



Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS
DEPARTAMENTO de FILOLOGÍA INGLESA
Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

WH- MOVEMENT IN CHILD ENGLISH

Cristina Arranz Gozalo

25 June 2015

Tutora: Raquel Fernández Fuertes

2014-2015

ABSTRACT

This dissertation deals with wh- movement and question formation, providing an empirical analysis of child data belonging to English L1 speakers from CHILDES in an attempt to study the acquisition of this syntactic phenomenon. The data are analyzed in terms of preference and complexity of wh- type and function of the wh- element, the complexity regarding pied-piping, adulthood, and the relation between the gradual acquisition of wh- movement (measured by the mean length of utterance) and the rate of correctness of the productions. The analysis of the data suggests that the usage and correctness of wh-movement in root questions closely mirrors the overall linguistic development along the acquisition process.

KEYWORDS: L1 acquisition, wh- movement, question formation, British English, corpus study, CHILDES

RESUMEN

Este trabajo trata sobre el movimiento wh- y la formación de preguntas, y en él se lleva a cabo un análisis empírico de datos de niños hablantes de inglés como primera lengua procedentes de CHILDES con objeto de estudiar la adquisición de este fenómeno sintáctico. Los datos son analizados en base a la preferencia y complejidad de cada forma wh y su función, la complejidad asociada al fenómeno denominado *pied-piping*, la madurez gramatical, y la relación entre la adquisición gradual del movimiento wh- (medida por la longitud media de los enunciados) y el índice de corrección de las producciones. Este análisis de datos sugiere que el uso y la gramaticalidad del movimiento wh- en las preguntas es un reflejo del desarrollo lingüístico a lo largo del proceso de adquisición.

PALABRAS CLAVE: Adquisición de la L1, movimiento wh-, formación de preguntas, inglés británico, estudio de corpus, CHILDES

Table of contents

1. Introduction.....	1
2. Literature review: theoretical and empirical accounts on wh- movement	2
2.1 Wh- movement: an overview	2
2.2 Types of wh- words.....	4
2.3 Functions of wh- words.....	6
2.4 Levels of complexity	8
2.5 Motivation for movement: the importance of features	8
2.6 Wh- movement and language acquisition	9
3. Research questions and hypothesis	13
4. Empirical analysis: wh- movement and question formation.....	15
4.1 Data selection.....	15
4.1.1 Corpus.....	15
4.1.2 Participants	17
4.2 Data classification criteria	20
4.3 Data analysis.....	22
4.3.1 Automatic searching: the CLAN programs	22
4.3.2 Grammatical analysis	24
5. Conclusion	37
6. Bibliography.....	39

[Note. To comply with the undergraduate dissertation regulations (part VII, section 10.1.a), the database elaborated to carry out this empirical analysis appears in a CD attached to the printed copy of the dissertation.]

1. Introduction

The different types of syntactic movement operations, and wh- movement in particular, have always been a matter of interest and, therefore, subjected to study among linguists. Although extensive research has been carried out throughout the past few decades with the aim to describe this phenomenon (not only from a theoretical perspective, but also through the analysis of child data in order to examine its acquisition by L2 (second language) speakers and compare it with other languages in cross-linguistic studies, it is true that there is not nearly as much investigation carried out in terms of L1 (first language) acquisition.

In an attempt to further explore the acquisition of wh- movement in child L1 English, and more specifically in British English, this dissertation presents an empirical study that aims to answer questions with regards to the nature of this syntactic phenomenon and its acquisition. The analyzed data provide information about the constructions preferred by children in terms of the following issues:

- The form and function of the wh- element involved.
- The complexity of this type of movement in terms of the preference of structures in which pied-piping is involved, and also the rate of non-adult-like versus adult-like productions.
- The relation between this complexity factor and the linguistic development as measured by the MLU of the children involved in the corpus used in this study.

The present dissertation is divided into four different parts. In the first section, wh-movement in questions is presented from a theoretical point of view, and so it is analyzed in terms of the typology of elements involved and their grammatical function, the complexity which the production of some of the structures under study entail, as well as the motivation for this type of syntactic movement. Furthermore, the process of acquisition of wh- movement is also addressed in this part, including references to previous empirical studies regarding data from both English L1 and L2 speakers. In the second section, the main research questions and the main hypothesis derived from these are presented as the elements guiding the empirical analysis that is carried out in this dissertation. In the third section, the empirical study is presented, including information

about the corpus selected and the participants, the software and programs used to analyze the data, as well as the data classification and closer examination. In the fourth and last section, the main conclusions reached after analyzing the data are laid out with a reference to the previous research questions posed and seeking for a confirmation for the initial hypothesis.

2. Literature review: theoretical and empirical accounts on wh-movement

2.1 Wh- movement: an overview

Wh- movement is a syntactic phenomenon that has interested a high number of linguists and grammarians, especially within the generative grammar tradition and, in particular, after Chomsky's theories on movement and Universal Grammar revolutionized the field of linguistics a few decades ago. Among his many works, one of the most relevant ones is *The Minimalist Program* (Chomsky, 1995), a collection of four articles that this linguist used to hypothesize, among many different issues, about the existence of a syntactic domain called phase and the capability of a constituent to move out of a phase (with the notion that two different phases form a sentence: VP and CP) as long as it first moves to the left part of the phase. This is referred to as the phase impenetrability condition (Chomsky, 1995).

In a few words, wh- movement can be defined as a compulsory syntactic movement in which a wh- constituent or wh- phrase moves to the front part of a clause or sentence (i.e. to the CP level). This type of movement can take place in relative clauses and questions, and the properties and motivations for the movement are the same in both types of structures. The difference between them has to do with their own grammatical properties. On the other hand, a relative clause, as shown in (1), is a dependent finite or non-finite clause, introduced by a relative pronoun and following a noun functioning as antecedent, which the relative clause modifies. On the other hand, questions involving wh- movement can be divided into two different types: direct and indirect questions, as shown in examples (2) and (3).

