Korean	5 (3)	1 (1)
Nepali	1 (1)	0 (0)
Polish	0 (0)	0 (0)
Portuguese	0 (0)	0 (0)
Punjabi	2 (1)	1 (1)
Romanian	0 (0)	0 (0)
Russian	3 (3)	0 (0)
Spanish	5 (2)	0 (0)
Tibetan	2 (2)	2 (1)
Ukrainian	0 (0)	0 (0)
TOTAL	*56 (30)*	*11 (10)*

2.3.2.2.3.1. Coding

Each embedded question was coded as either correct (native-like) or incorrect (nonnative-like) with respect to word order, verbal morphology and presence of relevant grammatical categories (subject, *wh*-word, lexical verb). The coding scheme was the same as the one used for the ICLE corpus. Incorrect questions were further coded into three categories:

- Subject-auxiliary inversion errors, as in (151).
- Auxiliary omission or errors of lack of morphology, as in (152).
- Other errors. Other errors included embedded questions with an incorrect *wh*-element, as in (153), and questions without a subject (154).

(151) I am wondering why would you like to apply for this job

(152) The interviewer asked how Mark get interested to the thing that way different with Mark’s major

(153) Would you like to tell me when do you come from, Mark?

(154) Mark is surprised, he doesn't know what should talk next

2.3.2.2.3.2. Results

The distribution of embedded yes/no questions across the coding categories is shown in Table 43 below.

Table 43: CUNY corpus writers’ production of embedded questions by coding category and question-type

Coding	**Question type**	
	***Wh*-questions**	**Yes/No questions**
Correct	45 (79%)	10 (91%)
Inverted	9 (15.8%)	0 (%)
No auxiliary/No morphology	1 (1.7)	0 (0%)
Other	2 (3.5%)	1 (9%)

As can be seen from Table 43, inversion errors only occurred in embedded *wh*-questions; but the difference between wh- and yes/no questions was not significant (χ^2 (1) = .85, p = .355). Utterances with subject-auxiliary inversion errors are listed below. Information about writers’ first languages and MTELP scores is given in parentheses:

(155) Oh, I am wondering why would you like to apply for this job. (L1 Korean; MTELP = 40)

(156) Tell me how are you better than others who have an accounting major. (L1 Hebrew; MTELP = 44)

(157) He was asked as how can having done major in psychology help him in finance world. (L1 Punjabi; MTELP = 41)

(158) He might be asked upon getting the job, how soon can he start working. (L1 Punjabi; MTELP = 41)

(159) Would you like to tell me about why would you like to work for our company since you were majoring in psychology. (L1 Chinese; MTELP = 41)

(160) Would you like to tell me what do you think your psychology background would bring to our company? (L1 Chinese; MTELP = 41)

(161) Do you mind to tell me why do you wanna work in our firm? (L1 Chinese; MTELP = 37)

(162) ok, so tell me a a [sic] little something about your day, and where do you come from.(L1 Chinese; MTELP = 45)

(163) The interviewer should ask him questions about how can he prove to the bank he can do this job even though he has no experience. (L1 Spanish; MTELP = 41)

Very little can be said about inversion errors in this corpus due to the low number of productions and errors. Embedded inversion in *wh*-questions was fairly frequent in this corpus, with inversion rates being similar to those found in the elicited production experiment. It is worth noticing that inversion errors were all produced by speakers whose MTELP scores were above average. No further analyses were thus conducted on this corpus.

2.3.3. Oral and Written Production of English Questions: Summary and Discussion

The oral and written production studies presented in this chapter had several aims. First, I wanted to quantify inversion errors in the production of main and embedded questions by adult L2 English speakers, filling a gap in the L2 literature. Taken together, the results of the oral and the written production studies show that accuracy and inversion patterns in intermediate /advanced L2 learners differ from those of native speakers.

The second aim of these studies was to investigate the effect of L1 transfer on English interrogative structures in both the oral and the written output modality. Taken together, the two studies show that inversion errors occur in speakers with a variety of L1 backgrounds. More importantly, the analyses conducted in the previous sections show that inversion errors cannot be accounted for by simply looking at the syntactic properties of learners' L1s. While the written production study indicates that L2 learners have an advantage acquiring the relevant L2 property if the same property is instantiated in the L1 (e.g., obligatory inversion in main *wh*-questions), both the written and the oral production studies suggest that the learning task might be especially difficult when the L1 is in a superset/subset relation with the L2 (e.g., if the L1 has optional inversion and the L2 has obligatory inversion). However, this was only true for main questions: there was no effect of L1 typology in either the oral or the written production of embedded questions. Notice, in fact, that Chinese displays no T-to-C movement in embedded question, while in Spanish T-to-C movement is the preferred option. If the learning task were especially difficult for L1 Spanish learners of English due to the fact that Spanish can be conceived as a superset of English with respect to T-to-C movement in questions, we would have expected L1 Spanish speakers to produce more inversion errors than L1 Chinese speakers in embedded questions. In contrast, we found that L1 Spanish and L1 Chinese learners of English did not differ in terms of inversion errors in the oral elicited production task, and that L1 Chinese learners made more errors than L1 Spanish learners in the written task. One way to account for the asymmetry between main and embedded questions would be to hypothesize that different learning mechanisms are at play for different types of structures: the L1 might be the basis for simple clauses that get acquired early in L2 development (showing an L1 superset/subset effect), while properties of L2 simple clauses might represent the basis for L2 complex ones (showing no L1 effect).

Another aim of this study was to investigate the effect of question-type and *wh*-words on inversion errors. The L2 literature presents contradictory results with respect to the effect of question type: for example, Eckman et al. (1989) found that acquisition of main yes/no questions implies acquisition of *wh*-questions, but Pienemann, Johnston and Brindley (1988) found that yes/no questions are acquired earlier than *wh*-questions.

The oral production study and the written production study present two slightly different pictures with respect to this effect in main questions: the oral production study showed that inversion errors occur more frequently in *wh*-questions than in yes/no questions, while in the ICLE corpus inversion rates did not differ for the two question-types.

It is possible that the difference between the oral and the written study and the inconsistencies in the literature are due to the different methodologies used to study learners' production. In fact, when context was controlled for (i.e., elicited production), inversion was more frequent in yes/no than in *wh*-questions, while this difference disappeared when context was not controlled for (i.e., semi-spontaneous written production). In other words, while inversion is particularly problematic for learners in main *wh*-questions, this difficulty is obscured in spontaneous production by the fact that pragmatic conditions for non-inversion are met. In a study of spontaneous conversation exchanges, Williams (1990) found that native speakers of American English produced 35% non-inverted yes/no questions in new information questions and 76% non-inversion in confirmation questions, thus suggesting that lack of inversion in yes/no question is the norm in confirmation questions in standard English.

While differences in context might explain some inconsistencies in the literature, it is important to note that the written production patterns of L1 Chinese and L1 Spanish learners showed an effect of question-type on correct and inverted responses (with yes/no questions being associated with higher correct inversion rates), indicating that, at least for these two groups, the overall pattern and extent of inversion errors in written production are similar to those identified in oral production.

With respect to the effect of question-type on embedded questions, the picture that emerges from the oral and the written production studies is pretty homogeneous: embedded inversion only happens in embedded *wh*-questions, and is virtually non-existent in yes/no questions. The structural explanation proposed independently for non-standard varieties of English that display embedded inversion accounts for this pattern by making the fairly uncontroversial assumption that embedded inversion can only apply if a structural position is available as the landing site of the inverted auxiliary; this position is not available when the complementizer *if* is present, because the two elements compete for the same position; as a result, no inversion errors occur in embedded yes/no questions. Notice, however, that there are some reasons to believe that the complementizer *whether* targets the same structural position as *wh*-elements (see footnote 9). The structural explanation should then predict inversion in embedded yes/no questions with *whether*, contrary to facts. In the written corpus, yes/no questions with *whether* were more common than embedded questions with *if*, but virtually no inversion errors were found, making the structural claim less compelling. I currenty do not have a good explanation for why inversion errors did not occur in embedded *whether*-questions in written production. However, it is important to keep in mind both that embedded inversion in this corpus was very low to begin with (around 4%), and that *whether* pertains to a higher, more formal register than *if*[47]. Future work should address the availability of inversion in embedded yes/no questions with *whether* via elicited production and acceptability judgment tasks.

An important fact that emerges when we only examine the written production of L1 Chinese and L1 Spanish writers is that inversion errors in embedded *wh*-questions are fairly low for L1 Spanish learners (5.5%) while they are high for L1 Chinese learners (20.5%). The two groups, on the other hand, did not differ in terms of inversion rates in the oral elicited production experiment. While it is possible that this difference stems from differences in proficiency between the L1 groups in the written corpus, this is hard to determine from the data at hand, because proficiency was not measured and controlled for in the ICLE corpus, and biographical data available for the speakers does not seem conclusive: on average the L1 Chinese writers had studied English in school for longer (11.5 years vs. 7.9 for the L1 Spanish group,) while the L1 Spanish writers had studied English in college for longer (3.4 vs. 1 year for the L1 Chinese

[47] I am not aware of data on inversion rates for *whether*-questions in non-standard varieties of English, probably because *whether* pertains to a fairly high register and is not frequent in spoken language.

group); finally, the two groups did not differ in terms of length of stay in an English-speaking country (2.3 months vs. 3.8 months for the L1 Chinese and L1 Spanish groups on average, respectively).

While the difference in accuracy between the L1 Chinese and L1 Spanish groups in the written study is not easily accounted for on the basis of available data, a crucial fact that emerges from the two studies taken together is that inversion errors in embedded questions were never higher for L1 Spanish speakers, confirming the conclusion that inversion errors in English embedded *wh*-questions are not due to simple transfer of L1 properties.

This study was also aimed at investigating whether different *wh*-words had an effect on inversion rates. In both the oral and the written production studies (and in line with the findings of many cross-linguistic and acquisition studies), we found that *why* was associated with lower inversion rates than other *wh*-elements, and we failed to show an argument/adjunct asymmetry. The existence of a *why*-effect on inversion rates in embedded questions, however, did not surface in the written production task. When we only look at the production of L1 Chinese and L1 Spanish speakers, the existence of a *why*-asymmetry in main questions is confirmed for L1 Chinese learners in main questions, while L1 Spanish learners in main questions exhibited low inversion rates for both *why*- and *where*-questions. Similarly, in embedded *wh*-questions, inversion rates were low for *where*- and *why*-questions for both L1 Chinese and L1 Spanish writers, suggesting the presence of an argument adjunct asymmetry in written production.

A final aim of this study was to compare L2 learners' oral and written production of English interrogative structures to determine whether inversion errors are a pervasive characteristic of the L2 production system or whether errors would not surface when learners are not under real time pressure and are allowed to revise their production. Overall, the picture that emerges from the production studies in this chapter is somewhat inconclusive: inversion errors in main and embedded questions occurred to some extent in both studies, and were more frequent in oral production than in written production. For example, inversion errors in main *wh*-questions in the oral task accounted for 11% of the relevant productions, while they accounted for 4% and 6% of the productions of main *wh*-questions in the ICLE and in the CUNY corpus, respectively. However, the extent of inversion errors in L1 Chinese and L1 Spanish writers was comparable to that seen in the oral production study (around 10%), indicating that the overall lower inversion rates seen in written production might be due to the contribution of language groups with native-like performance (e.g., L1 German, L1 Swedish).

With regards to embedded *wh*-questions, the overall difference between the oral and the ICLE written production study is more dramatic: inversion errors account for 34% of the total production of embedded *wh*-questions in the oral elicited production task, while these errors amounted to only 4% in the ICLE essays overall. Moreover, while L1 Chinese writers exhibited high rates of non-target inversion in the ICLE corpus as well, thus suggesting that inversion errors in L1 Chinese learners' productions are pervasive and not highly task-dependent, this was not the case for L1 Spanish writers, whose inversion rates were around 6% in the written corpus.

The CUNY corpus presents an intermediate picture: in this corpus, inversion errors in embedded *wh*-questions amounted to 15%. While more data needs to be collected in order to conclude something more definitive about the latter corpus, it is worth noticing that the writers from whom the essays in the CUNY corpus were collected were recruited from the same subject pool as the speakers in the elicited production study and had comparable English proficiency levels.

To sum up, inversion errors were present in both oral and written production, but were more frequent in the oral production task, especially in the case of embedded questions. This might be taken as an indication that second language learners' grammars are 'unstable' with respect to some grammatical properties, i.e., that some rules are not fixed in L2 grammatical representations and are particularly likely to surface under pressure. Similarly, it could be hypothesized that second language learners have access to multiple grammars: careful written production might allow speakers to access explicit grammatical rules, while real-time pressure might cause the production system to access non-target procedures/parameters. If it is true, as has been argued by Henry and Tangney (1999) and in slightly different terms by Tornyova and Valian (2009), that a grammatical system where the same rules apply to all relevant contexts (questions, in our case) is simpler than a grammatical system where different rules apply to similar contexts (main and embedded questions, in our case), it could be argued that the production system under pressure might end up accessing simplifying grammatical procedures/options (i.e., a grammar without obligatory movement). The obvious objection to this kind of hypothesis has to do with the fact that inversion rates are not the same in main and embedded questions, because L2 grammars are target-like most of the time (i.e., inversion is the norm in main questions and non-inversion is the norm in embedded questions). A variational approach such as the one proposed by Yang (2002, 2004) and further developed by Legate and Yang (2007), might be able to account for the type of accuracy profiles that emerged in these studies. In this model, different hypotheses compete to best account for the native adult input by being accessed probabilistically and then being punished or rewarded based on their success. To this, one might add that when the system is under pressure, simpler grammatical hypotheses are more easily accessed by the developing system under pressure.

Alternatively, the Acquisition by Processing Theory model (Truscott & Sharwood Smith, 2004), in which there are no dedicated language acquisition mechanisms and in which language development is seen a consequence of processing procedures, predicts the intermittent appearance of non-target productions in L2 speakers. According to this model, the production system is shared between the two (or more) languages of a speaker, and L1-transfer is a result of competition between L1 and L2 procedures. The appearance of non-target productions is thus an effect of the L1 procedure having won the competition, possibly due to its ease for the production system. The theory thus explains why non-target productions might still appear in the speech of advanced speakers. While the model was initially proposed as an alternative to the traditional view that L1-transfer is the result of erroneous parameter setting in L2, given its reliance on UG, it might be possible to extend it to explain intermittent non-target productions that cannot be imputed to the speaker's L1, but that might be due to UG defaults.

2.4. Present Study: L2 Acceptability Judgments of English Questions

A central issue in the study of second language acquisition concerns the distinction between implicit, procedural linguistic knowledge on one hand and explicit, declarative knowledge on the other hand. According to R. Ellis (2004), a number of features can be used to distinguish between implicit and explicit knowledge: metalinguistic awareness (high for explicit knowledge, low for implicit knowledge), type of knowledge (procedural for implicit, declarative for explicit), degree of systematicity in linguistic behavior (high for implicit, low for explicit knowledge), accessibility (automatic for implicit, controlled for explicit knowledge), self-report (non-verbalizable for implicit, verbalizable for explicit) and learnability (implicit knowledge is subject to maturational constraints, while explicit knowledge can be acquired throughout the lifespan).

As summarized in Bowles (2011), three main positions can be distinguished with respect to this debate in the L2 literature: the non-interface position (Hulstijn, 2002; Krashen, 1981), the strong interface position (Sharwood Smith, 1981; DeKeyser, 1998), and the weak interface position (R. Ellis, 1993; N. Ellis, 1994; Schmidt & Frota, 1986). According to the first hypothesis, the two types of knowledge are separate, develop autonomously, and are subserved by different cognitive mechanisms (Paradis, 1994; Ullman, 2004). According to the strong interface position, on the other hand, explicit knowledge can be proceduralized and become implicit through practice and usage. Finally, according to the weak interface hypothesis, explicit knowledge has a limited influence on implicit knowledge.

Regardless of what hypothesis turns out to better characterize the nature of L2 grammars, it is crucial for researchers in the field to know what kind of knowledge different performance tasks tap into. Surprisingly enough, few studies in the literature have investigated this question directly. Reasearch by R. Ellis (2004, 2005), recently replicated by Bowles (2011), showed that scores from different tasks load onto different factors: oral imitation, oral narration, and timed grammaticality judgments load one factor (hypothesized to be implicit knowledge), while scores in untimed grammaticality judgment and metalinguistic knowledge tasks loaded onto another factor (hypothesized to be explicit knowledge).

A central question that needs to be asked with respect to L2 learners' production patterns in English interrogative structures has to do with the nature of such errors, i.e., whether they stem from lack of knowledge or lack of automatization/implementation of target grammatical procedures. If we assume that L2 learners have access to complete explicit knowledge of word order patterns for English interrogative structures while their implicit knowledge might be influenced by non-target procedures, we expect them to be prone to non-target productions or judgments in tasks where learners draw heavily on implicit, rather than explicit, knowledge (i.e., oral production and timed acceptability judgments vs. written production and untimed grammaticality judgments).

However, this contraposition is likely to be unrealistically dichotomous, given that while it has been shown that the different performance tasks load onto different factors, this does not mean that they load *exclusively* onto them. In other words, it is unlikely that in written production and untimed acceptability judgments, learners will *only* access explicit grammatical knowledge, while in oral production and timed acceptability tasks they will only have access to implicit knowledge. The results from the oral and the written production studies seem to confirm this nuanced view: while overall inversion errors were somewhat higher in the oral production task, inversion errors also occurred systematically in the written production of L1 Chinese and L1 Spanish L2 learners of English. Moreover, similar qualitative patterns emerged from the two tasks, in that we found both a question-type and a *wh*-type asymmetry in L1 Chinese and L1 Spanish writers' main and embedded questions. I interpret these findings to mean that, by and large, intermediate/advanced L2 learners have acquired the English word order pattern, and that production errors are due to occasional use of non-target procedures that crucially do not derive from the learners' L1s. Oral and written production, when present, would thus differ only in how susceptible learners are to the appearance of these non-target procedures. If occasional errors in writing are to be interpreted as the temporary appearance of non-target procedures, it is expected that the specific error patterns in the oral and the written production studies will be similar, given that they are due to the same cause.

L2 learners' acceptability judgments are expected to resemble, by and large, those of native speakers in that inverted main questions should be considered more acceptable than non-inverted ones and non-inverted embedded questions more acceptable than inverted ones. If specific production errors are to be reflected in acceptability patterns, we also expect acceptability patterns to reflect a question-type asymmetry and a *wh*-type asymmetry. Based on the oral production results, we might thus expect:

a. non-inverted main *wh*-questions to be judged as more acceptable than non-inverted main yes/no questions, and/or inverted main *wh*-questions to be judged as less acceptable than inverted yes/no questions.[48]
b. non-inverted main *why*-questions to be judged as more acceptable than other non-inverted *wh*-questions, and/or inverted main *why*-questions to be judged as less acceptable than other inverted *wh*-questions
c. inverted embedded *wh*-questions to be judged as more acceptable than inverted yes/no questions, and/or non-inverted embedded *wh*-questions to be judged as less acceptable than non-inverted yes/no questions.
d. inverted embedded *why*-questions to be judged as less acceptable than other inverted *wh*-questions and/or non-inverted embedded *why*-questions to be judged as more acceptable than other non-inverted embedded questions.

Additionally, we expect timed acceptability judgments to be more prone to non-target judgments. This means that not all of the predictions outlined in (a-d) above might be borne out in the untimed acceptability judgment task.

Finally, L2 learners' judgments are expected to differ from those of native speakers in that only L2 learners' grammars should occasionally access non-target procedures; moreover, such non-target procedures should be more likely to surface in timed than in untimed acceptability judgments. This might entail that L2 learners' acceptability judgments will be less

[48] This asymmetry might not arise in acceptability judgments, given the fact that non-inverted main yes/no questions are grammatical in English under specific pragmatic circumstances.

sharp than those of native speakers, due to the intrusion of non-target procedures in L2 judgments. However, L2 learners' judgments might be less sharp than those of native speakers for other reasons (e.g., degree of certainty). I propose that a better way to investigate whether L2 learners occasionally apply non-target procedures in their grammars is to measure how often L2 learners, compared to monolinguals, will judge non-target utterances as acceptable and target utterances as non-acceptable, and to investigate whether the rates of nonnative-like judgments increases from untimed to timed judgments. Given that in the current studies a magnitude estimation task was used to measure acceptability patterns in L2 learners, the amount of non-target judgments was measured in terms of how often non-target questions received a score of 1 standard deviation (SD) or more below the subject's mean and how often target questions received a score of 1SD above the subject's mean.

2.4.1. Untimed Acceptability Judgments

2.4.1.1. Method

2.4.1.1.1. Participants

One of the aims of the elicited production study was to examine the role of L1 as one of the possible causes for inversion errors in the acquisition of English interrogatives. As such, participants included L2 learners of English whose L1s were either Chinese or Spanish.

In the oral elicited production of English main questions, L1 Spanish speakers produced overall more inversion errors than L1 Chinese speakers, while there was no effect of L1 on inversion errors in embedded questions and no interaction between L1 and other factors (question type and *wh*-type). In the two written corpus studies, a wider range of L1 backgrounds was examined. The effect of L1 was a significant predictor of inversion in the ICLE corpus, for both main and embedded questions. However, English proficiency was not controlled for in the corpus and it is thus not possible to disentangle the effects of proficiency from those of L1 background. Once speakers' L1s were grouped with respect to availability of subject-auxiliary inversion or subject-verb inversion in main and embedded questions, a more complex picture emerged. Obligatoriness of subject-auxiliary inversion in main questions correlated with higher rates of subject-auxiliary inversion in English main questions, while no such effect was found in embedded questions.

In the present experiment, speakers from a variety of L1 backgrounds were included and proficiency was measured via a shortened version of the Michigan Test of English Language Proficiency (MTELP).

Participants were recruited mainly through the Introduction to Psychology subject pools at Queens College and Hunter College. Participants received course credit for their participation in the study. One hundred participants (35 English monolingual speakers and 65 L2 speakers) were tested.[49] Participants were considered L2 speakers if they were not born in an English-speaking country and had moved to the US at or after age 14. Eight participants were excluded because they moved to the US after age 14 but were born in a country where English is an official language (e.g., India, Abu Dhabi, Pakistan). An additional five participants were

[49] An additional 160 participants were tested but were not included in the analyses because they either moved to an English-speakgin country before age 14, or were not monolingual native-speakers.

excluded because their MTELP scores were 2 or more standard deviations below the average of the other L2 participants' scores. Data from a total of 87 participants (35 English monolingual speakers and 52 L2 speakers) underwent further analyses. L2 participants were judged to be intermediate/advanced learners with respect to their English proficiency. Table 44 provides a summary of participants' proficiency scores, age, age of arrival and years of stay in the US. English controls were all monolingual English speakers. Their mean age was 22 (SD = 9.1) and their average MTELP score was 43/45 (SD = 1.8).

Table 44: Demographics of L2 participants – Untimed Judgment Experiment

L1	N	Average age (SD)	Average age of arrival (SD)	Average years of stay in US (SD)	Average MTELP (SD)
Albanian	1	26	17	9	44
Bengali	1	39	25	14	36
Burmese	1	23	16	7	39
Chinese	19	22.9 (4.2)	19.7 (3.8)	3.2 (3.7)	33.5 (6.8)
French	2	32	28	4	32
German	1	32	24	8	45
Hebrew	4	24.7 (6.6)	22.7 (7)	2 (1.6)	42.3 (2)
Italian	2	31	28.5	2.5	41.5
Japanese	1	27	21	6	43
Korean	4	23 (3)	17.5 (4.5)	5.5 (1.8)	39.8 (2.2)
Nepali	1	23	14	9	43
Polish	1	27	19	8	43
Portuguese	2	24.5	24.5	0	39
Romanian	1	24	23	1	34
Russian	5	35 (9.2)	22.4 (9.7)	12.6 (8.1)	40.4 (2.9)
Spanish	5	27.8 (5.2)	16.8 (3.8)	11 (5.5)	39.8 (5.5)
Tibetan	1	20	15	5	40
Ukranian	1	21	20	1	44
Total	*52*	*26 (6.3)*	*20.6 (4.9)*	*6.6 (6.5)*	*37.8 (6.4)*

2.4.1.1.2. Materials

The study consisted of a computerized magnitude estimation acceptability task, administered through the E-prime 2.0 software (Psychology Software Tools, Inc.). Magnitude estimation is a technique adapted from psychophysics where it is used to measure the perception of light or sound, among other phenomena. Magnitude estimation was initially adapted to linguistic research by Bard, Robertson, and Sorace (1996) and has been used extensively since. In a typical magnitude estimation task, participants are asked to judge stimuli with respect to one another (and to a modulus[50]) by using numbers selected by the participants. Magnitude estimation presents some advantages over other types of acceptability scales. For example, participants' judgments can be more fine-grained and are not bound to a predetermined scale. Moreover, while the dependent variable in standard Likert or 1–7 point scale experiments is considered to be ordinal, the dependent variable in magnitude estimation experiments is considered to be an interval or ratio variable, making it possible to use standard parametric tests.

Four experimental lists were constructed, each containing 32 main questions, 32 embedded questions, and 64 fillers. Half of the fillers were relative clauses (subject and object, with and without the complementizer *that*) and 32 were long distance questions (subject and non-subject, with and without the complementizer *that*). Each sentence frame (item) occurred in the four experimental lists (either as an inverted yes/no question, a non-inverted yes/no question, an inverted *wh*-question or a non-inverted *wh*-question), but each participant saw only one version of each experimental item. Each participant was randomly assigned to one of the four

[50] A modulus is an initial stimulus, judged by the experimenter to be of middle acceptability.

experimental lists and was presented with 128 sentences. The two fully *within* factors were question type (yes/no vs. *wh*-) and inversion (inverted vs. non-inverted). The type of *wh*-word in *wh*-questions was also manipulated. There were four types of *wh*-words: arguments *what* and *who* and adjuncts *why* and *where.*[51] Type of *wh*-word was a within-subjects factor but a between-items factor. Each participant was presented with a total of 32 main and 32 embedded questions,[52] as follows:

- 16 yes/no:
    - 8 with inversion (4 with *be*; 4 with *have*)
    - 8 without inversion (4 with *be*; 4 with *have*)

- 16 *wh*-questions
    - 8 with inversion (4 with *be*; 4 with *have*)
        - 2 *who*, 2 *what*, 2 *where*, 2 *why*
    - 8 without inversion (4 with *be*; 4 with *have*)
        - 2 *who*, 2 *what*, 2 *where*, 2 *why*

[51] While *where* can function either as an argument (e.g., where did you put the book?) or an adjunct (e.g., where did you have dinner?) depending on the verb, it always functioned as an adjunct in our experimental materials.

[52] Half of the inverted embedded yes/no questions contained *if*, while half did not. This condition was meant to investigate whether L2 learners would allow inversion in embedded yes/no questions when the complementizer was absent, as is the case in non-standard varieties of English (e.g., AAVE, Hiberno English, etc.). This condition could not be mirrored in the non-inverted yes/no questions, given that a yes/no question without inversion and without *if* would be interpreted as a declarative without *that* (e.g., She didn't know John had left).

Materials for the four experimental lists are provided in Appendix B.

2.4.1.2. Procedure

After reading and signing the consent forms, participants sat in front of a computer. The instructions for the experimental task followed closely those given in Johnson (2008). The concept of numerical magnitude estimation was introduced by using line length. Participants were instructed to make estimates of line length relative to the first line they would see, the reference line. As an example, they were told to assign the reference line the value of 150, and to assign a number to each following line based on this value.

After this, participants were told to perform the same task with language materials. The concept of linguistic acceptability was defined as "sounds good to you". An example of a sentence judged to be of middle acceptability ("Jamie gave to her friend a big brown sweater") was provided. The example sentence was assigned the value of 150. Participants were told that they could use any number they wanted as long as they judged sentences with respect to each other. Participants were also instructed to use a wide range of numbers, in order to distinguish as many degrees of acceptability as possible. It was also emphasized that there were no 'correct' answers and that they should base their judgments on first impressions, not spending too much time to think about any given sentence. They were also told that the experiment contained no spelling errors and that they should rate each sentence based on the 'order of the words in the sentence'. Participants were instructed to always press 'enter' to move on to the next slide/sentence. They could not revisit previous sentences or change their responses on previous items. There was no time limit for sentence presentation or response. Items were presented in random order, with a new randomization being generated for each participant.

At the end of the experimental session, participants were administered the MTELP. The study took approximately forty-five minutes.

2.4.1.3. Data analysis

As is standard for magnitude estimation analyses, raw judgments were first log-transformed. This transformation has the effect of transforming a multiplicative scale into an additive scale. Subsequently, the average judgment score was calculated for each participant. Log-transformed judgment scores were then transformed into Z-scores by subtracting the participant's mean score computed over all materials from the score assigned by that participant to each sentence and dividing this number by the standard deviation for that participant. This way of expressing scores has the effect of standardizing subjects' judgment scores with mean = 0 and standard deviation = 1. Z-scores above 3 or below −3 were excluded from further analysis.

Reaction times for each sentence were recorded (the RTs for each sentence consisted of the time needed to read the sentence and to judge it). Stimuli for which RTs were less than 1 second or more than 40 seconds were excluded from the analyses. All the figures in this section display means of normalized log-transformed judgments.

2.4.1.4. Main Questions: Results and Interim Discussion

The data were first inspected to see whether inverted main questions were consistently rated higher than non-inverted questions across L1s. Overall, participants showed a clear preference for inverted over non-inverted main questions (F_1 (1,85) = 178, p < .0001; F_2 (1,31) = 339, p <

.0001)[53]. There was also a significant interaction between the language groups (L2 learners vs. monolinguals) (F_1 (1,85) = 31, p < .0001; F_2 (1,31) = 90, p < .0001), indicating that monolinguals' judgments were sharper than those of L2 learners. Monolinguals judged inverted questions significantly higher than L2 learners (t_1 (85) = 3.9, p < .0001; t_2 (31) = 8.8, p < .0001) and non-inverted ones significantly lower (t_1 (85) = −3.7, p < .0001; t_2 (31) = −5.9, p < .0001). On the other hand, there wasn't a clear pattern with respect to RTs. For some language groups, inverted questions were associated with longer RTs, while for others the opposite was true. The overall difference in RTs between inverted and non-inverted main questions was not significant.

Table 45 presents the average Z-scores and RTs for inverted and non-inverted main questions for the different language groups.

---

[53] Only the L1 Burmese and L1 Ukranian speakers preferred, on average, non-target main questions to target ones.

Table 45: Z-scores and RTs for main questions by L1 and question-type – Untimed Judgment Experiment

L1	Average Z-scores for inverted questions	Average Z-scores for non-inverted questions	Average RTs (ms) for inverted questions	Average RTs (ms) for non-inverted questions
Monolingual	0.62	−0.37	5416	5448
Albanian	0.72	−0.70	6881	5724
Bengali	0.43	0.36	10754	10865
Burmese	**0.00**	**0.16**	4273	3988
Chinese	0.32	−0.01	6623	6748
French	0.09	−0.12	8217	9091
German	0.68	−0.57	6021	5010
Hebrew	0.32	−0.24	8800	8799
Italian	0.46	−0.03	8945	8637
Japanese	0.78	−0.77	9789	8668
Korean	0.12	0.06	6209	6139
Nepali	−0.70	−1.08	5763	5332
Polish	−0.92	−0.41	2252	2437
Portuguese	0.60	−0.12	9495	7871
Romanian	0.45	0.40	5958	5527
Russian	0.53	0.04	10013	10764
Spanish	0.57	−0.14	6188	6524
Tibetan	0.13	0.08	4200	3981
Ukrainian	**0.37**	**0.50**	6625	6411
Total L2	*0.33*	*−0.08*	*7175*	*7166*

2.4.1.4.1. Question-type

An issue that I wanted to investigate in this experiment was whether L2 learners' acceptability judgments would mirror the production patterns seen in oral production, where participants were inverting more frequently in yes/no questions than in *wh*-questions despite the fact that lack of inversion is present in the English native input for yes/no questions.

Table 46 and Table 47 present the average Z-scores and RTs for L2 learners and monolingual English speakers for inverted and non-inverted *wh*- and yes/no questions.

Table 46: Average Z-scores for main questions by question-type, inversion and L1 – Untimed Judgment Experiment

Language Group	Question-Type			
	Wh-		Yes/No	
	inverted	non-inverted	inverted	non inverted
L2	0.29	−0.15	0.38	0.00
Monolingual	0.64	−0.77	0.61	0.04
Total	*0.43*	*−0.40*	*0.47*	*0.01*

Table 47: Average RTs for main questions by question type, inversion and language group – Untimed Judgment Experiment

Language Group	Question-Type			
	Wh-		Yes/No	
	inverted	non-inverted	inverted	non inverted
L2	7378	7418	6968	6912
Monolingual	5398	5152	5434	5740
Total	**6581**	**6519**	**6343**	**6442**

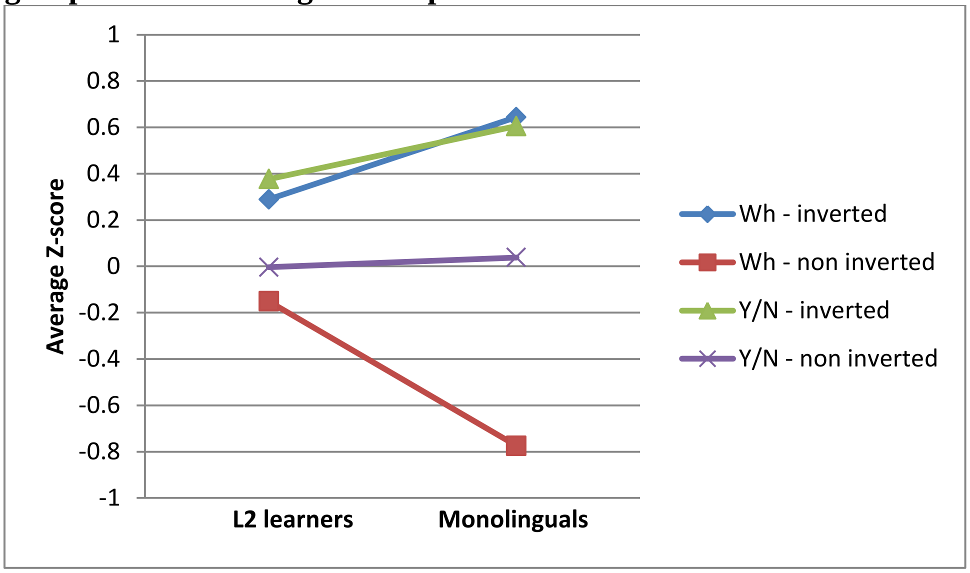
Acceptability Judgments

A first 2 (language group) x 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Question-type had a significant effect on Z-scores (F_1 (1,85) = 39, p < .0001; F_2 (1,31) = 76, p < .0001), and so did inversion (F_1 (1,85) = 181, p < .0001; F_2 (1,31) = 316, p < .0001). Moreover, all the interactions were significant: there was a significant two-way interaction between language group and question-type (F_1 (1,85) = 9.4, p = .003; F_2 (1,31) = 16, p < .0001), a significant two-way interaction between language group and inversion (F_1 (1,85) = 31, p < .0001; F_2 (1,31) = 82, p < .0001), a significant two-way interaction between question-type and inversion (F_1 (1,85) = 33, p < .0001; F_2 (1,31) = 56, p <

.0001), and a significant three-way interaction between language group, inversion, and question-type (F_1 (1,85) = 25, p < .0001; F_2 (1,31) = 34, p < .0001).

The three way interaction indicates that the interaction between question-type and inversion depended on language group. The interaction between question-type and inversion was in fact significant only for monolinguals (monolinguals: F_1 (1,34) = 19, p < .0001; F_2 (1,31) =58, p < .0001; L2 learners: all *F*s < 1). While monolingual speakers judged inverted *wh*- and yes/no questions as equally acceptable, they judged non-inverted yes/no questions as significantly more acceptable than non-inverted *wh*-questions (t_1 (34) = 6.8, p < .0001; t_2 (31) = 10.4, p < .0001). Even though the interaction was not significant for L2 learners, I decided to investigate whether this group would also judge non-inverted yes/no questions as more acceptable than non-inverted *wh*-questions. This difference was only marginally significant in the subject analysis but was significant in the item analysis (t_1 (51) = 1.9, p = 0.6; t_2 (31) = 3.3, p = .002). The pattern of acceptability is illustrated in Figure 10.

Figure 10: Average Z-scores for main questions by question-type, inversion and language group – Untimed Judgment Experiment

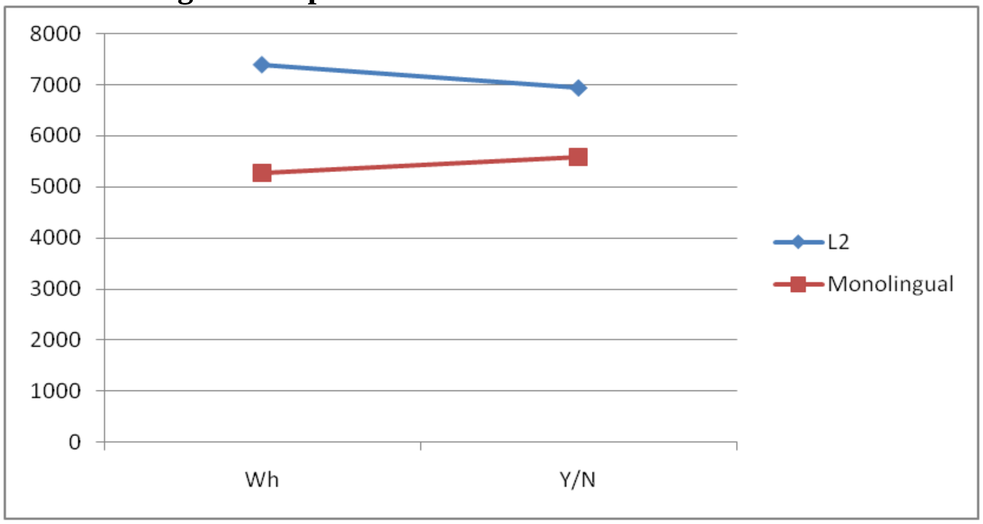


L2 learners' judgments are significantly less sharp than those of monolinguals. However, as discussed in the introduction to this section, this fact might reflect that L2 learners are less certain about their judgments and are thus less likely to use a large range of values to distinguish differences in acceptability, or that L2 learners' judgments are noisier than those of monolingual speakers, in that L2 learners are more prone to use non-target procedures to judge English sentences. A way to investigate whether L2 judgments are occasionally non-target-like is, as suggested in the introduction, to measure how often they judge non-target sentences as acceptable and vice versa. To do this, I measured how often L2 learners and monolingual native speakers judge non-inverted main questions with a positive score (1SD or more above the average) and how often they judge inverted main questions with a negative score (1SD or more below the average). L2 learners judged non-target non-inverted questions with a score above the average 7% of the time, while monolingual speakers did so 5% of the time. This difference was not significant in the subject analysis but significant in the item analysis (t_1 (83) = 1, n.s.; t_2 (31) = 2, p = .051). On the other hand, L2 learners judged target-like inverted main questions with a negative score 10% of the time, while monolingual speakers did so only 3% of the time. This difference was significant (t_1 (76) = 3.2, p = .002; t_2 (31) = 4.4., p < .001). While it is clear that this question should be pursued further by using a different type of task (e.g., forced choice paradigm or categorical grammaticality task) the present results indicate that L2 learners differ from native speakers in that they are more prone to non-target judgments of English main questions.

Reaction Times

A 2 (language group) x 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with RTs as the dependent variable was conducted. The main effect of language group was significant (F_1 (1,85) = 10.6, p = .002; F_2 (1,31) = 182, p < .0001), indicating that, overall, monolingual speakers' reaction times were faster than those of L2 learners. The interaction between language group and question-type was also significant (F_1 (1,85) = 7.9, p = .006; F_1 (1,31) = 5.5, p = .025), indicating that the difference in reaction times between monolinguals and L2 speakers was more pronounced for *wh*-questions than for yes/no questions. Nothing else was significant (all *F*s < 1).

Figure 11: Average RTs for main questions by question-type and language group – Untimed Judgment Experiment



2.4.1.4.2. Wh-type

The second issue that I wanted to investigate in this study was whether L2 learners' acceptability judgments would mirror the production patterns seen in elicited production, where participants were inverting less frequently with *why* relative to other *wh*-elements.

Lee (2008) found that acceptability patterns for non-inverted main questions in L2 learners differed depending on whether the *wh*-word was an argument or an adjunct. Specifically, non-inverted adjunct questions were judged as significantly more acceptable than non-inverted argument questions.

The results from the present elicited oral production study and the written corpus study, on the other hand, did not show the existence of an argument/adjunct distinction in production. Table 48 and Table 50 present the average Z-scores for L2 learners and monolingual English speakers, respectively, for inverted and non-inverted *wh*-words.