- (1) I know the person [**whom** Thelma will meet after lunch].
- (2) **What** does he do?
- (3) Do you know [**what** he does]?

Example (1) includes a relative clause introduced by the relative pronoun *whom*, the wh- element that moves from the direct object position of the verb *meet* to the specifier position of the CP level. It is, therefore, a declarative sentence, in which the speaker is giving information about the object of *know*, i.e. *the person*. This relative clause is functioning as the post-modifier of the person, and together with it, it makes up the direct object of *know*. The example in (2) includes a direct question introduced by the wh- pronoun *what*. The wh- element also moves to the specifier position of the CP level and, regarding its function, it is the direct object of the verb *do*. Example (3) involves an indirect wh- interrogative sentence, which is the direct object of the main verb *know*. Both examples (2) and (3) involve wh- questions and these demand a different type of answer than yes/no, that is, they involve a content answer in which the content of the wh- word (the direct object of *do*, in these two cases) is provided.

The grammatical representation of wh- movement in direct and indirect questions (examples 2 and 3) is done taking into account a series of movement theory concepts that will be briefly explained below (Haegeman and Guéron, 1999: 172).

- Moved element: the element which undergoes the movement operation, which in this case, as we are concerned with wh- movement, is the wh- element. This moved element is located in the so-called extraction-site.
- Trace: an empty category that occupies the syntactic position previously occupied by the moved element. This empty category is represented with a *t*.
- Extraction-site: the position from which the element is moved.
- Landing-site: the position to which the element is moved.
- Co-indexation: the way of connecting both elements involved in the wh- movement process (i.e. moved element and trace) as well as the positions affected by the movement operation (i.e. extraction site and landing site). This is done by assigning the same sub-index to both elements. As traditionally established, the first letter used as a sub-index is the letter *i*.

Taken these concepts as a starting point, the grammatical representation of the wh-movement operations in the examples (1), (2) and (3) above would be as in (4), (5) and (6):

- (4) I know the person [**whom**_i Thelma will meet t_i after lunch].
- (5) **What**_i does he do t_i?
- (6) Do you know [**what**_i he does t_i]?

The three examples above show wh- movement has taken place at two different levels: a main clause, as in (5); and in a dependent clause, as in (4) and (6). In these examples, the wh- elements *whom* and *what* have moved out from their extraction positions (as objects of their respective verbs) and have landed in the specifier position of the CP they belong to (their respective landing sites). A trace is left in the extraction site which is co-indexed with *whom* and *what*. This process illustrates the characteristics of wh- movement (Haegeman and Guéron (1999: 212):

- Movement leaves a trace.
- Movement is towards the left.
- Movement is towards the specifier position.

As explained in this section above, relative clauses and wh- questions share a number of characteristics, including wh- forms. Although the aim of the present dissertation is to focus on wh- movement in questions (from a theoretical and empirical perspective), it is important to point that wh- constituents in both (wh- questions and wh- relative clauses) are subjected to the same constraints (Haegeman and Guéron, 1999: 176).

2.2 Types of wh- words

According to Huddleston (1984), the interrogative words used in order to create questions are 9, and they are grouped as follows:

- Who, whom and whose: these three forms are in fact different instantiations of the pronoun *who* and they differ in their case specification: *who* is nominative, *whom* is accusative, dative or ablative and *whose* is genitive. *Who* and *whom* are usually fused in *who*, as the double option in example (7) suggests. Moreover, *who* constructions incorporate the [+ human] feature, in opposition to *what*

constructions. *Whose* can occur as a determiner in an NP as in (8), but it can appear with or without a following head noun, as in (9a) versus (9b).

- (7) a. **Who** are you talking to?
- b. **Whom** are you talking to?
- (8) **Whose** entry won the prize?
- (9) a. **Whose** team won?
- b. **Whose** won?

- What: it can either be a pronoun (as in 10), as a determiner in an NP (as in 11), or as a complement to the possessive clitic (as in 12). When appearing as a pronoun, *what* accounts for the [- human] feature (as in 10), in contrast with *who* constructions.

- (10) **What** caused that?
- (11) **What** book are you reading?
- (12) **What** schoolchild's imagination could fail to be simulated by such a challenge?

- Which: it can also be a pronoun (as in 13) or a determiner (as in 14) and it is neutral to the [+ human] contrast (as in 15).

- (13) **Which** do you want?
- (14) **Which** candidates do you support?
- (15) **Which** of the versions shall we use?

- When and where: both of these wh- types can occur as adverbs (as in 16) or pronouns as complements of a preposition (as in 17).

- (16) **When** did he arrive?
- (17) **Where** does he come from?

- Why: this wh- type appears as an adverb, as it can be seen in (18).

- (18) **Why** did she leave?

- How: it can occur either as a degree adverb modifying an adjective (as in 19) or as an adjective (as in 20), where it functions as a subject complement.

- (19) **How** big is it?
- (20) **How** are you?

2.3 Functions of wh- words

Huddleston (1984) also offers a classification of wh- words in terms of the function they play in the sentence. In particular, and taking the classification above in terms of wh- word types as a point of departure, he distinguishes the following functions of wh- elements, among others: subject, object, subject complement and adjunct. The distribution of functions in terms of wh- word types has been made in terms of the functions that can be found in child data, and it is as follows:

- Who, whom and whose: As explained in section 2.2, *who* and *whom* can be fused into *who*. *Who* usually functions as a subject (as in 21) or as the complement of a preposition (as in 22). *Whose* appears as a determiner of an NP, which can function as direct object (as in 23) or as subject complement.

(21) **Who** is she?

(22) **Whom** were they talking to?

(23) **Whose** wallet did they steal?

- What: this wh- type can either appear as a pronoun or determiner (as seen in section 2.2). When appearing as a pronoun, *what* functions as subject or direct object (as in 24 and 25). However, when it appears as a determiner, it is part of an NP which can function as subject complement (as in 26), direct object (as in 27) and adjunct (as in 28).

(24) **What** happened?

(25) **What** did you buy?

(26) **What** time is it?

(27) **What** movie are we watching?

(28) **What** time are they coming?

- Which: this wh- type can appear as a pronoun or as a determiner, and it can function as subject complement and direct object or as part of an NP which is a subject complement (as in 29) or a direct object (as in 30).

(29) **Which** one is John?

(30) **Which** one do you want?

- When, where, why and how: the main function of *when*, *where* and *why* is that of an adjunct (as seen in 31). However, *where* may also function as subject

complement (as in 32). Likewise, *how* may function as adjunct or as subject complement (as in 33 and 34).

- (31) **Why** are you going to the park?
- (32) **Where** are you?
- (33) **How** did you make the cake?
- (34) **How** are you?

Although there are several types of direct questions, they can be distinguished in terms of the form and the function of the *wh-* element, it is important to point out that subject questions are constructed in a different way than most *wh-* questions, as seen in (35) and (36).

- (35) [CP **who**_i [IP t_i annoyed him]
- (36) [CP **who**_i [IP did she marry to t_i]

In (20b) *who* functions as the subject of the clause, and the question is formed without an auxiliary verb. Instead, the tense of the verb is a simple past tense. In (21b), however, *who* functions as direct object of the main verb, and the question is formed – aside from moving the *wh-* element to the specifier position of CP – with the auxiliary *do* in past tense and the main verb in infinitive form (without *to*).

When discussing vacuous subject movement (VSM) in English and the fact that Chomsky (1986: 50) argued that “the language learner assumes that there is syntactic movement only where there is overt evidence for it”, Trotta (2004) concludes after carrying out an empirical analysis that this supposition implies that language learners are unaware of other analogous types of movement operations which, according to him “would lead them to a different (tacit) treatment of the structure in question” (Trotta, 2004: 15). Furthermore, he adds that there are no empirical data nor theoretical foundation which contradict the existence of a moved *wh-* subject. Apart from Trotta’s conclusions, it is important to point out that subjects are never originally in the CP level. Instead, they are in the IP level. That is, a *wh-* subject moves from the IP to the CP level, too.

2.4 Levels of complexity

When studying wh- movement in question formation, the level of complexity in terms, not only of the function played by the wh- element, but also in terms of the effects of movement must be taken into account. In particular, the effects of wh- movement are related to the phenomenon of pied-piping. This term refers to the fact that when a wh- word is moved, other elements within the phrase may be dragged along with the wh- element to the front. Pied-piping occurs when the wh- word is part of a prepositional phrase, as in (37a) or when the wh- word is a modifier within an NP, as in (38a). In the first case, pied-piping is optional, as the contrast between example (37a) and example (37b) shows. In the second case, pied-piping is obligatory, as the ungrammaticality of (38b) reflects.

- (37) a. **To whom**_i did you speak t_i?
b. **Whom**_i did you speak to t_i?
(38) a. **Which book**_i are you reading t_i?
b. ***Which**_i are you reading t_i book?

2.5 Motivation for movement: the importance of features

Carnie (2007: 362) explains that the motivation for wh- movement lies in the fact that wh- phrases move to the specifier position of CP to be near the [+WH] feature present in the head level of the CP (i.e. in C). In other words, they move in order to check this feature. Therefore, Interrogative sentences containing a wh- element have a [+WH] feature, and this [+WH] feature attracts any wh- constituent within the sentence, thus triggering the movement.

Furthermore, interrogative sentences have a [+Q] feature also located in C. However, other types of structures (e.g. relative clauses) have a [-Q] feature. As explained above wh- movement takes place when the wh- phrase moves to be near the [+WH] feature. This can take place both in questions ([+Q]) and in non-questions ([-Q], e.g. in relative clauses, in negative clauses, in statements).

As a result, interrogative wh- sentences have a double feature in the CP leave: [+Q] (as questions versus statements) and [+wh] (as wh- questions versus yes-no questions).

2.6 Wh- movement and language acquisition

Language acquisition has also been a subject of interest and study among many linguists, and extensive work has been done in regards to this topic also within the generative tradition.

When studying linguistic development, Yule (2014: 171-180) divided the language acquisition process into at least 5 different stages: the first one is referred to as babbling and it goes from 6 to 11 months of age; the second one, the one-word-stage, starts at 12 months and continues until 18 months; this second stage is followed by the two-word stage, which takes place at 18 – 20 months of age. Next, there comes the stage named as telegraphic speech, which extends from 21 to 24 months (until the second birthday). During the telegraphic speech stage, the child is able to create strings of words in sentences (e.g. this shoe all wet). The telegraphic speech stage is followed by the multiple-word stage that usually begins at the age of 2 or 3 years of age, depending on the children. Within the multiple-word stage, further morphological, syntactic and semantic developments take place. These developments and, therefore, the multiple-word stage, last until the child is five years old or even older than that until the child finally reaches the adult grammar in all linguistic domains.

Yule (2014: 177) expressed his interest in two particular grammatical structures acquired within the multiple-word stage that “seem to be acquired in a regular way by most English-speaking children” questions and negatives. As this dissertation focuses on question formation, the process of acquiring negatives will be left out. Therefore, the information below only refers to interrogative sentences.

Yule (2014) distinguished three different stages involved in the production of interrogative (and negative) sentences, occurring during the multiple speech stage and pertaining to syntactic development.

- First stage: two different procedures are involved in the first stage when building up interrogative sentences: the first one entails adding a wh- form at the beginning of the sentence, as seen in (39) and (40); and the second one implies a raise in intonation at the end of the sentence, as illustrated in (41) and (42), where no wh- form is present.