Table 48: L2 learners' average Z-scores for main questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	0.37	-0.11
where	0.33	-0.24
who	0.09	-0.30
why	0.36	**0.04**
Total	*0.29*	*-0.15*

Table 49: L2 learners' average RTs for main questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	7653	7717
where	6784	7748
who	7833	7572
why	7235	6627
Total	*7175*	*7166*

Figure 12: L2 learners' average Z-scores for main questions by *wh*-word and inversion – Untimed Judgment Experiment

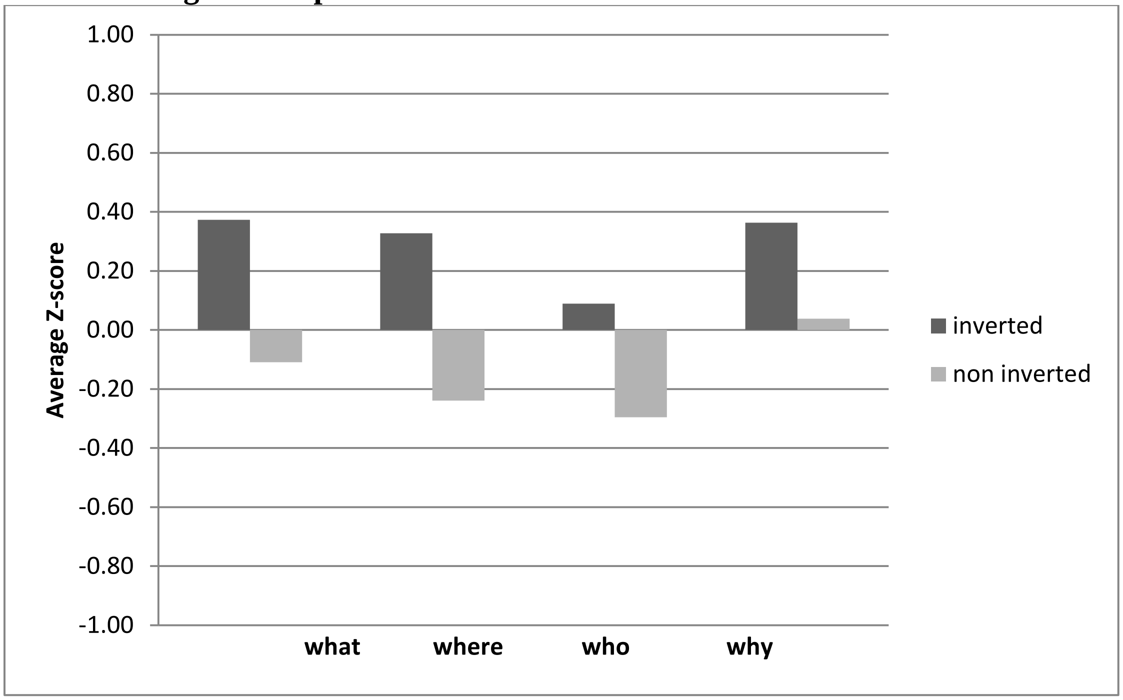


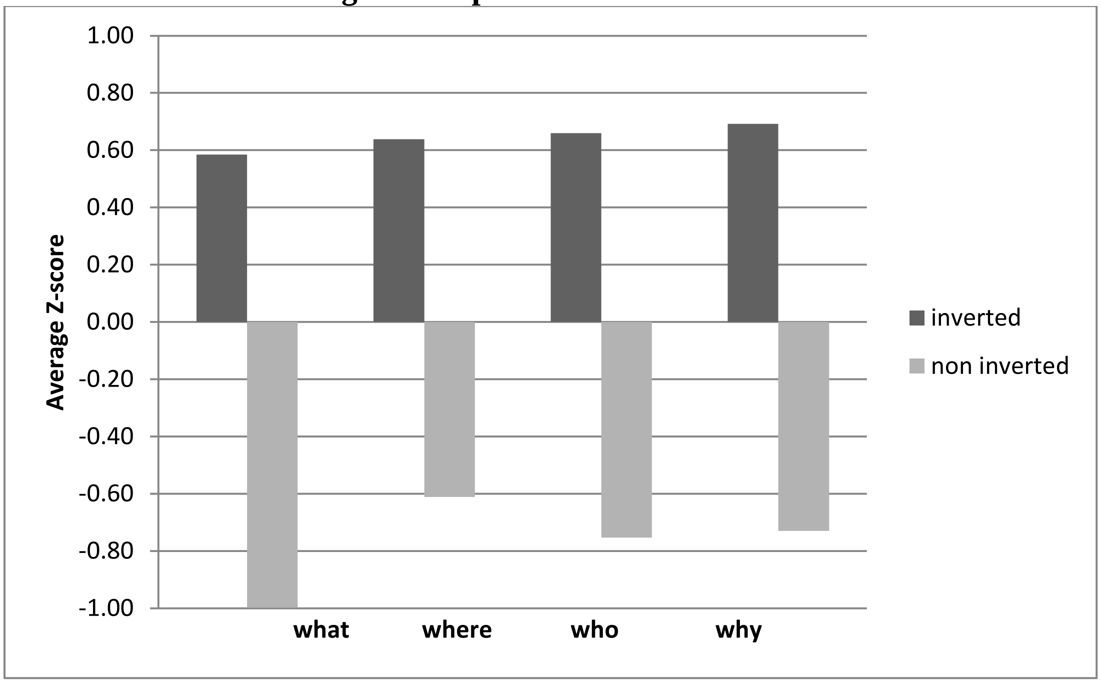
Table 50: Monolingual speakers' average Z-scores for main questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	0.59	−1.00
where	0.64	−0.61
who	0.66	−0.75
why	0.69	−0.73
Total	*0.64*	*−0.77*

Table 51: Monolingual speakers' average RTs for main questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	5594	5391
where	5150	5467
who	5611	4845
why	5236	4876
Total	*5416*	*5448*

Figure 13: Monolingual speakers' average Z-scores for main questions by *wh*-word and inversion – Untimed Judgment Experiment



Acceptability Judgments

A 2 (language group) x 4 (*wh*-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,82) = 193, p < .0001; F_2 (1,28) = 436, p < .0001), showing that inverted questions were judged significantly higher than non-inverted ones. The effect of *wh*-type was marginal in the subject analysis and non-significant in the items analysis (F_1 (3,246) = 2.6, p = .06; F_2 (1,28) =1.2, p = .3). The only significant interactions were between language group and inversion (F_1 (1,82) = 55, p < .0001; F_2 (1,28) = 123, p < .0001), and between *wh*-type and language group (F_1 (3,246) = 3.8, p = .01; F_2 (3,18) = 6.4, p = .002). The former interaction indicates that monolinguals' acceptability patterns are sharper than those of L2 learners for both inverted and non-inverted questions, while the second interaction shows that the two language groups behaved differently with respect to different *wh*-words.

Although the three-way interaction between language type by group by inversion was not significant in the subject analysis, (F_1 (3,246) = 1.4, p = .2; F_2 (1,31) = 82, p < .0001), visual inspection suggested that L2 learners treat inverted *who*-questions differently from other inverted *wh*-questions and, more importantly, non-inverted *why* differently from other *wh*-words, as predicted on the basis of the oral production data. For this reason, I compared acceptability judgments for inverted and non-inverted *wh*-questions in L2 learners. For inverted sentences, there was a significant difference between *who* and other *wh*-words (subject analysis: *who* vs. *what*, p = .016; *who* vs. *where*, p = .02, *who* vs. *why*, p = .017; item analysis: *who* vs. *what*, p = .027; *who* vs. *why*, p = .036).[54] However, these differences disappeared once Bonferroni correction for multiple comparisons was applied. In non-inverted sentences, there was a significant difference between *why* and *where* and between *why* and *who* (subject analysis: *why* vs. *where*, p = .009; *why* vs. *who*, p = .02; item analysis: *why* vs. *where*, p = .045; *why* vs. *who*, p = .015). The only difference that remained significant once Bonferroni correction was applied was that between adjuncts *where* and *why* in the subject analysis (p = .05).

---

[54] This pattern might be explained in a similar fashion as the high number of subject responses obtained with *who* in the elicited production experiment. L1 learners have been shown to follow a canonical bias, interpreting surface NVN sentences as agent-action-theme sequences. If L2 learners have the same bias, they will tend to interpret the first N in the sentence (*who*) as the subject/agent. This will lead to a garden path once they find the second NP in the inverted questions, causing them to judge object *who*-questions as less acceptable than other non-subject *wh*-questions. This would not happen in *what*-questions, because *what* refers to inanimate objects and, as such, tends to be non-agentive.

In order to investigate whether L2 judgments are occasionally non-target-like, I measured how often L2 learners and monolingual native speakers judge non-inverted main *wh*-questions with a positive score (1SD or more above the average) and how often they judge inverted main questions with a negative score (1SD or more below the average). L2 learners judged non-target non-inverted *wh*-questions with a score above the average 7% of the time, while monolingual speakers did so 3.7% of the time. The overall difference was not significant in the subject analysis and marginal in the item analysis (F_1 (1, 84) = 2.1, p = .15; F_2 (1,28) = 3.6, p = .07); there was no significant effect of *wh*-word and no interaction (all *Fs* < 1). L2 learners judged target-like inverted main *wh*-questions with a negative score around 10% of the time, while monolingual speakers did so only 2.5% of the time. While the overall difference was significant (F_1 (1,84) = 8.1, p = .005; F_2 (1,28) = 23.4., p < .001), there was no significant effect of *wh*-word and no interaction (all *F*s < 1).

Reaction Times

A 2 (language group) x 4 (*wh*-type) x 2 (inversion) mixed deisgn ANOVA with RT as the dependent variable was conducted. The effect of language group was significant (F_1 (1,83) = 13.4, p < .0001; F_2 (1,28) = 106.7, p < .0001), indicating that, on average, monolinguals' RTs were faster than those of L2 learners. Nothing else was significant (all *Fs* < 1).

The present experiment investigated whether production patterns of inversion would be reflected by acceptability judgments. Second language learners differed from English native speakers in that they occasionally produced non-inverted main questions, while native speakers never did. This pattern was particularly pronounced in the oral production, a task that is considered to tap into implicit knowledge, and somewhat reduced in written production, a task in which the effects of explicit linguistic knowledge might be more robust.

With respect to acceptability judgments, the main difference between L2 learners and English monolingual speakers was that the judgments of the latter group were sharper (higher for inverted and lower for non-inverted questions).This difference might be due to the fact that L2 learners' judgments are occasionally non-target, in that L2 learners might sometimes find a non-inverted question acceptable or an inverted question un-acceptable. This question does not receive an immediate, direct answer from this study due to the very nature of magnitude estimation, a task used to probe into relative rather than absolute judgments. However, by looking at how often L2 learners and monolingual speakers' Z-scores for non-inverted question are 1SD or more above the speaker's average judgment and how often theirs scores for inverted questions are 1SD or more below the speaker's average judgment, we have found that L2 learners differ from monolingual native speakers in that they sometimes assign a low value to a grammatical inverted question and vice versa. This result suggests that, on occasion, L2 learners might apply non-target procedures while judging English sentences, even in an untimed task, which is traditionally thought to load more heavily on explicit linguistic knowledge.

With respect to the effect of question-type, L1 Chinese and L1 Spanish learners' inversion rates in the oral and writtern production studies were significantly higher for yes/no than for *wh*-questions. This pattern is surprising if we consider that English allows non-inverted yes/no questions under specific pragmatic circumstances, while non-inverted *wh*-questions are never grammatical. The present acceptability judgment study showed that while both learners and monolingual speakers judged inverted yes/no and inverted *wh*-questions as equally acceptable, both groups judged non-inverted yes/no questions as more acceptable than non-inverted *wh*-questions. This pattern is expected given the properties of English, and it suggests that the question-type asymmetry seen in the oral production study might be production-specific to speech.

In the elicited production experiment and in the written corpus study, we found that *why*-questions differ significantly from other *wh*-questions in that they are associated with significantly lower inversion rates. This result was partially replicated in the present experiment: L2 learners, but, crucially, not monolingual speakers, showed a tendency to judge non-inverted *why*-questions as more acceptable than other non-inverted *wh*-questions. This difference reached significance only when *why*-questions were compared with *where*-questions. No argument/adjunct asymmetry was found in this experiment.

2.4.1.5. Embedded questions: Results and Interim Discussion

The data were first inspected to see whether non-inverted embedded questions were consistently rated higher than inverted embedded questions across L1s. Overall, participants showed a clear preference for non-inverted over inverted embedded questions (F_1 (1,85) = 200, $p < .0001$; F_2 (1,31) = 605, $p < .0001$); all groups judged inverted embedded questions as less acceptable than non-inverted ones. There was also a significant interaction between language group and inversion (F_1 (1,85) = 13.5, $p < .0001$; F_2 (1,31) = 29, $p < .0001$). Monolinguals judged inverted questions significantly lower than L2 learners (t_1 (85) = −4.9, $p < .0001$; t_2 (31) = −9.3, p <

.0001) while there was no difference with respect to grammatical non-inverted ones (t_1 (85) = −.1, p = .9; t_2 (31) = −.1, p = .9).

The difference in RTs between inverted and non-inverted embedded questions was significant (F_1 (1,85) = 7.9, p = .006; F_2 (1,31) = 5.4, p = .03), indicating that, in general, RTs for grammatical non-inverted embedded questions were slower than those for inverted ungrammatical ones. The effect of language group was significant (F_1 (1,85) = 11.2, p < .0001; F_2 (1,31) = 129, p < .0001), indicating that monolinguals' RTs were faster than L2 learners'. There was no significant interaction between inversion and language group (all Fs ≤ 1).

Table 52 presents the average Z-scores and RTs for inverted and non-inverted embedded questions for the different language groups.

Table 52: Average Z-scores and RTs for embedded questions by inversion and L1 – Untimed Judgment Experiment

L1	Average Z-scores for inverted embedded questions	Average Z-scores for non-inverted embedded questions	Average RTs (ms)for inverted embedded questions	Average RTs (ms) non-inverted embedded questions
Monolingual	−0.71	0.33	6324	6958
Albanian	−1.20	0.52	5907	9573
Bengali	−0.58	0.21	13931	14780
Burmese	−0.26	0.16	5862	5583
Chinese	−0.27	0.28	7147	7982
French	−0.33	0.37	10072	7713
German	−0.83	0.62	5442	8567
Hebrew	−0.43	0.28	9421	9504
Italian	−0.69	0.39	13400	12207
Japanese	−0.89	0.46	10502	10783
Korean	−0.07	0.17	7217	7757
Nepali	0.14	0.40	5786	4698
Polish	0.16	0.46	2654	2815
Portuguese	−0.45	0.60	10632	12095
Romanian	−0.44	0.36	6763	7360
Russian	−0.22	0.43	13439	12452
Spanish	−0.34	0.42	8042	7753
Tibetan	−0.08	0.16	4710	4281
Ukrainian	−0.03	0.52	8017	7224
Total L2	*−0.32*	*0.34*	*8352*	*8653*

2.4.1.5.1 Question-type

An issue that I wanted to investigate in this experiment was whether L2 learners' acceptability judgments would mirror the patterns seen in the two production studies, where participants were inverting more frequently in embedded *wh*- than embedded yes/no questions. Table 53 and Table 54 present the average Z-scores and RTs, respectively, for L2 learners and monolingual speakers.

Table 53: Average Z-scores for embedded questions by question-type, inversion and language group – Untimed Judgment Experiment

Language Group	**Question-Type**			
	Wh-		**Yes/No**	
	inverted	**non-inverted**	**inverted**	**non inverted**
L2	−0.14	0.33	−0.51	0.34
Monolingual	−0.63	0.41	−0.79	0.25
Total	*−0.33*	*0.37*	*−0.62*	*0.31*

Table 54: Average RTs for embedded questions by question-type, inversion and language group – Untimed Judgment Experiment

Language Group	**Question-Type**			
	Wh-		**Yes/No**	
	inverted	**non-inverted**	**inverted**	**non inverted**
L2	8677	8335[55]	8023	8978
Monolingual	6207	6914	6443	7003
Total	*7687*	*7764*	*7390*	*8175*

Acceptability Judgments:

A 2 (language group) x 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Question-type had a significant effect on Z-scores (F_1 (1,85) = 21.3, p < .0001; F_2 (1,31) = 25.2, p < .0001), and so did inversion (F_1 (1,85) = 262, p < .0001; F_2 (1,31) = 610, p < .0001). There was a significant two-way interaction between

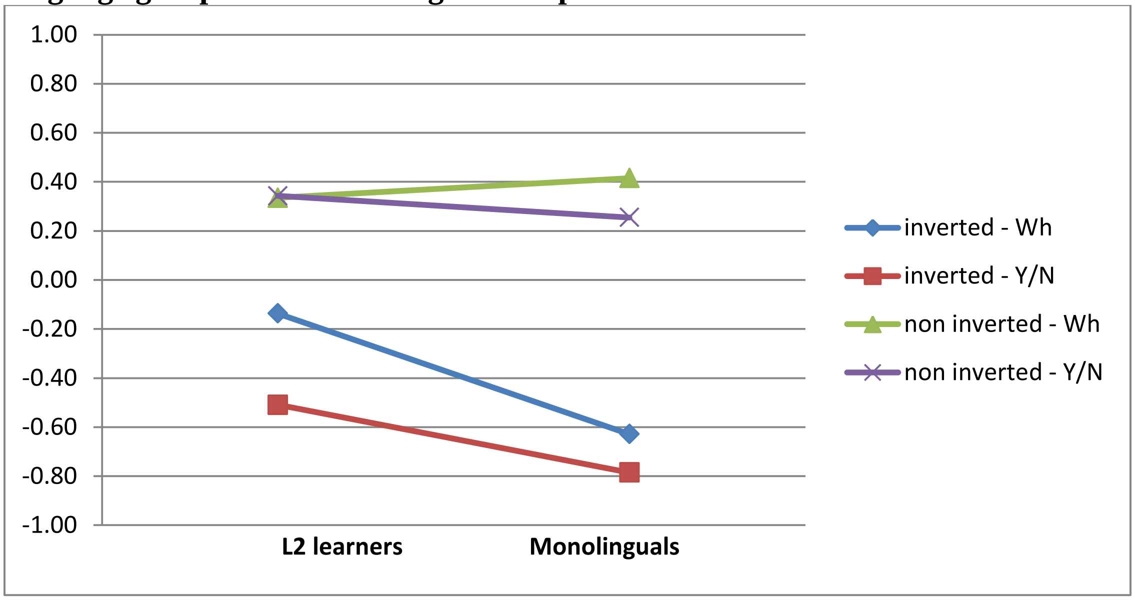
[55] Overall, as noticed above, RTs tended to be faster for ungrammatical than for grammatical sentences. From this result, we might infer that the anomaly in word order is fairly noticeable for participants, which causes them to judge these sentences faster. However, the opposite is true for L2 speakers when judging embedded *wh*-questions. I wonder if one reason for this could be that, in order to judge inverted embedded *wh*-questions, L2 learners resort to a different type of grammatical knowledge (e.g., explicit vs. procedural).

language group and inversion (F_1 (1,85) = 12.5, p = .001; F_2 (1,31) = 26, p < .0001), a significant two-way interaction between question-type and inversion (F_1 (1,85) = 10.2, p = .002; F_2 (1,31) = 4.2, p = .05) and a significant three-way interaction between language group, inversion and question-type (F_1 (1,85) = 10.4, p = .002; F_2 (1,31) = 8.2, p = .007).

The three way interaction indicates that the interaction between question-type and inversion depended on the language group. The interaction between question-type and inversion was in fact significant only for L2 learners (L2 learners: F_1 (1,51) = 33, p < .0001; F_2 (1,31) = 14, p = .001; monolinguals: all *F*s < 1). In other words, while L2 learners judged non-inverted *wh*- and yes/no questions as equally acceptable, they judged inverted *wh*-questions as being more acceptable than inverted yes/no-questions (t_1 (51) = 5.8, p < .0001; t_1 (31) = 5.8, p < .0001).[56] The pattern of acceptability is illustrated in Figure 14.

[56] L2 learners judged inverted embedded yes/no questions without *if* as less acceptable than those with *if*. The difference between *wh*- and yes/no questions was significant for both yes/no structures with and without *if*, so they were combined for these analyses.

Figure 14: Average Z-scores for embedded questions by question-type, inversion and language group – Untimed Judgment Experiment



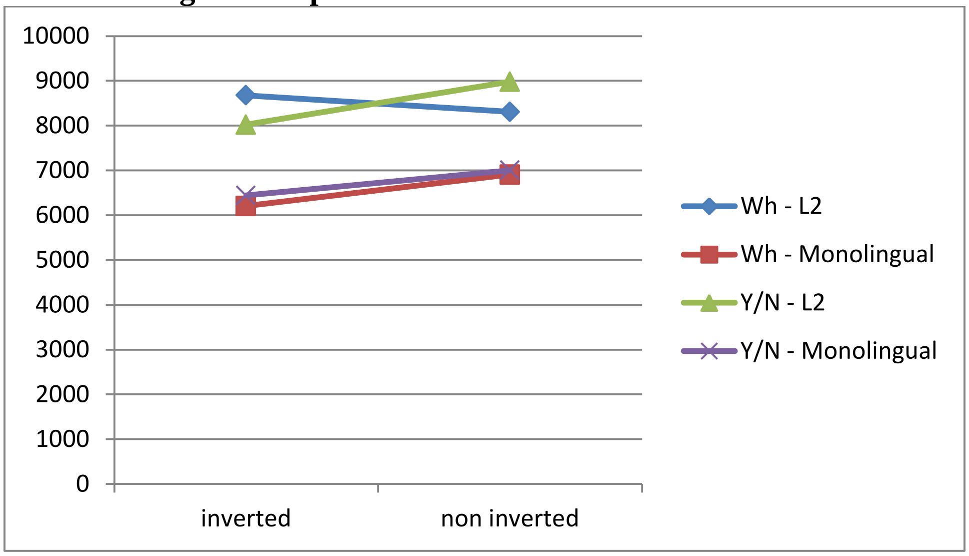
In order to investigate whether L2 judgments are occasionally non-target-like, I measured how often L2 learners and monolingual native speakers judged non-target inverted embedded questions with a positive score (1SD or more above the average) and how often they judged target non-inverted embedded questions with a negative score (1SD or more below the average). L2 learners judged non-target inverted *wh*-questions with a score 1 SD above the average 4% of the time, while monolingual speakers did so 1.7% of the time. The overall difference was significant (F_1 (1, 85) = 3.9, p = .05; F_2 (1,31) = 7.7, p = .009). L2 learners judged target non-inverted embedded *wh*-questions with a negative score around 7% of the time, while monolingual speakers did so 6% of the time. This difference was not significant (all *F*s < 1).

Reaction Times

A 2 (language group) x 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with RTs as the dependent variable was conducted. Inversion had a significant effect on RTs (F_1 (1,85) = 7.6, p = 007; F_2 (1,30) = 5.3, p = .03), indicating that, overall, RTs for non-target inverted embedded questions were faster than those for target non-inverted ones. There was a significant effect of language group, indicating that, on average, monolinguals' RTs were faster than those of L2 learners (F_1 (1,85) = 11.5, p = .001; F_2 (1,30) = 133, p < .0001). There was a marginally significant two-way interaction between inversion and question-type (F_1 (1,85) = 3.6, p = .062; F_2 (1,30) = 3.4, p = .07), and a significant three-way interaction between language group, inversion and question-type (F_1 (1,85) = 4.2, p = .04; F_2 (1,30) = 5.5, p = .026), indicating that, for L2 learnrers, while RTs for inverted yes/no embedded questions were faster than those for non-inverted ones, the opposite was true for *wh*-questions. Nothing else was significant.

The three way interaction indicates that the interaction between inversion and question-type depended on the language group. Only for L2 learners was the interaction between inversion and question-type significant (F_1 (1,51) = 7.4, p = .009; F_2 (1,30) = 8.3, p = .007) indicating that L2 learners' RTs for inverted yes/no embedded questions were faster than those for inverted *wh*-questions, while the opposite was true for non-inverted embedded questions. On the other hand, monolinguals showed a consistent pattern of slightly higher RTs for yes/no than *wh*-questions in both inverted and non-inverted questions.

Figure 15: Average RTs for embedded questions by question-type and language group – Untimed Judgment Experiment



2.4.1.5.2. Wh-type

The second issue that I wanted to investigate in this experiment was whether L2 learners' acceptability judgments would mirror the production patterns seen in oral production, where participants' inversion errors were most frequent with *who* and least frequent with *why*. Thus, in this section, I investigate whether *wh*-type had an effect on acceptability patterns. Table 55 and Table 57 present the average Z-scores for L2 learners and monolingual English speakers, respectively, for inverted and non-inverted *wh*-words. Table 56 and Table 58 present L2 learners and monolinguals' RTs by *wh*-word and inversion, respectively.

Table 55: L2 learners' average Z-scores for embedded questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	−0.05	0.42
where	−0.08	0.39
who	−0.32	0.10
why	−0.10	0.43
Total	*−0.14*	*0.33*

Table 56: L2 learners' average RTs for embedded questions by *wh*-word and inversion – Untimed Judgment Experiment

***Wh*-word**	**inverted**	**non-inverted**
what	8403	7348
where	9246	8359
who	8422	9192
why	8636	8440
Total	*8677*	*8307*

Figure 16: L2 learners' average Z-scores for embedded questions by *wh*-word and inversion – Untimed Judgment Experiment

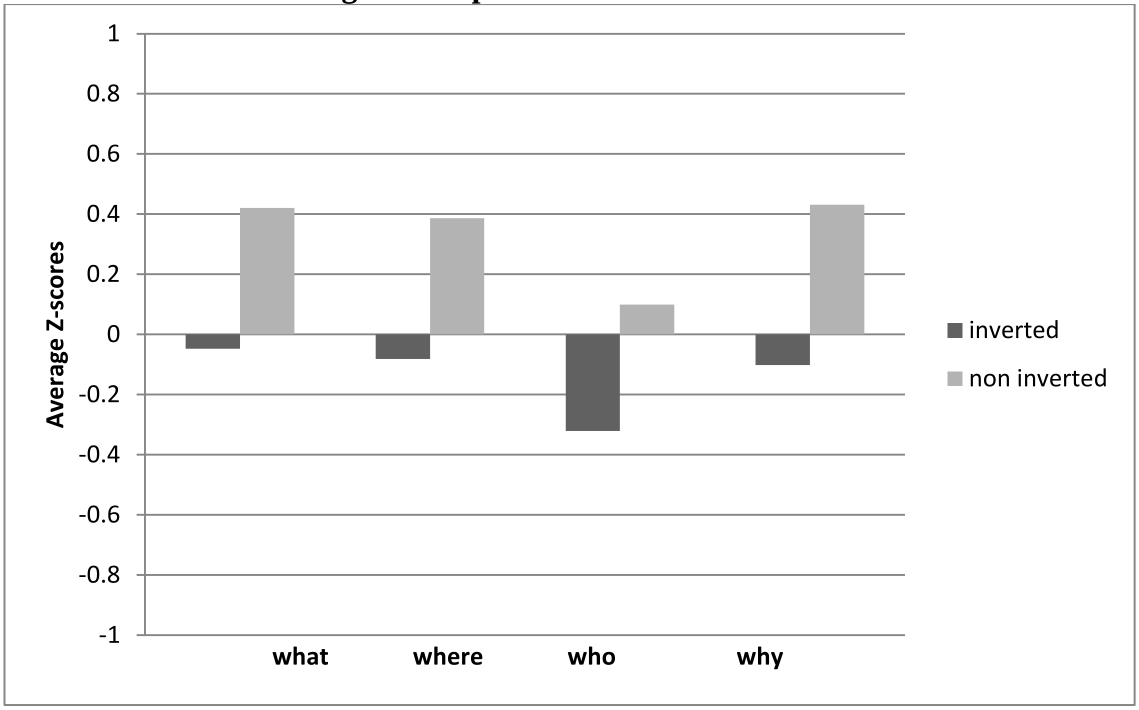


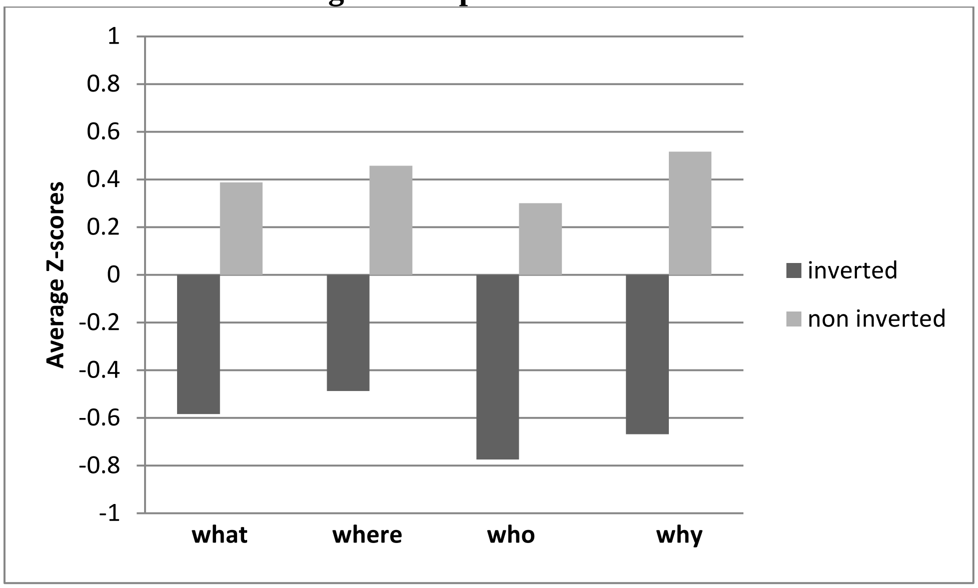
Table 57: Monolinguals' average Z-scores for embedded questions by *wh*-word and inversion- Untimed Judgment experiment

***Wh*-word**	**inverted**	**non-inverted**
what	−0.58	0.39
where	−0.49	0.46
who	−0.78	0.30
why	−0.67	0.52
Total	*−0.63*	*0.41*

Table 58: Monolinguals' average RTs for embedded questions by *wh*-word and inversion- Untimed Judgment experiment

***Wh*-word**	**inverted**	**non-inverted**
what	5772	5815
where	6285	7781
who	6022	7129
why	6764	6888
Total	*6207*	*6906*

Figure 17: Monolinguals' average Z-scores for embedded questions by *wh*-word and inversion- Untimed Judgment Experiment



Acceptability Judgments

A 2 (language group) x 4 (*wh*-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,84) = 210, p < .0001; F_2 (1,27) = 29, p < .0001), showing that non-inverted *wh*-questions were judged as significantly more acceptable than inverted ones. The effect of *wh*-type was significant (F_1 (3,252) = 6.5, p < .0001; F_2 (1,28) = 3.4, p = .03). The only significant interaction was the one between language group and inversion (F_1 (1,84) = 33, p < .0001; F_2 (1,28) = 45, p <.0001), indicating that monolinguals' acceptability patterns are sharper than those of L2 learners, particularly for ungrammatical inverted questions.

The effect of *wh*-word was mainly due to the fact that embedded *who*-questions, both inverted and non-inverted, were judged as being less acceptable than other embedded *wh*-questions. This pattern was only significant for L2 learners: when corrected for multiple comparisons (Bonferroni), scores for *who*-questions were significantly different from those of other *wh*-words (subject analysis: *who* vs *what*, p = .03; *who* vs. *where*, p = .06, *who* vs. *why*, p = .03; item analysis *who* vs. *what*, p = .02; *who* vs. *where*, p = .05, *who* vs. *why*, p = .05). The difference between non-subject *who*-questions and other embedded *wh*-questions might be due to processing reasons. If L2 learners have a bias to interpret the first (animate) N in the sentence as the subject/agent (see footnote 17), they might be led down a garden path once they encounter the second NP, causing them to judge non subject *who*-questions as being less acceptable than other non-subject *wh*-questions overall.

In order to investigate whether L2 judgments are occasionally non-target-like, I measured how often L2 learners and monolingual native speakers judge non-target inverted embedded *wh*-questions with a positive score (1SD or more above the average) and how often they judge target non-inverted *wh*-embedded questions with a negative score (1SD or more below the average). L2 learners judged non-target inverted *wh*-questions with a score 1 SD above the average 4% of the time, while monolingual speakers did so 1% of the time. The overall difference was significant (F_1 (1,84) = 5.2, p = .025; F_2 (1,28) = 7, p = .01), but there was no effect of *wh*-type and no interaction (all *F*s < 1). L2 learners judged target non-inverted main *wh*-questions with a negative score around 7% of the time, while monolingual speakers did so 4% of the time. This difference was not significant (F_1 (1,84) = 3, p = .09; F_2 (1,27) = 2.5, p = .13) and there was no effect of *wh*-type and no interaction (all *F*s $\leq$ 1).

Reaction Times

A 2 (language group) x 4 (*wh*-type) x 2 (inversion) mixed design ANOVA with RTs as the dependent variable was conducted. The effect of inversion was significant (F_1 (1,83) = 31, $p < .0001$; F_2 (1,28) = 42, $p < .0001$). The effect of language group was also significant (F_1 (1,83) = 10.4, $p = .002$; F_2 (1,28) = 45, $p < .0001$), indicating that, on average, monolingual speakers' RTs were faster than those of l2 learners. The interaction between inversion and language group was also significant (F_1 (1,83) = 4.1, $p = .05$; F_2 (1,27) = 4.6, $p = .04$), indicating that the difference between monolinguals and L2 learners in terms of reaction times was more extreme for inverted than for non-inverted questions. Finally, the effect of *wh*-type was significant (F_1 (3,249) = 4, $p = .008$; F_2 (3,28) = 3.9, $p = .02$); the effect was due to the fact that RTs for *what* were significantly faster than those for other *wh*-questions (subject analysis: *what* vs. *where*, $p < .0001$; *what* vs. *why*, $p = .015$; *what* vs. *where*, $p = .003$; *what* vs. *why*, $p = .018$).Nothing else was significant.

The present experiment investigated whether L2 learners' inversion patterns in the production of embedded questions would be reflected in their acceptability judgments. Second language learners differed from English native speakers in that they produced a consistent number of inverted embedded questions, while native speakers never did. With respect to acceptability judgments, the main difference between L2 learners and English monolingual speakers was that the judgments of the latter group were sharper in the case of ungrammatical inverted embedded questions. With respect to the effect of question-type, L2 learners were shown to produce significantly more inverted structures in embedded *wh*-questions than in embedded yes/no questions, where inversion errors were extremely sparse. This pattern was mirrored in L2 learners' acceptability judgments of inverted embedded questions: L2 learners judged inverted *wh*-questions significantly better than inverted yes/no questions. This pattern was expected given the elicited production data. Finally, in the elicited production experiment, *why*-questions were associated with significantly lower non-target inversion rates than other *wh*-questions. This result was not replicated in the present experiment. On the other hand, L2 learners judged *who*-questions lower than other *wh*-questions, both in the inverted and the non-inverted conditions. This difference might be due to processing biases.

2.4.2. Timed Acceptability Judgments

Following R. Ellis (2005) and Bowles (2011), the maximum time allowed for reading and judging each sentence was established by increasing by 20% (and rounding to the next second) monolinguals' average response times for the two experimental conditions (main and embedded questions) in the previous experiment. While in the previous studies (Ellis, 2005; Bowles, 2011), the maximum time allowed varied for each item, in this study, the maximum time allowed for each item was kept constant (7000 milliseconds). That is, the maximum time allowed was around 1000 ms. shorter than average L2 learners' RTs in the previous experiment.

2.4.2.1. Method

2.4.2.1.1. Participants

Participants from a variety of L1 backgrounds were included and proficiency was measured via a shortened version of the Michigan Test of English Language Proficiency (MTELP). Participants were recruited mainly through the Introduction to Psychology subject pools at Queens College and Hunter College and received course credit for their participation in the study.

The main objective of this study was to investigate whether similar acceptability patterns would surface in timed and untimed acceptability judgments. The studies in the literature that have compared L2 learners' performance on timed and untimed grammaticality judgment tasks have used accuracy as the dependent variable and have shown accuracy rates to decrease dramatically in timed tasks. However, an interesting question has to do with whether different qualitative patterns emerge in the two tasks, and whether different structures are differently susceptible to task manipulation. For example, under the hypothesis that timed judgments are more likely than untimed ones to tap into speakers' implicit knowledge, we might expect timed acceptability judgments to reflect production patterns more closely than untimed ones. Specifically, similarly to oral production data and different from untimed judgments, we could expect a clearer *wh*-type effect to surface in both main and embedded questions. For this reason, and because the present study was meant to be exploratory, only L2 learners were recruited for this study. Twenty five L2 learners participated in this experiment. Two participants were excluded because their RTs were lower than 1000 ms for more than 80% of the items. Participants were considered L2 speakers if they were not born in an English-speaking country and had moved to and English-speaking country at or after age 12. Table 59 provides a summary of participants' proficiency scores, age, age of arrival and years of stay in the US.

Table 59: Demographics of L2 participants – Timed Judgment Experiment

L1	N	Average age (SD)	Average age of arrival (SD)	Average years of stay in US (SD)	Average MTELP (SD)
Burmese	1	21	14	7	42
Chinese	8	22.3 (2.9)	19.9 (3.5)	3.7 (1.3)	36.3 (5.8)
Georgian	1	34	21	13	44
German	1	54	14	40	44
Hebrew	1	33	28	5	45
Korean	2	22 (3)	19.5 (4.5)	2.5 (1.5)	40.5 (4.5)
Russian	5	33.2 (9.5)	20.4 (3.3)	12.8 (9.6)	42 (1.6)
Spanish	2	25.5 (2.5)	12.5 (.5)	13 (2)	41 (2)
Tibetan	1	18	15	3	42
Vietnamese	1	18	15	3	n.a.[57]
Total	*23*	*27.2 (9.3)*	*18.4 (4.4)*	*8.8 (9.4)*	*40 (4.8)*

2.4.2.1.2. Materials

The same materials as in the Untimed Experiment were used.

2.4.2.1.3. Procedure and Data Analysis

The same Procedure and Data Analysis as in the Untimed Experiment were used.

2.4.2.2. Main questions: Results and Interim Discussion[58]

The data were first inspected to see whether inverted main questions were consistently rated higher than non-inverted questions across L1s. Overall, participants showed a preference for inverted over non-inverted main questions (F_1 (1,22) = 7.6, p = .011; F_2 (1,31) = 31.3, p < .0001), with sixteen out of twenty-three participants (70%) preferring inverted over non-inverted main

[57] The MTELP for this speaker was lost due to program malfunctioning.
[58] Only Acceptability Judgments were examined in this experiment. RTs were not examined, given that this was a timed experiment in which the maximum time allowed to read and judge a sentence was fixed at 7000 ms.

questions. The participants that preferred non-inverted questions came from a variety of L1 backgrounds (two were L1 Chinese, one was L1 Korean, two were L1 Russian, one was L1 Spanish and one L1 Vietnamese). Table 60 presents the average Z-scores and RTs for inverted and non-inverted main questions for the different language groups.

Table 60: L2 learners' average Z-scores for main questions by inversion and L1 – Timed Judgment Experiment

L1	Average Z-score for inverted questions	Average Z-score for non-inverted questions
Burmese	0.70	-0.85
Chinese	0.15	0.01
Georgian	0.58	-1.09
German	0.72	-0.30
Hebrew	0.55	-0.10
Korean	0.18	0.15
Russian	0.22	0.04
Spanish	0.13	0.27
Tibetan	0.66	-0.28
Vietnamese	0.36	0.50
Total	*0.29*	*-0.05*

2.4.2.2.1. Question-type

The first issue under investigation was whether L2 learners' acceptability judgments would mirror the production patterns seen in the oral production study, where participants were inverting more frequently in yes/no questions than in *wh*-questions.

Table 61 presents the average Z-scores for inverted and non-inverted *wh*- and yes/no questions in this experiment and compares them with the average Z-scores for L2 learners obtained in the previous untimed experiment. As can be seen by the table, L2 learners' judgments are very similar across the two experiments.

Table 61: L2 learners'average Z-scores for main questions by question-type, inversion and experiment

Experiment	**Question-Type**			
	Wh		**Yes/No**	
	inverted	**non-inverted**	**inverted**	**non inverted**
Timed	0.27	-0.10	0.30	0.01
Untimed	0.29	−0.15	0.38	0.00

A 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,22) = 7.8, p = .01; F_2 (1,31) = 33, p < .0001). Nothing else was significant (all *F*s < 1). Similarly to what we saw for L2 learners in the untimed experiment, there was no interaction between inversion and question type. However, differently from the previous experiment, in this experiment L2 learners did not judge non-inverted yes/no questions as more acceptable than non-inverted *wh*-questions (all *F*s < 1), indicating that they might not have fully acquired the native asymmetry.

In order to investigate whether L2 learners' judgments would be less target-like in a timed than in the untimed version of the experiment, I measured how often L2 learners in this experiment judged non-inverted main questions with a positive score and how often they judge inverted main questions with a negative score compared to the previous untimed experiment. L2 learners in the timed experiment judged non-target non-inverted questions with a score above their own mean 3% of the time, while L2 learners in the untimed experiment did so 7% of the time. L2 learners in the timed experiment judged target-like inverted main questions with a negative score 10% of the time, exactly like L2 learners in the untimed experiment. These data indicate that L2 learners' judgments do not become less target-like under time pressure, contrary to what is suggested by the existing literature.

2.4.2.2.2. Wh-type

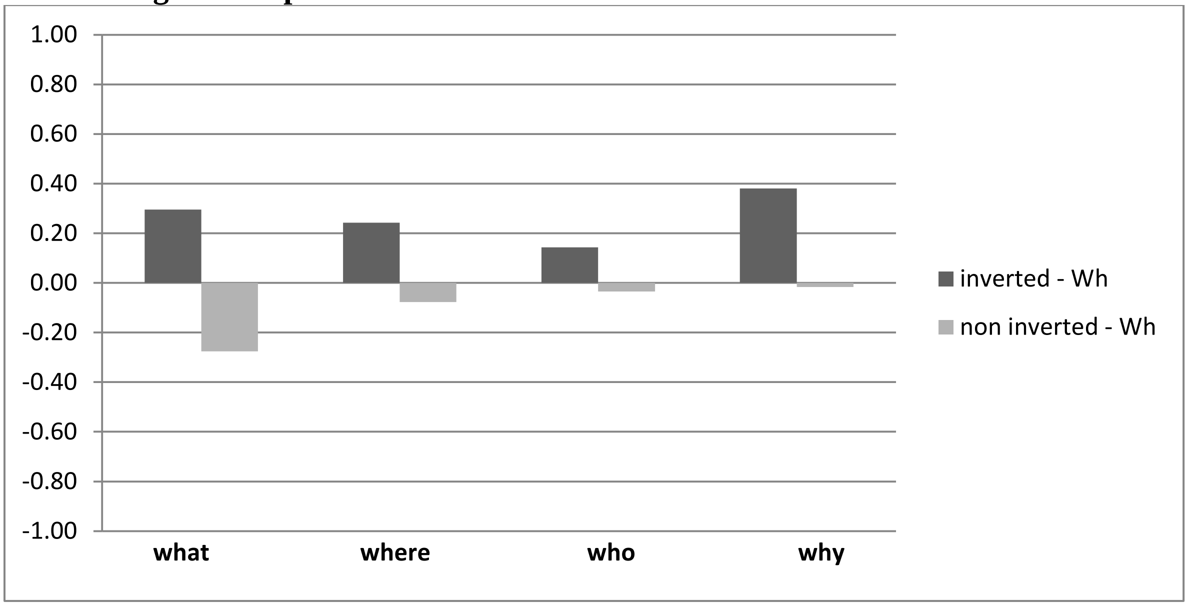
The second issue that I wanted to investigate in this study was whether L2 learners' acceptability judgments would mirror the production patterns seen in elicited and written production, where participants were inverting less frequently with *why* relative to other *wh*-elements.