- (39) **Where** kitty?
- (40) **Where** horse go?
- (41) Doggie?
- (42) Sit chair?

- Second stage: this stage is characterized by the wider range of wh- expressions used to form questions, as shown in (43) and (44). It is also noticeable that the raising intonation is still used.

- (43) **What** book name?
- (44) **Why** you smiling?

- Third stage: in this stage, subject-verb inversion and *do* insertion appear for the first time, as examples (45), (47) and (48) show. Although subject-verb inversion or *do* insertion do not always happen in the production of questions at this stage, as there is alternation between adult-like forms (as in 47) and non-adult-like forms (as in 46, where the auxiliary verb *is* is missing), these productions are close to adult-like examples, as seen in (47) and (48).

- (45) Can I go?
- (46) **How** that opened?
- (47) **What** did you do?
- (48) Will you help me?

It can be therefore established that there is a gradual progression in the process of acquiring questions, which is determined by the types of constructions produced and the level of complexity of these constructions. Both of these factors can also be linked to the fact that, not only do children produce more complex questions, but also the rate of adult-like utterances is higher and higher.

It is important to keep in mind that this is a part of the whole language acquisition process, and that it does not occur in isolation, but simultaneously along with other morphological, syntactic and semantic developments. Moreover, these developmental stages show that language and, in particular, the grammatical properties of languages are acquired progressively by children acquiring their first language (L1).

Various studies have aimed to explore question formation in data from L1 speakers. Gavrusseva and Thornton (1999) focused on the acquisition of questions with wh-

possessor phrases, as seen in examples (49a), (49b) and (49c). These authors conducted an experimental analysis based on data from L1 English children. The 12 participants involved in this study (9 females and 3 males) were in day-care at the University of Maryland and their ages ranged from 4;5 to 6;0.

The type of *wh*- structure analyzed by Gavrusseva and Thornton (1999) is illustrated in (49).

- (49) a. **Whose** hat did he take?
b. ***Whose** did he take hat?
c. ***Who** did he take 's hat?

The example in (49a) shows the adult structure in which *wh*- movement of the *wh*- word (*whose*) has taken place and in which *wh*- movement requires the obligatory pied-piping of the possessed noun (*hat*). The *wh*- word is the possessor and as such it is morphologically marked (*whose*) and thus differs from the non-possessor counterpart (*who*) (i.e. *whose* in the genitive case, as explained in section 2.2 above). The lack of pied-piping in (49b and (49c), and the correct morphological marking in the *wh*- word in (49c) lead to non-adult-like productions.

The technique used to elicit the target structure involved the use of a puppet to engage children in a guessing game, and two researchers to carry out the two experiments. These two procedures required one of the researchers to act out short stories using a puppet, and the other to use a puppet that appeared at the end of each story to guess things about the stories. Children were expected to help the first researcher quiz the marionette after watching the stories. The researcher first gave the children a lead-in statement, and then he asked a question to the children. Approximately 20 questions with *whose* phrases were asked by each child during the two sessions in the study, with an interval of 2 – 3 weeks in between sessions. Each session entailed 6 to 8 short stories.

The data obtained were classified in terms of the three possessive structures illustrated in the example (49) above. This appears in table 1.

Table 1. Questions with extracted *who* and *whose* vs. adult-like questions

Children	Extraction of <i>who</i>	Extraction of <i>whose</i>	Adult-like	Total	Other
Mandy	24	1	0	25	10
Gab	7	-	3	10	7
Sandra	8	-	8	16	16
Jane	6	1	4	11	48
Matt	3	1	0	4	41
Peter	4	1	1	6	46
Meg	3	-	17	20	7
Tori	2	4	9	15	14
Kate	1	1	10	12	14
Sage	1	-	8	9	22
Tonya	-	2	11	13	10
Mary	-	-	13	13	15
Total	59 (38%)	11 (7%)	84 (55%)	154	250

Source: Gavrusseva and Thornton's (1999: 166) table 3.

By analyzing the data obtained in these sessions, as in table 1, they reached to two different conclusions: As the total row suggests, half of the *whose* structures produced by these 12 children were adult-like (55%), as in (49a). In the case of the non-adult-like cases, most corresponded to the movement of *who* (38%), as in (49c). This suggests that *whose* questions are more difficult than *who* questions, given their complexity associated to both morphological marking and pied-piping of the possessed noun.

Another study carried out by Van Valin (1998) aimed to give answers regarding the order of acquisition of *wh*- questions (in terms of *wh*- type and function) and also to explain the appearance of structures that children had never been exposed to before.

Van Valin's study is based on Stromswold's (1995) analysis regarding the production of *wh*- questions by 12 children, whose data were extracted from the CHILDES database (MacWhinney, 2000). Through the analysis of the data, Stromswold observed that object *wh*- questions start being produced earlier than subject *wh*- questions. Furthermore, she studied the emergence of long-distance questions in complex sentences and found that, regarding this, object questions start being produced earlier, too.

The first important conclusion reached by Van Valin is that, although subject *wh*- questions appear to be simpler than other types of *wh*- questions (since there is no subject-auxiliary inversion and they are formed by simply replacing the subject with a

wh- form), they are not the first type of wh- questions that children comprehend and produce. According to this author, the explanation for this lies in the fact that subject questions are narrow-focused (in contrast with object questions). In English, the unmarked focus position in the clause is the last position. Van Valin's explanation as to why subject questions are narrow-focused is based on the assumption that "children's first questions involving multiple-argument verbs contain simple transitive rather than ditransitive verbs, object position correlates with the least marked narrow focus position and subject position with the most marked narrow focus position" (Van Valin, 1998: 14).