Table 62 presents the average Z-scores for L2 learners in the present, timed experiment and compares them with the Z-scores obtained in the previous untimed experiment.

Table 62: L2 learners' average Z-scores for main questions by *wh*-word, inversion and experiment

***Wh*-word**	**Timed Experiment**		**Untimed Experiment**	
	inverted	**non-inverted**	**inverted**	**non-inverted**
what	0.30	-0.28	0.37	-0.11
where	0.24	-0.08	0.33	-0.24
who	0.14	-0.04	0.09	-0.30
why	0.38	-0.02	0.36	0.04
Total	*.27*	*-0.10*	*0.29*	*-0.15*

Figure 18: L2 learners' average Z-scores for main questions by *wh*-word and inversion – Timed Judgment Experiment



A 4 (*wh*-type) x 2 (inversion) mixed-design ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,20) = 5.1, p = .035; F_2 (1,28) = 26.2, p < .0001), showing that inverted questions were judged significantly higher than non-inverted ones. Nothing else was significant (all *F*s < 1). L2 learners in this experiment judged non-inverted main *wh*-questions (except *what*-questions) with values close to the mean. However, there was no effect of *wh*-type in either inverted or non-inverted questions (all *F*s < 1).

L2 learners in the timed experiment judged non-target non-inverted *wh*-questions with a score 1 SD above the average 3% of the time, while L2 learners in the untimed experiment did so 7% of the time. Both groups of L2 learners judged target-like inverted main *wh*-questions with a negative score around 10% of the time. As is evident from Table 63, acceptability judgments obtained in the timed experiment were, if anything, more target-like than those obtained in the untimed experiment.

Table 63: L2 learners' average non-target judgments for main questions by *wh*-word, inversion and experiment

	Timed Experiment		Untimed Experiment	
Wh-word	**% Judgments 1 SD below average for inverted Qs**	**% Judgments 1 SD above average for non-inverted Qs**	**% Judgments 1 SD below average for inverted Qs**	**% Judgments 1 SD above average for non-inverted Qs**
what	0.05	0.00	0.07	0.08
where	0.12	0.02	0.11	0.07
who	0.17	0.05	0.13	0.04
why	0.07	0.04	0.12	0.10
Total	0.10	0.03	0.10	0.07

2.4.2.3. Embedded questions: Results and Interim Discussion

The data was first inspected to see whether non-inverted embedded questions were consistently rated higher than inverted embedded questions across L1s. Overall, participants showed a preference for non-inverted over inverted embedded questions (F_1 (1,22) = 16.8, $p < .0001$; F_2 (1,31) = 40.4, $p < .0001$). Nineteen of the twenty-three participants judged target non-inverted embedded questions as more acceptable, on average, than inverted questions. Four participants (2 L1 Chinese, 1 L1 Korean, 1 L1 Russian) showed the opposite pattern.

Table 64 presents the average Z-scores for inverted and non-inverted embedded questions for the different language groups.

Table 64: L2 learners' average Z-scores for embedded questions by inversion and language group – Timed Judgment Experiment

L1	Average Z-scores for inverted embedded questions	Average Z-scores for non-inverted embedded questions
Burmese	-0.80	0.70
Chinese	-0.20	0.18
Georgia	-0.78	0.84
German	-1.00	0.22
Hebrew	-0.54	0.19
Korean	-0.08	-0.03
Russian	-0.06	0.11
Spanish	0.11	0.18
Tibetan	-0.81	0.27
Vietnam	-0.06	0.10
Total	*-0.25*	*0.20*

2.4.2.3.1. Question-type

An issue that I wanted to investigate in this experiment was whether L2 learners' acceptability judgments would mirror the patterns seen in the two production studies, where participants were inverting more frequently in embedded *wh*- than embedded yes/no questions. Table 65 compares L2 learners' acceptability judgments for inverted and non-inverted *wh*- and yes/no questions in the timed and the untimed experiments. As can be seen by the table, L2 learners' judgments are similar across the two experiments, but judgments in the timed experiment seem to be less sharp than those in the untimed experiment.

Table 65: L2 learners' average Z-scores for embedded questions by question-type, inversion and experiment

Language Group	Question-Type			
	Wh		**Yes/No**	
	inverted	**non-inverted**	**inverted**	**non-inverted**
Timed	−0.20	0.19	−0.31	0.21
Untimed	−0.14	0.33	−0.51	0.34

A 2 (question-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,22) = 16.5, p = .001; F_2 (1,31) = 40.2, p < .0001). Nothing else was significant (all *F*s < 1). In particular, differently from what we saw for L2 learners in the untimed experiment, there was no interaction between inversion and question type, indicating that L2 learners did not judge inverted embedded *wh*-questions as more acceptable than inverted yes/no questions (*F* < 1).

In order to investigate whether L2 learners' judgments would be less target-like in a timed than in the untimed version of the experiment, I measured how often L2 learners in this experiment judged inverted embedded questions with a positive score and how often they judged non-inverted embedded questions with a negative score compared to the previous untimed experiment. L2 learners in the timed and untimed experiment judged non-target inverted questions with a score SD above the average 4% of the time. L2 learners in the timed experiment judged target-like non-inverted embedded questions with a negative score 10% of the time, exactly like L2 learners in the untimed experiment. These data indicate that L2 learners' judgments do not become less target-like under time pressure, contrary to what is suggested by the existing literature.

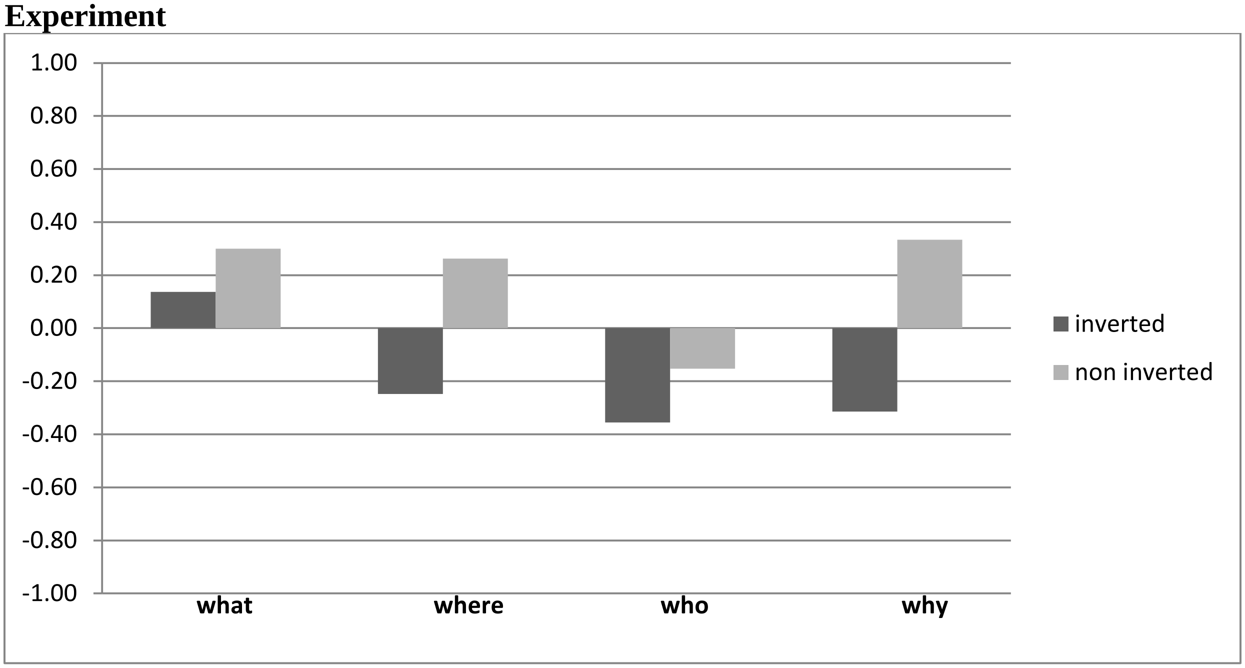
2.4.2.3.2. *Wh*-type

Another issue that I wanted to investigate in this experiment was whether L2 learners' acceptability judgments would mirror the production patterns seen in elicited production, where participants' inversion errors were most frequent with *who* and least frequent with *why*. Thus, in this section, I investigate whether *wh*-type had an effect on acceptability patterns. Table 66 presents the average Z-scores for L2 learners in the timed and untimed experiments, respectively, for inverted and non-inverted *wh*-words.

Table 66: L2 learners' average Z-scores for embedded questions by *wh*-word, inversion and experiment

Wh-word	Timed Experiment		Untimed Experiment	
	inverted	non-inverted	inverted	non-inverted
what	**0.14**	0.30	−0.05	0.42
where	-0.25	0.26	−0.08	0.39
who	-0.36	-0.15	−0.32	0.10
why	-0.31	0.33	−0.10	0.43
Total	-0.20	0.19	*−0.14*	*0.33*

Figure 19: L2 learners' average Z-scores by *wh*-word and inversion –Timed Judgment Experiment



A 4 (*wh*-type) x 2 (inversion) mixed deisgn ANOVA with Z-scores as the dependent variable was conducted. Inversion had a significant effect on Z-scores (F_1 (1,21) = 7.3, p = .013; F_2 (1,28)

= 14.9, p = .001), showing that non-inverted embedded questions were judged significantly higher than inverted ones. The effect of *wh*-type was significant (F_1 (3,63) = 3.3, p = .036; F_2 (1,28) = 4.8, p = .008). The interaction between inversion and *wh*-type was not significant (F_1 (1,63) = 1.8, p = .15; F_2 (3,28) = 1.2, p = .32).

While the effect of inversion was significant across *wh*-words (i.e., the non-target inverted version was considered significantly worse than its non-inverted counterpart), inverted *what*-questions were judged as being on average more acceptable than all other inverted embedded *wh*-questions. When corrected for multiple comparisons (Bonferroni), scores for *what*-questions were significantly different from those for *who*-questions only (subject analysis: *what vs. who*, p = .015; item analysis *what* vs. *who*, p = .005). It should be pointed out that *who*-questions (both non-inverted and inverted) were assigned particularly low scores in this experiment (and judgments scores for target non-inverted *who*-questions were lower than those for all other *wh*-questions). While this pattern has come up in the untimed experiment as well, it was particularly pronounced in this experiment; this might be due to the fact that under time pressure L2 learners have more trouble recovering from the garden-path initial subject/agent interpretation of sentence-initial *who*.

In order to investigate whether L2 learners' judgments were target-like in the timed than in the untimed version of the experiment, I measured how often L2 learners in this experiment judged inverted embedded questions with a positive score and how often they judged non-inverted embedded questions with a negative score compared to the previous untimed experiment. L2 learners in the timed and untimed experiment judged non-target inverted *wh*-questions with a score above the average 4% and 5% of the time, respectively. L2 learners in the timed experiment judged target-like inverted main questions with a negative score 11% of the time, while L2 learners in the untimed experiment did so 7% of the time. As is evident from Table 67, the pattern of these non-target judgments in the two experiments was similar: non-inverted *who*-questions were associated with the highest non-target judgments in both experiments, and the pattern was sharper in the timed experiment, while non-target judgments were distributed similarly across the other *wh*-elements.

Table 67: Average non-target judgments for embedded questions by *wh*-word, inversion and experiment

	Timed Experiment		**Untimed Experiment**	
Wh-word	**% Judgments below average for non-inverted Qs**	**% Judgments above average for inverted Qs**	**% Judgments below average for non-inverted Qs**	**% Judgments above average for inverted Qs**
what	0.07	0.05	0.03	0.05
where	0.07	0.05	0.07	0.06
who	0.24	0.02	0.13	0.04
why	0.09	0.05	0.06	0.04
Total	*0.11*	*0.04*	0.07	0.05

2.4.3. Acceptability Judgments of English Questions: Summary and Discussion

The present acceptability judgment experiments investigated whether L2 learners' inversion patterns in production of main and embedded questions would be reflected by acceptability judgments. Previous research (N. Ellis, 2004, 2005; Bowles, 2011) suggests that untimed acceptability judgments, together with written production tasks, might be more likely to reflect learners' explicit, metalinguistic knowledge of their L2, while timed acceptability judgments, together with oral production tasks, might be more likely to reflect learners' implicit knowledge.

Taken together, the results from the two acceptability tasks showed that, by and large, advanced intermediate L2 learners' judgments of English interrogative structures closely resemble those of native speakers, in that target grammatical sentences are consistently assigned higher scores than non-target structures. However, the acceptability judgment experiments showed that L2 learners' acceptability patterns differ from those of native speakers in a number of ways.

First, native speakers show a clear distinction between non-inverted main yes/no and *wh*-questions, judging the former as significantly more acceptable than the latter. L2 learners in the untimed experiment only showed a trend towards this distinction, and L2 learners in the timed experiment did not show a distinction altogether. This might be taken as an indication that L2 learners do not yet fully master all the nuances of the complex system they are acquiring.

Second, compared to L2 learners, native speakers' judgments tend to be sharper than those of L2 learners, in that native speakers assign higher scores to grammatical sentences and lower scores to ungrammatical sentences. This difference is likely to reflect both the fact that L2 learners are less certain about their judgments, and thus less likely to use extreme values to distinguish differences in acceptability, and the fact that L2 learners' judgments are noisier than those of monolingual speakers, in that low scores are occasionally assigned to grammatical sentences, and (less often) high scores are assigned to ungrammatical sentences.

Third, L2 learners' judgments show grammatical distinctions that are not present in the judgments of native speakers. For example, both in the timed and untimed experiments L2 learners distinguish between ungrammatical *wh*-questions, judging some as more acceptable than others; importantly, in the untimed experiment, L2 learners judged ungrammatical inverted embedded *wh*-questions as significantly more acceptable than embedded inverted yes/no questions, in this mirroring a pattern seen in the production studies.

A central aim of this study was to investigate whether specific production errors would be reflected by acceptability patterns. Specifically, I wanted to see whether acceptability patterns would reflect the question-type and *wh*-type asymmetries seen in production.

With respect to main questions, acceptability patterns did not show a clear asymmetry between yes/no and *wh*-questions; if anything, the asymmetry went in the opposite direction from what we expected based on the oral production data, in which lack of inversion was significantly higher in *wh*- than yes/no questions.

With respect to main *wh*-questions, the untimed acceptability judgments showed a trend towards a *why*-distinction in the same direction as the production experiments: *why* was associated with lower inversion rates in production tasks and non-inverted *why*-questions were judged as more acceptable than other non-inverted *wh*-question.

With respect to embedded questions, the untimed acceptability judgments confirmed the pattern seen in the production studies: L2 learners produced inversion errors in embedded *wh*-questions but not in embedded yes/no questions, and they judged inverted embedded *wh*-questions as more acceptable than inverted yes/no questions. This pattern did not reach significance in the untimed acceptability judgments. On the other hand, acceptability judgments did not mirror the *why*-asymmetry seen in the oral production experiment.

Finally, it should be noted that this study did not confirm the prediction based on the literature according to which timed acceptability judgments would be more likely than timed ones to reflect oral production patterns. This might be partially due to the lower statistical power in the timed experiment, where data from 23 L2 participants were analyzed as opposed to data from 51 in the untimed experiment. Once more participants are tested, the results from the timed experiment might look more similar to those of the untimed experiment. However, at present, it seems unlikely that the patterns of the judgments in the timed experiment will mirror the oral production pattern more directly than those obtained via the untimed experiment. While this result might be surprising, it is important to bear in mind that previous experiments, which found that untimed judgments were more likely to reflect explicit knowledge, were concerned with accuracy rates and not with specific qualitative patterns. The present experiment, on the other hand, was a magnitude estimation experiment, which, by its nature, was concerned with degrees and patterns of acceptability and not with absolute judgments. From what we can infer from these results, then, timed acceptability judgments are not more likely to reflect production patterns than untimed ones.

3. First Language Acquisition of English Questions

3.1. Introduction

In Chapter 2, I focused on the acquisition of English questions by adult second language learners of English. The elicited production studies in Section 2.3.1. showed that simple transfer of L1 properties is arguably not the cause of L2 learners' non-target productions in main and embedded questions: L1 Spanish speakers were less accurate at producing subject-auxiliary inversion than L1 Chinese speakers in main questions and did not differ from L1 Chinese speakers in terms of non-inversion rates in embedded questions, while these pattern were reversed in the written production study.

Moreover, L1 Spanish and L1 Chinese L2 learners showed a clear question-type asymmetry in their oral and written production of main and embedded questions, with *wh*-questions being associated with significantly higher non-target productions than yes/no questions. As we have seen, this pattern does not immediately follow from L1 properties or from properties of the L2 input.

The fact that L1 and L2 input properties are arguably not the main sources of non-target productions in L2 learners leaves open the possibility that these phenomena are the product of biases specific to the language learning system. In other words, due to the fact that non-target productions in L2 learners cannot be imputed to either L1 or L2 properties, it seems reasonable to hypothesize that they could stem from UG defaults. If this is true, we then predict that children learning English as their first language will entertain the same hypotheses about their target language and produce the same type of UG-driven errors as adults learning their second language.

In this chapter, I investigate the extent to which first language learners' production patterns of English interrogatives are affected by the question type and by the *wh*-word present in the structure they are producing. In particular, I will examine whether children's non-target productions resemble those seen in adult L2 learners (i.e., higher accuracy rates in main and embedded yes/no questions compared to *wh*-questions, and presence of an argument/adjunct asymmetry and/or of a *why*-effect).

From the first days of modern psycholinguistics (e.g., Klima & Bellugi, 1966), it has been known that children learning English as their first language often use the wrong word order when producing a main question, failing to front the auxiliary, especially if a *wh*-word is present, as in (164).

(164) *What you are eating?

As detailed in the literature review presented in the next section, many issues concerning the acquisition of subject-auxiliary inversion by children learning English as their first language remain open. To this day, no experimental study that I am aware of has systematically investigated the acquisition of yes/no and *wh*-questions in the same population via the same experimental protocol. Moreover, a large body of developmental studies in the 1990s (e.g., de Villiers, 1991; Sarma, 1991; Stromswold, 1990 among others) suggested that children treat argument and adjunct *wh*-words differently, while other researchers have suggested that the reported asymmetry is restricted to individual *wh*-words (e.g., *why*) and not to the whole adjunct class.

In this study, I compare children's production of main yes/no questions with their production of non-subject main *wh*-questions. Additionally, I compare accuracy and inversion rates for different *wh*-questions: arguments *what* and *which* and adjuncts *when* and *why*. Similar to the L2 production study, I have decided to focus on argument and adjunct *wh*-words. Different from the L2 study, I focus here on *which* and *when* instead of *who* and *where*. Object *which* was chosen instead of object *who* because *who* gave rise to a large number of non-target subject responses both in the L2 studies and in the first Pilot experiment with children. Additionally, *which* was chosen because it is an argument, but, in contrast to *what* and *who*, it is low in frequency. This fact will allow us to see whether the reported argument/adjunct asymmetry in the literature is frequency-independent. An additional reason for choosing argument *which* is that some recent proposals in the acquisition literature (Roeper, 2011, see Section 3.2.1.2.1.4.) have argued that inversion rates in children might depend on whether the proposition associated with a question is presupposed; moreover, it has recently been proposed in the theoretical literature (Eilam, 2011, Section 3.1.1.2.1.4.) that argument *wh*-questions are weakly presuppositional, while adjuncts are strongly presuppositional. Argument *which* is an exception to this generalization, in that it is an argument but it is strongly presuppositional. In order to investigate if the reported argument/adjunct distinction in the literature could be reduced to a difference in presuppositional content, I compared inversion rates for argument *which* with inversion rates for argument *what* and presupposed adjunct *why*. *When* was elicited in this study instead of *where* because it is more similar to *why*, in that they both are sentential modifier (as opposed to *where*, which, in most instances, is a VP modifier), and because the few studies that investigated inversion rates with *when* have found that they are comparable to or lower than those for *why*. By systematically comparing *why* with an element that, differently from *where*, is unambiguously an adjunct and is similar to *why*, in being a TP-modifier, we should be able to see whether the reported low inversion rates in child English are common to sentential adjuncts or restricted to *why*.

While the literature on the acquisition of main questions in child English is abundant, albeit inconclusive, very little research has been done on the acquisition of embedded questions, and the results in the literature are contradictory: Stromswold (1990) found that children produce a substantial number of word order errors in embedded *wh*-questions, while Sarma (1991) found that children do not produce subject-auxiliary inversion in embedded *wh*-questions.

A secondary aim of this study is to investigate whether the production patterns observed in L1 learners can be explained by patterns in the adult input. More specifically, I will investigate the constructivist hypothesis, according to which the frequency of specific word combinations in the input predicts children's target and non-target productions in interrogative structures. While investigating linguistic input for L2 learners is complicated due to the fact that input is likely to be extremely variable across individuals and L1 groups, the CHILDES database (MacWhinney, 2000) makes it relatively easy to examine quantitative patterns in parental input for children learning English as their L1.

With respect to the question-type distinction in main questions, if children's production patterns were to resemble those reported by a number of studies in the L1 literature and those observed for second language learners in the present study (i.e., higher accuracy rates for yes/no questions than for *wh*-questions), we would expect adult input data to present child learners with abundant evidence for inversion in main yes/no questions; on the other hand, we would expect the adult input to present potentially ambiguous evidence with respect to inversion in *wh*-questions. At first sight this prediction seems unlikely to be confirmed by the data because non-inverted main yes/no questions are acceptable in standard English while non-inverted main *wh*-questions are not, but the prediction might turn out to be accurate if we were to discover that non-inverted main yes/no questions are infrequent in the input or overall less frequent than non-inverted *wh*-structures. Assuming that non-inverted embedded *wh*-questions, among other structures, could be taken by the child as evidence for the fact that *wh*-questions do not necessarily invert, the constructivist hypothesis predicts non-inverted embedded *wh*-questions to be more frequent than non-inverted yes/no questions.

The same type of reasoning might explain *wh*-word effects on subject-auxiliary inversion. According to the constructivist hypothesis, we expect children not to make mistakes with word combinations that are highly frequent in the input (e.g., *what*+*is*) but to make errors with word combinations that are not frequent in the input (e.g., *when*+*is*). Additionally, we might expect children to be accurate with *wh*-elements for which the relative evidence of inversion outnumbers the evidence for non-inversion. Children should thus make few mistakes in main questions with *wh*-elements that occur frequently in main questions and not frequently in embedded questions, i.e. with *wh*-elements for which the frequency in main questions is high compared to the frequency in embedded questions. With respect to embedded questions, constructivist accounts, in line with structural accounts, predict non-target inversion to occur in embedded *wh*-questions due to the fact that *wh*-words appear followed by an auxiliary in more frequent main *wh*-questions. Embedded yes/no questions, on the other hand, differ from their main counterpart in that they contain an overt element, *if* or *whether* (see Section 2.1.), and thus might be treated separately by learners who compute frequencies for different bigrams. A possible difference between constructivist and structural accounts is that the former predict non-target inversion in embedded *wh*-questions to be more likely with *wh*-elements that are frequently inverted in the input and/or with *wh*-elements for which the ratio of frequency in main compared to frequency in embedded is high. In contrast, structural accounts, according to which the cause of inversion errors in embedded contexts is overgeneralization, predict a correlation between inversion rates in main and embedded questions in children's *production*, irrespective of what the frequency of these elements is in the input.

These predictions will be tested in a subset of corpora from the CHILDES database. Six corpora of American English will be examined: Bates, Bloom 70, Clark, Gleason, Snow, and Valian. Two of these corpora (Clark and Snow) are longitudinal corpora of parent-child interactions with a single child. Bates, Bloom, Clark and Snow are longitudinal corpora, while Gleason and Valian are cross-sectional. Taken together, the corpora examined here contain parental interactions with 77 children; the age range of the children in the corpora is 1;4-3;9, and the total number of utterances in the corpora is 270,621 (of which 167,757 were produced by adults).

The present study is aimed at contributing to the research on the acquisition of word-order in main and embedded questions by eliciting main and embedded yes/no and *wh*-questions in a group of children acquiring English as their first language. Additionally, the present study aims at comparing production patterns in first language acquisition with results from second language learners of English and at investigating whether children's productions can be derived from input patterns in a clear, predictable way.

This chapter is organized as follows. In Section 3.2, I review the literature on L1 acquisition of main and embedded questions in some detail. In Section 3.3, I present the findings from the two production studies and the CHILDES search. In Section 3.4, I discuss the findings of the experimental studies in light of evidence from adult input data and compare these results with the results obtained from the production studies conducted with adult L2 learners.

3.2. Previous Studies of L1 Acquisition of English Questions

3.2.1. L1 Acquisition of Main Questions

3.2.1.1. Experimental and Quasi-Experimental Data

The acquisition of English main questions by first language learners has been the focus of much interest in the generative and non-generative literature in the past 40 years. Nonadult-like forms in English questions mainly involve errors of auxiliary omissions, as in (165), lack of subject-auxiliary inversion, as in (166) or tense-raising, as in (167); and double marking of tense, as in (168):

(165) What she eating?

(166) What she can do?

(167) What she likes?

(168) What does she likes?

The data from different studies on the acquisition of English main questions are often hard to compare, and, as a result, the theoretical explanations that derive from these data are hardly conclusive. The existing data on auxiliary use in questions are inconsistent with respect to the frequency with which children omit or invert auxiliaries and with respect to whether auxiliary inclusion and inversion rates differ according to question type (e.g., *wh-* or *yes/no* questions), *wh*-word (e.g., *what* vs. *why*), or auxiliary type (e.g., *be* vs. *have*). These inconsistencies are due partly to differences in coding and definition of errors and in the methodologies employed (parental diaries, spontaneous production samples,[59] elicited imitation, elicited production and grammaticality judgments). For example, some researchers (e.g., Thornton, 2008) count questions with an omitted auxiliary as non-inverted, while others either do not include these productions or code them in a separate category (e.g., Rowland & Pine, 2000). For instance, Valian, Lasser and Mandelbaum (1992) used four different coding categories: *inverted, non-inverted* (for questions where the auxiliary followed the subject, as in (166)), *declarative word order* (for questions lacking an auxiliary but with morphology on the verb, as in (167)) and *plus verb minus auxiliary* (for questions without an auxiliary and with no morphology on the main verb, as in (165)). In Valian et al.'s corpus study, there were 130 *inverted wh*-questions, 5 *non-inverted wh*-questions, 37 *declarative word order wh*-questions, and 27 *plus verb minus auxiliary*

[59] It has been argued that spontaneous production data tend to inflate the rate of correct responses because they only show what children can actually ask (Snyder, 2007). Erreich (1984), for example, noticed that subject auxiliary inversion in questions is more consistent in spontaneous than in elicited production. On the other hand, Sarma's (1991) results show a similar pattern of inversion errors in spontaneous (23/68, 38%) and elicited production (15/44, 33%).

questions. Differences in coding criteria cause dramatic changes in the number of non-inverted questions: according to Valian et al.'s coding scheme, the percent of non-inverted questions in this corpus is 2.5%, while according to Thornton's coding scheme, it would be 34.7%. This points up the importance of researchers working on the acquisition of questions adopting a common coding scheme, or at least providing an explicit justification of the coding scheme adopted and a discussion of experimental findings in light of other coding schemes..

Early studies in the literature (Klima & Bellugi, 1966, Cazden, 1970) proposed a developmental sequence for the acquisition of English questions, and have linked this developmental sequence with MLU stages. Table 68 presents a summary of the developmental stages proposed for the acquisition of English structures by first language learners.

Table 68: Stages in L1 acquisition of English questions

Klima and Bellugi's period	**MLU**	**Yes/No questions**	***Wh- q*uestions**
A 28 months	2.0	Interrogative force expressed by intonation only: Sit here?	Non-productive use of routines: What x doing? Where x go?
B 35 months	2.5	Non-productive use of *don't* and *can't*: You can't fix it?	More complex questions, (aux-omission): Where my bag? Why you waking me up?
C 38 months	3.5	Subject-auxiliary inversion: Can you help me? Does Daddy go?	Auxiliary verb inclusion. Failure of subject auxiliary inversion: What you ate? Why she is going away?
D-F 42-54 months	4+	Tag questions: That's funny, isn't it?	Mature system: Why is the cat sleepy?

However, not all studies have confirmed this developmental path, and subsequent research has questioned the existence of some of the above stages. While there is a considerable amount of disagreement among researchers with respect to the details of developmental profiles and pathways for English interrogative structures, some qualitative patterns and strong tendencies with respect to order of acquisition and rate and type of nonadult-like behaviors seem nonetheless to emerge.

a) Some *wh*-words (e.g. *what*, *who*) tend to be acquired earlier than others (e.g. *why*, *when*) and to display higher rates of correct use (Smith, 1933; Klima & Bellugi, 1966; Bloom et al., 1982; Kuczaj & Brannick, 1979). In a single case study, Bloom et al. (1982) found that *wh*-words like *where*, *what* and *who* were acquired earlier (around 26 months of age) than *why*, *when*, *whose* and *which* (after 35 months). These findings were confirmed in a longitudinal study with 12 children (ages 1;8-2;0 at the beginning of the study and 2;9-3;0 at the end of the study) by Rowland and Pine (2003). Bloom and colleagues interpreted this developmental sequence in terms of syntactic-semantic saliency: the *wh*-elements that appear early in development replace major VP-related sentence constituents as opposed to sentential ones.[60]

[60] This explanation is clearly not tenable for *which* and *whose*, which are not sentential modifiers. Their delay is perhaps due to the fact that they are NP-specifiers and discourse-related elements.

b) Some *wh*-elements are associated with higher error rates than others: *what* is the first *wh*- to be used correctly (Bloom et al., 1982), while *why*, which is acquired later, displays a higher rate of inversion errors (e.g., de Villiers, 1991; Thornton, 2004; 2008; Rowland & Pine, 2000) and gives rise to high error rates in comprehension (de Villiers et. al., 2008). For example, in a corpus study with 14 children, Stromswold (1990) found high inversion rates for *who* (100%), *how* (97%), *what* (94%) and *where* (95%), followed by *why* (87%), *which* (79%) and *when* (77%). In a study using both elicited production and spontaneous speech samples of 18 children (aged 2;5-3;0), Erreich (1984) found that inversion was higher for *which* (100%), *who* (94%), followed by *what* (76%), *where* (72%), and low for *how* (28%), *why* (23%) and *when* (8%). Two single case studies report overall lower inversion rates: Labov and Labov (1978) report higher inversion rates for *how* (89%), followed by *which* (88%), *who* (87%), *where* (78%), *what* (66%), *when* (56%) and *why* (15%), while Rowland and Pine (2000) report high inversion rates for *who* (100%), followed by *how* (85%), *what* (79%), *where* (73%), *which* (67%) and exceptionally low rates for *why* (8%). The fact that *why* is associated with exceptionally low inversion rates has been confirmed by a number of studies: in a study with 16 children between 1;11 and 4;25, Berk (2003) found higher non inversion rates for questions with *why* as opposed to all other *wh*-elements. Inversion rates for *why* and all other *wh*-elements were 38.5% and 100%, respectively, in children between 2;6 and 3;5, and 89.5% and 100% in children between 3;6 and 4;5 years of age. Similarly, Thornton (2008) reports data from a child whose inversion rate with *why* was 37% at age 2-2;6 and 50% at 5:-5;6.[61]The only two studies that do not report a similar trend are Kuczaj and Brannick (1979) and Ambridge et al. (2006). The latter was an elicited production study with 28 children aged 3;6-4;6; the average inversion error rates for *why* (7%) and *how* (10%) were not significantly different from those for *what* (11%) and *who* (15%).[62]

Leaving aside idiosyncratic differences, a fairly robust trend seems to emerge from the studies reviewed above: *what*, *who*, and *where* tend to associate with high inversion rates, while *why* and *when* are associated with low inversion rates; the status of *which* and *how* is more controversial.

c) Some studies have found that inverted yes/no questions appear earlier than inverted *wh*-questions (Klima & Bellugi, 1966), and that, in general, yes/no questions are associated with higher inversion rates than *wh*-questions (Bellugi, 1971; Klima & Bellugi, 1966; Kuczaj and Maratsos, 1975; Rowland 2007).[63] However, these findings are not robust:

---

[61] The extremely low inversion rate at this later stage might be partially due to the high number of negated questions in the corpus. Negated questions (e.g. 'Why don't you want to go out?'), as is well documented in the literature (Guasti et al. 1995, among others), give rise to a high number of non-inversion errors (e.g. at 5;0-5:6, AL still fails to invert 83% of the time in negated *why*-questions).

[62] The fact that the only studies not reporting a non-inversion pattern for *why* are also two of the three elicited production studies in the relevant literature suggests that the difference between *why* and other *wh*-words could be an artifact of the different syntactic contexts in which *why* naturally occurs (e.g. combination with auxiliary type, negation, verb form and morphology). More experimental work is needed to explore this question.

[63] Depending on the study, inversion error rates range from 0% to 55% in *wh*-questions and from 0% to 51% in *yes/no* questions.

for example, Rowland (2007) found that *wh*-questions exhibited more inversion errors than *yes/no* questions, but this effect disappeared when *why*-questions were removed from the analysis. Other studies have found either no difference between inversion rates for yes/no and *wh*-questions (Derwing & Smith, 1988; Stromswold, 1990; Tyack & Ingram, 1977 for children at Bellugi's stages C-F), or the opposite tendency[64] (Tyack & Ingram, 1977 for children at Bellugi's stage A-B[65]; Ingram & Tyack, 1979; Erreich, 1984; Valian et al, 1992). In an elicited imitation and production study, Theakston and Lieven (2005) found no auxiliary omission and no inversion errors in yes/no questions.

Some of the empirical inconsistencies in this domain are due to coding criteria. Non-inverted yes/no questions are sometimes considered adult-like, while non-inverted *wh*-questions are not, causing accuracy rates for *yes/no* to be higher than accuracy rates for *wh*-questions. Moreover, as Rowland (2007) notes "no study has explicitly compared error rates across yes/no questions and *wh*-questions" in the same sample of the population. Despite the fact that her study was designed to fill this gap in the literature, she excluded non-inverted yes/no questions "because it is often difficult to determine from the transcript whether a true question was asked or whether the utterance had been given a question mark in error" (Rowland, 2007:116).

[64] With the exception of Stromswold (1990), none of these earlier studies reports statistical analyses. Hence, it is possible that the reported inconsistencies among studies are just due to random sampling error and not to statistically significant differences.

[65] Note that this is a diary study and that no statistical analyses are reported.

d) A fairly uncontroversial finding in the literature is that negated questions are associated with high rates of inversion errors (Erreich, 1984: Bellugi, 1971; Stromswold, 1990; Guasti et al. 1995; Van Valin, 2001; Rowland & Pine, 2000). Bellugi argued that this is due to the fact that negative sentences are syntactically more complex than positive ones. Rowland and Pine (2000) notice that *all* of Adam's negated questions are non-inverted. Van Valin (2001) proposes that the presence of negation obscures tense marking, making it hard for the child to locate the tensed element which must be placed *core-initially* (see Section 3.2.1.2.3.). In an elicited production study, Guasti and colleagues showed that children's non-adult negative questions take the form of auxiliary-doubling (40%), as in (169), negation and aux-doubling (7%), as in (170), and plain inversion errors (23%), as in (170):

(169) What did he didn't wanna bring to school?

(170) What didn't Miss Piggy don't like to do?

(171) What she doesn't want for her witch's brew?

e) English-speaking children have been shown to produce tense or auxiliary doubling errors (as in (172)-(173) and (174), respectively) in yes/no and non-subject *wh*-questions (Hurtford, 1975; Mayer, Erreich & Valian, 1978; Maratsos & Kuczaj, 1978):

(172) Does she eats it?

(173) Did she went there?

(174) Will she will come?

Stromswold reports that doubling errors in questions are infrequent in child spontaneous speech. In a study of 2 children's spontaneous speech Maratsos and Kuczaj found, on the other hand, that doubling errors are frequent in questions involving do-support (around 15% with *does* and 10% with *do*), but rare in questions with auxiliary *be* or with modals. Hattori (2003)'s CHILDES search of the Brown corpus confirmed Maratsos and Kuczaj's results: about 15% of yes/no questions requiring *do*-support contained tense doubling errors, while they rarely occurred in other types of yes/no questions.

f) Auxiliary DO and copula BE have been reported to give rise to more errors than other auxiliaries (Hurtford, 1975; Klee, 1985 and Hattori, 2003 for double marking errors; Rowland, 2007, for commission errors with DO in *yes/no* questions; and Santelmann et al., 2002). Stromswold (1990), Santelmann et al., (2002) and Hattori (2003) argue that errors are due to these verbs' language specific behavior. Santelmann and colleagues studied the productions of 65 children between 2;1 and 5;3 who were asked to imitate 10 declarative clauses and 10 questions each. Their results showed no main effect of sentence type (declarative vs. question) on repetition accuracy for sentences with modals or auxiliary BE, while there was a significant effect of sentence type, but only for the younger group, in sentences with DO and copula BE. Children's incorrect responses were coded as either word order mismatches (12% of the total mismatches) or inflection mismatches (80%). Significantly more word order mismatches occurred in questions than in declarative clauses, while inflection mismatches did not differ by sentence type. This empirical finding is, however, controversial. Labov and Labov (1978)'s single-case study showed higher rates of correct inversion for auxiliary DO than for modal CAN. Similarly, Stromswold (1990) found higher rates of inversion for auxiliary DO, followed by auxiliary BE, HAVE and modals, while copula BE inverted the least.[66] Rowland & Pine (2000) also found that *do, did* and *does* have a strong tendency to invert. Rowland et al. (2005) analyzed the longitudinal spontaneous production of 13 children. Differently from Santelmann et al., they found significantly higher rates of correct use for copula BE than for auxiliary BE and HAVE, but no significant difference between DO and modal auxiliaries or between auxiliaries BE and HAVE.[67] Moreover, the authors found significant differences between the forms of copula BE (are, is), auxiliary BE (are, is) and auxiliary HAVE (have, has), with the morphologically marked 3rd person singular forms used more accurately than the unmarked ones. On the other hand, Theakston and Lieven

---

[66] Stromswold explains the higher rates of inversion for auxiliary DO as follows: "the reason non-negated do is inverted 100% of the time is that the only reason do is in the question is in order to invert. If there is no inversion, there is no need for do". (Stromswold, 1990: 202). The difference between Stromswold's and Santelmann et al.'s results might stem from the different methodologies used, i.e., elicited imitation in Santelmann et al.'s study and spontaneous production in the case of Stromswold's study.

[67] In this study, correct use is calculated over contexts that require an auxiliary and not over contexts in which both an auxiliary and a subject are present. The only excluded contexts were those in which both the subject and the auxiliary had been omitted. The authors looked at inversion errors in their corpus but report that 'the Manchester corpus was not dense enough to perform reliable analyses on these errors.' The data on inversion only come from one child (Lara) for which they had parental diaries. In Lara's corpus, questions requiring DO or modals, but not questions requiring copula BE, were associated with significantly more errors than questions requiring auxiliary HAVE and BE.

(2005), in a study that combined elicited imitation and production, found that auxiliary BE displayed higher rates of correct use than auxiliary HAVE in both declarative clauses and *yes/no* questions,[68] but found no significant differences between the marked and unmarked forms of each auxiliary (*is* vs. *are*, and *have* vs. *has*, respectively). Rowland (2007) studied the spontaneous production of *yes/no* and non-subject *wh*-questions in 10 English-speaking children between 2 and 5 years of age. The prediction that children would be less accurate in their production of questions with auxiliary DO than with other auxiliaries was borne out in this study (questions with DO were associated with more errors than questions with modals), but only for *yes/no*-questions (*yes/no* questions with a modal exhibited the lowest error rates). Moreover, Rowland (2007)'s results revealed a main effect of question type (with *wh*-questions being associated with higher error rates than *yes/no*) and an interaction between question and auxiliary type.[69] The main effect of question type, however, disappeared when *wh*y-questions were excluded from the analysis, suggesting that these questions were responsible for a great number of non-inversion errors.[70]

---

[68] In declarative clauses, auxiliaries did not differ with respect to omission errors, but only with respect to commission errors (agreement errors, incorrect auxiliary selection and incorrect verb form selection). Interestingly, there were no omissions or inversion errors in either imitated or elicited questions.

[69] One possible source of difference could be that uninverted yes/no questions were excluded from the analyses.

[70] On the other hand, a main effect of question type was still present when negative questions were factored out, suggesting that negation affects error rates in both yes/no and *wh*-questions.

3.2.1.2. Theoretical Accounts

Despite the many inconsistencies in the empirical literature on the acquisition of main questions, there are some fairly robust findings that all accounts need to explain: children produce errors for a prolonged period of time, and errors are more likely to occur with some items (adjunct *wh*-words, copula BE, negative questions) than with others (argument *wh*-words, auxiliary BE, positive questions). In the following section, I will discuss how the most influential theoretical accounts explain the acquisition data summarized above.

3.2.1.2.1. Generative Accounts

3.2.1.2.1.1. Optionality

Valian, Lasser and Mandelbaum (1992) noticed that the correlation between rate of inversion and *wh*-status (argument vs. adjunct) is weak and that inversion errors are largely due to the *wh*-words *where* and *why*, but also include argument *what*. They argue that failure of inversion is not related to argument status and that inversion operates optionally in children's grammars, requiring significant positive evidence for it to become productive for each *wh*-word. Children's errors could be due to parameter mis-setting (non-inversion is a grammatical option in languages like French), to the presence in the input of optionally non-inverted yes/no questions, and/or to the presence of non-inverted *wh*-questions (subject *wh*-questions, *how come*-questions, embedded *wh*-questions). According to this proposal, the task that English children face is that of learning that inversion is obligatory with main *wh*-questions.

3.2.1.2.1.2. Syntactic Accounts

A number of researchers working in the generativist tradition have focused on the finding that inversion rates tend to be higher with adjuncts than with arguments. Most of the early analyses are formulated within the P&P framework and make use of Empty Category Principle (ECP) to explain the results. For example, according to Stromswold (1990), the difference between arguments and adjuncts stems from the theta criterion: argument *wh*-words leave a trace in their base generation position in order to satisfy the theta criterion and traces need to be governed by an element in Spec, CP in order to satisfy the ECP. Adjunct *wh*-words, on the other hand, can, in principle, be base-generated in clause-initial position[71] (but crucially, not in Spec, CP) without violating the theta criterion, given that, by definition, they are not subcategorized by the verb. In this case, the ECP would not be violated because there would be no *wh*-trace at all in the structure. By assuming that inversion is triggered by the presence of an element in Spec, CP, the inversion difference between arguments and adjuncts is borne out: argument *wh*-words in Spec, CP trigger subject-auxiliary inversion, while adjunct *wh*-words in pre-Spec do not. In a similar fashion, Sarma (1991) attributes the lower inversion rates found with adjunct *wh*-words to ECP violations. According to this analysis, children's grammars initially assume that only heads can properly govern traces. In the case of non-inverted *wh*-adjunct questions, the *wh*-trace is governed by the empty head in C^0, which is co-indexed with the *wh*-trace via Spec-Head agreement between the empty head and the *wh*-element in Spec, CP. However, in adjunct *wh*-

[71] Stromswold assumes this position to be a pre-Spec position because in general, elements in Spec, CP are operators binding variables, while *wh*-adjuncts in this case are not operators (there is no *wh*- trace in VP).

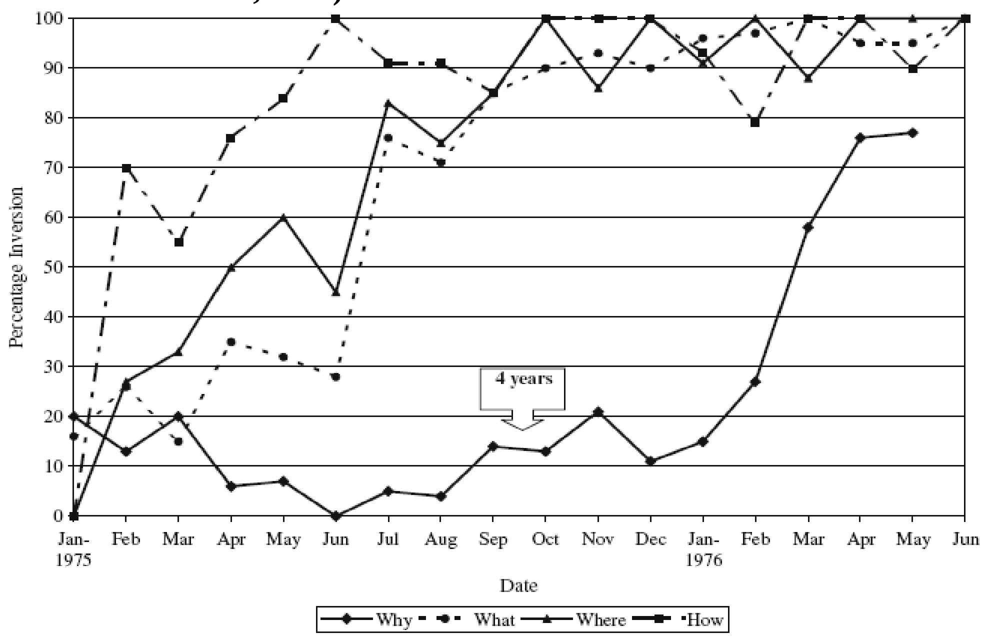
questions in which the auxiliary has raised and adjoined to C, the *wh*-trace in its base-generation position is not properly governed because segmented heads[72] do not count as proper governors (Tiedeman, 1989). In order for children to arrive at the adult grammar, they need to realize that XPs can be proper governors. On the other hand, Sarma argues that argument traces move cyclically through Spec, VP. This allows V^0 to be co-indexed with the *wh*-argument trace, which is in turn properly governed. More simply, de Villiers (1991) proposes that children's initial questions are IPs and that children start out by base-generating adjunct *wh*-words as IP-specifiers. In this configuration, the auxiliary stays in its base-generation position in I because there is no functional head for it to move to. In order to reanalyze main questions as CPs, children need unambiguous evidence from embedded questions, where the *wh*-element has moved to a position higher than IP. In support of this proposal, de Villiers shows that there is a correlation between children's use of embedded questions with a particular *wh*-word and their inversion rates for that *wh*-word. Thus, the idea is that for each *wh*-word, children gather piecemeal evidence of its position in CP. The late acquisition of inversion for main *why*-questions is the consequence of the low frequency of embedded *why*-questions in the adult input.

As noted above, one of the most robust findings in the literature is that *why* is associated with low inversion rates in children's productions (Labov & Labov, 1978; Berk, 2003; de Villiers, 1991; Rowland & Pine, 2000, among others). In Figure 20, it is apparent that inversion with *why* is not reliable until at least age 4;6.

---

[72] Sarma follows Tiedeman's (1989) proposal, according to which, if β adjoins to α, β counts as a segment of α iff α and β are coindexed.

Figure 20: L1 Acquisition of inversion for different *wh*-words (Thornton, 2008 adapted from Labov and Labov, 1978)



Some researchers have tried to explain these results by appealing to the particular semantics of *why*. For example, Kay (1980) argues that *why* differs semantically from all other *wh*-elements: answers to other *wh*-questions specify the location, time or manner of the main event, while answers to *why*-questions call into play a cause or an expected outcome. That is, w*hy*-questions refer to a secondary event, external to the main one (Why did you go to the store? I went to the store *because John asked me*), while other *wh*-words just add information that refers to the same main event (Where did you go? I went *to the store*). Moreover, *why*-questions always presuppose a sentential answer, while answers to other *wh*-elements can minimally involve a noun phrase or a prepositional phrase. While it seems that *why* might indeed be more complex than other *wh*-elements because it requests an explanation based on causal relations and its answer always requires a full CP declarative sentence, it is unclear how these facts can be used in a cogent way to explain low inversion rates; for instance, it seems problematic to me to explain inversion rates in questions based on the complexity of the answer they require. Furthermore, it has been noted that *why*-questions are problematic in adult second language acquisition of English as well (Lee, 2008, this study). The fact that adult learners produce low inversion rates with *why* seems to rule out a conceptual complexity explanation and suggests that we should look for a more fine grained syntactic explanation for this phenomenon. An easier way to explain the low error rates with *why* would be to propose some sort of semantic uniformity principle, according to which learners initially assume that elements with similar semantics have similar syntactic behavior (see Berk, 2003). Thus, given that English *how come*-questions do not allow inversion, learners would initially extend this property to *why* as well. This proposal faces some problems. First of all, for the proposal to work, it would be important to show that *how come* questions are frequent or salient in the input. Additionally, if learners initially assume that *why* behaves like *how come*, one would expect their inversion rates in learners' productions to overlap, i.e. they should both never invert or invert to a similar extent. To the best of my knowledge, only three studies in the literature specifically examined inversion in *how come* questions: Kuczaj and Brannick (1979), Berk (2003) and Stromswold (1990). Kuczaj and Brannick found that children incorrectly imitate non-inverted *how come*-questions with inversion between 3% and 38% of the time.[73] In Berk's corpus, there were only 6 *how come* questions, 3 of which were correctly non-inverted and 3 of

---

[73] Group I inverted in 38.3% of their imitations of uninverted *how come* model sentences, while Group II inverted in 20% of imitations, Group III inverted in 28.3% of imitations, and Group IV inverted 3.3% of the time. Conversely, when presented with ungrammatical inverted *how come* questions, children in Group I correctly produced uninverted sentences 5.8% of the time, while children in Group II, III and IV did so 18.3%, 11.7% and 24.2% of the time, respectively.

which lacked an auxiliary altogether. Stromswold found that incorrect inversion with *how come*-questions ranged from 0% to 14%, but that *how come*-questions were overall extremely rare in child speech. Finally, in a grammaticality judgment task, Stromswold found that children dislike *how come*-questions, regardless of inversion. The low MLU group judged ungrammatical inverted *how come* questions as grammatical 50% of the time, while they judged grammatical non-inverted ones as grammatical 35% of the time. On the other hand, children in the medium and high MLU groups preferred grammatical non-inverted *how come* questions to ungrammatical inverted ones (65% and 40% for the medium MLU and 63% and 30% for the high MLU group, respectively).

In order to explain the particular behavior of *why*-questions, Thornton (2008) proposes that *why* in English child grammar behaves like its adult Italian counterpart *perché*.[74] As opposed to other *wh*-words, both elements are compatible with contrastive focus elements, can be followed by topic phrases and do not consistently require inversion. Following Rizzi (2001), Thornton proposes that *why* in child grammar can be optionally base-generated in Spec,IntP. IntP is higher than CP and is endowed with a [+wh] interrogative feature, thus not requiring movement of the [+wh] auxiliary to C. Alternatively, *why* can also be base-generated in TP and

[74] *Why* elements show a behavior that distinguishes them from other *wh*-words in many languages, including Spanish (Torrego, 1984), French (Rizzi, 1990), Irish (McCloskey, 2006), Korean, Japanese and Chinese (Ko, 2005). In Spanish, adjunct *wh*-elements do not give rise to obligatory V-to-C movement, while in French, *pourquoi* does not allow ‘stlylistic inversion’. On the basis that these elements are not sensitive to negative islands, Rizzi (1990) hypothesized that this is due to the fact that *why* elements in different languages are base-generated in Spec, CP. Ko (2005) proposes a similar analysis for *why* elements in Korean, Japanese and Chinese, due to the lack of intervention errors when *why*-elements are preceded by scope-bearing elements.

needs to move to Spec,FocP, thus requiring T to C movement of the [+wh] feature on the auxiliary. An interesting pattern noticed by Conroy (2007) and Thornton (2008) is that AL does not consistently invert in new information questions, but does so in long distance *why*-questions, as in (175), suggestions, as in (176), and rhetorical questions, as in (177):

(175) Why do you think Santa's not coming this year?

(176) Why don't you use this as a magic wand?

(177) Why would any witch not do spells?

According to Rizzi (2001), long distance *why* questions, in which the *wh*-element is base-generated in the subordinate clause, obligatorily invert in Italian, too, because the *wh*-element cannot be directly merged in the matrix Spec,IntP, but moves successively cyclically from the subordinate clause into Spec,FocP of the matrix clause.[75] Thus, the idea is that English-speaking

[75] Conroy (2006) challenged this judgment, claiming that lack of inversion is also possible in long distance embedded questions in Italian. According to this proposal, long distance non-inverted *why*-questions as in (vii) should be ambiguous between a matrix and an embedded reading. On the other hand, according to Rizzi (2001), the long distance reading should only be available if the verb and the subject have undergone inversion, as in (viii):

(vii) Perche' Gianni ha detto che si e' dimesso?
Why Gianni has said that *pro* resigned

(viii) Perche' ha detto Gianni che si e' dimesso?
Why has said Gianni that *pro* resigned

In agreement with Conroy (2006), I find that, to the extent that the embedded reading is available in (viii), it is also available in (vii).

children, just like Italian adult speakers, need to invert in long distance *why*-questions because, in this case, the *wh*-element cannot be base-generated. According to Thornton (2008), additional evidence in support of the existence of a correlation between subject-auxiliary inversion and *wh*-movement comes from the fact that two-clause *how come*-questions only have the matrix reading, as noted by Collins (1991).[76] In other words, in a sentence like (178), *how come* can only modify the matrix verb *said* and not the embedded verb *resigned*:

(178) How come John said that Mark resigned?

However, as suggested in Conroy (2006), it is possible that some other factor[77] might be responsible for the lack of an embedded reading in sentences like (179), given that *why the hell* also lacks an embedded reading (den Dikken & Giannakidou, 2002) but requires inversion of the subject and the auxiliary:

---

[76] Another similarity between *how come*-questions and uninverted questions in AL's corpus is that lack of inversion with *why* happened only in new information questions, where the use *how come* is grammatical, while suggestions and rhetorical questions, which are contexts in which *how come* cannot appear (Collins, 1991), were always inverted:

(ix) Why don't we go out for a drink tomorrow? (suggestion)
(x) *How come we don't go out for a drink tomorrow? (suggestion)
(xi) W`hy would anybody buy this? (rhetorical question)
(xii) *How come anybody would buy this? (rhetorical question)

[77] In Conroy's (2006) analysis, the unavailability of an embedded reading is associated with the presence of a 'surprise operator'.

(179) Why the hell did John say that Mark resigned?

3.2.1.2.1.3. Morphological Accounts

A number of proposals have suggested that errors in main questions are due to the specific morpho-syntactic properties of English. Santelmann et al. (2002) claim that inversion, as a particular instance of movement operations provided by Universal Grammar, is available in English-speaking children's grammars from the first testable stages, but that children have to learn the idiosyncratic properties of English auxiliaries (i.e., copula BE and auxiliary DO). The behavior of copula BE is unique because it is the only lexical verb that gives rise to inversion in (American) English, while DO is different from other auxiliaries in that it is not present in the underlying (declarative) structure. Santelmann and colleagues found that children were as accurate at repeating yes/no questions as they were at repeating declarative clauses in sentences with modals or auxiliary BE, while there was a significant effect of sentence type, for the youngest groups, in sentences with DO and with copula BE. Children's incorrect responses were coded as either word order mismatches (12% of the total mismatches) or inflection mismatches (80%). Significantly more word order mismatches occurred in questions than in declarative clauses, while inflection mismatches did not differ by sentence type. The authors conclude that children have knowledge of inversion from the earliest testable age, as predicted if inversion is an option instantiated in UG. What English-speaking children need to learn are language-specific morpho-syntactic properties, such as do-support and raising of main verb *be*. Guasti (1996; 2000; 2002) also suggested that the reason why English-speaking children produce more errors in main questions than children learning other languages (e.g., Italian or German) is due to the idiosyncratic properties of English morpho-syntax: while auxiliary verbs/modals raise to I, main verbs remain in V and require *do*-support. According to Guasti, this difference might be the source of the delay with which English-speaking children create a general rule for question formation. In a similar vein, Hattori (2003) imputes the fact that doubling errors are common in questions involving do-support, but are extremely rare in all other question types to the markedness of English do-support.

3.2.1.2.1.4. Syntax/Semantics Accounts: Presupposition

Roeper (2011a) proposed an account of T-to-C movement based on presuppositions. According to this analysis, the presence of a tensed element in a proposition has the effect of asserting the proposition. T-to-C movement, on the other hand, has the effect of blocking the assertion of that proposition. Roeper presents a series of facts in support of this hypothesis. For example, the difference between an inverted and a non-inverted yes/no question has to do with the presence of a presupposed proposition:

(180) Can you play baseball? -- no presupposition

(181) You can play baseball? -- presupposition: you can play baseball

Roeper argues that this proposal is able to account for a series of findings in child language acquisition. For example, Potts and Roeper (2005) suggest that children's verbless utterances are non-propositional and non-deniable, much like adult verbless exclamation like (182):

(182) You idiot!

If T-in-C causes the projection of an assertion, then children would be using T-drop in order not to project a proposition. Similarly, Roeper suggests that this analysis can account for the observation made by Van Valin (2001), among others, that children learning English tend to invert first with overtly tensed elements (e.g., *is*, *are*, *was*, *do*, *did*, *has*) and only later with non-overtly tensed elements (e.g., *can*, *could*, *may*, *must*, *might*, *shall*, *should*, *will*). Roeper (2011b) argues that this pattern fits with the theory summarized above, assuming that children realize at some point that T needs to be moved to C in the adult grammar in order for TP not to be asserted. The idea, then, is that children will first invert elements that are clearly marked for tense, and only later will start inverting elements that are not overtly tensed.

According to this proposal, double auxiliary questions in child English might be analyzed as a device to project the proposition under question: if T is left in C, the production should not be interpreted as a question, but as an assertion. As evidence for this analysis, Roeper reports that double tensed questions in child English are often uttered in contexts where the desired response has the same propositional context as the TP:

(183) [$_{CP}$ Do [$_{TP}$ you don't want to go outside]]?

Differences in presuppositional content, as already noted by Fitzpatrick (2005), distinguish *how come* from *why*-questions in adult English: while *how come*-questions are strongly presuppositional (and crucially are not associated with subject-auxiliary inversion), *why* questions are not. As evidence for this, Fitzpatrick notes that *how come*-questions, differently from *why*-questions, cannot be used in suggestions and do not license NPIs, as the contrast in (184)-(185) shows:

(184) * How come John ever said anything?

(185) Why did John say anything?

In a parallel fashion, Conroy (2006) observes that children fail to invert in presuppositional information *why*-questions, while they always invert in non-presuppositional suggestions and rhetorical questions. According to Roeper's analysis, presuppositional differences play a role in English embedded questions too: both in Belfast English and in AAE, T-to-C movement can only occur when the proposition expressed by the CP complement is not presupposed, and it is ungrammatical when the proposition is presupposed, i.e. when the main verb is factive:

(186) He didn't know could he do it

(187) *He knew could he do it

(188) *I remember clearly how many people did they arrest

I find the proposed account quite compelling, and I believe that a number of additional facts about the acquisition of questions might follow from this analysis.[78] For example, the reported argument-adjunct asymmetry in child English follows quite easily from an in depth analysis of the proposition associated with questions. Many authors have been concerned with the status of the proposition associated with a question (i.e., associated proposition or AP). A long tradition in the literature has argued that the associated proposition is presupposed (Katz & Postal 1964; Comorovski 1996; Beaver, 2001), while other researchers have analyzed the AP as a conversational implicature (Groenendijk & Stokhof, 1984; Ginzburg, 2004). In particular, while it has been argued that *why*-questions might not presuppose the proposition they modify (Fitzpatrick, 2005), the existence of a strong bias in *why*-questions has been argued for by a number of authors (Lawler, 1971; Tomioka, 2009, among others). For example, notice that presuppositions, in general, can be cancelled by adding *ever* to a *wh*-question, but that this is not possible in *why*-questions:[79]

(189)	When did she help you?	presupposition: she helped you (at some point)
(190)	When did she ever help you?	no presupposition
(191)	Why did she help you?	presupposition: she helped you (for some reason)
(192)	Why did she ever help you?	presupposition: she helped you

[78] What seems somewhat problematic in the account is that sometimes T-to-C movement is hypothesized to result in non-projection of an assertion and other times in the non-projection of a presupposition. Assertion and presupposition are clearly distinct, given that a regular declarative sentence asserts a proposition but does not presuppose it.
[79] This argument hasn't been made before in the literature, to the best of my knowledge.

Notice that this pattern interacts with mood and that only in the conditional mood can the presupposition of a *why*-question be lifted:

(193) Why would she (ever) help you?

Recent proposals in the literature (Brandtler, 2008; Eilam, 2011) argue that different types of *wh*-questions are associated with different types of APs. In Eilam's analysis, the AP of argument questions is an epistemic bias, while the AP of adjunct questions and cleft questions is a true presupposition.[80] The main difference between an epistemic bias and a presupposition is that, while an epistemic bias is "a speaker's belief, not necessarily shared by the hearer, that the probability that a proposition is true is greater than the probability that it is false" (Eilam, 2011:77), a presupposition needs to be satisfied in the common ground, that is, shared by the participants in the conversation. Eilam discusses two tests that can be used to distinguish between biased and presupposed AP: *too*-test and compatibility with negative answers. The first test discussed by Eilam is the ability of questions to serve as antecedents for *too*. While presupposed APs can serve as antecedents for *too*, biased APs cannot, as shown by (196):

(194) Q: Who is it that had a meeting with the dean yesterday?

[80] This account does not immediately account for d-linked argument *wh*-questions (e.g., *which*). I will return to this issue in the discussion of the experimental findings in Section 3.4.

A: I don't know, but I did, too.

(195) Q: When did John give his talk at Penn?

A: I don't know, but he’s talking tomorrow, too.

(196) Q: Who had a meeting with the dean yesterday?

A: #I don't know, but I did, too.

The second test is compatibility with negative answers. Felicity of negative answers in questions depends on question type. They are infelicitous with clefted *wh*-questions and adjunct questions, while they are felicitous with argument *wh*-questions:

(197) Q : Who is it that came ?

A : #No one

(198) Q : When did John buy that book ?

A : #Never

(199) Q : Who came ?

A : No one

The idea is that only questions associated with a presupposition are felicituous with *too* and questions associated with an epistemic bias are felicitous with negative answers.

It should be clear at this point that Eilam’s findings about the presuppositional differences of individual *wh*-words can be easily combined with Roeper’s account of inversion in child English: if English-speaking children leave T-in-TP when a proposition is presupposed, they will invert with argument *wh*-questions but not with adjunct *wh*-questions. Moreover, if children use T-in-TP to indicate that a proposition is presupposed, it follows that they should invert consistently when the proposition is not presupposed, e.g., in neutral yes/no questions. Finally, the proposal seems to be able to account for another very robust finding in the acquisition of questions, i.e., the fact that negated questions show a high rate of doubling errors and inversion errors. Given that negated questions exhibit a strong presupposition towards a negative answer, they are thus expected not to invert consistently in child English.

To summarize, the presuppositional account hypothesizes that early English productions use inversion as a device to encode differences in meaning that are not consistently encoded in the adult standard grammar. While in the adult standard grammar lack of inversion is optionally used to indicate a confirmation bias towards a proposition with the same content as the question in yes/no questions and is obligatorily used in strongly presuppositional *how come* questions, in child English lack of inversion is hypothesized to surface when the proposition associated with a *wh*-question is presupposed. The proposal is that children invert consistently when there is no bias or a weak bias towards a presupposition, and do not invert when the associated proposition is presupposed. The differences in inversion rates between different adjunct *wh*-elements might then reflect the strength of the AP: low inversion rates with strongly presuppositional adjuncts (e.g., *why*), and higher rates with less presuppositional adjuncts.[81]

It seems to me that this analysis could be easily accommodated in a model like Rizzi's (2001). For example, it could be hypothesized that different *wh*-elements target different

---

[81] See Lawler (1971) for the proposal that *when* is less strongly presuppositional than *wh*y.

positions in the structure according to the type of proposition they are associated with; for instance, adjunct *wh*-elements might target Spec,IntP and not require T-to-C movement (IntP is inherently associated with a *wh*-feature), while argument *wh*-elements would target Spec,FocP and require T-to-C movement in order to satisfy the *wh*-criterion, as is standard for all *wh*-elements in the adult grammar. This is all very speculative, but I find the proposal promising in terms of the breadth of phenomena it can account for.

3.2.1.2.2. Constructivist Accounts

The main claim of constructivist accounts is that children do not generalize over categorial variables like AUXILIARY or ARGUMENT/ADJUNCT, but acquire *wh*+auxiliary or auxiliary+noun combinations, hence basing their learning on frequent, lexically specific schemas (*how do*, *what are*, *where is*), which will eventually turn into more general/abstract schemas. In other words, constructivist accounts claim that children can produce adult-like questions without having access to grammatical rules (e.g., subject-auxiliary inversion) by reproducing specific high frequency formulae in the input. In a study of spontaneous production of 10 children aged 2-5, Rowland (2007) found a positive correlation between accuracy and frequency of specific frames in children's speech and also between accuracy in children's production and frequency of frames in the adult input, thus supporting the constructivist view that, at least initially, children can rely on memorized combinations to produce questions.

Constructivist accounts maintain that errors in children's productions are not linked to specific *wh*-words or auxiliaries but to combinations of particular auxiliaries and *wh*-forms. Rowland and Pine (2000), in their analysis of the production of *wh*-questions by a single child, did not find a systematic correlation between adjunct words and inversion errors: while only 9.3% of the child's *why* questions were inverted, *how* gives rise to more inversion than argument *what*. Contrary to the predictions of Valian et al. (1992)'s optional inversion rule, Rowland and Pine found that only 3 of the 46 combinations of *wh*-elements and auxiliaries occurred both in inverted and non-inverted form (*how*+*can, what*+*is, why*+*is*) in Adam's speech, lending some support to the constructivist claim that inversion and non-inversion are dependent upon individual *wh*-element+auxiliary combinations. They conclude that while individual *wh*-elements (e.g. *why*) and auxiliaries (*do, did, does*) may show a strong bias towards inversion or non-inversion, the identity of the auxiliary or *wh*-element alone fails to predict inversion errors. On the contrary, they show a significant correlation between frequency of *wh*-element+auxiliary combinations in maternal input and accuracy in the child's speech.

One serious problem that constructivist accounts need to handle is the fact that there is evidence that children's output does not always mirror the frequency seen in adult's input. So for example, Adam produced 33/36 non-inverted negative *why*-questions, despite the 25 instances of inverted negative *why* questions in the parental input. Rowland and Pine (2003) propose that this apparent discrepancy between input and output is due to the fact that for only 10 out of the 33[82] non-inverted *why*-questions attempted by the child was there a correspondent *wh*-word+auxiliary combination in the parental input, and that 6/10 were instances of *why*+don't. The authors argue that Adam could have picked up a *why*+*don't*+*you* formula from parental input and have failed to

[82] 10/33 is the count reported by Rowland and Pine. From Table 3 on page 172 of their paper, the count is actually 9/33.

generalize to a productive *why+don't* template. Once the 6 instances of non-inverted *why+don't* are removed from parental input, positive evidence was present in the input for only 4/33 (12.1%) of the child's non-inverted questions. The argument seems rather weak: first, 5/6 *wh+ don't* productions in Adam's speech used subjects other than *you*, thus making it unlikely that Adam was employing a *why+don't+you* formula. Second, the correlation between number of instances in the input and accuracy in Adam's questions is far from complete: the combination *how+does* occurs only twice in parental input, while *how+does* questions are inverted 100% (14/14) of the time in Adam's speech. Third, as noticed by Thornton (2008), questions addressed toward children include a very high number of questions with a second person singular subject, but this does not prevent generalization: maternal input to Adam contained 89 instances of *what+do+you* and only 2 of *what+do+they*. Although the input does not present many instances of *what+do* followed by a range of DP subjects, Adam shows evidence of having generalized the rule, since 27/27 *what+do* questions in Adam's production were inverted. Finally, as reported in Rowland and Pine (2003), maternal input from the same transcripts used to study Adam's questions presented 29 instances of *why+did* and 167 instances of *why+don't*, suggesting that input frequency cannot be the sole cause for the Adam's poor performance with *why+don't* and for his perfect performance with *why+did* questions. This fact is explicitly acknowledged by Ambridge and colleagues: "a third possible characterization of the relationship between input frequency and children's language acquisition is that input frequency (though perhaps not simple input frequency) is one of a number of many different factors that influence acquisition, quite possibly in a complex and interacting fashion" (Ambridge et al., 2006:545).

3.2.1.2.3. Functionalist Accounts

Van Valin (2001) proposes an account of inversion errors in L1 acquisition framed in the Role and Reference Grammar Theory. According to this analysis, English signals declarative illocutory force by placing a tense bearing element clause-internally; imperative force by lack of tense; and interrogative force by means of having a tense-bearing element clause-initially. Van Valin hypothesizes that children gradually develop this rule: they first start by placing explicitly tensed elements in clause initial position in questions (e.g., *is*, *are*, *did*, *has*), then by placing clause-initial elements that are not overtly tensed (e.g., *might*, *can*), and, finally, elements that do not end in a tense morpheme (e.g., *didn't, haven't*).

3.2.1.3. Methodologies

3.2.1.3.1. Spontaneous Production

One of the most widely used methods to gather child language data is the collection of naturalistic speech samples. Some of the earlier studies used parental diaries (Ingram & Tyack, 1979; Klee, 1985; Labov & Labov, 1978), while most spontaneous production studies use transcriptions of recorded speech samples (Bellugi, 1971; Dabrowska, 2000; Rowland, 2007; Rowland & Pine, 2000; Rowland et al. 2005; Stromswold, 1990; Valian et al., 1992).

The CHILDES database (MacWhinney, 2000) provides an impressive amount of data and is an ideal source for studies looking at the development of structures over time and for comparing children's productions with adult input. One of the obvious drawbacks of spontaneous speech studies is that spontaneous production, by its very nature, can only show what children use and know, and, hence, it might be mute with respect to what children do not know. Moreover, in the study of spontaneous speech, it is often hard to distinguish productive rules from formulae. Spontaneous production can also underestimate a child's competence: one might be tempted to infer from the absence of particular structures in children's speech samples that children are actively avoiding such structures because, if they knew them, they would use them in the appropriate contexts. However, unless an obligatory context for the production of a given structure is provided and the child attempts to produce it (unsuccessfully), we cannot argue that the child has not mastered the structure; absence of evidence is not evidence of absence! Like any other task, spontaneous speech might be affected by processing limitations (e.g., attention, memory, need to control fast motor responses) and by the interaction of processing limitations with factors that are likely to occur in an uncontrolled, non-experimental setting (e.g., presence of distractions, background noise, multiple interlocutors).

3.2.1.3.2. Elicited Imitation

In elicited imitation experiments, participants are presented with linguistic stimuli and are asked to repeat them verbatim. The dependent measure is the participant's repetition (with respect to some feature of interest) of the target utterance. Elicited imitation (EI) studies are based on the assumption that participants do not 'parrot' what is being said to them but actively reconstruct it based on the grammatical rules available to them. A clear advantage of EI over other production techinques is that the lexical load is lessened and that, differently from elicited production, it does not require elaborate experimental contexts to elicit a particular construction.

Some studies have shed doubt on the validity of EI as an experimental methodology, in that it has been found that children sometimes accurately reproduced what they did not produce spontaneously, or, more worrisomely, failed to reproduce what they had accurately produced spontaneously (Potts et al., 1979; Fraser, Bellugi & Brown, 1963; Hood & Lightbown, 1978; Kuczaj & Brannick, 1979). For example, contrary to the predictions of 'active reconstruction', children who were able to produce correctly non-inverted *how come* questions repeated the ungrammatical inverted ones verbatim (Kuczaj and Brannick, 1979), and children who consistently inverted *wh*-questions in their spontaneous and elicited speech repeated ungrammatical non-inverted questions verbatim (Sarma's (1991).

To sum up, it seems that EI is maximally informative in cases when researchers show that children are consistently unable to repeat some features of a construction verbatim, rather than when they show correct verbatim repetition. The studies on the acquisition of English main questions that have used EI are Kuczaj and Brannick (1979), Santelmann et al. (2002) and Valian and Casey (2003).

3.2.1.3.3. Elicited Production

Differently from EI, accurate production in an elicited production study should be taken as evidence of knowledge by the child. This is because, assuming that the experimenter chose experimental items that are unlikely to be rote-learned formulae, the only way for the child to produce the correct structure is to have knowledge of that structure (however this knowledge is to be represented). Elicited production (EP) experiments aim at creating contexts that are uniquely felicitous for the production of a given structure. In the case of *wh*-questions, a uniquely felicitous context is one in which the child does not know the answer to a question, while someone else can provide the child with this information. As already pointed out by Klee

(1985) and Sarma (1991), a crucial problem with some of the EP studies in the literature is that they used embedded questions in order to elicit main questions. The problem lies in the fact that in English embedded questions, the order of the subject and the auxiliary is exactly the opposite of the order found in main questions. The surprisingly high rate of inversion errors in Bellugi's (1971) and Erreich's (1984) studies could then be due partially to this methodological pitfall.

This is an extremely important methodological point and some evidence that the methodology used to elicit main questions can have serious repercussion on the results can be illustrated by looking at Sarma's results. In Experiment 1, she tested 16 children and successfully elicited 161 main questions (argument and adjunct), finding only 8 non-inversion errors, produced by two children in *where* and *why* questions. Crucially, the experimental protocol in Experiment 1 did not use embedded questions to elicit main questions:

> *Experimenter: In this game there is a bed, some French fries, and Mickey Mouse. Mickey Mouse is holding a bottle of ketchup. OK, Mickey, shut your eyes! Now, Mickey Mouse is looking for a place to pour the ketchup*[83]*. He sees the bed and thinks, 'Ugh, that's a lousy place to pour ketchup'. Then he sees the French fries and thinks 'Wow, that's the perfect place!' So there he is, pouring the ketchup on the fries-pssst! So, he's not pouring it on the bed, but he's pouring it somewhere. Ask Mickey where.* (Sarma, 1991: 74)

---

[83] It is unclear to me how Mickey can have his eyes shut and at the same time be 'seeing' all the props. While it is possible that there were two different puppets in this experimental set-up, it seems unlikey that they would be called Mickey and Michely Mouse.

In Experiment 3, which was conducted with 6 children, Sarma finds higher non-inversion rates: 23 non-inverted questions over a total of 58. These non-inverted sentences are produced by 4 children. A closer look at the protocol used in Experiment 3 shows that in this case, non-inverted embedded questions were used in the prompt:

> *In this story we have a dog named Pluto. Now, Pluto has all kinds of ways to go to his friend's house. He can go by truck, or he can go on a bike, or he can go on roller skates. Today, he wants to go on roller skates. So, <u>we know which way Pluto can go</u> to his friend's house. Ask Mickey which way.* (Sarma, 1991: 94)

However, it should be noted that Ambridge et al.'s EP experiment used embedded questions as prompts for main questions but found overall fairly low rates of non-inversion errors across *wh*-words (~10%). The studies that have used EP to study the acquisition of English questions are Ambridge et al. (2006), Bellugi (1971), Erreich (1984) and Sarma (1991).

3.2.1.3.4. Grammaticality Judgments

To the best of my knowledge, only two studies have investigated children's knowledge of inversion via grammaticality judgments: Stromswold (1990) and Grinstead et al. (2009).

In Stromswold's study, children listened to a dog puppet producing a sentence and got to feed him a bone or a rock, depending on whether they thought the dog's sentence sounded good or not. In Grimstead's study, on the other hand, children listened to two different puppets saying a variation on the same sentence and got to award the puppet who said it better by feeding him.

The advantage of these studies is that, like all grammaticality judgments, they aim at limiting performance factors and processing demands. However, these studies are hardly conclusive: in a study with 22 children aged 3-6 (mean age 4;6), Stromswold found that children prefer inverted main questions (accepted 76.5% of the time) over non-inverted main questions (accepted 51.5% of the time), and non-inverted embedded questions (accepted 77.3 % of the time) to inverted embedded questions (accepted 62.1% of the time). While it seems clear that children prefer grammatical to ungrammatical interrogative structures, the high rate of acceptance for ungrammatical structures casts some doubt on children's ability to perform the task or on what they thought the purpose of the task was.

3.2.2. L1 Acquisition of Embedded Questions

While first language acquisition of English main questions has attracted much research interest and has been the focus of heated debates, the acquisition of embedded questions is a surprisingly understudied phenomenon. With the exception of Stromswold (1990) and Sarma (1991), few researchers even mention the existence of non-standard subject-auxiliary inversion in first language acquisition. de Villiers (1991)'s account of non-inversion in main questions capitalizes on the acquisition of embedded questions: as discussed above, according to de Villiers, children start inverting consistently in main questions once they start producing embedded questions because they realize that *wh*-elements are located in Spec, CP. The reasoning is not completely transparent: while de Villiers seems to claim that children only use embedded questions to learn about the position of *wh*-elements in the left periphery (Spec, CP), one would need to ensure that they wouldn't also use embedded questions as triggers for word order, thus producing non-inverted main and embedded questions.[84]

It has been proposed in the literature (Brown & Hanlon, 1970; Maratsos et al.1979; Valian et al., 1992; Tornyova & Valian, 2009) that inversion errors in main *wh*-questions could be partly due to misleading data from embedded *wh*-questions. With respect to this, Erreich (1984) points out that "a problem with this account is that it cannot explain the unidirectional nature of the overgeneralization: if the non-inversion pattern in embedded questions is over-extended to matrix *wh*-questions because the two constructions are considered similar, there ought to be an equal amount of overextension of the inverted pattern from matrix *wh*-questions to embedded ones" (Erreich, 1984:590). It seems to me that not only is the unidirectionality of the overgeneralization problematic, but it also goes in the opposite direction than one might expect, i.e., the least frequent pattern is over-generalized to the most frequent pattern. Moreover, the existence of inversion in embedded questions is controversial: Erreich (1984) notes that only one study has found instances of such overgeneralization, *I know what is that* (Menyuk, 1969), and that, in her own study, there were only four such errors produced by two children. de Villiers, on the other hand, writes: "No satisfactory account is provided here of why children should not invert in embedded clauses, and in fact, they do, at first […]" (de Villiers, 1991:171).

---

[84] It seems to me that the *Penthouse Principle* (Ross, 1973), according to which 'more goes on upstairs than downstairs' (i.e., movement operations may only apply to main clauses, but not only in embedded clauses), should guide the learner against generalizing word order facts from main to embedded clauses, but not vice-versa. For a similar idea, see Roeper & Weissenborn (1991), while for the idea that children learn only from unembedded binding domains, see Lightfoot (1989, 2004).

A larger body of data on the acquisition of word order in embedded clauses is available for other Germanic languages. Westergaard and Bentzen (2007) present a summary of the relevant literature. The early findings on the acquisition of word order in German is that correct placement of the verb in verb final position is in place from the moment when embedded clauses start appearing in spontaneous speech (Clahsen & Smolka, 1986). On the other hand, Penner's (1996) study on one child learning Swiss German shows that a first phase of correct word order (through age 3;2) was followed by a short period in which V2 alternated with the target verb-final order, while Schönenberger (2001) found a consistent V2 pattern through age 4;11 in two children learning Swiss German. Verb placement errors for monolingual German and bilingual German-English children are also reported by Gawlitzek-Maiwald, Tracy and Fritzenshaft (1992) and Dopke (1998). Håkansson and Dooley-Collberg (1994) found that Swedish-speaking children produce non-target word-order patterns in embedded clauses, allowing modals to move across negation and adverbs. Westergaard & Bentzen (2007), in a corpus study of three children aged 1;9-3;3, found that Norwegian children learning the Norwegian variety spoken in Tromsø produced non-target word order, with the verb moving across negation and adverbs, in embedded declarative clauses but not in embedded questions. The children in their corpus produced 108 embedded questions, all of which occurred with target-consistent non-V2 order. The authors also ran a pilot study where they elicited embedded questions in two children (ages 5;9 and 8). The older child produced all target consistent embedded questions (12/12), while the younger child placed the verb before negation or the adverb in 7/8 contexts, but never past the subject. By comparing the relative frequencies of embedded declaratives (0.5% in the mother's input and 0.44% in the investigator's) and embedded interrogative structures (1.04 % in the mother's input and 2.5% in the investigator's), the authors reject the hypothesis that input frequency is the cause of errors in the former and lack of errors in the latter. They argue that the difference in input frequency is not large enough to explain ungrammatical/disfavored verb movement in embedded declaratives and lack of V2 order in embedded clauses. Conversely, they hypothesize that children are guided by economy principles and that they avoid movement unless there is evidence for it in the input.

The CHILDES corpus that Stromswold (1990) investigated consisted of 55,700 sentences containing an auxiliary. Twelve monolingual English-speaking children (age range 0;11-2;10 when recording began, 0;11-7.10 when recording ended)[85] were included. Overall, Stromswold's data show that children invert 91.5% of the time (range: 50.6-97.4; median: 94%) in main questions. In embedded contexts[86], inversion was around 10% in embedded *wh*-questions (36/364[87]) but non-existent in the production of embedded yes-no questions by 8 children (0/46)[88]. Stromswold investigated whether there was a correlation between inversion rates in

[85] The amount of available speech per child also varied greatly. The range was 1,500 to 38,000 lines (6 to 139 transcripts per child).
[86] The age at which children produce embedded inversion in this corpus ranges from 2;6 to 4:11 and varies greatly across children: Peter inverted in embedded questions between 2;6 and 2;10, Sarah between 4;2 and 4;10, Adam between 3;0 and 4;10, Mark between 3;8 and 4;8, Ross between 3;5 and 4;11, Shem between 2;7 and 2;11, Nina between 2;10 and 3;3, while Naomi showed embedded inversion only in the transcripts from age 2;11 years.
[87] 50 additional inverted embedded w*h*-questions were not counted in the analysis because they were hesitations or contained a contracted auxiliary.
[88] All 46 yes/no questions were introduced by the complementizer *if*. There were no embedded yes/no questions introduced by *whether*.

main and embedded questions for different *wh*-words.[89] She found a significant correlation ($r = .875$, $p < .01$). Interestingly, she also found instances of inversion with embedded clauses that were 'distinctly un-question like' (Stromswold, 1990:162):[90]

(200) I know what time is it [Adam41 (3;11)]

(201) Look at what are they building? [Adam47]

(202) I know where de is de 'A'. [Adam36]

(203) Look how good boy am I [Adam29]

(204) Dis is how (will) it go [Sarah120]

(205) This how shall be begin[91] [Adam52]

Given that these data come from a spontaneous production corpus, some instances of embedded inversion could just be improperly transcribed examples of quotative questions. This possibility is hard to test, given that only four of the inverted embedded questions were introduced by the verb *wonder*, which is the only verb disallowing quotative questions in the adult grammar. Some of the examples ended with a question mark, suggesting perhaps that the prosody was that of a main question:

---

[89] Eight of the thirty-six inverted embedded questions were introduced by *don't know*, ten by *know,* one by *don't understand*, two by *see*, two by *say*, one by *don't say*, two by *look* and four by *tell*. Three instances contained *who*, twenty-three *what*, six *where*, eight *how*, and three *why*.

[90] As a matter of caution, it should be noted that 4/5 of these utterances come from Adam's corpus.

[91] There were two such tokens in the Adam52 transcript.

(206) I don't xxx what is dis? [Adam24]

(207) I wonder what are dese for? [Adam44]

(208) Look at what are they building? [Adam47]

(209) And he said what are you named? [Mark63]

(210) Tell me what do you wanna do this morning? [Nina55]

(211) I don't know what is his name? [Ross57]

(212) You - you- you tell me what is it? [Shem30a]

(213) Know where'd is it? [Peter 18]

(214) Yeah. No dat's where is it? [Shem32]

(215) I said how is Rinny doing? [Adam52]

(216) I don't understand why is the grass poisoned up? [Mark78]

The second source on the acquisition of embedded questions available in the literature is Sarma's (1991) Ph.D. dissertation. She designed an experiment to see whether the embedded inversion phenomenon Stromswold documented in spontaneous production was also apparent in elicited production. In order to elicit embedded questions she presented children with the following set-up:

> *Experimenter: This time we're going to have a box, a chair and a horse. Now this horse wants to hide somewhere. He can't hide under a chair-he's too big! But look, he can hide*

in the box. So we know that the horse can hide in one of these places, right? Ask Mickey if he knows where.[92]

Child (target): Do you know where the horse can hide? (Sarma, 1991:88)

After the child produced an appropriate question, the puppet would answer and the child would get to feed the puppet. The materials consisted of 10 prompts aimed at eliciting embedded questions: two subject questions (1 *what*, 1 *which*), 4 non-subject argument questions (2 *what*, 2 *where*), and 4 non-subject adjunct questions (2 *why*, 2 *where*). The verbs used in the prompt were *know* (used once), *guess* (used twice), *remember* (used 4 times) and *show* (used 3 times). The participants were eighteen children whose ages ranged from 3;0 to 5;7 (mean age: 4;6).[93] Children's accuracy was 100% and there were no instances of embedded inversion. Children also produced some embedded questions in their spontaneous speech: a total of 13 embedded questions were produced, 2 of which were inverted. Sarma attributed this to performance factors, but it is interesting that, although virtually nothing can be inferred from such a small number of spontaneous utterances, the relative number of errors in spontaneous production resembles that in Stromswold's spontaneous production study. A potential problem in this study is the use of the verb *show*, which does not seem to have interrogative force and seems more likely to be

[92] Note that Mickey always knew the answer to the child's question; this fact is potentially problematic because it might give rise to a number of non-*wh*-questions (Do you know that the horse can hide in the box) and because embedded inversion has been found to be associated with contexts in which the experiencer of the main verb (e.g., know) does not know the answer to the embedded question (see page 259).