The second conclusion reached by Van Valin is that Role and Reference Grammar (RRG) can account for the children's examples and structures that they have never been exposed to before. RRG assumes that "children are born with a rich cognitive endowment, which makes language learning and other types of learning possible" (Van Valin and LaPolla, 1997). On this account, Crain, Goro and Thornton (2005) state that "child languages can differ from the local adult language only in ways that adult languages can differ from each other" (Crain, 1991; Crain and Pietroski, 2002; Goodluck, 1991; Pinker, 1984) and that children are always trying out and testing constructions and can at any time be speaking a possible human language, which does not necessarily have to be a language they have been exposed to before or spoken around them.

The studies above suggest that (i) there is a developmental path in the acquisition of wh- questions by L1 English children so that is it not an all-or-nothing scenario; and (ii) during this acquisition process, there are non-adult-like structures that need to be accounted for and explained using general grammatical rules such as syntactic complexity.

3. Research questions and hypothesis

The present study deals with wh- movement in root questions in child English and its acquisition. It focuses on question formation in British English. Moreover, it aims to provide a close analysis regarding the acquisition of wh- movement in root questions

and the existence of this type of obligatory syntactic movement by using real data from children recorded in a naturalistic setting.

Taking as a point of departure the different studies presented in section 1 dealing with both the grammatical properties of wh- movement and the acquisition of this structure, this analysis strives to answer the following questions:

1. Is there an order of acquisition in terms of the type of wh- constituent used in root questions? Are all of them acquired around the same time and with the same level of easiness/difficulty?
2. Is there a preference with regards to the grammatical function of the wh- element within the interrogative sentence? If so, can we establish that wh- subject questions are more difficult for children to acquire because they are not formed using the regular pattern for question formation?
3. Does grammatical complexity as measured by the presence of pied-piping play a role in the acquisition process? If so, which structures are favored: those with or without pied-piping? Is this so in both optional and obligatory pied-piping contexts?
4. In the case of non-adult-like constructions, is language development and complexity playing a role in all the above cases? That is, in the case of development, do non-adult-like constructions appear in the initial stages of acquisition in all children? And, in the case of grammatical complexity, do complex wh- forms (as opposed to simple ones), complex wh- functions (as opposed to simple ones) and non-pied-piping (as opposed to pied-piping) have a higher error rate if compared to their less complex counterparts?
5. Is linguistic development along the acquisition process (as measured by the mean length of utterance (MLU)) closely related to the usage of wh- movement in root questions?

Based on the research carried out by previous authors and discussed in previous sections, the initial hypothesis of the present dissertation is that wh- movement and the different properties it entails (as outlined in the research questions above) would be gradually acquired and, therefore, there would be a correlation between MLU increase

and increase of general rate of correctness, regardless of grammatical complexity (in terms of form, function and pied-piping).

4. Empirical analysis on wh- movement and question formation

An empirical analysis has been carried out in an attempt to provide an answer to the research questions above and to test the hypothesis previously described. This analysis consists of several steps which will be presented next: a description of the corpus used to select data, information about the participants whose data form this corpus, the data classification criteria, and the data analysis.

4.1 Data selection

In order to study wh- movement and question formation in British English, a monolingual British English corpus has been selected. The necessary data have been extracted from a system called TalkBank, and more specifically, CHILDES (MacWhinney, 2000). TalkBank is a project that provides databases for communication study and research purposes. Coordinated by Brian MacWhinney, the online website for this project offers several different L1 and L2 browsable and downloadable databases. CHILDES (Child Language Data Exchange System), as its name already indicates, deals with child language, that is, it is the child component of TalkBank. The corpus chosen for this analysis was extracted from the Eng-UK-MOR section, and it includes transcriptions of the recorded conversational interactions of British children. This corpus is the Wells corpus.

4.1.1. The Wells corpus

As previously stated, the corpus used for this empirical study features data from British children, and it has been analyzed in its entirety. Under the name of Wells and directed by Gordon Wells and other colleagues, this corpus formed by 299 files contains data from 32 children. The age of the 16 male and 16 female children ranges from 1 year and 6 months to 5 years of age. The information that follows is taken from the database manual section available in CHILDES.

A summary of the main features of the Wells corpus appears in table 2.

Age range	# of participants	# of utterances	# of wh- sentences	# of wh- questions
1;6 to 5;0	32	33050	1783	702
		(100%)	(5.3%)	(2.1%)

In table 2, it is shown that out of the total number of utterances produced by these 32 children (33050), 2.1% of these are wh- questions (702), which are the ones under analysis in the present study.

The data compiled in the Wells corpus belong to the project entitled “The Bristol Language Development Study: language development in preschool children” (1973), in which children were recorded in a lifelike setting ten times at three-monthly periods. Moreover, all the samples are of 90 seconds of length, after which the recorder used to tape the conversations automatically stopped recording. Twenty four examples at twenty minute intervals between 9 a.m. and 6 p.m. were recorded, and neither the parents nor the child had knowledge about the exact time in which the samples would be recorded. This involved a complex logistic operation that included a microphone attached to the child’s clothes and a radio receiver, a tape recorder and a programmed timing mechanism. The microphone was placed when the child was getting dressed up in the morning and, as the reception was good in 100 meters, it allowed the child to move around with total freedom around the house. Apart from the obvious attempt to make the target child unaware of being taped, the span between recordings was irregular in order to avoid the parents planning any kind of activity beforehand.

Although this procedure is rather innovative in terms of a researcher not being present while the child is being recorded in order to make notes and comments about the context and so forth, this type of information was retrieved thanks to the parents. They were able to provide further details about the location, participants and context of the conversation after listening to the sample recorded the very same day. Therefore, potentially useful information in terms of context was not lost as one could have presumably assumed. Instead, these observations were as naturalistic as possible, considering the lack of planning and the fact that the child was completely unaware of being recorded.