[93] 4 children were under 4 (ages: 3;0, 3;2, 3;8, 3;11), 8 were between 4 and 5, 6 were between 5 and 7.

followed by a free relative. A more crucial problem has to do with the fact that both who was asking the question (the child) and who was answering it (the puppet) already knew the answers to the embedded questions. As McCloskey (2006) notes, embedded inversion in the English varieties that allow it seems to be restricted to contexts that denote "[…] a question act, understood as a certain kind of context change potential. The question act is appropriate in this context only if the issue it raises is unresolved for the individual denoted by the experiencer argument of the verb" (McCloskey, 2006:35). It could then be that the lack of embedded inversion is due to an experimental artifact, i.e. the fact that the embedded question did not have real interrogative force, and that the issue was not unresolved for the subject of the main verb, the puppet. A final concern has to do with the paucity of experimental items (given that subject questions are mute with respect to inversion facts).

In conclusion, the data in the literature do not provide conclusive evidence of the extent of embedded inversion in English child language, or of the elements that are more likely to trigger it (individual *wh*-words, verb types, etc.). An elicited production experiment should also address the question of whether children, like speakers of non-standard varieties of English, have the option of omitting the yes/no complementizer (*if*, *whether*) and consequently producing embedded inversion in yes/no questions. Ideally, the same children should be tested on both main and embedded questions, to see whether there is a correlation between inversion in main and embedded questions.

3.3. Present Study: L1 Production of English Questions

3.3.1. Pilot Studies

As detailed in Section 3.2.1.1., many issues regarding the acquisition of subject-auxiliary inversion by children learning English as their first language remain open. For example, there is an ongoing debate with respect to whether main yes/no and *wh*-questions behave differently in acquisition. One could expect inversion rates to be lower in yes/no questions, given that non-inverted yes/no questions are grammatical in the adult input (Crowley & Rigsby, 1987; van Herk, 2000). Conversely, one could expect inversion rates to be higher in yes/no questions, given that in this structure, the auxiliary (and the fact that it occurs before the subject) is salient due to its often being sentence-initial (Gleitman, Newport & Gleitman, 1984), i.e., while the auxiliary can never be the first word in a *wh*-question, the auxiliary is more often than not the first word in a yes/no question.

Early work (e.g., Klima & Bellugi, 1966) suggested that mastery of yes/no questions is obtained earlier than mastery of *wh*-questions, while subsequent work (e.g., Erreich, 1984: Valian, Lasser & Mandelbaum, 1992) has failed to replicate this result. To this day, no experimental study that I am aware of has systematically investigated the acquisition of yes/no and *wh*-questions in the same population via the same experimental protocol. Moreover, a large body of studies in the 1990s (de Villiers, 1991; Sarma, 1991; Stromswold, 1990) suggested that children may treat argument and adjunct *wh*-words differently, while other work suggests that the reported asymmetry is restricted to individual *wh*-words (e.g., *why*) and not to the whole adjunct class.

Similarly, there are many open issues in the acquisition of embedded questions. Very little research has been done on this topic, and the results in the literature are contradictory: Stromswold (1990) found that children produce a substantial number of word order errors in embedded *wh*-questions, while Sarma (1991) found that they are at ceiling with respect to their mastery of non-inversion in embedded *wh*-questions.

The present study is aimed at contributing to the research on the acquisition of word-order patterns in main and embedded questions by eliciting main and embedded yes/no and *wh*-questions in a group of children acquiring English as their first language. Three pilot studies were conducted before settling on the current experimental protocol.

3.3.1.1. Pilot Study 1

The first pilot experiment involved an interaction with 3 puppets and two experimenters. In order to elicit main questions, one experimenter held a curious but shy puppet (Elmo) who wanted to know 'things', while a second experimenter held another puppet (Bert) who knew all the answers. The child was prompted by Elmo to ask a question to Bert.

To elicit embedded questions, one experimenter held Elmo, who would ask questions to another puppet (Oscar the Grouch), played by the second experimenter. Oscar was in his trashcan, not able to hear Elmo's question. After Elmo had asked his question, Oscar would come out of the trashcan and ask the child: "What did Elmo want to know?"

The idea was to create a pragmatically natural context for the production of an embedded clause. This protocol was piloted with 3 children. The presence of three puppets and two experimenters seemed to provide too much distraction and to be somewhat confusing for the children, who did not seem to understand the task.

3.3.1.2. Pilot Study 2

The experimental protocol was similar to that of Pilot 1, but it only involved the presence of the two experimenters: in the case of main questions, the first experimenter (instead of Elmo) would prompt the child to ask questions to the second experimenter (instead of to Bert).

In the case of embedded questions, the first experimenter would ask questions to the second experimenter (who was covering her ears), and the second experimenter would then ask the child: “What did [Experimenter 1] want to know?” This protocol was piloted with two children. Children seemed confused by the task and did not produce many target structures.

3.3.1.3. Pilot Study 3

The new protocol was based on the successful format used in a standardized test (SPELT-P2) to elicit main questions and was adapted to elicit embedded ones. This new protocol only required one experimenter and two books. In the first part of the experiment, the aim was to familiarize the child with the task of asking questions. The experimenter showed the child a box and told the child to guess what object was hidden in the box by asking questions (‘I have something in this box. You can ask me questions about it’). The idea was that of collecting a corpus of spontaneous main questions.[94]

---

[94] A similar task was successfully used in Klee (1985).

In the second part of the experimental session, the experimenter showed the child a picture book and said that they would play an 'asking' game. All pictures in the book depicted a young girl (Katie) in the act of asking her mother a question. There was a thought bubble next to the girl, with the picture of the object used in the prompt. The experimenter held an index card with the prompt typed on it. The experimenter would then prompt the child to ask a question for Katie ("Katie's dog is eating something. Katie wants to find out WHAT. She says…".). After the child produced a question, the experimenter would give the index card to the child and asked her to put it in an envelope. No answer was produced in response to the child's question.

In the third part of the experimental session, the experimenter showed the child a second picture book and said that they would play a 'remembering' game. All the pictures in the book depicted a young girl in the act of asking her brother a question. There was a speech bubble next to the girl, displaying the text of the question. The experimenter read the question to the child. The child was told to remember the question. After reading Katie's question to the child, the experimenter would ask the child: "What did Katie want to know?" The child was instructed to always start her answer with 'Katie wanted to know…', in order to make stand alone embedded answers (e.g., 'What Jim was eating') and direct quotation of the prompt less likely.

This protocol was fairly successful at eliciting main and embedded questions. However, a number of problems emerged: children did not produce many spontaneous questions in order to discover what was in the box. They seemed confused by the task and would either only ask "What is it?" or they would try to guess "It's a ball! It's a frog!" The protocol used to elicit main questions was successful with eight of the children, but five children were confused by the task and did not produce any questions. I conjectured that part of the problem was the lack of feedback to the child's question, which made the task unnatural. The protocol used to elicit embedded questions was fairly successful too, but it was also unnatural in that the experimenter both read the question that Katie was asking her brother and subsequently prompted the child by asking what Katie has asked. Through discussion of the protocol with the parents, it emerged that some of the older children could read and hence were probably not remembering the prompt but reading it directly. This might cause the children who can read to produce a high number of quotative questions.

3.3.1.4. The Final Protocol

The final experimental protocol is a modification of the protocol used in Pilot 3. A number of changes were made:

1) The first part of the experiment was taken out because it was unsuccessful at eliciting a range of spontaneous questions.

2) The same pictures developed for main and embedded questions were used in the final protocol, but they were presented on a computer screen through the aid of Microsoft Power Point and not on paper.

3) In the main question experiment, children were prompted to ask a question to Katie's mother as before. The main difference is that appropriate answers to the child's target question were pre-recorded for each experimental trial. After asking a question, the child would now click on a button and hear the pre-recorded answer to her question. After hearing the answer, the child received the index card from the experimenter and would put it in a box.

4) In the embedded question experiment, children would not be read Katie's question by the experimenter but would play the recording by clicking on a button. A research assistant recorded all the prompts. The recordings were manipulated using the software Audacity. Formant frequencies were modified so that the questions sounded as though they were uttered by a child. The experimenter would then prompt the child to remember Katie's question ('What did Katie want to know?'). Crucially, the brother did not answer Katie's question. This was done to ensure that the child would not answer to the experimenter by proving the information to the question (e.g., 'Katie wanted to know *that*...'). After producing a response to the experimenter, the child would put an index card in the box.

5) During Pilot 3, it was noticed that alternation between yes/no and *wh*-questions was giving rise to a high number of non-target responses, in that children would perseverate with one type of question and would produce yes/no questions in place of *wh*-questions or vice versa. I thus decided to present *wh*- and yes/no questions in two different blocks (8 *wh*-questions followed by 8 yes/no questions or vice versa as described below).

3.3.2. Experimental Investigation

3.3.2.1. Experiment 1: Main questions

3.3.2.1.1. Method

3.3.2.1.1.1. Participants

Child participants were all monolingual English-speaking children. They were recruited through local day-cares, personal contacts and Experian[95]. At the end of the experimental session, children were administered a standard expressive language test normed on preschool children (SPELT-P2). This test is normed on English-speaking children aged 3-5;11. Only children that scored within the normal range for their age were included in the analyses. A total of 35 English monolingual participants were tested. Two participants were excluded: one was excluded because she was diagnosed with auditory processing impairment and was receiving speech-language therapy at the time of testing, and one was excluded because his SPELT score was more than 2 SD below the mean for his age group, suggesting possible language impairment. The average age of the children included in the analyses was 4;3 (Range: 3;3-5;9, SD = 8 months; Median: 4;2) and the average SPELT score was 32.3/40 (SD = 3).

[95] Experian is a credit information group that collects and provides information on people and businesses.

3.3.2.1.1.2. Materials

Two experimental lists were constructed, each containing six practice items and sixteen experimental items. Type of question (yes/no vs. *wh*-) was a fully within factor, while type of *wh*-word was a within subject but between item factor. The two lists differed with respect to the fully within variable: the items that appeared as prompts for yes/no questions in List 1 appeared as prompts for *wh*-questions in List 2 and vice versa. Yes/no questions and *wh*-questions were elicited in blocks and did not alternate. This was done because, during pilot testing, it emerged that children had a strong tendency to perseverate with the question-type used in the previous item, thus producing a high number of non-target responses

To control for order effects, two additional lists (1B and 2B) were created by switching the order of the first four and last four items in each block (i.e., the last four items in the first block in List A appeared as the first four items of that block in List B). *Wh*-questions were thus elicited in the first block in list 1A and list 2B, while yes/no prompts appeared in the first block in list 1B and 2A. Each block of experimental questions was preceded by three practice items, for a total of six practice items. There were 4 argument *wh*-words (object *what* and *which*) and 4 adjunct *wh*-words (*why*, *when*). Participants were assigned randomly to the experimental lists.

All stimuli contained a transitive verb and were compatible with all *wh*-words. Each verb occurred in both lists (either in a prompt aimed at eliciting a yes/no question or in a prompt aimed at eliciting a *wh*-question). All items contained the appropriate form of auxiliary *be* in the prompt in order to reduce the likelihood of the child not producing an auxiliary. The experimental items were in present tense, while the practice items were in past tense. Half of the prompts were aimed at eliciting a question with a second person subject and half of the items were aimed at eliciting a question with a 3rd person subject. Verbs were matched for number of syllables and frequency in child's speech (based on the MacArthur Communicative Development Inventory). The experimental items were pseudo-randomized so that no more than 2 consecutive experimental sentences shared any of the features relevant to the investigation (i.e., yes/no and *wh*-type). See Appendix F for the four experimental lists.

3.3.2.1.1.3. Procedure

Each child was tested individually in a quiet room. The child was seated in front of a computer screen and the experimenter would explain the game to the child:

> *This is an asking game. Look, this is Katie and this is her mom. Katie wants to know some things. We are going to help her ask her mom questions. Let's play!*

The experimenter would then move on to the practice items. During the practice phase, if the child did not produce a target question[96], the experimenter would say the target question and ask the child to repeat it. However, no explicit corrective feedback was given to the child. After the practice trials, the child would move on to the first eight experimental trials. If the child forgot the prompt, was distracted, or produced a non-target question (e.g., a *what*-question instead of a

[96] I considered as 'target' a production where the question-type (*wh*- or yes/no) and the *wh*-word (*what*, *which* or *who*) were the same as the prompt. No feedback with respect to subject-auxiliary inversion was provided. For example, if the child produced a yes/no question without inversion or a subject question with inversion, the experimenter would accept it and move on to the next practice item.

why-question) during the experimental phase, the experimenter prompted the child again. Each child received a prompt a maximum of two times. If, after being prompted twice, the child still did not produce the target question, the experimenter would move on to the next prompt. After the first eight experimental trials, the child was started on the second block of questions and was presented with three practice items and eight additional experimental items. The experiment took approximately 10-12 minutes and was recorded on a digital recorder.

3.3.2.1.1.4. Transcription and Coding

Each experimental session was transcribed by either me or a research assistant, and then checked by a second researcher. When there was a relevant discrepancy between the transcribers (e.g., presence vs. absence of a subject or auxiliary), a third researcher listened to the relevant production and discussed it with the primary transcriber. All discrepancies were resolved through discussion. All utterances were then coded and the coding was then checked by a second researcher. Disagreements were resolved by discussion.

The same coding scheme used for the L2 production studies was used in this experiment. Each production was coded as either correct (adult-like) or incorrect (non-adult-like) with respect to word order, verbal morphology and presence of target lexical items (e.g., subject and *wh*-elements).[97] Following Ambridge et al.'s (2006) coding scheme, incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors, as in (217), or raising errors as in (218).

[97] Lack of inversion in main yes/no questions was coded as incorrect for ease of comparison with *wh*-questions.

- Double tense/double auxiliary errors, as in (219)–(220).
- Omitted auxiliary errors, as in (221), or errors that, due to lack of morphology, were ambiguous between non-inversion and omitted auxiliary errors, as in (222).
- Other errors. Other errors included questions that differed in type from the target (yes/no instead of *wh*-questions and vice versa), subject *wh*-questions instead of object *wh*-questions, as in (223), productions that differed from the target in the lexical items used; questions without a subject, questions with VP movement, as in (224), and questions where the *wh*-word differed from the target one, as in (225). Skipped items and unintelligible productions were also coded in this category.

(217) Why you are calling dad?

(218) When my brother sings a song?

(219) What does my brother builds?

(220) What are you are cooking?

(221) Which cat you brushing?

(222) Why you call dad?

(223) Which one is brushed?

(224) Which toy was chewing the dog?

(225) Why are you feeding the doll? (target: which doll are you feeding?)

3.3.2.1.2. Results and Interim Discussion

The main goal of this experiment was to examine the relative contribution of question type and *wh*-type to the production of subject-auxiliary inversion in embedded English questions by first language learners of English. As was the case with the L2 studies, two sets of analyses were performed: the first set of analyses used arcsine transformed mean percent *correct* productions as the dependent variable. This was calculated by dividing the number of correct responses by the *total* number of productions. The second set of analyses used arcsine transformed mean percent *inversion* as the dependent variable, and this was calculated by dividing the number of correct responses by inverted and non-inverted responses; productions that provided no evidence one way or another with respect to inversion or that contained errors unrelated to subject-auxiliary inversion (i.e., auxiliary omission, different structure, morphological errors, etc.) were thus excluded from this second set of analyses.

3.3.2.1.2.1. Question-type

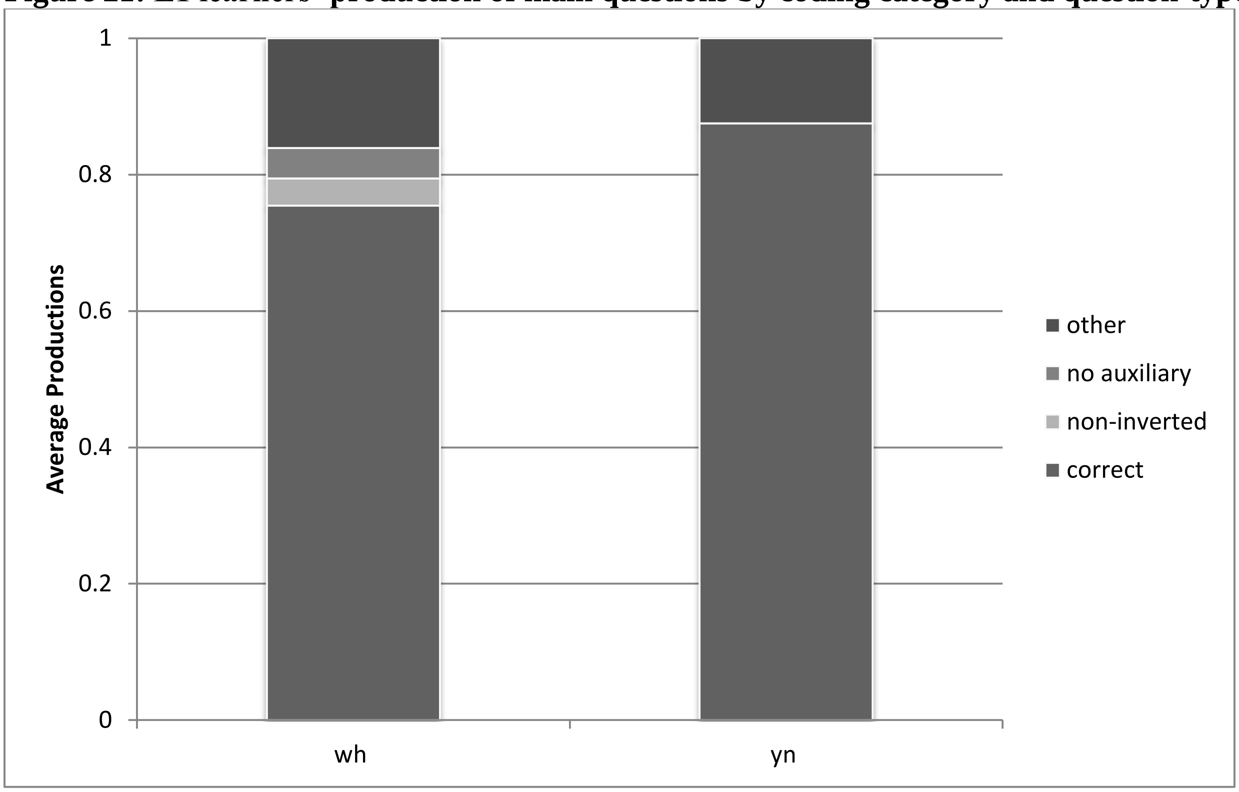
Table 69 reports the raw number (and percentage) of productions in each coding category by question type and Figure 21 summarizes the distribution of productions. In order to investigate whether the same pattern of errors emerged in younger and older children, children were further divided in two groups based on their age. The groups were created based on the median age (4;2). There were 16 children in the first group of younger (age below 4;2) children and 17 children in the older (age above 4;2) group. The average SPELT score for the children in the first group was 31.3/40 (Range: 22-36; SD = 3.4) and the average score for the older group was 32.7 (Range: 24-37, SD = 3.6). Table 70 and

Table 71 report the raw number (and percentage) of productions in each category by question-type, for children in the younger and in the older group, respectively.

Table 69: L1 learners' production of main questions by coding category and question-type

Coding	Question Type	
	Yes/No-	***Wh-***
Correct	230 (87.5%)	195 (73.9%)
Non-inverted	1 (0.4%)	13 (4.9%)[98]
Double aux/tense	0	1 (0.4%)
No auxiliary	0	11 (4.2%)
Other	33 (12.1%)	44 (16.6%)
Total	*264*	*264*

Figure 21: L1 learners' production of main questions by coding category and question-type



[98] Notice that while the average rates of non-inversion errors in this study is somewhat lower than the values reported in other studies, if we collapse non-inversion and no aux errors, as many studies in the literature have done, the error rates for yes/no questions are comparable to those reported in the literature.

Table 70: L1 learners' production of main questions by coding category and question-type – Children under 4;2

Coding	Question Type	
	Yes/No	***Wh-***
Correct	110 (87.7%)	87 (68%)
Non-inverted	0	6 (4.7%)
Double tensaux/tense	0	0
No auxiliary	0	6 (4.7%)
Other	18 (13.3%)	29 (22.7%)
Total	*128*	*128*

Table 71: L1 learners' production of main questions by coding category and question-type – Children above 4;2

Coding	Question Type	
	Yes/No	***Wh-***
Correct	120 (88.2%)	108 (79.4%)
Non-inverted	1 (0.7%)	7 (5.1%)
Double teaux/tense	0	1 (0.7%)
No auxiliary	0	5 (3.7%)
Other	15 (11%)	15 (11%)
Total	*136*	*136*

A first 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *correct* as the dependent variable showed a significant effect of question type (F_1 (1,31) = 15.2, p < .0001; F_2 (1,15) = 9, p = .009), no effect of age group (F_1 < 1; F_2 (1,15) = 2.7, n.s.), and no interaction between question type and age group (F_1 (1,31) = 1.5, n.s.; F_2 < 1).

A second 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *inversion* as the dependent variable showed a significant effect of question type F_1 (1,30[99]) = 12.5, p = .001; F_2 (1,15) = 5.4, p = .035) and no effect of age group (F_1 < 1; F_2 (1,15) = 1.8, n.s.) and no interaction (all *Fs* < 1).

Overall, yes/no questions were associated with higher rates of correct responses and with higher rates of inverted responses in both age groups. Older children did not produce higher rates of correct or inverted responses.

3.3.2.1.2.2. Wh-type

Table 72 reports the raw number (and percentage) of productions in each coding category by *wh*-type and Figure 22 summarizes the distribution of productions. In order to investigate whether the same pattern of errors emerged in younger and older children, children were divided in two groups based on their age on the basis of the median age (4;2).

[99] One child exclusively produced non-target yes/no questions and was thus excluded from this analysis.

Figure 22: L1 learners' production of main questions by coding category and *wh-type*

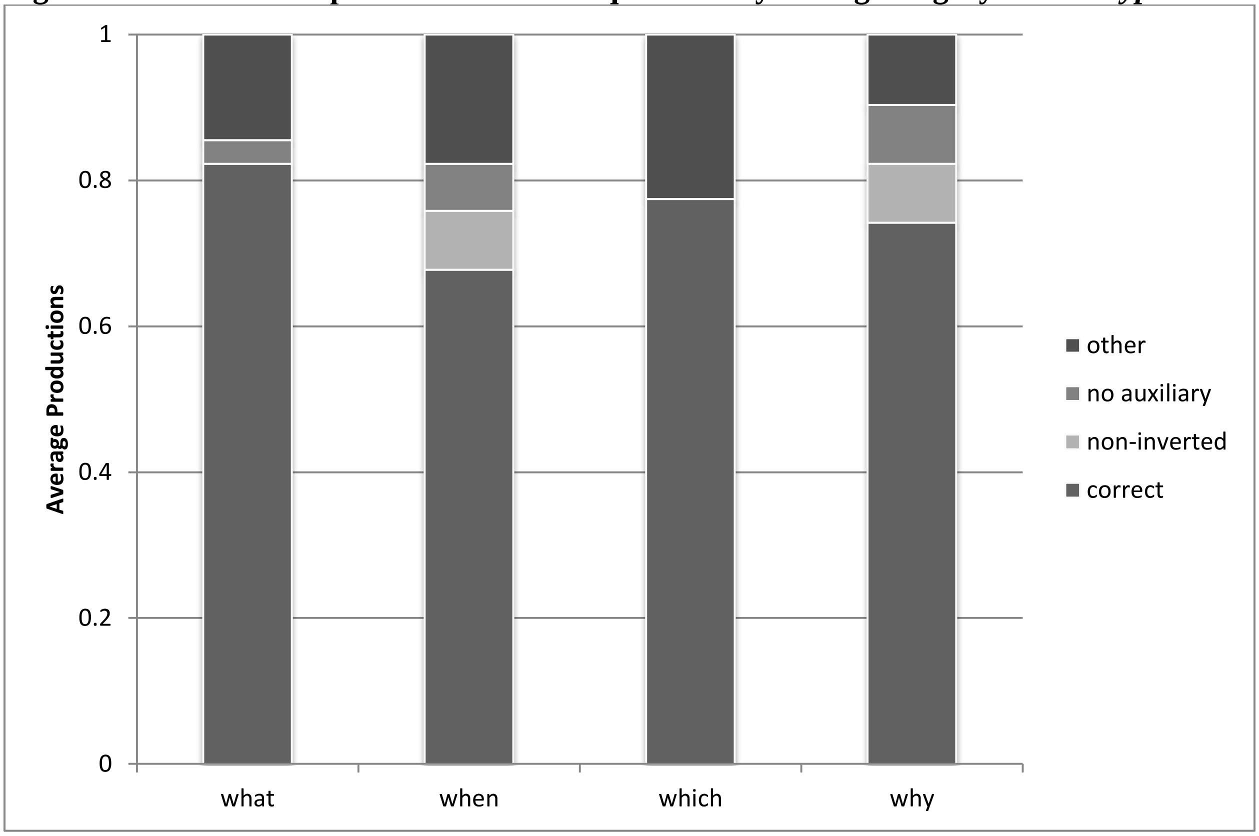


Table 73 and Table 74 report the raw number (and percentage) of productions in each category by question type for children in the younger and in the older group (4;2), respectively.

Table 72: L1 learners' production of main questions by coding category and *wh*-type

Coding	*Wh*-type			
	Argument *wh*-		**Adjunct *wh*-**	
	What	***Which***	***When***	***Why***
Correct	53 (80.3%)	50 (75.8%)	45 (68.2%)	47 (71.2%)
Non-inverted	0	0	7 (10.6%)	6 (9.1%)
Double aux/tense	1 (1.5%)	0	0	0
No auxiliary	2 (3%)	0	3 (4.5%)	6 (9.1%)
Other	10 (15.2%)	16[100] (24.2%)	11 (16.7%)	7 (10.6%)
Total	*66*	*66*	*66*	*66*

[100] 7/16 'other' responses in this category were subject *wh*-questions instead of object *wh*-questions. These subject *wh*-questions were all produced with a specific item in which the *wh*-element was animate (i.e., which cat). As noted before (see footnote 17 and Section 2.4.1.5.2. 2.4.1.5.2. Wh-type), this seems to suggest that learners have a bias towards producing (and interpreting) an initial animate noun phrase as the agent/subject.

Figure 22: L1 learners' production of main questions by coding category and *wh*-type

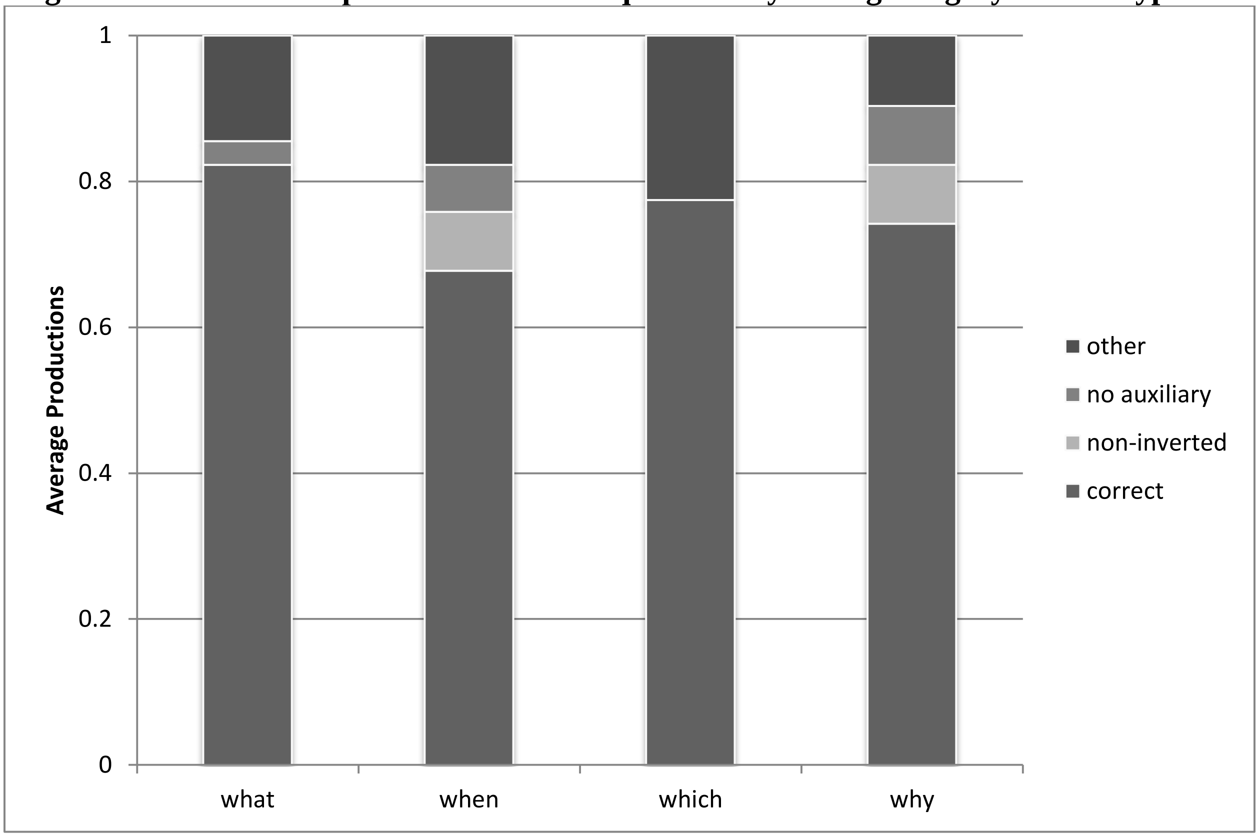


Table 73: L1 learners' production of main questions by coding category and *wh*-type – Children under 4;2

Coding	***Wh*-type**			
	Argument *wh-*		**Adjunct *wh-***	
	What	***Which***	***When***	***Why***
Correct	25 (78.2%)	23 (71.9%)	19 (59.4%)	20 (62.5%)
Non-inverted	0	0	3 (9.4%)	3 (9.4%)
Double aux/tense	0	0	0	0
No auxiliary	1 (3.1%)	0	2 (6.3%)	3 (9.4%)
Other	6 (18.7%)	9 (28.1%)	8 (25%)	6 (18.7%)
Total	*32*	*32*	*32*	*32*

Table 74: L1 learners' production of main questions by coding category and *wh*-type – Children above 4;2

Coding	*Wh*-type			
	Argument *wh*-		**Adjunct *wh*-**	
	What	***Which***	***When***	***Why***
Correct	28 (82.4%)	27 (79.4%)	26 (76.5%)	27 (79.5%)
Non-inverted	0	0	4 (11.8%)	3 (8.8%)
Double aux/tense	1 (2.9%)	0	0	0
No auxiliary	1 (2.9%)	0	1 (2.9%)	3 (8.8%)
Other	4 (11.8%)	7 (20.6%)	3 (8.8%)	1 (2.9%)
Total	*34*	*34*	*34*	*34*

Given that each child only was prompted to produce two questions per *wh*-word and that visual inspection suggested a very similar behavior for *what* and *which*-questions on one hand and *why* and *when*-questions on the other hand (i.e., visual inspection of the data seems to confirm the argument-adjunct distinction proposed in the literature), the analyses in this section investigate whether an argument/adjunct distinction is present in these data.

A 2 (argument type: argument vs. adjunct) x 2 (age group) mixed design ANOVA using arcsine transformed percent *correct* as the dependent variable showed only a trend towards an effect of argument type in the subject analysis (F_1 (1,31) = 3.2, p = .085; F_2 (1,14) = 1.2, p = .3), no effect of age group (F_1 (1,31) = 2.4, n.s.; F_2 < 1) and no interaction between argument type and age group (F_1 (1,31) = 1.4, n.s.; F_2 < 1).

A second 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *inversion* as the dependent variable showed a significant effect of argument type F_1 (1,28[101]) = 10.7, p = .003; F_2 (1,14) = 15, p = .002), no effect of age group (all $Fs \leq 1$) and no interaction (all $Fs \leq 1$). This analysis violated one crucial assumption of analysis of variance, given that inversion was at 100% for arguments. To overcome this problem, a non-parametric Wilcoxon test was performed on percent correct responses. The difference between accuracy rates for arguments and adjuncts was not significant for either the younger or the older children (z = -1.45, p = .15, and z = -.5, p = .6, respectively). A Wilcoxon signed ranks test was also performed on percent inverted responses. For the younger children, inversion rates were higher for arguments than for adjuncts (z = -2, p = .041, r = -.38), while for older children, this difference was only marginally significant (z = -1.89, p = .059, r = -.033).

Overall, an argument-adjunct asymmetry with respect to inversion rates was found in these data. While accuracy rates were comparable across argument and adjunct questions, both younger and older children produced significantly higher inversion rates with arguments than with adjuncts.

The data on inversion errors from the present production study are pretty straightforward and replicates some robust findings in the literature (i.e., yes/no questions are associated with higher accuracy and inversion rates than *wh*-questions). This pattern mirrors the findings from the elicited production studies in Section 2.3. with adult L2 learners, suggesting that inversion in yes/no questions is easier to acquire than in *wh*-questions.

---

[101] Three children produced only non-target argument questions and were thus excluded from this analysis.

No asymmetry between arguments and adjuncts was found with respect to accuracy rates, while the existence of an argument-adjunct asymmetry with respect to inversion rates was confirmed: argument *wh*-questions were associated with higher inversion rates than adjunct *wh*-questions, in line with a number of studies on the L1 acquisition of English questions.

The difference between the accuracy and the Analysis of Inverted Responses has important implications, in that it suggests that presence or absence of an asymmetry might be due to coding criteria and it emphasizes once again the importance of a common coding criterion for researchers working on the acquisition of English questions.

Additionally, the existence of an argument-adjunct asymmetry contrasts with the findings from L2 elicited production where we were not able to replicate the asymmetry seen in the literature. The reason for this asymmetry is likely to be due to the fact that the two experiments used different *wh*-adjuncts (*where* and *why* in the L2 study and *when* and *why* in the L1 study). Finally, it is important to point out that inversion errors, at least in this population, were far from widespread: while non-inversion rates were overall comparable to those reported in the literature (around 10%), errors of inversion were only present in a subset of children (10/33; range per child: 1-3 out of 8 questions).

Here, I would like to sketch what is at this point a very speculative account for the *why*-effect on L2 learners' inversion and the argument-adjunct asymmetry observed with this group of L1 learners. In the literature, it is customary to distinguish between higher sentence adverbials and lower verb-related adverbials (e.g., Jackendoff, 1972; Ernst, 2000; Parsons, 1990 among many others).

On the bases of German and English facts, Frey (2000, 2003) proposes that adverbials/adjuncts pertain to a number of different semantic classes; adverbs in these different classes are located in different positions in the syntax, as indicated in (226):

(226) sentence adverbials > frame and domain adverbials > event-external adverbials (e.g. causals) > highest ranked arguments > event-internal adverbials (e.g. locatives, instrumentals) > (internal arguments) > process-related adverbials (e.g. manner) > verb

The relevant distinction for the current purposes is between high adverbials (e.g., frame adverbials, event external adverbials) and low adverbials (event-internal adverbials). Causal adverbials are event-external, while temporal and locative adjuncts can be construed either as frame-setting adverbials (when they restrict the domain of the assertion of the entire proposition), or as event-internal adverbials, when they modify the event described by the VP.

In Frey's analysis, frame-setting adverbials and event-external adverbials are base-generated high in the clause and can be moved to a clause-initial position without any particular interpretative effect. Event-internal adverbials, on the other hand, are generated lower in the clause and can move to a clause-initial position only under a particular interpretation (e.g., when they bear a focus feature). According to this proposal, temporal and locative adjuncts can be construed either as frame-setting adverbials or as event-internal adverbials, while causal adverbials are part of the class of event-external adverbials and can never be constructed as low event-internal adverbials.

Let us now make the fairly uncontroversial assumption that *wh*-adjuncts are base-generated in the same position as their non-*wh* counterparts. Following Frey (2003), *why* will then be always generated in a high position in the clause, while adjuncts *where* and *when* could be generated either in a high position, when they are construed as frame-setting adverbials, or in in a low position, when they are construed as event-internal adverbials, while argument *wh*-words will always be generated low in the clause. Let us also make the further assumption that this distinction is not something that learners need to learn about the grammar, but is specified by UG.

My proposal for the inversion errors in the acquisition of English main questions runs as follows: learners are aware that English *wh*-elements are located clause-initially; the fact that this is a salient property of the language is confirmed by the fact that English learners (virtually) always place *wh*-words in a clause-initial position (Batmanian, Sayehli, & Valian, 2008; White et al., 1991; Kellerman, 1979; Eckman, Moravcsik, & Wirth, 1989).

Because arguments and event-internal adverbials are base-generated low in the clause, the fact that their *wh*-counterparts are located in a clause-initial position in questions is evidence that they bear a particular interpretative feature (focus, in this case) and that they have moved to the left periphery of the clause. Learners will thus correctly hypothesize that *wh*-arguments and *wh*-event internal adverbials need to raise to Spec,FocP in questions; in turn, the presence of a *wh*-element in Spec,FocP triggers T-to-C movement (e.g., due to the wh-criterion, Rizzi, 1996).

On the other hand, because event-external adverbials (causal adjuncts) and frame-setting adverbials (temporal and locative) are generated high in the clause and can be moved to a clause-initial position without any special interpretation, the fact that their *wh*-counterparts are located clause-initially is not evidence for the fact that they have moved to the CP domain or that they bear focus features. In other words, the fact that these elements appear clause-initially is not taken by learners as unambiguous evidence for their having moved to the CP, because they could be in a clause-initial TP position. If clause-initial *wh*-event-external adverbials and *wh*-frame-setting adverbials are analyzed as being located in TP, no T-to-C movement occurs. Unambiguous evidence for the fact that these adverbials are not located in TP comes from T-to-C movement, but this evidence might be obscured for the learner by the fact that auxiliaries in post-*wh* position are often contracted and thus not salient.

The idea is then that the learner might initially analyze and produce these *wh*-adverbials in a TP clause-initial position, with consequent lack of T-to-C movement and that even when the learner has gathered enough evidence for the fact that these adverbials are located in the left periphery, they might occasionally leave them in TP.

Crucially, this proposal entails that differences in inversion rates between arguments and adjuncts depend on where a given *wh*-element is generated in the clause. Adjuncts *when*, *where* and *how* can be construed, depending on the semantics, as either high or low adverbials in the clause, while *why* is always base-generated high and arguments are always generated low. This might in turn explain some data in the literature. First, the fact that *why* is the only element that seems to be associated with consistent low inversion rates; according to the present proposal the reason is that *why* is the only *wh*-element that is always unambiguously high in the clause, differently from other *wh*-adjuncts. Second, the inconsistency among studies with respect to some *wh*-elements (*where* and *how*, in particular); according to the present proposal, only when these *wh*-adverbials are constructed high in the clause can movement to the left periphery not occur, while if they are base generated low, movement is enforced. Given that different studies used different sentence materials and that most studies in the literature are studies of spontaneous production, it is very likely that the semantics of these *wh*-elements differed greatly, within a child and across children. Future research should examine the semantics of the materials in these studies more closely and control experimental materials for subtle semantic differences.

3.3.2.2. Experiment 2: Embedded questions

3.3.2.2.1. Method

3.3.2.2.1.1. Participants

The same participants from Experiment 1 participated in Experiment 2**.** However, two participants only completed Experiment 1 and were thus excluded from the analysis of Experiment 2. The two experiments were administered during the same session. Experiment 2 was always administered after Experiment 1, given that main questions were used to elicit embedded questions. A total of 31 participants completed Experiment 2. The average age of children in this experiment was 4;3 (SD = 7 months) and the average SPELT score was 32.3 (SD=3).

3.3.2.2.1.2. Materials

Two experimental lists were constructed, each containing six practice items and sixteen experimental items. Type of question (yes/no vs. *wh*-) was a fully within factor, while type of *wh*-word was a within-subject but a between-item factor. The two lists differed with respect to the fully within variable: the items that appeared as prompts for yes/no questions in List 1 appeared as *wh*-questions in List 2, and vice versa. Yes/no questions and *wh*-questions appeared in blocks and did not alternate.

To control for order effects, two additional lists (1B and 2B) were created by switching the order of the first and last 4 sentences in each block. Embedded yes/no questions were thus elicited in the first block in list 1A and list 2B, while *wh*-questions appeared in the first block in lists 1B and 2A. Each block of experimental questions was preceded by three practice items, for a total of six practice items. There were 4 argument *wh*-words (object *what* and *which*) and 4 adjunct *wh*-words (*why*, *when*). The same main questions that participants were prompted to produce in Experiment 1 were now used to elicit embedded questions.[102] Participants were assigned to lists so that in Experiment 2, they would be presented with a different version of the item they were prompted to produce in Experiment 1 (e.g. if a participant had produced item 1 as a *wh*-question in Experiment 1, they would hear item 1 as a yes/no question in Experiment 2).