Not only was this corpus selected because of the fact that it featured English-speaking L1 children, but also because of the age range, which happened to be optimal for this study, considering the fact that question formation and wh- movement appear at the multiple-word stage (which starts at the age of around 2 years; see section 2.6). As explained above, wh- elements begin to appear in the first stage of this process. However, it is not until the third stage of the process of question formation suggested by Yule (2014) when a different kind of syntactic movement, a head movement in the form of subject-auxiliary inversion is produced alongside wh- movement (Cook and Newson, 2007) and, therefore, the sentences produced by children can be classified as adult-like. Hence why the selected corpus is optimal to study the acquisition of wh- movement in L1 speakers: it covers a period in which the first questions emerge (between one year and a half and two years) and it goes until the age of five when adult-like production is expected.

4.1.2 Participants

In order to select the participants involved in this study, researchers approached the families of the approximately 1000 names drawn at random from the birth record held by the City Medical Officer. Needless to say, children whose parents refused to take part in this study were not considered. After that, they proceeded to rule out children whose parents were not English native speakers, children with any kind of disability and lastly, children who were in daycare full-time. Wells himself explains the reason as to why this selection was made:

“These categories were excluded, not out of any lack of interest in the problems that such children might be expected to encounter, but because their numbers in a sample of this size could not be expected to be large enough to permit meaningful comparisons to be made with the “normal” population” (Wells, 1981).

Once this selection was carried out, the participants as well as a list of reserves were chosen at random. The total number of participants added up to 129 children. However, many of these families dropped out of the project the following four years (Wells, 1985: 15) leaving a final number of 32 children being part of it.

Table 3 provides information about the participants in this corpus, which includes the following: the name, the age of the children in the first and last recording, the MLU of each child in the first and last sessions, the total number of utterances produced by each child, the total number of wh- sentences and the total number of wh- questions. The data in table 3 have been arranged in terms of the number of wh- questions, from highest to lowest.

Table 3. The Wells corpus: participants' information					
Name	Age range	MLU range	# of utterances	# of wh-sentences	# of wh- questions
Jonathan	1;6.5 - 4;7.14	1.307- 4.485	2109	215 (10.1%)	92 (4.3%)
Ellen	1;5.26 - 4;9.22	1.800 - 3.701	2277	140 (6.1%)	50 (2.1%)
Gavin	1;6.21 - 4;9.18	1.918 - 3.225	1567	127 (8.1%)	49 (3.1%)
Jack	1;5.26 - 4;9.1	1.649 - 4.064	1882	164 (8.7%)	44 (2.3%)
Iris	1;6.0 - 4;8.4	1.167 - 3.940	1201	76 (6.3%)	41 (3.4%)
Geoffrey	1;6.0 - 3;3.9	1.093 - 3.437	1136	78 (6.8%)	40 (3.5%)
Frances	1;6.1 - 4;10.8	1.068 - 3.090	1275	78 (6.11%)	39 (3%)
Abigail	1;5.28 - 4;8	1.417 - 4.827	1072	68 (6.3%)	29 (2.7%)
Elspeth	1;5.30 - 5;0.3	1.431 - 3.631	1132	66 (5.8%)	27 (2.3%)
Gary	1;6.0 - 4;9.0	1.777 - 3.743	1742	72 (4.1%)	26 (1.4%)
Gerald	1;6.6 - 4;9.5	1.216 - 4.389	1424	65 (4.5%)	25 (1.7%)
Debbie	1;6.9 - 1;11.25 ¹	1.359 - 4.098	1208	102 (8.4%)	25 (1.7%)
Darren	1;6.2 - 4;10.6	1.571 - 3.880	1461	52 (3.5%)	25 (1.7%)
Neville	1;5.25 - 3;5.27	1.426 - 4.380	981	55 (5.6%)	22 (2.2%)
Neil	1;6.04 - 3;6.01	1.250 - 4.329	617	29 (4.7%)	21 (3.4%)
Harriett	1;6.2 - 4;10.3	1.527 - 4.442	1240	49 (3.9%)	20 (1.6%)
Tony	1;5.26 - 3;6.8	1.250 - 3.213	594	38 (6.3%)	18 (3%)
Benjamin	1;5.21 - 5;0.24	1.376 - 4.418	1361	49 (3.6%)	18 (1.3%)
Betty	1;6.3 - 4;11.2	1.593 - 3.572	1178	30 (2.5%)	14 (1.1%)
Olivia	1;6.0 - 3;5.22	1.582 - 3.582	768	28 (3.6%)	10 (1.3%)
Penny	1;6.9 - 3;5.26	1.538 - 3.689	697	21 (3%)	9 (1.2%)
Jason	1;6.0 - 5;0.19	1.208 - 3.797	1153	37 (3.2%)	8 (0.6%)
Stella	1;6.8 - 3;5.30	1.200 - 4.267	648	31 (4.7%)	7 (1%)
Simon	1;5.21 - 3;5.22	1.087 - 2.729	425	12 (2.8%)	7 (1.6%)
Sheila	1;11.25 ² - 3;6.25	1.250 - 2.509	611	9 (1.4%)	7 (1.1%)
Rosie	1;5.29 - 3;6.11	1.167 - 3.333	240	15 (6.2%)	7 (2.9%)
Martin	1;5.26 - 3;5.28	1.574 - 2.944	616	22 (3.5%)	6 (0.9%)
Laura	1;6.1 - 3;6.2	1.219 - 4.495	823	23 (2.7%)	4 (0.4%)
Samantha	1;6.6 - 3;6.11	1.333 - 3.686	591	14 (2.3%)	1 (0.1%)
Lee	1;5.28 - 3;5.29	1.000 - 3.096	371	7 (1.8%)	1 (0.2%)
Sean	1;6.11 - 3;6.9	1.533 - 1.729	270	7 (2.5%)	0 (0%)
Nancy	1;6.2 - 3;3.03	1.875 - 3.800	380	4 (1%)	0 (0%)
TOTAL	1;5.21- 5;0.19	1.000 - 4.827	33050 - (100%)	1783 - (5.3%)	702 - (2.1%)

The information from the 32 children, as provided in table 3, shows that differences across children appear. For example in terms of the number of utterances produced, some children are more productive than others (e.g. Rosie's 240 utterances versus

¹ The age of the file debbie21 appears as 1;11.25? (with a question mark included) in both the file itself and in the British English manual available at the CHILDES website. However, the MLU of this file suggests that the age is erroneous, as it is virtually impossible to find an MLU rate of 4.094 at that age.