All stimuli contained a transitive verb and were compatible with all *wh*-words. Each verb occurred in both lists (either in a prompt aimed at eliciting a yes/no question or in a prompt aimed at eliciting a *wh*-question). All items contained the auxiliary *be*. The experimental items were in present tense, while the practice items were in past tense. Half of the question prompts contained a second person subject and half contained a third person subject. All verbs were

[102] However, the practice items in Experiment 2 were not the same as in Experiment 1 because it was noticed that some children showed considerable difficulties transforming the second person pronoun in the prompt into a third person subject. I thus decided that all practice items in Experiment 2 (where pronoun change was crucial) would contain a second person pronoun, so that the experimenter had as many opportunities as possible to ensure that the child paid attention to this feature.

matched for number of syllables and frequency (based on the MacArthur Communicative Development Inventory). The experimental items were pseudo-randomized so that no more than 2 consecutive experimental sentences shared any of the features relevant to the investigation (i.e., yes/no, *wh*-type). See Appendix G for the four experimental lists.

3.3.2.2.1.3. Procedure

Each child was tested individually in a quiet room. Once the child had completed Experiment 1, the experimenter would ask the child to play a different game:

> *This time, we are going to play a remembering game. It's called 'Katie wanted to know'. Can you say 'Katie wanted to know'? Katie is going to ask her brother some questions. Listen to her questions and remember them. You always start with 'Katie wanted to know'. Let's play!*

The experimenter would then move on to the practice items. If the child did not produce a target question during the practice items, the experimenter would say the target embedded question and ask the child to repeat it. However, no explicit corrective feedback was provided. After the practice trials, the child would move on to the eight experimental trials. If the child forgot the prompt, was distracted or produced a non-target structure (e.g., a *what*-question instead of a *why*-question), the experimenter would play the recording again. Each child received a prompt a maximum of three times. If the child still did not produce the target structure, the experimenter would move on to the next item. After the eight experimental trials, the child was started on the second block of questions and was presented with three practice items and eight experimental items. The experiment took approximately 12-14 minutes and was recorded on a digital recorder.

3.3.2.2.1.4. Transcription and Coding

The same coding and checking procedures used in Experiment 1 were used in Experiment 2. Each production was coded as either correct (adult-like) or incorrect (non-adult-like) with respect to word order, verbal morphology, and presence of target lexical items (e.g., subject and *wh*-element). Incorrect questions were further coded into four categories:

- Subject-auxiliary inversion errors, as in (227).
- Double tense/double auxiliary errors, as in (228)
- Omitted auxiliary/omitted morphology errors, as in (229)–(230).
- Other errors. ‘Other errors’ included questions that differed in type from the target (e.g., a yes/no instead of a *wh*-question and vice versa), subject *wh*-questions instead of object *wh*-questions , as in (231), productions that differed from the target in the lexical items used, questions without a subject, questions where the *wh*-word differed from the target one, and skipped sentences. Sentences where a second person pronoun in the prompt failed to be substituted with a third person pronoun in the child’s output, as in (232)–(233), were coded as ‘other,’ independently of the relative order of the subject and the auxiliary.

(227) She wanted to know what was her brother cooking.

(228) Katie wanted to know why is mom is washing the dog.

(229) She wanted to know what her brother cooking.

(230) Katie wanted to know what her mom draw.

(231) She wanted to know which cat was being brushed

(232) Katie wanted to know if you were cooking pizza today.

(233) Katie wanted to know what are you cleaning.

3.3.2.2.2. Results

The main goal of this experiment was to examine the relative contribution of question type and *wh*-type to the production of subject-auxiliary inversion in main English questions by first language learners of English. Moreover, I wanted to investigate whether there was a relationship between inversion rates in main and embedded questions. I hypothesized that, if inversion in embedded questions is to be imputed to overgeneralization of inversion from main questions, there should be a correlation between inversion rates in main and embedded questions, above and beyond input frequency.

Two sets of analyses were performed: the first set of analyses used arcsine transformed mean percent *correct* productions as the dependent variable. This was calculated by dividing the number of correct responses by the *total* number of productions. The second set of analyses used arcsine transformed mean percent *non-inversion* as the dependent variable, and this was calculated by dividing the number of correct responses by inverted and non-inverted responses; productions that provided no evidence one way or another with respect to inversion or that contained errors unrelated to subject-auxiliary inversion (i.e., auxiliary omission, different structure, morphological errors, etc.) were thus excluded from this second set of analyses.

3.3.2.2.2.1. Question-type

Table 75 reports the raw number (and percentage) of productions in each coding category by question type. In order to investigate whether the same pattern of errors emerged in younger and older children, children were further divided in two groups based on their age. The groups were created based on the median age (4;3). There were 16 children in the first group of younger children and 15 children in the older group. The average SPELT score for the children in the first group was 31.7/40 and the average score for the older group was 32.9/40. Table 76 and Table 77 report the raw number (and percentage) of productions in each category by question type for children in the younger and in the older group, respectively.

Table 75: L1 learners' production of embedded questions by coding category and question-type

Coding	Question Type	
	Yes/No	***Wh-***
Correct	202 (81%)	112 (45.2%)
Inverted	2 (.8%)	56 (22.6%)
Double aux/tense	0	1 (.4%)
No auxialiry	7 (2.8%)	7 (2.8%)
Other	38 (15.3%)	72 (29%)
Total	*248*	*248*

Table 76: L1 learners' production of embedded questions by coding category and question-type - Children below 4;3

Coding	Question Type	
	Yes/No	***Wh-***
Correct	99 (77.3%)	57 (44.5%)
Inverted	0	23 (18%)
Double aux/tense	0	1 (.8%)
No auxilairy	1 (.8%)	4 (3.1%)
Other	28 (21.9%)	43 (33.6%)
Total	*128*	*128*

Table 77: L1 learners' production of embedded questions by coding category and question-type - Children above 4;3

Coding	Question Type	
	Yes/No	***Wh-***
Correct	102 (85%)	55 (45.8%)
Inverted	2 (1.7%)	33 (27.5%)
Double aux/tense	0	0
No auxiliary	6 (5%)	3 (2.5%)
Other	10 (8.3%)	29 (24.2%)
Total	*120*	*120*

The experimental task was difficult for children, as can be inferred from the high number of 'other' responses produced. The high number of non-target responses is partially due to children failing to transform a second person pronoun into a third person pronoun, suggesting that children had a tendency to directly quote the question in the prompt. 30% of 'other' responses in the *wh* condition and 31% of the other responses in the yes/no condition were productions with a second person pronoun. A concern regarding whether the high number of inverted responses in this experiment is an experimental artifact (i.e., due to children directly quoting the prompt, given that only half of the experimental items contained a second person pronoun and as such could be excluded if the child did not transform it into a third person subject) needs to be raised at this point. While I think the concern is valid, it seems unlikely to me that the data in the experiment is biased: only two children failed to transform a second person pronoun more than 50% of the time, and embedded inversion in these two children was not higher than the average (25% and 33%, respectively).

Overall, children produced a higher number of 'other' responses in the *wh* condition than in the yes/no condition (t_1 (30) = 2.8, p = .009), suggesting that *wh*-questions present an additional difficulty to children. Older children produced, on average, fewer 'other' responses, but the effect of age group was only significant in the item analysis (F_1 (1,29) = 1.1, n.s.; F_2 (1,15) = 8.7, p= .01). Finally, there was no interaction between age group and condition (all Fs < 1).

A 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *correct* as the dependent variable showed a significant effect of question type (F_1 (1,29) = 42.2, p < .0001; F_2 (1,15) = 153, p < .0001), no effect of age group (all Fs < 1) and no interaction between question type and age group (all Fs < 1).

A second 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *inversion* as the dependent variable showed a significant effect of question type F_1 (1,28) = 39.3, p < .0001; F_2 (1,15) = 35, p < .0001), a marginal effect of age group in the subject analysis (F_1 (1,29) = 3.7, p = .065.; F_2 (1,15) = 2.9, n.s.) and no interaction (F_1 < 1; F_2 (1,15) = 1.7, n.s.).

Overall, children's productions of embedded yes/no questions were significantly more accurate than productions of embedded *wh*-questions. Children produced a significantly higher number of non-target inverted responses when producing embedded *wh*-questions compared to embedded yes/no questions. Younger and older children did not differ significantly in terms of correct and inverted responses. While there was no significant correlation between accuracy rates in main and embedded *wh*-questions across subjects (r (31) = .17, p = .34) or items (r (16) = .33, p = .21), there was a trend towards a *negative* correlation between inversion in main and non-inversion in embedded across subjects (r (30) = -.33, p = .076). That is, the more children were inverting in main *wh*-questions, the more they were inverting in embedded *wh*-questions, indicating perhaps that inversion in embedded *wh*-questions comes from overgeneralization of inversion from main *wh*-questions. On the other hand, the correlation between inversion rates in main questions and non-inversion rates in embedded questions was not significant but positive across items in main and embedded questions (r (16) = .33, p = .21), indicating that items for which children produced higher inversion rates in main questions were items for which children would produce lower inversion rates in embedded questions.

3.3.2.2.2.2. Wh-type

Table 78 reports the raw number (and percentage) of productions in each coding category by *wh*-type. Table 79 and Table 80 report the raw number (and percentage) of productions in each category by *wh*-type for children in the younger and in the older group (4;3), respectively.

Table 78: L1 learners' production of embedded questions by coding category and *wh*-type

Coding	***Wh*-Type**			
	What	***Which***	***When***	***Why***
Correct	29 (46.8%)	26 (41.9%)	29 (46.8%)	28 (45.2%)
Inverted	17 (27.4%)	9 (14.5%)	13 (21%)	17 (27.4%)
Double aux/tense	0	0	0	1 (1.6%)
No auxilliary	1 (1.6%)	1 (1.6%)	3 (4.8%)	2 (3.2%)
Other	15 (24.2%)	26 (41.9%)	17 (27.4%)	14 (22.6%)
Total	*62*	*62*	*62*	*62*

Table 79: L1 learners' production of embedded questions by coding category and *wh*-type - Children below 4;3

Coding	***Wh*-Type**			
	What	***Which***	***When***	***Why***
Correct	15 (46.9%)	11 (34.4%)	13 (40.6%)	18 (56.3%)
Inverted	8 (25%)	5 (15.6%)	7 (21.9%)	3 (9.4%)
Double aux/tense	0	0	0	1 (3.1%)
No auxilairy	1 (3.1%)	1 (3.1%)	1 (3.1%)	1 (3.1%)
Other	8 (25%)	15 (46.9%)	11 (34.4%)	9 (28.1%)
Total	*32*	*32*	*32*	*32*

Table 80: L1 learners' production of embedded questions by coding category and *wh*-type - Children above 4;3

Coding	*Wh*-Type			
	What	***Which***	***When***	***Why***
Correct	14 (46.7%)	15 (50%)	16 (53.3%)	10 (33.3%)
Inverted	9 (30%)	4 (13.3%)	6 (20%)	14 (46.7%)
Double aux/tense	0	0	0	0
No auxiliary	0	0	2 (6.7%)	1 (3.3%)
Other	7 (23.3%)	11 (36.7%)	6 (20%)	5 (16.7%)
Total	*30*	*30*	*30*	*30*

A 2 (argument vs. adjunct) x 2 (age group) mixed design ANOVA using arcsine transformed percent *correct* as the dependent variable showed no effect of argument type (all *F*s < 1), no effect of age group (all *F*s < 1) and no interaction between argument type and age group (F_1 (1,29) = 1.4, n.s.; F_2 < 1).

A 2 (question type) x 2 (age group) mixed design ANOVA using arcsine transformed percent *inversion* as the dependent variable showed no effect of argument type (all *F*s < 1), no effect of age group (all *F*s < 1), but a significant interaction between argument type and age only in the subject analysis (F_1 (1, 25) = 5, p = .03; F_2 < 1). While it is hard to infer anything definitive from a series of null effects, this analysis seems to suggest that children's production of embedded *wh*-questions was not affected by the type of *wh*-element produced and do not differ significantly among younger and older children.

To sum up, the data from the production of embedded questions mirrors Stromswold's (1990) spontaneous production findings and the L2 elicited production findings reported in Section 2.3.1.2 in that inversion is significantly higher in embedded *wh*-questions than in embedded yes/no questions. However, differently from Stromswolds's findings, from the L2 production data and from the data on L1 production in main *wh*-questions in section 3.3.1, the present data do not indicate the existence of an argument-adjunct asymmetry on accuracy and inversion rates, indicating that inversion is a generalized phenomenon in embedded *wh*-questions in child English.

3.3.3. Input analysis

As already mentioned in the introduction to this chapter, one of the aims of this study was that of investigating whether constructivist accounts would be successful at predicting children's inversion rates in main questions. For this reason, the experimental findings from Experiment 1 were examined in light of adult input data.

Adult input to children in six corpora of American English was examined. The total number of input utterances to children in the corpora is 167,757. Table 81 gives an overview of the different corpora examined:

Table 81: Distribution of utterances in the CHILDES corpora

Corpus	# Children	Age Range	Child Utterances	Adult input Utterances	Total Utterances
Bates	27	1;8-2;4	42,86	11,274	15,560
Bloom 70	3	1;4-2;10	31,334	40,385	71,719
Clark	1	2;3-3;2	22,539	32,349	54,888
Gleason	24	2;1-5;2	17,459	37,698	55,157
Snow	1	2;3-3;9	13,152	19,801	32,953
Valian	21	1;9-2;8	14,094	26,250	40,344
Total	*77*	*1;8-5;2*	*102,864*	*167,757*	*270,621*

Experiment 1 in Section 3.3.1 showed that children invert in main yes/no questions at significantly higher rates than in main *wh*-questions and that children's inversion rates are significantly higher for argument *wh*-questions than for adjunct *wh*-questions. In this section, I wanted to investigate whether these findings could be predicted from the frequency with which different questions appeared in the adult input.

The first analysis that I conducted investigated whether children's inversion patterns could be predicted by looking at the *absolute* frequency of adults' inverted main questions. This is the standard analysis used by constructivist accounts. In this analysis, neither the question-type asymmetry nor the argument-adjunct asymmetry was borne out.

The second analysis investigated whether children's inversion patterns could be predicted by looking at the *relative* frequency of adults' inverted questions over non-inverted questions. In this analysis, neither the question-type asymmetry nor the argument-adjunct asymmetry was borne out.

Finally, the third analysis investigated whether children's inversion patterns could be predicted by looking at the *relative* frequency of adults' inverted *strings* over non-inverted *strings*. In other words, I investigated whether children's production errors could be predicted if we hypothesize that children use non-inverted *wh*-structures (embedded *wh*-questions, *wh*-exclamatives, and free relatives) indiscriminately as evidence for the fact that subject-auxiliary inversion is not obligatory in *wh*-structures[103]. In this analysis, while the question-type

[103] Notice, however, that *wh*-exclamatives, embedded *wh*-questions and free relatives differ from main *wh*-questions in terms of prosody (falling intonation vs. rising intonation) and that exclamatives and free relatives also do not exhibit interrogative semantics. If children were thus to ignore prosody and semantics and only pick up word order patterns form the input, evidence of inversion in non *wh*-structures (main yes/no questions) should be computed not only by calculating the frequency of inverted yes/no questions over inverted and non-inverted yes/no

asymmetry was not borne out, the argument-adjunct asymmetry could be derived from the input alone.

3.3.3.1. Absolute Frequency Analysis

The experimental prompts in Experiments 1 and 2 contained either the auxiliary *is* or the auxiliary *are* and *wh*-elements *what*, *which*, *when* and *why*. In order to best compare the experimental data with the input data, all sequences in the input that contained one of the four experimental *wh*-elements and the auxiliary *is* or *are* (and their contracted versions) were extracted.[104]

In this analysis, only main questions with interrogative semantics were included.[105] In Table 82, I present the distribution of inverted questions in the different corpora.

---

questions, but over all sequences that present either the order aux+DP and DP+aux, with the result of including all declarative clauses in the denominator.

[104] While *are* and *is* in the experimental investigation were always instances of the auxiliary and not of the copula, I decided to extract all instances of *is* and *are* because it is unclear to me whether constructivist accounts would distinguish between the two.

[105] Given that constructivist theories look at sequences of word strings, I decided to count subject questions as inverted structures because they do offer evidence for the combination *wh*+auxiliary.

Table 82: Distribution of inverted main questions by type and corpus

Corpus	**Question-Type**				
	What[106]	***Which***	***When***	***Why***	***Yes/No***
Bates	735	5	3	19	258
Bloom 70	1339	20	12	33	1056
Clark	1092	16	2	125	299
Gleason	1008	27	4	27	487
Snow	800	29	5	21	59
Valian	1423	36	6	14	581
Total	*6397*	*133*	*32*	*239*	2,740

Table 82 shows that, overall, main *what*-questions were by far the most common type of question in the input, followed by yes/no questions, *why*-, *which*- and *when*-questions.

Based on these frequencies, one would correctly expect *what*-questions and yes/no questions to be highly inverted in children' productions, and similar low inversion rates for argument *which* and adjunct *why*, contrary to facts.

3.3.3.2. Relative Frequency Analysis[107]

3.3.3.2.1. Interrogative Structures

In this second analysis, inverted and non-inverted structures with interrogative semantics were included. Table 83 presents the distribution of non-inverted questions in the different corpora. In

[106]With respect to *what*-questions, I only included strings in which *what* occurred alone and not as a modifier of another noun (e.g., "what color is…"), for two reasons: first, in the experiment, *what* never occurred as a modifier; second, these strings could potentially be considered ambiguous, depending on how the linear string is parsed.
[107] Relative frequencies of inverted questions and strings are expressed as odds throughout this section.

order to keep things as simple as possible, tag questions, as in (234), were not included in the count because they are made up of a non-inverted and an inverted part. Furthermore, non-inverted non-*wh*-strings ending with *right*, *okay*, or *huh*, as in (235), were also excluded because, based on their prosody, might reasonably be considered as declarative sentences followed by a reduced question, and not as non-inverted yes/no questions.

(234) Jogging is fun, isn't it?

(235) She's your favorite teacher, right? → She is your favorite teacher. (Is that) right?

Table 83: Distribution of non-inverted questions by type and corpus

Corpus	***What***	***Which***	***When***	***Why***	***non-inverted Yes/No***
Bates	25	0	0	2	78
Bloom 70	89	3	1	2	391
Clark	98	3	3	7	367
Gleason	197	0	0	6	285
Snow	23	3	0	4	11
Valian	176	2	4	3	338
Total	608	11	8	24	1,470

Overall, the odds of inverted questions are much higher for *wh*-questions than for yes/no questions (10.5 vs. 1.9, respectively). Moreover, the odds of inverted structures are similar for *what*-, *which*- and *why* questions (10.5, 12.1, and 10 respectively), and particularly low for *when*-questions (ratio = 4). Based on these characteristics of the input, one would incorrectly expect inversion rates to be lower for yes/no questions than for any *wh*-question, and inversion rates to be similarly high for *what*-, *which*- and *why*-questions, contrary to facts.

3.3.3.2.2. All Structures

This analysis grouped all inverted and non-inverted sequences irrespective of their semantics. In Table 84, I present the distribution of inverted strings in the different corpora, while in Table 85, I present the distribution of non-inverted strings in the different corpora. In other words, in this analysis, all strings in which a *wh*-word (i.e., *what*, *which*, *when* and *why*) was followed directly by *is* or *are* were considered as 'inverted *wh*-strings', while strings in which *is* or *are* were directly followed (but not preceded) by a DP were considered as 'inverted non-*wh*-strings'. Conversely strings in which a *wh*-word was directly followed by a DP and then directly followed by *is* or *are* were considered as 'non-inverted *wh*-strings' and strings in which a DP was directly followed by *is* or *are* were considered 'non-inverted non-*wh*-strings.'

Table 84: Distribution of inverted strings by type and corpus

Corpus	**String-Type**				
	What	***Which***[108]	***When***	***Why***	**inverted non-*wh***[109]
Bates	735	5	3	19	262
Bloom 70	1339	20	12	33	1126
Clark	1092	16	2	125	316
Gleason	1008	27	4	27	541
Snow	800	29	5	21	76
Valian	1423	36	6	14	617
Total	6397	160	32	239	2938

[108] Included in the count are relative clauses (e.g., "We're going on vacation, which is good") and subject questions.
[109] Included in the count are structures in which an auxiliary is not preceded by a DP and is followed by *is* or *are*. Presentational sentences (e.g., "Here is the cat") are thus included in this count.

Table 85: Distribution of non-inverted strings by type and corpus

Corpus	**String-Type**					
	What	***Which***	***When***	***Why***	**non-inverted Yes/No questions**	**non-inverted non-*wh***
Bates	42	0	4	4	78	727
Bloom 70	128	3	33	4	391	4295
Clark	121	3	41	14	367	2792
Gleason	245	0	41	10	285	4115
Snow	30	3	2	4	11	1192
Valian	193	3	40	10	338	3151
Total	759	12	161	46	1,470	16,272

Overall, inversion in *wh*-strings was predominant: inverted *wh*-strings were 7 times more frequent than non-inverted *wh*-strings. Inversion in non-*wh*-strings, on the other hand, was not predominant, given the numerical significance of declarative structures: the odds of inverted non-*wh*-strings were .16.

On the other hand, the odds of inverted structures are similar for *what*- and *which*-strings (8.4 and 13.3, respectively), lower for *why*-strings (5.19), and particularly low for *when*-strings (.2). Based on these odds, we would incorrectly expect inversion rates to be lower for yes/no questions than for any *wh*-question. However, the argument-adjunct asymmetry could be predicted based on these frequencies, given that the ratio of inverted over non-inverted strings is higher for arguments and lower for adjuncts.

A last point should be made with respect to input frequency. If children were to compute frequencies from the input for *wh*-questions together without distinguishing between main and embedded contexts, we would expect inversion rates in main and embedded questions to be comparable, contrary to fact. The elicited production experiment has in fact shown that while non-inversion accounts for about 10% of children's productions in main *wh*-questions, children only invert about 20% of the time in embedded *wh*-questions. While it is possible that at an earlier stage of development children indeed treat main and embedded questions alike, a challenge for theories that hypothesize that lack of uniformity in the input causes the learner to have unstable, noisy representations is that of explaining exactly how children get to distinguish between the two types of clauses, producing the opposite type of error to a similar extent (10% vs. 20%).

3.4. L1 Production of English Questions: Discussion

One aim of the present study was to quantify the production of inversion in main and embedded questions by children learning English as their first language via a controlled elicited production experiment, and to compare the qualitative patterns of non-target productions in L1 and L2 learners. The results from the experimental investigation presented in this section show that accuracy and inversion patterns in 3-5 year olds are similar to those seen in intermediate to advanced L2 learners, in that non-target inversion occurs in the two groups at similar rates for the same structures: non-inversion errors occur at around 10% in main *wh*-questions in both groups, and inversion errors occur at around 20% in embedded *wh*-questions in both groups.

Another aim of the present study was to assess the contribution of linguistic factors (e.g., question-type and *wh*-type) to inversion rates. In main questions, L1 learners, similarly to L2 learners, produced significantly higher inversion rates in yes/no questions than in *wh*-questions. The fact that yes/no questions are associated with higher rates of inversion confirms findings from earlier studies on first language acquisition research of English questions. This result is particularly important because other studies had failed to replicate this asymmetry or had found opposite trends. By using the same protocol and a within-subject design, the present study confirms that the question-type asymmetry is a robust phenomenon in child and adult learners of English.

The present finding of a question-type asymmetry cannot be directly predicted from patterns in the adult input, as discussed in Section 3.3.3, for two reasons. First, the absolute frequency of the relevant main *wh*-questions is overall higher than that of inverted main yes/no questions. Second, the frequency of inverted *wh*-strings over non-inverted *wh*-strings is higher than the ratio of inverted non-*wh*-strings over non-inverted non-*wh*-strings. The same pattern emerges when we focus only on structures with interrogative semantics: the ratio of inverted *wh*-questions over non-inverted *wh*-questions (embedded *wh*-questions) is higher than the ratio of main inverted yes/no questions over main non-inverted ones.

With respect to the effect of question-type in embedded questions, L1 learners, similarly to L2 learners, were significantly more accurate in their production of yes/no questions than *wh*-questions, in that they produced inversion errors in *wh*-questions and virtually never in yes/no questions. This difference cannot be simply an effect of the input because embedded *wh*-questions are much more frequent than embedded yes/no questions in child directed speech: in the subset of corpora examined in the present study, there were 651 embedded *wh*-questions in the adult input (*what*, *which*, *when* and *why*-questions followed by a subject and by either *is* or *are*), while there was a total of 139 embedded yes/no questions (with *if* followed by a subject and by either *is* or *are*). However, according to input-based constructivist accounts, the lack of inversion errors in children's production of embedded yes/no questions could be derived by surface properties of the input: while *wh*-elements are often followed immediately by an auxiliary (e.g., in main *wh*-questions) in the input, *if* never is.

Alternatively, it could be argued that inversion errors in embedded *wh*-questions are the result of overgeneralization of inversion from main to embedded contexts. Similarly to what has been suggested for adult L2 learners, it could be hypothesized that L1 learners are sensitive to structural differences between embedded yes/no and *wh*-questions. Further, similarly to what has been found in non-standard varieties of English (e.g., AAVE, Scottish English, Hiberno English, and Appalachian English, among others) in which inversion is licit in embedded *wh*-questions and embedded yes/no questions that lack an overt complementizer (Labov, 1972; Henry, 1995; Filppula, 1999; Green, 2002), it could be that embedded inversion in child English cannot take place in embedded yes/no questions due the presence of an overt complementizer *if*[110] in the position targeted by inverted auxiliaries (C^0).

Under the hypothesis that embedded inversion in *wh*-structures stems from syntactic overgeneralization, we hypothesized to find a correlation between inversion rates in main and embedded *wh*-questions (see Stromswold, 1990). While there was no significant correlation between *accuracy* rates in main and embedded *wh*-questions across subjects (r (31) = .17, p = .34) or items (r (16) = .33, p = .21), there was a trend towards a *negative* correlation between inversion in main and non-inversion in embedded across subjects (r (30) = -.33, p = .076). That

[110] Children never produced embedded yes/no questions with *whether* and there were a total of only 5 *whether*-questions in the input to children in the subset of corpora examined.

is, the more children were inverting in main *wh*-questions, the more they were inverting in embedded *wh*-questions, indicating perhaps that inversion in embedded *wh*-questions comes from overgeneralization of inversion from main *wh*-questions. On the other hand, the correlation between inversion rates in main questions and non-inversion rates in embedded questions was not significant but positive across items in main and embedded questions (r (16) = .33, p= .21), indicating that items for which children produced higher inversion rates in main questions were items for which children would produce lower inversion rates in embedded questions. This trend is probably due to the influence of individual *wh*-items (recall that *wh*-type was a within-subject variable, but a between-items variable), given that, for example, main *which*-questions were associated with high inversion rates in main and low inversion rates in embedded contexts.

Another aim of this study was to investigate the existence of an argument-adjunct asymmetry with respect to inversion rates in main and embedded *wh*-questions. Differently from the findings with L2 speakers, these data showed a significant effect of argument-type in main questions, with argument *wh*-questions being associated with significantly higher rates of inversion than adjunct *wh*-questions.

This result replicates early findings in the literature (Stromswold, 1990; de Villiers, 1991) and is particularly important because two of the *wh*-elements included in this study (*which* and *when*) have not been extensively studied before. The fact that *which* patterns exactly like *what* with respect to inversion argues against an analysis of inversion in terms of presuppositional differences of the associated propositions. As we have seen in Section 3.2.1.2.1.4, Roeper (2011a,b) has argued that inversion errors could be due to children not inverting when the proposition modified by the *wh*-element is presupposed. We have also seen that some recent work (Eilam, 2011) has argued for the existence of a distinction between argument and adjunct *wh*-elements in terms of presuppositions, in that propositions associated with argument *wh*-questions do not seem to be presupposed, differently from that of adjunct *wh*-questions. However, *which* is special in that it is an argument, but its associated proposition is presupposed. The fact that *which*-questions in this study behave exactly like *what*-questions suggests that the argument-adjunct asymmetry reported in the literature is not due to differences in presuppositions between *wh*-elements.

As discussed in Section 3.3.3., the argument-adjunct asymmetry found in Experiment 1 might be partially explained by the adult input if we take into account the relative frequency of inverted *wh*-strings and compare them with the frequency of non-inverted *wh*-strings. Across the corpora examined in this study, in fact, the ratio of inverted *what* and *which*-sequences over non-inverted ones was higher than the ratio of inverted *when* and *why*-sequences over non-inverted ones. This is due to the fact that main subject *wh*-questions, which exemplify the inverted pattern (*wh*+aux) but are mute with respect to subject-auxiliary inversion, were counted as inverted. However, when only structures with interrogative semantics were taken into consideration, the existence of an argument-adjunct asymmetry was not supported by the input analysis, because the relative frequency of *why*-questions is comparable to that of *what*-questions. Finally, when we focus only on the absolute frequency of inverted main questions in child speech (as constructivist researchers normally do), the argument-adjunct asymmetry found in Experiment 1 does not follow from the input, given that *why*-questions are almost twice as frequent as *which*-questions.

Based on an independently motivated distinction between high and low adjuncts (e.g., Frey, 2003), according to which arguments are always base-generated low in the clause, some *wh*-adjuncts are always generated high in the clause (i.e., *why*), while other adjuncts (e.g., *where*, *when*) can be generated high or low in the clause (depending on their semantics), I hypothesized that the argument-adjunct asymmetry found in the present studies and reported in the literature depends on the base-generation position of different *wh*-elements. Simply put, the idea is that while *wh*-arguments and *wh*-adjuncts that are base-generated low in the sentence always move to the CP domain in learner English (triggering consequent T-to-C movement), *wh*-adjuncts that are generated high in the sentence can be moved to a clause-initial position within TP, without triggering T-to-C movement. Further research should look closely at the linguistic materials used in previous studies in the literature and carefully manipulate the semantic properties of the *wh*-adjuncts to see whether this very preliminary hypothesis provides a viable explanation for the empirical facts.

The results from Experiment 2 indicate a lack of an argument-adjunct asymmetry in embedded contexts, contrary to what could have been predicted on the basis of the results from Experiment 1 and differently from what has been found in L2 acquisition. While the effect of *wh*-type was not significant in this sample, non-target inversion rates were the lowest for *which* and the highest for *what*. This result does not follow from the input for two reasons. First, if we were to predict target non-inversion rates from the absolute frequency of embedded *wh*-questions in the input, we would erroneously predict embedded *what* questions to be associated with the lowest inversion rates, because they are by far the most frequent embedded question (608 vs. 11, 8 and 24 for *which*, *when* and *why*, respectively). Second, if we were to predict inversion rates from the relative frequency of non-inverted *wh*-structures over inverted ones, we would erroneously predict *when*-questions to have the lowest inversion rates.

While a comprehensive explanation of the inversion patterns found in Experiment 2 requires the findings to be replicated and extended by other studies, it seems that our conclusion should provisionally be that inversion rates in embedded *wh*-questions are not specific to particular *wh*-elements and do not mirror the patterns in the input.

By focusing on a structure that is associated with one of the rare word-order errors observed in the production of first and second language learners, the present study has replicated and extended some early findings in the literature, namely, the existence of a question-type asymmetry in main and embedded questions, and the existence of an argument-adjunct asymmetry in main *wh*-questions. By using a similar protocol and comparable materials, the findings from the elicited production studies with L2 learners (Sections 2.3.1.1. and 2.3.1.2.) and L1 learners (Sections 3.2.2.1. and 3.2.2.2.) strongly suggest the presence of similar patterns in child and adult learners, supporting the hypothesis that similar mechanisms might be at play in language acquisition across populations. A crucial issue that needs to be addressed by future research concerns the limits of this similarity, particularly the fact L2 learners, differently from L1 learners, often do not end up with a grammar that is indistinguishable from the grammars of adult native speakers.

4. Conclusions

In this dissertation, my goal was to provide the beginning of a systematic investigation of the extent, nature and causes of a word order error that has been reported to occur extensively in the production of first and second language learners of English. In order to investigate the nature of the L2 errors, i.e. whether they stem from lack of knowledge of the L2 system or from difficulties with the implementation of the target L2 procedure, I examined L2 learners' performance in a number of tasks aimed at tapping into learners' implicit and explicit knowledge (production and acceptability judgments). In order to investigate the causes of the error, i.e., whether the error could be attributed to the properties of the native language (in the case of adult L2 learners) or to quantitative and qualitative patterns in the adult input as (in the case of monolingual child learners) I compared L1 Chinese and L1 Spanish L2 learners of English in terms of their oral and written productions and I compared the results of two elicited production studies with monolingual 3-5 year olds with patterns in the input to children of comparable age. The results of these *interrogative investigations* are summarized below.

Second Language Acquisition

In this dissertation, I investigated the acquisition of English interrogatives in second language learners by looking at production (oral and written) and acceptability judgments (timed and untimed). Additionally, I compared oral and written production and timed and untimed acceptability judgments to see whether errors would surface differently in tasks where learners are pressed for time and in tasks where they have time to revise their productions/decisions. The findings and the implications of these different studies can be summarized as follows:

a. Elicited Oral Production: L1 Chinese and L1 Spanish L2 learners of English produce substantial rates of inversion errors in both main and embedded questions. While L1 Spanish speakers were less accurate than L1 Chinese speakers in their production of main questions, the two groups did not differ with respect to their error rates in embedded questions. Both groups of speakers produced significantly more inversion errors in *wh*-questions than in yes/no questions and, in particular, inversion errors for both groups were significantly lower for *why*-questions. The existence of inversion errors in embedded questions in both L1 groups suggests that these errors might be due to excess automatization of an L2 procedure (overgeneralization of inversion from main to embedded questions). The existence of a question-type asymmetry, which has also been documented in the L1 literature, might be due to the reduced saliency of the auxiliary in post-*wh* position (Gleitman, Newport & Gleitman, 1984), or to the greater complexity of constructing two chains (movement of the *wh*-word and of the auxialiry) compared to one (movement of the auxiliary in yes/no questions), as proposed by Klima and Bellugi (1966). In particular, the low inversion rates associated with *why*-questions might be explained by the fact that *why* is unique in that it targets a different position from other *wh*-elements in the syntax of a number of languages of the world (e.g., for Chinese, Korean, see Ko, 2005 inter alia; for Spanish, see Goodall, 1991; Italian, Rizzi, 2001). The fact that *why* has been argued to behave differently from other *wh*-elements in both

Chinese and Spanish does not allow us to decide whether this result is L1- or UG-driven. The existence, on the other hand, of a question-type asymmetry in embedded questions (not driven by any *wh*-element in particular) cannot be due to an L1 effect on word order given that neither Chinese nor Spanish distinguish between embedded yes/no and *wh*-questions in terms of word order. However, in the English native input, *if* is never followed by an auxiliary, but *wh*-elements are. This asymmetry could thus be explained by input-driven and UG-driven accounts alike.

b. Written Production: L2 learners of English with different L1 backgrounds produced inversion errors in both main and embedded questions. While a pattern of errors similar to that found in the elicited production experiment emerged with respect to the effect of *why* on inversion rates in main questions (lower inversion rates for *why* than for other *wh*-elements), the question-type asymmetry was not confirmed. This mght be due to the fact that in spontaneous written production, the discourse-pragmatic condition for non-inversion in yes/no questions was met. Additionally, error rates were particularly low in the ICLE corpus, suggesting that inversion errors might not be due to insufficient knowledge. However, when only the essays of L1 Spanish and L1 Chinese speakers were considered, results similar to those from the oral production task were obtained for main but not for embedded questions. This result might be taken as an indication that difficulties with word order in English main questions are best characterized in terms of differences in representation, rather than differences in implementation of target procedures, while difficulties with embedded questions might stem from overapplication of non-target procedures under time pressure and that difficulties with word order are pervasive in L2 learners' productionof main questions, regardless of output modality and time pressure. Further research is needed to replica the initial findings on written production of embedded questions and to determine the source of the present asymmetry between oral and written production of embedded structures.

c. Untimed Acceptability Judgments: L2 learners of English with different L1 backgrounds resembled native speakers in that they consistently preferred inverted main questions to non-inverted ones and non-inverted embedded questions to inverted ones. However, L2 learners' judgments differ from those of native speakers both quantitatively and qualitatively. The main quantitative difference between L2 learners' and monolinguals' judgments is that L2 learners' judgments tend to be less sharp than those of native speakers. This is probably due to L2 learners' judgments being occasionally non-target-like. L2 learners also differed from native speakers in that they do not exhibit some native distinctions (between non-inverted main yes/no and *wh*-questions), and, at the same time, show sensitivity to some non-native contrasts (non-inverted main *why*-questions are judged as less deviant than other non-inverted main *wh*-questions and inverted embedded *wh*-questions are judged as less deviant than inverted embedded yes/no questions by L2 learners). These non-native distinctions are particularly important because they mirror the asymmetries seen in production. In sum, the untimed acceptability judgment experiment confirmed that L2 learners prefer grammatical over ungrammatical structures but indicated that some distinctions (*why*-effect in main questions; question-type effect in embedded questions) are only operative in the L2 grammar system.

d. Timed Acceptability Judgments: Previous research based on L2 learners' performance on grammaticality judgment experiments has indicated that untimed tasks are more likely to reflect learners' explicit, metalinguistic knowledge of their L2, while timed tasks are more likely to reflect learners' implicit knowledge. However, these results are based on categorical acceptability judgments where the dependent variable was percent accuracy, and, for this reason, they probably just reflect that L2 learners are more accurate when not pressed for time. The present magnitude estimation experiment was aimed at investigating whether different patterns of acceptability would emerge in a timed acceptability experiment, compared to its untimed version. The results of the experiment do not indicate L2 learners' acceptability patterns to be less target-like or to be qualitatively more similar to production patterns: acceptability patterns in the timed experiment closely resembled the patterns seen in the untimed experiment, but did not show some of the critical asymmetries seen in the untimed version (*why*-effect in main questions; question-type effect in embedded questions). This difference might be simply due to the fact that statistical power in the timed experiment was reduced (23 L2 learners participated in the timed experiment vs. 51 in the untimed one) or to the fact that, faced with an unfamiliar task (magnitude estimation) under time pressure, L2 learners are less likely to express subtle distinctions that are part of their grammatical knowledge.

The L2 studies presented in this dissertation present a complex picture of the L2 acquisition of word order in English interrogative structures. The fact that inversion errors were more common in oral production than in the written production of L2 leaners (when speakers of all L1 backgrounds were included in the analyses) might be taken as an indication that this error does not stem from lack of knowledge but from implementation of L2 procedures when learners produce speech. In contrast, if we limit our comparison to the oral and written production of L1 Spanish and L1 Chinese speakers, a more homogeneous picture with respect to main questions emerges: inversion errors are of comparable magnitude across written and spoken production, and *wh*-questions are associated with higher rates of inversion errors than yes/no questions. The findings from the production of embedded questions are, on the other hand, less straightforward: while *wh*-questions are associated with higher error rates in both oral and written production, errors are of comparable magnitude across output modality only for L1 Chinese speakers, and are instead considerably lower for L1 Spanish writers.

An important issue that requires further study regards the numerical extent of inversion errors: in *wh*-structures, these errors occur at a rate between 10% and 30% in main *wh*-questions and embedded *wh*-questions, respectively. A challenge for any analysis based on parameter-settings and/or categorical grammatical representations is that non-target productions seem too consistent to be considered as noise, but might not be frequent enough to be considered genuinely grammatical options. A way to explain this is to assume a parametric model like the one proposed by Yang (2002) and Yang and Legate (2007) to deal with gradual changes in child language acquisition. In this model, different hypotheses compete to best parse the native adult input by first being accessed probabilistically and then punished or rewarded in terms of probability dependent upon their success. Non-target productions might thus be caused by parametric choices that have been considered by the learner and that will ultimately disappear because of their lack of success. Crucially, in this model, parametric options are provided by UG. Assuming this model for L2 acquisition would entail that adult learners have access to UG and predict similar development profiles for L2 and L1 learners.

Alternatively, the Acquisition by Processing Theory model (Truscott & Sharwood Smith, 2004), in which there are no dedicated language acquisition mechanisms and in which language development is seen as a consequence of processing procedures, predicts the intermittent appearance of non-target productions in L2 speakers. According to this model, the production system is shared between the two (or more) languages of a speaker and L1-transfer is a result of competition between L1 and L2 procedures. The appearance of non-target productions is thus an effect of the L1 procedure having won the competition, possibly due to its ease for the production system. The theory thus explains why non-target productions might still appear in the speech of advanced speakers. While the model was initially proposed as an alternative to the traditional view that L1-transfer is the result of erroneous parameter setting in L2, given its reliance on UG, it might be possible to extend it to explain intermittent non-target productions that cannot be imputed to the speaker's L1, but to UG-driven defaults.

First Language Acquisition

In this dissertation, I investigated the acquisition of English interrogatives in first language learners by conducting an elicited production study. The elicited production experiment conducted with first language learners of English had three main objectives: (1) to investigate whether child learners exhibit a question-type asymmetry and an argument-adjunct asymmetry in their production of main and embedded questions, (2) to compare the production patterns of first and second language learners in order to determine whether these two groups of learners are sensitive to the same factors, and (3) to investigate whether non-target productions could be predicted by reference to the linguistic input to children.

In order to investigate whether the results from the elicited production study could be derived from patterns in the input, I analyzed the frequency of inverted and non-inverted interrogatives in six CHILDES corpora. The results can be summarized as follows:

a. Main Questions: Child learners of English produced inversion errors in main questions. Similarly to L2 learners, children showed a question-type asymmetry, producing significantly more inversion errors in *wh*-questions than in yes/no questions. Moreover, an argument-adjunct asymmetry with respect to inversion rates was found, with adjunct questions being associated with lower inversion rates than argument questions. Inversion rates were similar for *which* and *what* on one hand, and for *when* and *why* on the other. Both results are incompatible with constructivist predictions, while they do confirm early findings in the generative literature. The existence of a question-type asymmetry, which has been documented in the L1 literature, might be due to the reduced saliency of the auxiliary in post-*wh* position (Gleitman, Newport & Gleitman, 1984), or to the greater complexity of constructing two chains (due to the movement of the *wh*-word and the auxiliary) compared to one (movement of the auxiliary in yes/no questions), as proposed by Klima and Bellugi (1966).