² Although the child's age in the file sheila02 is 1;11.25, the information found in the British English manual available at the CHILDES website (which contains information about the Wells corpus, among others) suggests that this age may be mistaken. Not only because it appears with an interrogative mark, but also because, taking into account the initial ages of the rest of the participants, it appears to be too high. Furthermore, it makes no sense that the age stated in the following file would be lower (sheila03's age – 1;9.2), since all the files within the corpus are chronologically ordered.

Ellen's 2277 utterances). This is part of the idiosyncrasy of each child. Likewise, table 3 also shows differences across children in terms of the number of wh- sentences produced (e.g. Nancy's 4 wh- sentences versus Jonathan's 2109 wh- sentences), therefore also being reflected in the number of wh- questions, which is the total number of wh- sentences minus the discarded examples, which include relative constructions and other examples discussed in section 3.2.

4.2 Data classification criteria

The data have been classified taking into account several different factors, which are described below. Considering the extensive number of participants analyzed (32) as well as the different grammatical criteria used, the data classification has been compiled in an excel sheet which appears in a CD alongside this document.

Due to the fact that this study focuses on wh- interrogative movement and language acquisition, adult data were neither classified nor analyzed. Therefore, only child data were examined.

The instances found in the data had to meet the following two requirements or else they were discarded: (i) show evidence of wh- movement in questions (wh- forms in isolation were, therefore, discarded, as it was not possible to detect whether movement has taken place or not, as in (50) and (51); and wh- movement in relative clauses was not considered either, as in (52); and (ii) not have been product of the imitation of an adult's previous utterance, or an already made sentence or unproductive language, such as Christmas carols and songs (as in 53).

(50) **What?** – Rosie09

(51) **Why** Mummie? – Tony07

(52) Look **what** I can do then – Geoffr09

(53) Ding dong bell the pussy's in the well **who** put her in? – Olivia09

Examples (50) and (51) show wh- questions in which the wh- elements (*what* and *why*) appear in isolation. Example (52) shows wh- movement in a relative clause. Example (53) shows a *who* construction that was part of the lyrics of a Christmas carol.

The criteria used in order to classify the data from the 32 children include the following variables: general inclusion criteria, wh- word types, wh- word function, pied-piping,

adult-like form, and MLU. All this considered, the total of instances analyzed was 702 (as in table 2).

- With respect to wh- types, nine wh- word forms were isolated: *what* (as in 54), *why* (as in 55), *when* (as in 56), *where* (as in 57), *how* (as in 58), *who* (as in 59), *whom* (as in 60), *whose* (as in 61) and *which* (as in 62).

- (54) **What** are you going to do? – Iris10
- (55) **Why** can't I do it? – Benjam06
- (56) **When** did you first say stop it then? – Jonath21
- (57) **Where** is it? – Debbie08
- (58) **How** do you do that? – Geoffr06
- (59) **Who** did that? – Frances07
- (60) **Whom** did you talk to?
- (61) And **whose** is that? – Ellen08
- (62) **Which** one shall I try? – Iris21

- In regards to wh- word function, six categories were distinguished: subject (as in 63), object (as in 64), subject complement (as in 65), adjunct (as in 66), determiner (as in 67) and premodifier (as in 68). Other functions mentioned in section 1.3 above do not appear in these data.

- (63) **Who** do that there? – Frances07
- (64) **What** did Daddy say? – Darren10
- (65) **What** is that? – Geoffr06
- (66) **Why** have you got to stay home? – Jason21
- (67) **Whose's** hammer is this? – Penny08
- (68) **Which** pub is it? – Neil10

- Concerning pied-piping, a division was made in terms of it being compulsory (as in 69) or optional (as in 70).

- (69) **Which** one is the highest? – Harrie10
- (70) To **whom** did you talk? / **Whom** did you talk to?

- The adult-like form is another property that was taken into account when classifying the data, dividing the selected data into adult-like examples (as in 71) and non-adult-like examples (as in 72).

- (71) **Who** bought that? – Nevill06
- (72) **Who** going to bath Iris? – Iris09

The reasons why a case may not be considered adult-like do not pertain to wh-movement itself, but rather to the grammar of the sentence: auxiliary verb missing (as in 73), subject missing (as in 74), lack of subject-verb agreement (as in 75), wrong verbal tense and wrong verbal tense form (as in 76) and incorrect use of prepositions (as in 77).

- (73) **What** you doing? – Penny04
- (74) **Where** go? – Darren08
- (75) **Where** is his feet? – Jonath10
- (76) **Who** taked that off? – Laura06
- (77) **Where**'s that to? – Darren09

This means that the operation of wh- movement along the terms described above (sections 2.1 and 2.5) was always done according to grammar in the data from these children.

- The MLU³ is an important factor in order to classify the data, as it measures linguistic development (Brown, 1973) and can, therefore, account for the increase of the rate of adult-like productions, which would explain the gradual process of acquisition of wh- questions, as hypothesized in section 2.

4.3 Data Analysis

This section is divided into two different parts: the first part deals with the computerized programs used to analyze the data, and the second part presents a grammatical analysis of the data regarding wh- type, wh- word function and complexity of the structure in terms of pied-piping, as well as adult-like forms and MLU correlations.