These results parallel to a great extent the results found in the elicited production study with L2 learners. In the L2 study, we did not find an argument-adjunct asymmetry, but a *why*-effect. The existence of an argument-adjunct asymmetry in L1 learners might be due to the specific *wh*-words used (*what, who, <u>where</u>* and *why* in the L2 experiment and *what, which, <u>when</u>* and *why* in the L1 experiment). While nothing conclusive about this can be said at this point, this difference could be explained by hypothesizing that some adjunct *wh*-elements (*why*, *when*) might be base-generated high in the clause and move to a TP internal position (with consequent lack of T-to-C movement) differently from arguments and low *wh*-adjuncts, which, based on their semantic properties, need to be base-generated VP-internally (see Section 3.3.2.1.2.2. for a discussion).

b. <u>Embedded questions</u>. The present study showed that child learners of English produce consistent inversion errors in embedded *wh*-questions, differently from what was found by Sarma (1991). Similarly to L2 learners of English, children showed a question-type asymmetry in their production of embedded questions, in that they produced significantly more inversion errors in *wh*-questions than in yes/no questions. This asymmetry could be explained by structural and constructivist accounts alike, given that the input presents no evidence for inversion in embedded yes/no contexts (i.e., the complementizer *if* is never followed by an auxiliary in the adult input), while the input presents evidence for inversion in *wh*-contexts (i.e. main *wh*-questions). The existence of inversion errors in embedded *wh*-questions in L1 and L2 learners' production suggests that these errors might be due to excess automatization of an English procedure (overgeneralization of inversion from main to embedded questions). The existence of a trend towards a negative correlation between children's inversion rates in main and non-inversion rates in embedded questions gives some initial support to this hypothesis.

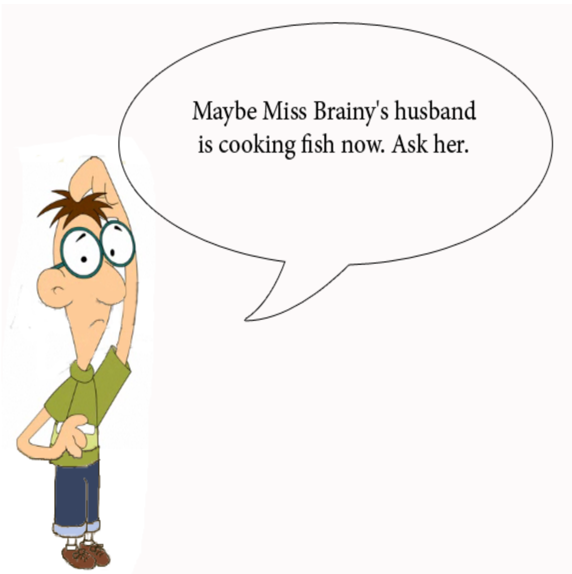
Taken together, the results from the elicited production experiments with L2 and L1 learners present a homogeneous picture of word order errors in English interrogative structures. The similarity between the performance of L1 and L2 learners suggests that first and second language learners use the same mechanisms and procedures while acquiring a first or a second language, and that non-target patterns seen in the production of second language learners are not due to their L1 and/or to adult learners using non-domain-specific cognitive mechanisms (Bley-Vroman, 1989; 2009) or a different cognitive system (declarative vs. procedural, Ullmann, 2004, 2005) for language acquisition. Furthermore, the comparison between the experimental investigation with L1 learners and the corpus study of interrogative structures in child directed speech strongly suggests that the patterns seen in the production of child learners do not derive in any direct way from the input.

5. Appendices

Appendix A.

Protocol for main yes/no questions:

Slide 1.



Slide 2.

Slide 3.

Protocol for main *wh*-questions:

Slide 1.

Slide 2.

Slide 3.

Appendix B.

Practice Items

Someone is yelling upstairs. Ask Miss Brainy who
Someone usually plays the piano next door. Ask Miss Brainy who.
Maybe Miss Brainy's son is painting a picture. Ask her.
Maybe Miss Brainy's husband usually drives her home. Ask her.
Somebody ran ten miles. Ask Miss Brainy who.
Maybe Miss Brainy's husband has left for work. Ask her.

List 1.

Maybe Gloria called Jim for advice. Ask Miss Brainy.
Maybe Miss Brainy's friends have invited her to dinner. Ask her.
The boss has complimented somebody. Ask Miss Brainy who.
The student read something. Ask Miss Brainy what.
Maybe Mary is carrying makeup in her bag. Ask Miss Brainy.
John is smoking a lot. Ask Miss Brainy why.
Laura usually visits someone in London. Ask Miss Brainy who.
Maybe Miss Brainy's brother washes his clothes downstairs. Ask her.
Miss Brainy's brother fired Mark. Ask Miss Brainy why.
Bill has forgotten his keys somewhere. Ask Miss Brainy where.
Maybe the nurse has fed the patient today. Ask Miss Brainy.
Maybe the waiter suggested pasta to Mary. Ask Miss Branny
Juan teaches Spanish somewhere. Ask Miss Brainy where.
Maybe Matt usually sees Sarah for brunch. Ask Miss Brainy.
Maybe Miss Brainy's husband is walking to work. Ask her.
Miss Brainy's husband is cooking something now. Ask her what.
Maybe Mark cleaned his office. Ask Miss Brainy.
Miss Brainy's husband has hired someone. Ask her who.
The teacher bought something. Ask Miss Brainy what.
Maybe the president has traveled to Taiwan. Ask Miss Brainy.
The math teacher normally helps someone. Ask Miss Brainy who.
The cat is sleeping outside. Ask Miss Brainy why.
Maybe the cat normally hides upstairs. Ask Miss Brainy.
Maybe Lee is making eggs for breakfast. Ask Miss Brainy.
Maybe the students learned Italian this year. Ask Miss Brainy.
John opened the window. Ask Miss Brainy why.
Maybe Miss Brainy's brother has met Luis in France. Ask her
The band has played somewhere. Ask Miss Brainy where.
Julia goes somewhere on vacation. Ask Miss Brainy where.

Maybe Lee is complaining about her job. Ask Miss Brainy.
Miss Brainy's brother is drinking something with dinner. Ask her what.
Maybe the baby-sitter drives Lucy to school. Ask Miss Brainy.

List 2.

Gloria has called Jim. Ask Miss Brainy why.
Maybe Miss Brainy’s friends invited her to dinner. Ask her.
The boss complimented somebody. Ask Miss Brainy who.
Maybe the student has read the textbook. Ask Miss Brainy.
Maybe Mary carries makeup in her bag. Ask Miss Brainy.
John smokes a lot. Ask Miss Brainy why.
Maybe Laura is visiting Bill in London. Ask Miss Brainy.
Miss Brainy’s brother is washing his clothes somewhere. Ask her where
Maybe Miss Brainy’s brother has fired Mark for stealing. Ask her.
Bill forgot his keys somewhere. Ask Miss Brainy where.
Maybe the nurse fed the patient today. Ask Miss Brainy.
The waiter has suggested something to Mary. Ask Miss Brainy what.
Maybe Juan is teaching Spanish in Manhattan. Ask Miss Brainy.
Mark is seeing somebody for brunch today. Ask Miss Brainy who.
Maybe Miss Brainy’s husband walks to work. Ask her.
Miss Brainy’s husband usually cooks something. Ask her what.
Mark has cleaned his office. Ask Miss Brainy why.
Miss Brainy’s husband hired someone. Ask her who.
Maybe the teacher has bought new clothes. Ask Miss Brainy.
Maybe the president traveled to Taiwan. Ask Miss Brainy.
Maybe the math teacher is helping Bill now. Ask Miss Brainy.
The cat sleeps outside. Ask Miss Brainy why
The cat is hiding somewhere now. Ask Miss Brainy where.
Maybe Lee usually makes eggs for breakfast. Ask Miss Brainy.
The students have learned something this year. Ask Miss Brainy what.
Maybe John has opened the window. Ask Miss Brainy
Maybe Miss Brainy’s brother met Luis in France. Ask her
The band played somewhere. Ask Miss Brainy where.
Maybe Julia is going to Europe this summer. Ask Miss Brainy.
Maybe Lee complains about her job. Ask Miss Brainy.
Miss Brainy’s brother usually drinks something with dinner. Ask her what.
The baby-sitter is driving somebody to school. Ask Miss Brainy who.

List 3.

Gloria called Jim. Ask Miss Brainy why.
Miss Brainy’s friends have invited her somewhere. Ask her where.

Maybe the boss has complimented Laura. Ask Miss Brainy.
The student read the textbook, I think. Ask Miss Brainy.
Mary is carrying something in her bag. Ask Miss Brainy what.
Maybe John is smoking a lot. Ask Miss Brainy.
Maybe Laura usually visits Bill in London. Ask Miss Brainy.
Miss Brainy's brother washes his clothes somewhere. Ask her where
Maybe Miss Brainy's brother fired Mark for stealing. Ask her.
Maybe Bill has forgotten his keys at school. Ask Miss Brainy.
The nurse has fed somebody today. Ask Miss Brainy who.
The waiter suggested something to Mary. Ask Miss Brainy what.
Maybe Juan teaches Spanish in Manhattan. Ask Miss Brainy.
Matt usually sees somebody for brunch. Ask Miss Brainy who.
Miss Brainy's husband is walking to work. Ask her why.
Maybe Miss Brainy's husband is cooking fish now. Ask her.
Mark cleaned his office. Ask Miss Brainy why.
Maybe Miss Brainy's husband has hired Tom. Ask her.
Maybe the teacher bought new clothes. Ask Miss Brainy.
The president has traveled somewhere. Ask Miss Brainy where
Maybe the math teacher normally helps Bill. Ask Miss Brainy.
Maybe the cat is sleeping outside. Ask Miss Brainy.
The cat normally hides somewhere. Ask Miss Brainy where.
Lee is making something for breakfast today. Ask Miss Brainy what.
The students learned something this year. Ask Miss Brainy what.
Maybe John opened the window. Ask Miss Brainy
Miss Brainy's brother has met someone in France. Ask her who.
Maybe the band has played in Manhattan. Ask Miss Brainy.
Maybe Julia goes to Europe in summer. Ask Miss Brainy.
Lee is complaining about her job. Ask Miss Brainy why.
Maybe Miss Brainy's brother is drinking wine with dinner. Ask her.
The baby-sitter drives somebody to school. Ask Miss Brainy who.

List 4.

Maybe Gloria has called Jim for advice. Ask Miss Brainy.
Miss Brainy's friends invited her somewhere. Ask her where.
Maybe the boss complimented Laura. Ask Miss Brainy.
The student has read something. Ask Miss Brainy what
Mary carries something in her bag. Ask Miss Brainy what.
Maybe John smokes a lot. Ask Miss Brainy
Laura is visiting someone in London now. Ask Miss Brainy who.
Maybe Miss Brainy's brother is washing his clothes downstairs. Ask her.
Miss Brainy's brother has fired Mark. Ask Miss Brainy why.
Maybe Bill forgot his keys at school. Ask Miss Brainy.

The nurse fed somebody today. Ask Miss Brainy who
Maybe the waiter has suggested the pasta to Mary. Ask Miss Brainy
Juan is teaching Spanish somewhere. Ask Miss Brainy where.
Maybe Matt is seeing Sarah for brunch today. Ask Miss Brainy.
Miss Brainy's husband walks to work. Ask her why.
Maybe Miss Brainy's husband usually cooks fish. Ask her.
Maybe Mark has cleaned his office. Ask Miss Brainy.
Maybe Miss Brainy's husband hired Tom. Ask her.
The teacher has bought something. Ask Miss Brainy what.
The president traveled somewhere. Ask Miss Brainy where
The math teacher is helping someone now. Ask Miss Brainy who.
Maybe the cat normally sleeps outside. Ask Miss Brainy.
Maybe the cat is hiding upstairs. Ask Miss Brainy
Lee usually makes something for breakfast. Ask Miss Brainy what.
Maybe the students have learned Italian this year. Ask Miss Brainy.
John has opened the window. Ask Miss Brainy why.
Miss Brainy's brother met someone in France. Ask her who
Maybe the band played in Manhattan. Ask Miss Brainy.
Julia is going somewhere on vacation. Ask Miss Brainy where.
Lee complains about her job. Ask Miss Brainy why.
Maybe Miss Brainy's brother usually drinks wine with dinner. Ask her.
Maybe the baby-sitter is driving Lucy to school. Ask Miss Brainy.

Appendix C.

List of all L2 learner participants, their mean percent inversion in *wh*- and yes/no questions and + and − signs for 80% and 90% accuracy threshold for mastery.

Participant	L1	% Inversion in *wh*-questions	% Inversion in yes/no questions	Inversion in *wh*-questions acquired (80%)	Inversion in yes/no questions acquired? (80%)	Inversion in *wh*-questions acquired (90%)	Inversion in yes/no questions acquired (90%)
1	Chinese	93.3	100	+	+	+	+
2	Chinese	100	100	+	+	+	+
3	Chinese	93.8	100	+	+	+	+
4	Chinese	87.5	100	+	+	−	+
5	Chinese	100	100	+	+	+	+
6	Chinese	100	100	+	+	+	+
7	Chinese	85.7	100	+	+	−	+
8	Chinese	93.8	100	+	+	+	+
9	Chinese	100	100	+	+	+	+
10	Chinese	80.0	100	−	+	−	+
11	Chinese	92.9	100	+	+	+	+
12	Chinese	91.7	100	+	+	+	+
13	Chinese	100	100	+	+	+	+
14	Chinese	100	100	+	+	+	+
15	Chinese	76.9	100	−	+	−	+
16	Chinese	71.4	100	−	+	−	+
17	Chinese	100	100	+	+	+	+
18	Chinese	91.7	100	+	+	+	+
19	Chinese	100	100	+	+	+	+
20	Chinese	91.7	100	+	+	+	+
21	Chinese	100	100	+	+	+	+
22	Chinese	100	100	+	+	+	+
23	Chinese	100	100	+	+	+	+
24	Chinese	71.4	100	−	+	−	+
25	Chinese	100	100	+	+	+	+
26	Chinese	100	100	+	+	+	+
27	Chinese	100	100	+	+	+	+
28	Chinese	54.5	100	−	+	−	+
29	Chinese	85.7	100	+	+	−	+

30	Chinese	100	100	+	+	+	+
31	Chinese	100	100	+	+	+	+
32	Chinese	100	100	+	+	+	+
33	Spanish	84.6	100	+	+	−	+
34	Spanish	100	100	+	+	+	+
35	Spanish	100	100	+	+	+	+
36	Spanish	57.1	100	−	+	−	+
37	Spanish	91.7	85.7	+	+	+	−
38	Spanish	100	100	+	+	+	+
39	Spanish	100	100	+	+	+	+
40	Spanish	46.2	92.3	−	+	−	+
41	Spanish	100	100	+	+	+	+
42	Spanish	100	100	+	+	+	+
43	Spanish	58.3	93.3	−	+	−	+
44	Spanish	100	100	+	+	+	+
45	Spanish	100	100	+	+	+	+
46	Spanish	100	100	+	+	+	+
47	Spanish	50.0	92.3	−	+	−	+
48	Spanish	20.0	100	−	+	−	+
49	Spanish	100	100	+	+	+	+
50	Spanish	61.5	100	−	+	−	+
51	Spanish	84.6	100	+	+	−	+
52	Spanish	100	100	+	+	+	+
53	Spanish	100	100	+	+	+	+
54	Spanish	100	100	+	+	+	+
55	Spanish	54.5	76.9	−	−	−	−
56	Spanish	57.1	92.9	−	+	−	+
57	Spanish	100	100	+	+	+	+
58	Spanish	66.7	100	−	+	−	+
59	Spanish	100	100	+	+	+	+
60	Spanish	100	100	+	+	+	+
61	Spanish	40.0	64.3	−	−	−	−
62	Spanish	100	100	+	+	+	+
63	Spanish	100	100	+	+	+	+
64	Spanish	50.0	80.0	−	+	−	−

Appendix D.

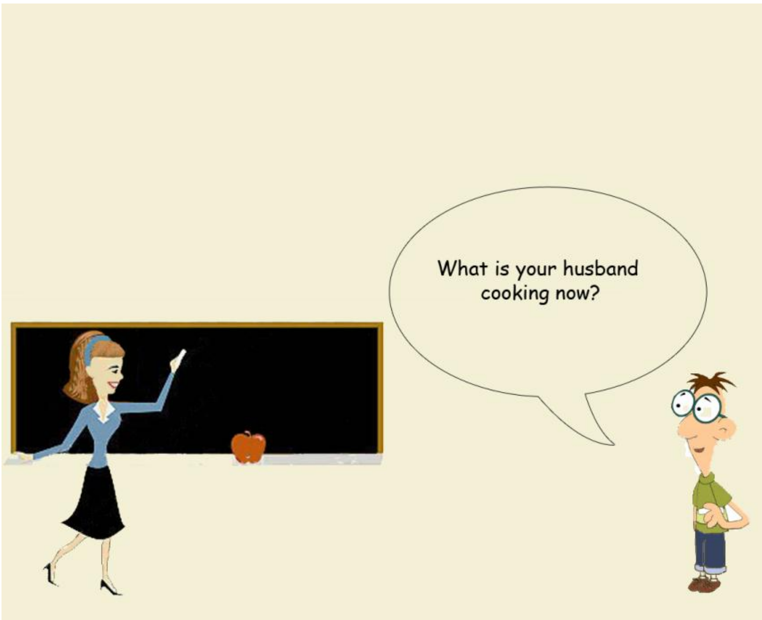
Protocol for embedded yes/no questions:

Slide 1.

Slide 2.

Protocol for embedded *wh*-questions:

Slide 1.



Slide 2.

Appendix E.

Embedded questions materials.

List 1.

Practice items:

Who is yelling upstairs?
Who usually plays the piano next door?
Is your son painting a picture?
Does your husband usually drive you home?
Who ran ten miles ?
Has your husband left for work?

List 1.

Did Gloria call Jim for advice?
Have your friends invited you to dinner?
Who has the boss complimented?
What did the student read?
Is Mary carrying makeup in her bag?
Why is John smoking a lot?
Who does Laura usually visit in London?
Does your brother wash his clothes downstairs?
Why did your brother fire Mark?
Where has Bill forgotten his keys?
Has the nurse fed the patient today?
Did the waiter suggest pasta to Mary?
Where does Juan teach Spanish?
Does Matt usually see Sarah for brunch?
Is your husband walking to work?
What is your husband cooking now?
Did Mark clean his office?
Who has your husband hired?
What did the teacher buy?
Has the president traveled to Taiwan?
Who does the math teacher normally help?
Why is the cat sleeping outside?
Does the cat normally hide upstairs?
Is Lee making eggs for breakfast today?

Did the students learn Italian this year?
Why did John open the window?
Has your sister met Luis in France?
Where has the band played?
Where does Julia go on vacation?
Is Lee complaining about her job?
What is your brother drinking with dinner?
Does the baby-sitter drive Lucy to school?

List 2.

Why has Gloria called Jim?
Did your friends invite you to dinner?
Who did the boss compliment?
Has the student read the textbook?
Does Mary carry makeup in her bag?
Why does John smoke a lot?
Is Laura visiting Bill now in London?
Where is your brother washing his clothes?
Has your brother fired Mark for stealing?
Where did Bill forget his keys?
Did the nurse feed the patient today?
What has the waiter suggested to Mary?
Is Juan teaching Spanish in Manhattan?
Who is Matt see for brunch today?
Does your husband walk to work?
What does your husband usually cook?
Why has Mark cleaned his office?
Who did your husband hire?
Has the teacher bought new clothes?
Did the president travel to Taiwan?
Is the math teacher helping Bill now?
Why does the cat sleep outside?
Where is the cat hiding now?
Does Lee usually make eggs for breakfast?
What have the students learned this year?
Has John opened the window?
Did your brother meet Luis in France?
Where did the band play?
Is Julia going to Europe this summer?
Does Lee complain about her job?
What does your brother drink with dinner?
Who is the baby-sitter driving to school?

List 3.

Why did Gloria call Jim?
Where have your friends invited you?
Has the boss complimented Laura?
Did the student read the textbook?
What is Mary carrying in her bag?
Is John smoking a lot?
Does Laura usually visit Bill in London?
Where does your brother wash his clothes?
Did your brother fire Mark for stealing?
Has Bill forgotten his keys at school?
Who has the nurse fed today?
What did the waiter suggest to Mary?
Does Juan teach Spanish in Manhattan?
Who does Matt usually see for brunch?
Why is your husband walking to work?
Is your husband cooking fish now?
Why did Mark clean his office?
Has your husband hired Anna?
Did the teacher buy new clothes?
Where has the president traveled?
Does the math teacher normally help Bill?
Is the cat sleeping outside?
Where does the cat normally hide?
What is Lee making for breakfast today?
What did the student learn this year?
Did John open the window?
Who has your brother met in France?
Has the band played in Manhattan?
Does Julia go to Europe on vacation?
Why is Lee complaining about her job?
Is your brother drinking wine with dinner?
Who does the baby-sitter drive to school?

List 4.

Has Gloria called Jim for advice?
Where did your friends invite you?
Did the boss compliment Laura?
What has the student read?
What does Mary carry in her bag?

Does John smoke a lot?
Who is Laura visiting now in London?
Is your brother washing his clothes downstairs?
Why has your brother fired Mark?
Did Bill forget his keys at school?
Who did the nurse feed today?
Has the waiter suggested pasta to Mary?
Where is Juan teaching Spanish?
Is Matt seeing Sarah for brunch today?
Why does your husband walk to work?
Does your husband usually cook fish?
Has Mark cleaned his office?
Did your husband hire Anna?
What has the teacher bought?
Where did the president travel?
Who is the math teacher helping now?
Does the cat sleep outside?
Is the cat hiding upstairs?
What does Lee usually make for breakfast?
Have the students learned Italian this year?
Why has John opened the window?
Who did your brother meet in France?
Did the band play in Manhattan?
Where is Julia going this summer?
Why does Lee complain about her job?
Does your brother drink wine with dinner?
Is the baby-sitter driving Lucy to school.

Appendix F.

List of all L2 learner participants, their overall inversion rates in main and embedded *wh*-questions and + and − signs for 80% and 90% accuracy criterion for mastery.

Participant	L1	MQ inverted	EQ non-inverted	MQ acquired (80%)	EQ acquired (80%)	MQ acquired (90%)	EQ acquired (90%)
1	Chinese	93.3	100.0	+	+	+	+
2	Chinese	100.0	100.0	+	+	+	+
3	Chinese	93.8	100.0	+	+	+	+
4	Chinese	87.5	88.9	+	+	−	−
5	Chinese	100.0	90.0	+	+	+	−
6	Chinese	100.0	85.7	+	+	+	−
7	Chinese	85.7	35.7	+	−	−	−
8	Chinese	93.8	100.0	+	+	+	+
9	Chinese	100.0	93.8	+	+	+	+
10	Chinese	80.0	80.0	−	−	−	−
11	Chinese	92.9	27.3	+	−	+	−
12	Chinese	91.7	92.9	+	+	+	+
13	Chinese	100.0	90.9	+	+	+	+
14	Chinese	100.0	53.8	+	−	+	−
15	Chinese	76.9	0.0	−	−	−	−
16	Chinese	71.4	33.3	−	−	−	−
17	Chinese	100.0	16.7	+	−	+	−
18	Chinese	91.7	70.0	+	−	+	−
19	Chinese	100.0	33.3	+	−	+	−
20	Chinese	91.7	28.6	+	−	+	−
21	Chinese	100.0	21.4	+	−	+	−
22	Chinese	100.0	14.3	+	−	+	−
23	Chinese	100.0	70.0	+	−	+	−
24	Chinese	71.4	91.7	−	+	−	+
25	Chinese	100.0	87.5	+	+	+	−
26	Chinese	100.0	100.0	+	+	+	+
27	Chinese	100.0	9.1	+	−	+	−
28	Chinese	54.5	80.0	−	−	−	−
29	Chinese	85.7	92.3	+	+	−	+
30	Chinese	100.0	93.3	+	+	+	+
31	Chinese	100.0	100.0	+	+	+	+

32	Chinese	100.0	100.0	+	+	+	+
33	Spanish	84.6	60.0	+	−	−	−
34	Spanish	100.0	75.0	+	−	+	−
35	Spanish	100.0	13.3	+	−	+	−
36	Spanish	57.1	87.5	−	+	−	−
37	Spanish	91.7	41.7	+	−	+	−
38	Spanish	100.0	75.0	+	−	+	−
39	Spanish	100.0	50.0	+	−	+	−
40	Spanish	46.2	80.0	−	−	−	−
41	Spanish	100.0	53.8	+	−	+	−
42	Spanish	100.0	66.7	+	−	+	−
43	Spanish	58.3	75.0	−	−	−	−
44	Spanish	100.0	87.5	+	+	+	−
45	Spanish	100.0	100.0	+	+	+	+
46	Spanish	100.0	53.3	+	−	+	−
47	Spanish	50.0	81.8	−	+	−	−
48	Spanish	20.0	90.0	−	+	−	−
49	Spanish	100.0	100.0	+	+	+	+
50	Spanish	61.5	76.9	−	−	−	−
51	Spanish	84.6	45.5	+	−	−	−
52	Spanish	100.0	55.6	+	−	+	−
53	Spanish	100.0	6.3	+	−	+	−
54	Spanish	100.0	33.3	+	−	+	−
55	Spanish	54.5	44.4	−	−	−	−
56	Spanish	57.1	86.7	−	+	−	−
57	Spanish	100.0	93.8	+	+	+	+
58	Spanish	66.7	33.3	−	−	−	−
59	Spanish	100.0	60.0	+	−	+	−
60	Spanish	100.0	75.0	+	−	+	−
61	Spanish	40.0	72.7	−	−	−	−
62	Spanish	100.0	15.4	+	−	+	−
63	Spanish	100.0	80.0	+	−	+	−
64	Spanish	50.0	55.6	−	−	−	−

List of all L2 learner participants, their overall inversion and non-inversion rates in main and embedded yes/no questions and + and − signs for 80% and 90% accuracy criterion for mastery.

Participant	L1	MQ inverted	EQ non-inverted	MQ acquired (80%)	EQ acquired (80%)	MQ acquired (90%)	EQ acquired (90%)
1	Chinese	100.0	100.0	+	+	+	+
2	Chinese	100.0	100.0	+	+	+	+
3	Chinese	100.0	100.0	+	+	+	+
4	Chinese	100.0	100.0	+	+	+	+
5	Chinese	100.0	100.0	+	+	+	+
6	Chinese	100.0	100.0	+	+	+	+
7	Chinese	100.0	50.0	+	−	+	−
8	Chinese	100.0	100.0	+	+	+	+
9	Chinese	100.0	100.0	+	+	+	+
10	Chinese	100.0	100.0	+	+	+	+
11	Chinese	100.0	100.0	+	+	+	+
12	Chinese	100.0	100.0	+	+	+	+
13	Chinese	100.0	100.0	+	+	+	+
14	Chinese	100.0	100.0	+	+	+	+
15	Chinese	100.0	90.0	+	+	+	−
16	Chinese	85.7	100.0	+	+	−	+
17	Chinese	100.0	100.0	+	+	+	+
18	Chinese	100.0	100.0	+	+	+	+
19	Chinese	100.0	100.0	+	+	+	+
20	Chinese	100.0	100.0	+	+	+	+
21	Chinese	100.0	76.9	+	−	+	−
22	Chinese	100.0	87.5	+	+	+	−
23	Chinese	100.0	100.0	+	+	+	+
24	Chinese	100.0	100.0	+	+	+	+
25	Chinese	100.0	85.7	+	+	+	−
26	Chinese	92.3	100.0	+	+	+	+
27	Chinese	100.0	100.0	+	+	+	+
28	Chinese	100.0	100.0	+	+	+	+
29	Chinese	100.0	100.0	+	+	+	+
30	Chinese	100.0	100.0	+	+	+	+
31	Chinese	100.0	100.0	+	+	+	+
32	Chinese	93.3	100.0	+	+	+	+
33	Spanish	100.0	100.0	+	+	+	+

34	Spanish	100.0	62.5	+	–	+	–
35	Spanish	100.0	100.0	+	+	+	+
36	Spanish	100.0	100.0	+	+	+	+
37	Spanish	100.0	100.0	+	+	+	+
38	Spanish	92.3	100.0	+	+	+	+
39	Spanish	100.0	100.0	+	+	+	+
40	Spanish	100.0	100.0	+	+	+	+
41	Spanish	100.0	93.3	+	+	+	+
42	Spanish	100.0	100.0	+	+	+	+
43	Spanish	100.0	100.0	+	+	+	+
44	Spanish	100.0	100.0	+	+	+	+
45	Spanish	100.0	100.0	+	+	+	+
46	Spanish	100.0	100.0	+	+	+	+
47	Spanish	100.0	87.5	+	+	+	–
48	Spanish	100.0	100.0	+	+	+	+
49	Spanish	100.0	85.7	+	+	+	–
50	Spanish	76.9	90.9	–	+	–	+
51	Spanish	92.9	100.0	+	+	+	+
52	Spanish	100.0	85.7	+	+	+	–
53	Spanish	100.0	100.0	+	+	+	+
54	Spanish	100.0	100.0	+	+	+	+
55	Spanish	100.0	100.0	+	+	+	+
56	Spanish	100.0	100.0	+	+	+	+
57	Spanish	100.0	100.0	+	+	+	+
58	Spanish	64.3	100.0	–	+	–	+
59	Spanish	100.0	87.5	+	+	+	–
60	Spanish	100.0	100.0	+	+	+	+
61	Spanish	100.0	100.0	+	+	+	+
62	Spanish	80.0	62.5	–	–	–	–
63	Spanish	100.0	100.0	+	+	+	+
64	Spanish	100.0	100.0	+	+	+	+

Appendix G.

List 1.
What Tom is drinking with dinner?
The band has played in Manhattan?
Is Lee complaining about her job?
Where the neighbors have invited Mary?
The math teacher is helping Bill now?
Has John opened the window?
Where is Julia going this summer?
What have the students learned this year?
Why is the cat sleeping outside?
Paul has met Luis in France?
Is Lee making eggs for breakfast today?
Who has the professor hired?
The cat is hiding upstairs now?
Has the teacher bought new clothes?
Where Juan is teaching Spanish?
Why has Mark cleaned his office?
What is Dianne's husband cooking now?
The president has traveled to Taiwan?
Is Ted walking to work?
Where has Bill forgotten his keys?
Matt is seeing Sarah for brunch today?
Has the boss fired Mark for stealing?
Who Laura is visiting in London?
What the waiter has suggested to Mary?
Why John is smoking a lot?
The nurse has fed the patient today?
Is Mary carrying makeup in her bag?
Who the boss has complimented?
Neil is washing his clothes downstairs?
Has the student read the textbook?
Who is the baby-sitter driving to school?
Why Gloria has called Jim?
Sarah doesn't know what Mary is wearing to the Opera.
Sarah doesn't know if the company has built a school in Bolivia.
Mary doesn't know if is Todd cleaning his bedroom.
Sue doesn't know where Helen is performing tonight.
Mary doesn't know if Sue is watching her friend on TV.
Bob doesn't know if has the dog dug a hole.
Phil doesn't know where is Mary taking her car.

Liz doesn't know what has Liam painted.
Mark doesn't know why is his colleague writing a complaint.
Bob doesn't know if Mary has brought her son to the party.
Erin doesn't know is Helen knitting a scarf.
Eric doesn't know who has the dog chased.
Tom doesn't know if Erica is studying in the library.
Peter doesn't know has the dog chewed a bone.
Sandra doesn't know where John is planting a tree.
Lucy doesn't know why has the boss hugged Martha.
Sue knows what is Mary having for breakfast.
Sheila knows if Alexandra has put her bag in the car.
Norah knows if is Sue throwing out her mattress.
Bob knows where has Mary sent her applications.
Sally knows if the police are stopping the thief.
John knows if have the children pulled the cat's tail.
Martin knows who the audience is applauding.
Nina knows what Jessica has baked.
Nancy knows why Bill is pushing John.
Joe knows if the dog is licking the cat.
Jill knows is her daughter drawing a flower.
Janet knows who her husband has seen for lunch.
Sean knows if his sister is eating at home.
Eva knows has the cat broken the vase.
Mindy knows who is the professor helping.
Seth knows why Sally has hit Luke.
What did Mary say that broke the vase?
Who did Sarah exclaim the fire burned?
Where did Sarah lament that her documents got lost?
What did Nancy think exploded because of the heat?
Who did Julia believe that called her boss?
What did John brag the boss gave him?
Why did Matt complain that his brother was laid off?
Who did the detective conclude escaped to Europe?
What did you assume that hit the window?
What did the teacher observe the new method improved?
Where did Matt emphasize that the battle took place?
What did you claim crashed because of the storm?
Who did the boss decide that would invite the client?
What did the professor remark the president signed?
Why did Mark stress that Diana left?
Who did the police declare was arrested in France?
What did Nate mention that injured his wife?
What did Nate mention injured his wife?

Why did you suggest that the flowers froze?
Why did you suggest the flowers froze?
Who did Lisa comment that stole the money?
Who did Lisa comment stole the money?
Where did the detective suspect that the thief would hide?
Where did the detective suspect the thief would hide?
What is Seth sorry that offended his host?
What is Seth sorry offended his host?
Why did the scientist figure that the liquid reacted?
Why did the scientist figure the liquid reacted?
Who did Susan regret that bought her boat?
Who did Susan regret bought her boat?
Where did you suppose that Phil arrived?
Where did you suppose Phil arrived?
The man that helped the couple was very nice.
The tourists the agency called were rude.
The friends that Erin visited were from Italy.
The cat attacked the dog was scared.
The people that invited Mary were happy.
The woman Bob hired was pleased.
The lawyer that Mark fired was sorry.
The friend introduced Joe to Vicky left early.
The friend that brought Nina to the party knew everybody.
The man Kate kissed was handsome.
The woman that John hugged is a dancer.
The man shot the policemen was arrested.
The woman that married Bill's brother is a designer.
The manager his employee accused won the cause.
The friend that Sue complimented on her dress was an artist.
The person criticized Luis was his boss.
The student that praised Mary got a good grade.
The woman Mark saw was a history professor.
The man that Sally dated left the country.
The people trusted the manager were upset.
The family that hosted Fiona gave her a present.
The students Jill met at the station did not speak English.
The person that Sue blamed was her coworker.
The woman interviewed the psychologist was upset.
The man that provoked Bill outside the pub was drunk.
The professor the principal irritated resigned.
The woman that Ana trained for the competition did a great job.
The man drove Susanna to the airport was extremely nice.
The woman that served John was nice.

The man the scouts guided through the cave was tired.
The friend that Mark distrusted is Nick.
The woman promoted Jessica was hired by another company.

List 2
What is Tom drinking with dinner?
Where the band has played?
Lee is complaining about her job?
Have the neighbors invited Mary to dinner?
Is the math teacher helping Bill now?
Why has John opened the window?
Where Julia is going this summer?
The students have learned Italian this year?
Why the cat is sleeping outside?
Who Paul has met in France?
Lee is making eggs for breakfast today?
Has the professor hired Anna?
Is the cat hiding upstairs now?
What the teacher has bought?
Where is Juan teaching Spanish?
Mark has cleaned his office?
What Dianne's husband is cooking now?
Where has the president traveled?
Ted is walking to work?
Has Bill forgotten his keys at school?
Is Matt seeing Sarah for brunch today?
Why the boss has fired Mark?
Who is Laura visiting in London?
The waiter has suggested pasta to Mary?
Why is John smoking a lot?
Who has the nurse fed today?
Mary is carrying makeup in her bag?
Has the boss complimented Laura?
Is Neil washing his clothes downstairs?
What has the student read?
Who the baby-sitter is driving to school?
Gloria has called Jim for advice?
Sarah doesn't know what is Mary wearing to the Opera.
Sarah doesn't know where the company has built a school.
Mary doesn't know if Todd is cleaning his bedroom.
Sue doesn't know if is Helen performing in Tokyo tonight.
Mary doesn't know if is Sue watching her friend on TV.
Bob doesn't know why has the dog dug a hole.

Phil doesn't know where Mary is taking her car.
Liz doesn't know if Liam has painted a picture.
Mark doesn't know why his colleague is writing a complaint.
Bob doesn't know who Mary has brought to the party.
Erin doesn't know if Helen is knitting a scarf.
Eric doesn't know has the dog chased the postman.
Tom doesn't know is Erica studying in the library.
Peter doesn't know what the dog has chewed.
Sandra doesn't know where is John planting a tree.
Lucy doesn't know if the boss has hugged Martha.
Sue knows what Mary is having for breakfast.
Sheila knows where has Alexandra put her bag.
Norah knows if Sue is throwing out her mattress.
Bob knows if has Mary sent her applications to graduate school.
Sally knows if are the police stopping the thief.
John knows why the children have pulled the cat's tail.
Martin knows who is the audience applauding.
Nina knows if Jessica has baked a cake.
Nancy knows why is Bill pushing John.
Joe knows who is the dog licking.
Jill knows if her daughter is drawing a flower.
Janet knows has her husband seen Matt for lunch.
Sean knows is his sister eating at home.
Eva knows what has the cat broken.
Mindy knows who the professor is helping.
Seth knows if Sally has hit Luke.
What did Mary say broke the vase?
What did Sarah exclaim that burned Jill?
Where did Sarah lament her documents got lost?
Why did Nancy think that the building exploded?
Who did Julia believe called her boss?
Who did John brag that gave him a present?
Why did Matt complain his brother was laid off?
Where did the detective conclude that the criminal escaped?
What did you assume hit the window?
What did the teacher observe that improved learning?
Where did Matt emphasize the battle took place?
Why did you claim that the boat crashed?
Who did the boss decide would invite the client?
Who did professor remark that signed the treaty?
Why did Mark stress Diana left?
Where did the police declare that the murderer was arrested?
What did you expect that would cause a fight?

What did you expect the misunderstanding would cause?
Where did Sarah note that the ring disappeared?
What did Sarah note disappeared at the party?
Who did the secretary maintain that the boss fired?
Who did the secretary maintain fired the consultant?
Why did Paul convey that the neighbors sold the house?
Who did Paul convey sold the house because of debts?
What did the professor imagine that would stop the fighting?
What did the professor imagine that the new policy would stop?
Where did the witness swear that the gun fight happened?
What did the witness swear happened in the shop?
Who did the student feel that pleased the professor?
Who did the student feel his friend pleased?
Why is Edmund sad that Dan got sick?
Who is Edmund sad got sick because of the weather?
The man helped the couple was very nice.
The tourists that called the agency were rude.
The friends Erin visited were from Italy.
The cat that the dog attacked was scared.
The people invited Mary were happy.
The woman that hired Bob was pleased.
The lawyer Mark fired was sorry.
The friend that Joe introduced to Vicky left early.
The friend brought Nina to the party knew everybody.
The man that kissed Kate was handsome.
The woman John hugged is a dancer.
The man that the policeman shot was arrested.
The woman married Bill's brother is a designer.
The manager that accused his employee won the cause.
The friend Sue complimented on her dress was an artist.
The person that Luis criticized was his boss.
The student praised Mary got a good grade.
The woman that saw Mark was a history professor .
The man Sally dated left the country.
The people that the manager trusted were upset.
The family hosted Fiona gave her a present.
The students that met Jill at the station did not speak English.
The person Sue blamed was her coworker.
The woman that the psychologist interviewed was upset.
The man provoked Bill outside the pub was drunk.
The professor that irritated the principal resigned.
The woman Ana trained for the competition did a great job.
The man that Susanna drove to the airport was extremely nice.

The woman served John was nice.
The man that guided the scouts through the cave was tired.
The friend Mark distrusted is Nick.
The woman that Jessica promoted was hired by another company.

List 3
Is Tom drinking wine with dinner?
Where has the band played?
Why is Lee complaining about her job?
The neighbors have invited Mary to dinner?
Who the math teacher is helping now?
Why John has opened the window?
Julia is going to Europe in summer?
Have the students learned Italian this year?
Is the cat sleeping outside?
Who has Paul met in France?
What Lee is making for breakfast today?
The professor has hired Anna?
Where the cat is hiding now?
What has the teacher bought?
Juan is teaching Spanish in Manhattan?
Has Mark cleaned his office?
Is Dianne's husband cooking fish now ?
Where the president has traveled?
Why Ted is walking to work?
Bill has forgotten his keys at school?
Who is Matt seeing for brunch today?
Why has the boss fired Mark?
Laura is visiting Bill in London?
Has the waiter suggested pasta to Mary?
Is John smoking a lot?
Who the nurse has fed today?
What is Mary carrying in her bag?
The boss has complimented Laura?
Where is Neil washing his clothes?
What the student has read?
The baby-sitter is driving Lucy to school?
Has Gloria called Jim for advice?
Sarah doesn't know if is Mary wearing a gown to the Opera .
Sarah doesn't know where has the company built a school.
Mary doesn't know why is Todd cleaning his bedroom.
Sue doesn't know if Helen is performing in Tokyo tonight.
Mary doesn't know who Sue is watching on TV.

Bob doesn't know why the dog has dug a hole.
Phil doesn't know if Mary is taking her car to the mechanic.
Liz doesn't know if has Liam painted a picture.
Mark doesn't know is his colleague writing a complaint.
Bob doesn't know who has Mary brought to the party.
Erin doesn't know what Helen is knitting.
Eric doesn't know if the dog has chased the postman.
Tom doesn't know where Erica is studying .
Peter doesn't know what has the dog chewed.
Sandra doesn't know if John is planting a tree in the backyard.
Lucy doesn't know has the boss hugged Martha.
Sue knows if is Mary having eggs for breakfast.
Sheila knows where Alexandra has put her bag.
Norah knows why Sue is throwing out her mattress.
Bob knows if Mary has sent her applications to graduate school.
Sally knows who are the police stopping.
John knows why have the children pulled the cat's tail.
Martin knows if the audience is applauding the pianist.
Nina knows if has Jessica baked a cake.
Nancy knows is Bill pushing John.
Joe knows who the dog is licking.
Jill knows what is her daughter drawing.
Janet knows if her husband has seen Matt for lunch.
Sean knows where is his sister eating.
Eva knows what the cat has broken.
Mindy knows if the professor is helping Tom.
Seth knows has Sally hit Luke.
What did Mary say that the wind broke?
What did Sarah exclaim burned Jill?
What did Sarah lament that got lost in Egypt?
Why did Nancy think the building exploded?
Who did Julia believe that her boss called?
Who did John brag gave him a present?
Who did Matt complain that was laid off because of the cuts?
Where did the detective conclude the criminal escaped?
What did you assume that the ball hit?
What did the teacher observe improved learning?
What did Matt emphasize that took place in China?
Why did you claim the boat crashed?
Who did the boss decide that the manager would invite?
Who did the professor remark signed the treaty?
Who did Mark stress that left because of the fight?
Where did the police declare the murderer was arrested?

Who did Nate mention the car injured?
Who did Nate mention that the car injured?
What did you suggest froze because of the cold?
What did you suggest that froze because of the cold?
What did Lisa comment the contractor stole?
What did Lisa comment that the contractor stole?
Who did the detective suspect would hide in the shed?
Who did the detective suspect that would hide in the shed?
Who is Seth sorry his behavior offended?
Who is Seth sorry that his behavior offended?
What did the scientist figure reacted because of the heat?
What did the scientist figure that reacted because of the heat?
What did Susan regret she bought?
What did Susan regret that she bought?
Who did you suppose arrived in Paris?
Who did you suppose that arrived in Paris?
The man that the couple helped was very nice.
The tourists called the agency were rude.
The friends that visited Erin were from Italy.
The cat the dog attacked was scared.
The people that Mary invited were happy.
The woman hired Bob was pleased.
The lawyer that fired Mark was sorry.
The friend Sal introduced to Vicky left early.
The friend that Nina brought to the party knew everybody.
The man kissed Kate was handsome.
The woman that hugged John is a dancer.
The man the policeman shot was arrested.
The woman that Bill's brother married is a designer.
The manager accused his employee won the cause.
The friend that complimented Sue on her dress was an artist .
The person Luis criticized was his boss.
The student that Mary praised got a good grade.
The woman saw Mark was a history professor.
The man that dated Sally left the country.
The people the manager trusted were upset.
The family that Fiona hosted gave her a present.
The students met Jill at the station did not speak English.
The person that blamed Sue was her coworker.
The woman the psychologist interviewed was upset.
The man that Bill provoked outside the pub was drunk.
The professor irritated the principal resigned.
The woman that trained Ana for the competition did a great job.

The man Susanna drove to the airport was extremely nice.
The woman that John served was nice.
The man guided the scouts through the cave was tired.
The friend that distrusted Mark is Nick.
The woman Jessica promoted was hired by another company.

List 4
Tom is drinking wine with dinner?
Has the band played in Manhattan?
Why Lee is complaining about her job?
Where have the neighbors invited Mary?
Who is the math teacher helping now?
John has opened the window?
Is Julia going to Europe this summer?
What the students have learned this year?
The cat is sleeping outside?
Has Paul met Luis in France?
What is Lee making for breakfast today?
Who the professor has hired?
Where is the cat hiding now?
The teacher has bought new clothes?
Is Juan teaching Spanish in Manhattan?
Why Mark has cleaned his office?
Dianne's husband is cooking fish now?
Has the president traveled to Taiwan?
Why is Ted walking to work?
Where Bill has forgotten his keys?
Who Matt is seeing for brunch today?
The boss fired Mark for stealing?
Is Laura visiting Bill in London?
What has the waiter suggested to Mary?
John is smoking a lot?
Has the nurse fed the patient today?
What Mary is carrying in her bag?
Who has the boss complimented?
Where Neil is washing his clothes?
The student has read the textbook?
Is the baby-sitter driving Lucy to school?
Why has Gloria called Jim?
Sarah doesn't know if Mary is wearing a gown to the Opera .
Sarah doesn't know if has the company built a school in Bolivia.
Mary doesn't know why Todd is cleaning his bedroom.
Sue doesn't know where is Helen performing tonight.

Mary doesn't know who is Sue watching on TV.
Bob doesn't know if the dog has dug a hole.
Phil doesn't know if is Mary taking her car to the mechanic.
Liz doesn't know what Liam has painted.
Mark doesn't know if his colleague is writing a complaint.
Bob doesn't know has Mary brought her son to the party.
Erin doesn't know what is Helen knitting.
Eric doesn't know who the dog has chased.
Tom doesn't know where is Erica studying .
Peter doesn't know if the dog has chewed a bone.
Sandra doesn't know is John planting a tree in the backyard.
Lucy doesn't know why the boss has hugged Martha.
Sue knows if Mary is having eggs for breakfast.
Sheila knows if has Alexandra put her bag in the car.
Norah knows why is Sue throwing out her mattress.
Bob knows where Mary has sent her applications.
Sally knows who the police are stopping.
John knows if the children have pulled the cat's tail.
Martin knows if is the audience applauding the pianist.
Nina knows what has Jessica baked.
Nancy knows if Bill is pushing John.
Joe knows is the dog licking the cat.
Jill knows what her daughter is drawing.
Janet knows who has her husband seen for lunch.
Sean knows where his sister is eating.
Eva knows if the cat has broken the vase.
Mindy knows is the professor helping Tom.
Seth knows why has Sally hit Luke.
What did Mary say the wind broke?
Who did Sarah exclaim that the fire burned?
What did Sarah lament got lost in Egypt?
What did Nancy think that exploded because of the heat?
Who did Julia believe her boss called?
What did John brag that the boss gave him?
Who did Matt complain was laid off because of the cuts?
Who did the detective conclude that escaped to Europe?
What did you assume the ball hit?
What did the teacher observe that the new method improved?
What did Matt emphasize took place in China?
What did you claim that crashed because of the storm?
Who did the boss decide the manager would invite?
What did the professor remark that the president signed?
Who did Mark stress left because of the fight?

Who did the police declare that was arrested in France?
What did you expect would cause a fight?
What did you expect that the misunderstanding would cause?
Where did Sarah note the ring disappeared?
What did Sarah note that disappeared at the party?
Who did the secretary maintain the boss fired?
Who did the secretary maintain that fired the consultant?
Why did Paul convey the neighbors sold the house?
Who did Paul convey that sold the house because of debts?
What did the professor imagine would stop the fighting?
What did the professor imagine that the new policy would stop?
Where did the witness swear the gun fight happened?
What did the witness swear that happened in the shop?
Who did the student feel pleased the professor?
Who did the student feel that his friend pleased?
Why is Edmund sad Dan got sick?
Who is Edmund sad that got sick because of the weather?
The man the couple helped was very nice.
The tourists that the agency called were rude.
The friends visited Erin were from Italy.
The cat that attacked the dog was scared.
The people Mary invited were happy.
The woman that Bob hired was pleased.
The lawyer fired Mark was sorry.
The friend that introduced Joe to Vicky left early.
The friend Nina brought to the party knew everybody.
The man that Kate kissed was handsome.
The woman hugged John is a dancer.
The man that shot the policeman was arrested.
The woman Bill's brother married is a designer.
The manager that his employee accused won the cause.
The friend complimented Sue on her dress was an artist.
The person that criticized Luis was his boss.
The student Mary praised got a good grade.
The woman that Mark saw was a history professor.
The man dated Sally left the country.
The people that trusted the manager were upset.
The family Fiona hosted gave her a present.
The students that Jill met at the station did not speak English.
The person blamed Sue was her coworker.
The woman that interviewed the psychologist was upset.
The man Bill provoked outside the pub was drunk.
The professor that the principal irritated resigned.

The woman trained Ana for the competition did a great job.
The man that drove Susanna to the airport was extremely nice.
The woman John served was nice.
The man that the scouts guided through the cave was tired.
The friend distrusted Mark was Nick.
The woman that promoted Jessica was hired by another company.

Appendix H.

List 1A

Practice

Someone was watching the birds. Katie wants to find out who. She says…
One of the dogs was running. Katie wants to find out which dog. She says…
Something was flying in the sky. Katie wants to find out what. She says…

Experimental

Katie's mom is cleaning something. Katie wants to find out what. She says…
Katie's brother is washing the dog. Katie wants to find out why. She says...
Katie's mom is singing a song. Katie wants to find out when. She says?
Katie's dog is chewing one of the toys. Katie wants to find out which toy. She says...
Katie's brother is doing his homework soon. Katie wants to find out when. She says...
Katie's mom is calling dad. Katie wants to find out why. She says...
Katie's brother is building something. Katie wants to find out what. She says...
Katie's mom is feeding one of the dolls. Katie wants to find out which doll. She says...

Practice

Maybe Katie's dog was barking, but maybe not. Katie wants to find out. She says…
Maybe Katie's mom was drinking juice, but maybe not. Katie wants to find out. She says…
Maybe Katie's brother was jumping, but maybe not. Katie wants to find out. She says…

Experimental

Maybe Katie's mom is reading a book, but maybe not. Katie wants to find out. She says...
Maybe Katie's mom is drawing a flower, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is eating cookies, but maybe not. Katie wants to find out. She says...
Maybe Katie's mom is cooking pizza, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is buying ice-cream now, but maybe not. Katie wants to find out. She says...
Maybe Katie's mom is writing a letter, but maybe not. Katie wants to find out. She says...
Maybe Katie's dog is chasing the cat, but maybe not. Katie wants to find out. She says...
Maybe Katie's dog is digging a hole, but maybe not. Katie wants to find out. She says...

List 1B

Practice

Maybe Katie's dog was barking, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum was drinking juice, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother was jumping, but maybe not. Katie wants to find out. She says…
Experimental

Maybe Katie's brother is buying ice-cream now, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is writing a letter, but maybe not. Katie wants to find out. She says...
Maybe Katie's dog is chasing the cat, but maybe not. Katie wants to find out. She says…
Maybe Katie's dog is digging a hole, but maybe not. Katie wants to find out. She says…
Maybe Katie's mum is reading a book, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is drawing a flower, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is eating cookies, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is cooking pizza, but maybe not. Katie wants to find out. She says...

Practice

Someone was watching the birds. Katie wants to find out who. She says...
One of the dogs was running. Katie wants to find out which one. She says…
Something was flying in the sky. Katie wants to find out what. She says…

Experimental

Katie's brother is doing his homework soon. Katie wants to find out when. She says...
Katie's mum is calling dad. Katie wants to find out why. She says...
Katie's brother is building something. Katie wants to find out what. She says...
Katie's mum is feeding one of the dolls. Katie wants to find out which doll. She says...
Katie's mum is cleaning something. Katie wants to find out what. She says…
Katie's brother is washing the dog. Katie wants to find out why. She says...
Katie's mom is singing a song. Katie wants to find out when. She says...
Katie's dog is chewing one of the toys. Katie wants to find out which toy. She says...

List 2A

Practice

Someone was watching the birds. Katie wants to find out who. She says...
One of the dogs was running. Katie wants to find out which one. She says...
Something was flying in the sky. Katie wants to find out what. She asks…

Katie's brother is buying ice-cream soon. Katie wants to find out when. She says...
Katie's mum is writing a letter. Katie wants to find out why. She says...

Katie's dog is chasing one of the cats. Katie wants to find out which one. She says...
Katie's dog is digging a hole. Katie wants to find out why. She says...
Katie's mum is reading a book. Katie wants to find out which book. She says...
Katie's mum is drawing something. Katie wants to find out what. She says...
Katie's brother is eating something. Katie wants to find out what. She says...
Katie's mum is cooking pizza. Katie wants to find out when. She says...

Practice

Maybe Katie's dog is barking, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is drinking juice, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is jumping, but maybe not. Katie wants to find out. She says...

Experimental

Maybe Katie's brother is doing his homework now, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is calling dad, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is building a sand castle, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is feeding the doll, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is cleaning her desk, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is washing the dog, but maybe not. Katie wants to find out. She says...
Maybe Katie's mom is singing a song, but maybe not. Katie wants to find out. She says...
Maybe Katie's dog is chewing his toy, but maybe not. Katie wants to find out. She says...

List 2B

Practice

Maybe Katie's dog was barking, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum was drinking juice, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother was jumping, but maybe not. Katie wants to find out. She says...

Maybe Katie's mum is cleaning her desk, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is washing the dog, but maybe not. Katie wants to find out. She says...
Maybe Katie's mom is singing a song, but maybe not. Katie wants to find out. She says...
Maybe Katie's dog is chewing his toy, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is doing his homework now, but maybe not. Katie wants to find out. She says...
Maybe Katie's mum is calling dad, but maybe not. Katie wants to find out. She says...
Maybe Katie's brother is building a sand castle, but maybe not. Katie wants to find out. She says...

Maybe Katie's mum is feeding the doll, but maybe not. Katie wants to find out. She says...

Practice

Someone was watching the birds. Katie wants to find out who. She says...
One of the dogs was running. Katie wants to find out which one. She says...
Something was flying in the sky. Katie wants to find out what. She asks…

Katie's mum is reading a book. Katie wants to find out which book. She says...
Katie's mum is drawing something. Katie wants to find out what. She says...
Katie's brother is eating something. Katie wants to find out what. She says...
Katie's mum is cooking pizza. Katie wants to find out when. She says...
Katie's brother is buying ice-cream soon. Katie wants to find out when. She says...
Katie's mum is writing a letter. Katie wants to find out why. She says...
Katie's dog is chasing one of the cats. Katie wants to find out which one. She says...
Katie's dog is digging a hole. Katie wants to find out why. She says...

Appendix I.

List 1A

Practice
Were you jumping?
Were you drinking juice?
Were you laughing?

Experimental

Are you cleaning the desk?
Is mom washing the dog?
Are you singing a song?
Is the dog chewing his toy?
Is mom doing her homework?
Are you calling dad?
Is mom building a sand castle?
Are you feeding the doll?

Practice

Who was playing with your toys?
What was on your bed?
Which teacher was helping you?

Experimental

Which book are you reading?
What are you drawing?
What is mom eating?
When are you cooking pizza?
When is mom buying ice-cream?
Why are you writing a letter?
Which cat is mom brushing?
Why is the dog digging a hole?

List 1B

Practice

Who was playing with your toys?
What was on your bed?
Which teacher was helping you?

Experimental

When is mom buying ice-cream?
Why are you writing a letter?
Which cat is mom brushing?
Why is the dog digging a hole?
Which book are you reading?
What are you drawing?
What is my mom eating?
When are you cooking pizza?

Practice

Were you jumping?
Were you drinking juice?
Were you laughing?

Experimental

Is mom doing her homework?
Are you calling dad?
Is mom building a sand castle?
Are you feeding the doll?
Are you cleaning your desk?
Is mom washing the dog?
Are you singing a song?
Is the dog chewing his toy?

List 2A

Practice

Were you jumping?
Were you drinking juice?
Were you laughing?

Experimental

Is mom buying ice-cream now?

Are you writing a letter?
Is mom brushing the cat?
Is the dog digging a hole?
Are you reading a book?
Were you drawing a flower?
Is my mom eating cookies?
Are you cooking pizza today?

Practice

Who was playing with your toys?
What was on your bed?
Which teacher was helping you?

Experimental

When is mom doing her homework?
Why are you calling dad?
What is mom building?
Which doll are you feeding?
What are you cleaning?
Why is mom washing the dog?
When are you singing a song?
Which toy is the dog chewing?

List 2B

Practice

Who was playing with your toys?
What was on your bed?
Which teacher was helping you?

Experimental

What are you cleaning?
Why is mom washing the dog?
When are you singing a song?
Which toy is the dog chewing?
When is mom doing her homework?
Why are you calling dad?
What is mom building?
Which doll are you feeding?

Practice

Were you jumping?
Were you drinking?
Were you laughing?

Experimental

Are you reading a book?
Are you drawing a flower?
Is mom eating cookies?
Are you cooking pizza today?
Is mom ice-cream now?
Are you writing a letter?
Is mom brushing the cat?
Is the dog digging a hole?

6. References

Antón-Méndez, I. (2010). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, 1–14.

Ambridge, B., Rowland, C., Theakston, A., & Tomasello, M. (2006). Comparing different accounts of inversion errors in children's non-subject *wh*-questions: What experimental data can tell us? *Journal of Child Language, 33*, 519–557.

Baauw, S. (1998). Subject-verb inversion in Spanish *wh*-questions: Movement as symmetry breaker. In R. Bezooijen, & R. Kager (Eds.), *Linguistics in the Netherlands 15* (1–12). Amsterdam: John Benjamins.

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude Estimation of Linguistic Acceptability. *Language, 72*(1), 32–68.

Batmanian, N., Sayehli, S., & Valian, V. (2008). Inclusion of auxiliaries: Competence and performance differences in early learners. *EUROSLA Yearbook, 8*, 215–234.

Beaver, D.I. (2001). *Presupposition and assertion in dynamic semantics*. Stanford, CA: CSLI Publications.

Bellugi, U. (1971). Simplification in children's language. In R. Huxley & E. Ingram (Eds.), *Language Acquisition: Models and methods* (95–119). New York: Academic Press.

Berk, S. (2003). Why "why" is different. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of BUCLD 27* (127–137). Somerville, MA: Cascadilla Press.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and development of language* (279–362). New York: Wiley.

Bhatt, R.M., & Hancin-Bhatt, B. (2002). Structural minimality, CP and the initial state in second language acquisition. *Second Language Research, 18*, 348–392.

Bialystock, E. (1982). On the relationship between knowing and using linguistic forms. *Applied Linguistics, 3*, 181–206.

Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3–49.

Bley-Vroman, R. (1997). *Features and patterns in foreign language learning*. Paper presented at *SLRF 17*, University of Michigan, East Lansing, MI.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175–198.

Bley-Vroman, R., & Masterson, D. (1989). Reaction time as a supplement to grammaticality judgments in the investigation of second language learners' competence. *University of Hawaii Working Papers, 9*, 207–237.

Bloom, L., Merkin, S., & Wooten, J. (1982). *Wh*-questions: linguistic factors that contribute to the sequence of acquisition. *Child Development, 53*, 1084–1092.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 33*(2), 247–271.

Brandtler, J. (2008). Why we should ever bother about *wh*-questions: On NPI-licensing properties of *wh*-questions in Swedish. *Working Papers in Scandinavian Syntax, 81*, 83–102.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. Hayes (Ed.), *Cognition and the development of language* (1–53). New York: Wiley.

Cancino, H., Rosandy, E.J., & Shuman, J.H. (1975). The acquisition of the English auxiliary by native Spanish speakers. *TESOL Quarterly, 9*(4), 421–430.

Cazden, C. (1970). Children's questions: Their forms, functions and roles in education. *Young Children, 25*(4), 202–220.

Chaudron, C. (1983). Research on metalinguistic judgments: A review of theory, methods, and results. *Language Learning*, *33*(3), 343–377.

Clahsen, H., & Smolka, K.-D. (1986). Psycholinguistic evidence and the description of V2 phenomena in German. In H. Haider & M. Prinzhorn (Eds.), *Verb Second phenomena in Germanic languages* (137–168). Dordrecht: Foris Publications.

Clahsen, H., & Muysken, P. (1986). The availability of Universal Grammar in adult and child learners: A study of the acquisition of German word order. *Second Language Research, 2*, 93–119.

Clahsen, H., Meisel, J., & Pienemann, M. (1983). *Deutsch als Zweitsprache* [German as a second language]. Tübingen: Narr.

Collins, C. (1991). Why and how come. In L. Cheng and H. Demirdache (Eds.), *MIT Working Papers in Linguistics 15*: *More Papers on* Wh-*Movement* (31–45). Cambridge, MA: MITWPL.

Comorovski, I. (1996). *Interrogative phrases and the syntax-semantics interface*. Dordrecht: Kluwer.

Conroy, A. (2006). The semantics of how come: A look at how factivity does it all. *University of Maryland Working Papers in Linguistics*, *14,* 1–24.

Conroy, A., & Lidz, J. (2007). Production/comprehension asymmetry in children's why questions. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of GALANA 2007* (73–83). Somerville, MA: Cascadilla Press.

Crowley, T., & Rigsby, B. (1987). Question formation. In T. Shopen (Ed.), *Languages and their status* (153–207). Cambridge, MA: Winthrop.

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

de Villiers, J.G. (1991). Why questions? In T. Maxfield & B. Plunkett (Eds.), *University of Massachusetts Occasional Papers in Linguistics: The Acquisition of Wh* (155–174). Amherst, MA: GLSA.

de Villiers, J.G., Roeper, T., Bland-Stewart, L., & Pearson, B. (2008). Answering hard questions: *Wh*-movement across dialects and disorder. *Applied Psycholinguistics*, *29*, 67–103.

DeKeyser, R.M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–533.

DeKeyser, R.M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language learning* (313–348). Oxford: Blackwell.

den Dikken, M., & Giannakidou, A. (2002). From hell to polarity: Aggressively non-D-linked *wh*-phrases as polarity items. *Linguistic Inquiry, 33*(1), 31–61.
Derwing, T., & Smyth, R. (1988). Auxiliary placement revisited. *Journal of the Atlantic Provinces Association*, *10*, 1–15.
Dopke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language, 25*, 555–584.
Dyson, B. (2008). What can we learn from questions: ESL question development and its implications for language assessment. *Prospect, 23*, 16–27.
Eckman, F., Moravscik, E., & Wirth, J. (1989). Implicational universals and interrogative structure in the interlanguage of ESL learners. *Language Learning, 39*, 173–205.
Eilam, A. (2011). *Explorations in the informational component*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
Ellis, N.C. (Ed.). (1994). *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.
Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly, 27*, 91–113.
Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227–275.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Language, 27*, 141–172.
Epstein S.D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences, 1*, 677–758.
Epstein, S.D., Flynn, S., & Martohardjono, G. (1998). The strong continuity hypothesis: Some evidence concerning functional categories in adult L2 acquisition. In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The generative study of second language acquisition* (61–77). Mahwah, NJ: Lawrence Erlbaum.
Erreich, A. (1984). Learning how to ask: Patterns of inversion in yes-no and *wh*-questions, *Journal of Child Language, 11*, 592–597.
Escutia, M. (2002). Spanish learners' production of English embedded interrogative clauses. *Circulo de Linguistica Aplicada a la Comunicación*. Madrid: Universidad Complutense de Madrid.
Eubank, L. (1993). On the transfer of parametric values in L2 development. *Language Acquisition, 3*, 183–208.
Eubank, L. (1996). Negation in early German-English interlanguage: More valueless features in the L2 initial state. *Second Language Research*, *12*, 73–106.
Felix, S. (1976) *Wh*-pronouns in first and second language acquisition. *Linguistische Berichte*, *44*, 52–64.
Filppula, M. (1999). *The grammar of Irish English: Language in Hibernian style*. New York: Routledge.
Finegan, E. (1999). *Language: Its structure and use*. Ft Worth, TX: Harcourt.

Fitzpatrick, J. (2005). The whys and how comes of presupposition and NPI licensing in questions. In J. Alderete, C.-H. Han, & A. Kochetov (Eds.), *Proceedings of WCCFL 24* (138–145). Somerville, MA: Cascadilla Press.

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in comprehension, imitation and production. *Journal of Verbal Learning and Verbal Behavior, 2,* 121–135.

Flynn, S. (1996). A parameter-setting approach to second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (121–58). San Diego, CA: Academic Press.

Flynn S., & Martohardjono, G. (1994). Mapping from the initial state to the final state: The separation of universal principles and language specific principles. In B. Lust, M. Suñer, & J. Whitman (Eds.), *Syntactic theory and first language acquisition: Cross-linguistic perspectives, Vol. 1: Heads, projections and learnability* (319–335). Hillsdale, NJ: Lawrence Erlbaum.

Gawlitzek-Maiwald, I., Tracy, R., & Fritzenshaft, A. (1992). Language acquisition and competing linguistic representations: The child as arbiter. In J. Meisel (Ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition* (139–179). Dordrecht: Kluwer.

Ginzburg, J. (2004). A quasi-naive semantics for interrogatives and its implications. In J. Gutierrez-Rexach (Ed.), *Semantics: Critical concepts in linguistics* (353–373). London: Routledge.

Gleitman, L.R., Newport, E.L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11,* 43–79.

Goodall, G. (1991). On the Status of SPEC of IP. Paper presented at *WCCFL X*, Arizona State University, Tempe, AZ.

Green, L. (2002). *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.

Grinstead, J., Warren, V., Ricci, C., & Sanderson, S. (2009). Finiteness, subject-aux inversion and the Head Movement Constraint in Child English. Paper presented at *GALA 2009*, Lisbon, Portugal.

Groenendijk, J., & Stokhof, M. (1984). *Studies on the semantics of questions and the pragmatics of answers*. Doctoral dissertation, University of Amsterdam, Amsterdam.

Guasti, M.T. (1996). Acquisition of Italian interrogatives. In H. Clahsen (Ed.), *Generative perspectives on language acquisition* (241–270). Amsterdam/Philadelphia: John Benjamins.

Guasti, M.T. (2000). An excursion into interrogatives in early English and Italian. In M. A. Friedemann & L. Rizzi (Eds.), *The acquisition of syntax: Studies in comparative developmental linguistics* (103–128). London: Longman.

Guasti, M.T. (2002). *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.

Guasti, M.T., Thornton, R., & Wexler, K. (1995). Negation in children's questions: The case of English. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of BUCLD 19* (228–240). Somerville, MA: Cascadilla Press.

Haig, J. (1991). *Universal grammar and second language acquisition: the influence of task type on late learner's access to the subjacency principle*. TESL monograph, McGill University, Montreal, Quebec, Canada.

Håkansson, G., & Dooley-Collberg, S. (1994). The preference for Modal+Neg. An L2 perspective applied to L1 acquisition. *Second Language Research, 10*(2), 95–124.

Hattori, R. (2003). Why do children say did you went? The role of do-support. In A. Brugos, L. Micciulla, and C.E. Smith (Eds.), *Supplement to the Proceedings of BUCLD 28.* Somerville, MA: Cascadilla Press.

Hawkins, R. (2001). Second language syntax: A generative introduction. Oxford: Blackwell Publishers.

Hawkins, R., & Chan, Y. (1997). The partial availability of Universal Grammar in second language acquisition: the `failed functional features hypothesis'. Second language Research, 13(3), 187-226.

Haznedar, B. (2003). The status of functional categories in child second language acquisition: Evidence from the acquisition of CP. *Second Language Research, 19*, 1–41.

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford: Oxford University Press.

Henry, A., & Tangney, D. (1999). Functional categories and parameter setting in the second-language acquisition of Irish in early childhood. In A. DeGraff (Ed.), *Language creation and language change: Crealization, diachrony and development* (139–253). Cambridge MA: MIT Press.

Heycock, C., & Kroch, A. (1999). Pseudocleft connectedness: Implications for the LF interface level. *Linguistic Inquiry, 30*(3), 365–397.

Hood, L., & Lightbown, P. (1978). What do children do when asked to 'Say what I say'? Does elicited imitation measure linguistic knowledge. *Allied Health & Behavioral Sciences, 1*, 195–219.

Huang, J. (1971). A Chinese child's acquisition of English syntax. MA thesis, UCLA.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing, and acquisition of second language knowledge. *Second Language Research, 18*, 193–223.

Hurford, J. (1975). A child and the English question formation rule. *Journal of Child Language, 2,* 299–301.

Ingram, D., & Tyack, D. (1979). Inversion of subject NP and aux in children's questions. *Journal of Psycholinguistics Research, 4*, 333–341.

International Corpus of Learner English: Version 2. (2009). S. Granger, E. Dagneaux, F. Meunier, & M. Paquot (Eds). Louvain-la-Neuve: Presses Universitaires de Louvain.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection: Theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553–562.

Johnson, J.S. (1992). Critical period effects in second language acquisition: the effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning, 42*, 217–48.

Johnson, K. (2008). *Quantitative Methods in Linguistics*, Malden, MA: Blackwell.
Johnson, J.S., & Newport, E.L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39,* 215–258.
Johnston, M. (1985). *Syntactic and morphological progression in learner English.* Canberra, Australia: Department of Immigration & Ethnic Affairs.
Katz, J., & Postal, P. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.
Kay, P. (1980). On the syntax and semantics of early questions. *Linguistic Inquiry, 11*(2), 426–429.
Kayne, S. R., (1991). Romance Clitics, Verb Movement, and PRO, *Linguistic Inquiry* 22, pp. 647-686.
Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition, 2*, 37–57.
Klee, T. (1985). Role of inversion in children's question development. *Journal of Speech & Hearing Research, 28*, 225–232.
Klein, W., & Dittmar, N. (1979). *Developing grammars: The acquisition of German syntax by foreign workers*. Berlin: Springer.
Kleinmann, H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning, 27*, 93–107.
Klima, E.S., & Bellugi, U. (1966). Syntactic regularities in the speech of children. In J. Lyons & R. J. Wales (Eds.), *Psycholinguistic Papers: The Proceedings of the 1966 Edinburgh Conference* (183–219). Edinburgh: Edinburgh University Press.
Ko, H. (2005). Syntax of why-in-situ: Merge into [Spec, CP] in the overt syntax. *Natural Language & Linguistic Theory, 23*, 867–916.
Kolk, H. (2001) Does agrammatic speech constitute a regression to child language? A three- way comparison between agrammatic, child, and normal ellipsis, *Brain & Language*, 77, 340–350.
Krashen, S. (1981). *Second language acquisition and second language learning*. London: Pergamon.
Kuczaj, S.A., & Brannick, N. (1979). Children's use of the *wh*-question modal auxiliary placement rule. *Journal of Experimental Child Psychology, 28*, 43–67.
Kuczaj, S., & Maratsos, M. (1983). The initial verbs of yes–no questions: A different kind of general grammatical category. *Developmental Psychology, 19*, 440–444.
Labov, W. (1972). *Language in the inner city: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
Labov, W., & Labov, T. (1978). Learning the syntax of questions. In R. Campbell & P. Smith (Eds.), *Recent advances in the psychology of language*. New York: Plenum Press.
Lawler, J.M. (1971). Any questions? In R. Hendrik et al (Eds.), *CLS 7*, 163–173.
Lee, S.-Y. (2008). Argument-adjunct asymmetry in the acquisition of inversion in *wh* questions by Korean learners of English. *Language Learning, 58*, 625–663.
Legate, J.A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition, 14*, 315–344.

Lightbown, P. (1978). Question form and question function in the speech of young French L2 learners. In M. Paradis (Ed.), *Aspects of bilingualism.* Columbia, SC: Hornbeam.
Lightfoot, D. W. (1989). The child's trigger experience: Degree-0 learnability. Behavioral and Brain Sciences 12: 321-334.
Lightfoot, D. W. (1994). Degree-0 leearnability. In B. Lust, G. Herman & J. Kornfilt (Eds.), *Syntactic theory and first language acquisition: Crosslinguistic perspectives*, Vol. 2: *Binding, dependencies and learnability*. Lawrence Erlbaum: Hillsdale, NJ: 453-472.
Lin, J.W. (1992). The syntax of *zenmeyang* 'how' and *weishenme* 'why' in Mandarin Chinese. *Journal of East Asian Linguistics, 1*, 293–331.
Mackey, A., Pienemann, M., & Doughty, C. (1992). The value of interactional tasks in eliciting morphosyntactic structures. *Language Acquisition Research Center Occasional Papers, 2* (2).
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.
Maratsos, M. (1978). How to get from words to sentences. In D. Aaronson & R. Rieber (Eds.), *Perspectives in psycholinguistics* (285–356). Hillsdale, NJ: Erlbaum.
Maratsos, M., & Kuczaj, S. (1978). Against the transformationalist account: A simpler analysis of auxiliary overmarking. *Journal of Child Language, 5*, 337–345.
Maratsos, M., Kuczaj, S., Fox, D., & Chalkley, M. (1979). Some empirical studies on the acquisition of transformational relations: Passives, negatives and one past tense. In A. Collins (Ed.), *Children's language and communication, Vol. 12: The Minnesota Symposium on Child Psychology*. Minneapolis: University of Minnesota Press.
Martohardjono, G., & Flynn, S. (1995). Language transfer: What do we really mean? In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (205–217). Philadelphia: John Benjamins.
Mayer, J., Erreich, A., & Valian, V. (1978). Transformations, acquisition. *Cognition, 6*, 1–13.
McCloskey, J. (2006). Questions and questioning in a local English. In R. Zanuttini, H. Campos, E Herburger, & P. H. Portner (Eds.), *Crosslinguistic research in syntax and semantics: Negation, tense, and clausal architecture* (87–126). Washington, D.C.: Georgetown University Press.
Menyuk, P. (1969). *Sentences children use*. Cambridge, MA: MIT Press.
Mikkelsen, L.H. (2005). *Copular clauses: Specification, predication and equation*. Amsterdam/Philadelphia: John Benjamins.
Moro, A. (1997). Dynamic antisymmetry: Movement as symmetry breaking phenomenon. *Studia Linguistica, 51*, 50–76.
Murphy, V. (1997). The effect of modularity on a grammaticality judgment task. *Second Language Research, 13*, 34–65.
Oller, J., & Ziahosseiny, S. (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning, 20*, 183–189.
Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (393–419). San Diego, CA: Academic Press.

Park, H. (2000). *When*-questions in second language acquisition, *Second Language Research, 16*, 44–76.

Penner, Z. (1996). From empty to doubly-filled complementizers. A case study in the acquisition of subordination in Bernese Swiss German. *Arbeitspapier*, *77*. Konstanz: Fachgruppe Sprachwissenschaft der Universität Konstanz.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10,* 52–79.

Pienemann, M. (1998). Developmental dynamics in L1 and L2 acquisition: Processability theory and generative entrenchment. *Bilingualism: Language and Cognition, 1*(1), 1–20.

Pienemann, M. (Ed.). (2005). Cross-linguistic aspects of processability theory. Amsterdam: John Benjamins.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition, 10*, 217–224.

Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (45–141). Adelaide, Australia: National Curriculum Resource Centre, AMEP.

Pollock, J.-Y. (1989). Verb Movement, Universal Grammar and the Structure of IP. *Linguistic Inquiry 20, 365-424*

Potts, C., & Roeper, T. (2005). The narrowing acquisition path: From expressive small clauses to declaratives. In L. Progovac, K. Paesani, E. Casielles, & E. Barton (Eds.), *The syntax of non-sententials: Multi-disciplinary perspectives* (183–201). Amsterdam: John Benjamins.

Potts, M., Carlson, P., Cocking, R., & Copple, C. (1979). *Structure and development in child language*. Ithaca, NY: Cornell University Press.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research,* 16, 103–134.

Rayson, P. (2008). *Wmatrix: A web-based corpus processing environment*. Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/.

Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (281–337). Dordrecht: Kluwer.

Rizzi, L. (2001). On the position "Int(errogative)" in the left periphery of the clause. In G. Cinque & G. Salvi (Eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi* (287–296). Amsterdam: Elsevier.

Rizzi, L. (2005). On some properties of subjects and topics. In L. Brugé, G. Giusti, N. Munaro, W. Schweikert, & G. Turano (Eds.), *Proceedings of the XXX Incontro di Grammatica Generativa* (203–224). Venice: Cafoscarina.

Rizzi, L. (2006) Selective residual V-2 in Italian interrogatives. In P. Brandt & E. Fuss (Eds.), *Form, structure, and grammar*, 229–242. Berlin: Akademie Verlag.

Roeper, T. (2011). How the emergence of propositions separates strict interfaces from general inference. In I. Reich et al. (Eds.), *Proceedings of Sinn & Bedeutung 15* (1–21). Saarbrücken, Germany: Saarland University Press.

Roeper, T. (2011b). *Vacate phase and a symmetrical theory of movement: How the reconstruction of propositionality in child language reveals core grammar*. Ms., University of Massachusetts, Amherst, MA.

Roeper, Tom and Jürgen Weissenborn. 1990. How to make parameters work: Comments on Valian. Language processing and language acquisition, eds. Lyn Frazier and Jill de Villiers, 147-163. Dordrecht: Kluwer.

Ross, John R. (1973). The penthouse principle and the order of constituents. In C.T. Corum, T.C. Smith-Stark, & A. Weiser (Eds.), *You take the high node and I'll take the low node*, 397–422. Chicago: Chicago Linguistic Society.

Rowland, C. F., & Pine, J. M. (2003). The development of inversion in *wh*-questions: A reply to Van Valin, *Journal of Child Language*, *30*, 197–212.

Rowland, C. F., Pine, J. M., Lieven, E. V., & Theakston, A. L. (2005). The incidence of error in young children's *wh*-questions. *Journal of Speech, Language, & Hearing Research, 48*, 384–404.

Rowland, C.F. (2007). Explaining errors in children's questions, *Cognition, 104*, 106–134.

Rowland, C.F., & Pine, J.M. (2000). Subject-auxiliary inversion errors and *wh*-question acquisition: 'What children do know?', *Journal of Child Language*, *27,* 157-181.

Rutten, N.C.M.A. (1995). *Inversión en oraciones interrogativas Españolas: Un estudio sobre la accesibilidad de la GU para estudiantes de Español cuya lengua materna es el Holandés*. MA thesis, Utrecht University, Utrecht, The Netherlands.

Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language, 2*, 813–842.

Sarma, J. (1991). *The acquisition of* wh-*questions in English*. Doctoral dissertation, University of Connecticut, Storrs, CT.

Sorace, A. (1993). Incomplete vs. divergent representations of inaccusativity in native and non-native grammars of Italian, *Second Language Research, 9*, 22–47.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, *6*, 93–124.

Schmidt, R.W., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *"Talking to learn": Conversation in second language acquisition* (237–326). Rowley, MA: Newbury House.

Schönenberger, M. (2001). *Embedded V-to-C in child grammar: The acquisition of verb placement in Swiss German (Studies in Theoretical Psycholinguistics)*. Dordrecht: Kluwer.

Schwartz, B.D. (1998). The second language instinct. *Lingua*, *106*, 133–160.

Schwartz, B.D., & Sprouse, R.A. (1994). Word order and nominative case in nonnative language acquisition: A longitudinal study of (Ll Turkish) German interlanguage. In T. Hoekstra & B.D. Schwartz (Eds.), *Language acquisition studies in generative grammar: Papers in* *honor of Kenneth Wexler from the 1991 GLOW workshops* (317–368). Amsterdam: Benjamins.

Schwartz, B.D., & Sprouse, R.A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*, 40–72.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, *2*, 159–169.

Snyder, W. (2007). *Child language: The parametric approach*. New York: Oxford University Press.

Song, H.S., & Schwartz, B.D. (2009). Testing the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31(2)*, 323–361.

Spada, N., & Lightbown, P. (1999). Instruction, first language influence and developmental readiness in second language acquisition. *The Modern Language Journal, 83*, 1–22

Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries*. Doctoral dissertation, MIT.

Suñer, M. (1994). V-movement and the licensing of argumental *wh*-phrases in Spanish. *Natural Language & Linguistic Theory, 12(2)*, 335–372.

Taylor, B.P. (1975a). Adult language learning strategies and their pedagogical implications. *TESOL Quarterly, 9*, 391–399.

Taylor, B.P. (1975b). The use of overgeneralization and transfer learning strategies by elementary and intermediate students in ESL. *Language Learning, 25*, 73–108.

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

Thornton, R. (1991). *Adventures in long distance moving: The acquisition of complex* wh-*questions*. Doctoral dissertation, University of Connecticut, Storrs, CT.

Thornton, R. (2004). Why continuity. In A. Brugos, L. Micciulla, & C. Smith (Eds.), *Proceedings of BUCLD 28* (620–632). Somerville, MA: Cascadilla Press.

Thornton, R. (2008). Why continuity. *Natural Language & Linguistic Theory, 26*, 107–146.

Tiedeman, R. (1989). *Government and locality conditions on syntactic relations*. Doctoral dissertation, University of Connecticut, Storrs, CT.

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

Tomioka, S. (2009). Why questions, presuppositions, and intervention effects. *Journal of East Asian Linguistics, 18*(4), 253–271.

Toribio, J.A. (2000). Setting parametric limits on dialectic variation in Spanish. *Lingua, 110*, 315–341.

Tornyova, L., & Valian, V. (2009). The role of cross-linguistic variation in the acquisition of auxiliary inversion in *wh*-questions. In J. Crawford, K. Otaki, & M. Takahashi (Eds.), *Proceedings of GALANA 2008*, 282–290. Somerville, MA: Cascadilla Press.

Torrego, E. (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry, 15*, 103–130.

Truscott, J., & Smith, M. S. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition, 7*(1), 1-20.

Tsimpli, I.M. & Mastropavlou, M. (2007): Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language, 4*, 211–224.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270.

Ullman, M.T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative procedure model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (141–178). Washington, DC: Georgetown University Press.

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X-bar theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B.D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (265–316). Amsterdam: John Benjamins.

Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research, 12*, 7–39.

Vainikka, A., & Young-Scholten, M. (1996). The roots of syntax and how they grow: Organic grammar, the basic variety and processability theory". In Unsworth, S., T.Parodi, A. Sorace, & M.Young-Scholten (Eds.), *Paths of Development in L1 and L2 acquisition*, (77–106). Amsterdam, John Benjamins.

Vainikka, A. and M. Young-Scholten (2005). The roots of syntax and how they grow: Organic Grammar, the Basic Variety and Processability Theory. In S. Unsworth, T. Parodi, A. Sorace and M. Young-Scholten (Eds.), *Paths of Development*. Amsterdam: Benjamins.

Valian, V.V., & Casey, L. (2003). Young children's acquisition of *wh*-questions: The role of structured input. *Journal of Child Language, 30,* 117–143.

Valian, V.V., Lasser, I. & Mandelbaum, D. (1992). *Children's early questions*. Paper presented at *BUCLD 17*. Boston University, Boston, MA.

van Herk, G. (2000).The question question: Auxiliary inversion in early African American English. In S. Poplack (Ed.), *The English history of African American English* (175–197). Oxford: Blackwell Publishers.

van Valin, R.D. (2001). The acquisition of complex sentences: a case study in the role of theory in the study of language development. In J. Boyle, J.-H. Lee, & A. Okrent (Eds.), *CLS 36*, 14–39.

Westergaard, M., & Bentzen, K. (2007). The (non-)effect of input frequency on the acquisition of word order in Norwegian embedded clauses. In I. Gülzow & N. Gagarina (Eds.), *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept* (271–306). Berlin/New York: Mouton de Gruyter.

White, L. (1989). *Second language acquisition and Universal Grammar*. Amsterdam: John Benjamins.

White, L., Spada, N., Lightbown, P.M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, *12*(4), 416–432.

Williams, J. (1990). Another look at yes/no questions: Native speakers and non-native speakers. *Applied Linguistics, 11*(2), 159–182.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.
Yang, C. (2004). Universal Grammar, statistics, or both? *Trends in Cognitive Science, 8(10)*, 451–456.
Youhanaee, M. (2007). On the second language acquisition of L2 syntactic features and their projection domains. *Pazhuhesh-e zabanha-ye-Khareji, 32*(Special Issue: English), 187–210.
Zobl, H. (1979). Nominal and pronominal interrogation in the speech of adult francophone ESL learners: Some insights into the workings of transfer. *SPEAQ Journal, 3,* 69–93.
Zobl, H. (1995). Converging evidence for the 'acquisition-learning' distinction. *Applied Linguistics, 16*, 35–56.