4.3.1 Automatic searching: the CLAN programs

In order to study the data from the corpus, the CLAN (Computerized Language ANalysis) programs available in the CHILDES project to analyze conversational data were used. The specific CLAN programs used to analyze data for the present study are the following: MLU, FREQ and KWAL. Each of these programs is briefly described below:

³ The mean length of utterance (MLU) is calculated by dividing the number of morphemes/words by the number of utterances produced. The MLU is used as an indicator of language development, as proposed by Brown (1973).

The MLU program calculates the mean length of utterance, that is to say, the ratio of morphemes or words per utterances. A typical MLU output appears in (76).

```
(78)
mlu +t*CHI @
*****
From file <c:\WELLS\Geoffrey\geoffr04.cha>
MLU for Speaker: *CHI:
Number of: utterances = 62, morphemes = 101
Ratio of morphemes over utterances = 1.629
Standard deviation = 0.9
```

The MLU output in (78) includes the following information about a chosen speaker (Geoffrey, in this case) within a certain file or files (file <geoffr04>, in this case which appears as @ in the syntax line <mlu +t*CHI @>): the number of utterances (62), the number of words/morphemes (101), the ratio of morphemes over utterances (i.e. the actual MLU value: 1.6) and the standard deviation (0.9).

The FREQ program outputs the frequency in which a certain word (or words) is used. This means that it counts the number of times a word appears in a file. A typical FREQ output is shown in (79).

```
(79)
freq +t*CHI +s"where" @
*****
From file <c:\WELLS\Geoffrey\geoffr04.cha>
Speaker: *CHI:
 1 where
-----
 1 Total number of different item types used
 1 Total number of items (tokens)
```

The FREQ output (in 79), provides the number of times the word *where* has been uttered by Geoffrey in the file <geoffr04> and this amounts to only 1 occurrence.

The KWAL program outputs utterances that match certain requirements stated by the program user through a word search. Moreover, this program provides the user with the opportunity to view the context in which the utterance has been produced. A typical KWAL output is shown in (80).

```

(80)
kwal +t*CHI +s"where" -w2+w2 @
*****
From file <c:\WELLS\Geoffrey\geoffr04.cha>
-----
*** File "c:\WELLS\Geoffrey\geoffr04.cha": line 810. Keyword: where
*RAC: was it (.)?
*CHI: xxx.
*CHI: where be be that?
*CHI: feet up
*CHI: xxx.

```

In (80) KWAL was used to provide the context in which *where* has been produced by Geoffrey. In file <geoffr04>, KWAL also allows to broaden the context of the word to the utterances preceding or following the utterance in which the target word appears. In (80) KWAL shows the two utterances before (-w2) and two utterances following (+w2) the target utterance. If more context is needed, the corresponding commands in the syntax line can be changed in order to be shown more utterances produced before or after it.

After the corresponding analysis, this automatic search provided with the following information: the total number of wh- words used that fitted the inclusion criteria (a total of 702 instances); the total number of wh- words per word type; the total number of wh- words per function type; the different instances to be classified in terms of adult-like or non-adult-like cases; and finally, the corresponding MLU values showing the linguistic development. This information is the basis for the grammatical analysis shown next.

4.3.2 Grammatical analysis

In this section, the data belonging to the 32 children in table 2 (section 3.1.2) is analyzed in terms of wh- movement in question formation and accounting for five different factors: the type of wh- element used, the function of the wh- element, the complexity in terms of pied-piping, the adulthood of the examples and the MLU value. Apart from the analysis of all the participants altogether (as seen from table 4 to table 7), 6 children (those with the highest number of wh- questions produced, as shown in table 2 above) have been selected in order to take a closer look at the production

preferences of each one of them. The data belonging to these 6 selected children is shown in tables 8 to 13.

In the case of the data from the 32 children collapsed, table 4 shows a classification in terms of wh- words and distinguishing between adult-like and non-adult-like cases.

Table 4. Per wh- type – all participants										
	What	Why	When	Where	How	Who	Whom	Whose	Which	Total
Adult-like	170	63	10	51	25	29	0	4	8	360
Non-adult-like	139	36	7	94	23	30	0	4	9	342
Total	309 (44%)	99 (14.1%)	17 (2.4%)	145 (20.6%)	48 (6.8%)	59 (8.4%)	0	8 (1.1%)	17 (2.4%)	702

As the total row in table 4 shows, there is a preference for wh- questions with the wh- form *what* (44%), followed by *where* (20.6%). Moreover, it is important to point out that although structures with the form *who* entail a higher level of complexity (as they are not formed in the same manner most wh- questions are), they have a higher rate of production than other wh- words (e.g. *who*'s 8.4% versus *when*'s 2.4%). Furthermore, *whose* and *which* structures (which involve a higher level of difficulty in terms of pied-piping) are also less favored than other types of wh- questions (e.g. *whose*'s 1.1% versus *how*'s 6.8%). In the case of the accusative form of *who* (i.e. *whom*) no cases have been found in the data and in the case of *whose* they amount to 1.1% of the cases. This suggests that, among the different *who* forms (as explained in section 2.2 above) the nominative form *who* is the most frequent one (8.4%), possibly subsuming different grammatical functions (e.g. subject, object). With respect to correctness, the fact that the rate of adult-like productions is somehow higher than the non-adult like utterances, as seen in the last column, is also important as it shows that, in this respect, the adult grammar is being implemented. However, the non-adult-like forms still amount to 49% of the total number of cases, and therefore, a closer look at the non-adult-like forms will be shown later.

Table 5 shows a classification of the functions of the wh- elements in the wh- questions produced by the 32 children collapsed, distinguishing between adult-like and non-adult-like cases.

Table 5. Per wh- function – all participants				
	Subject	Object	Adjunct	Subject complement
Adult-like	38	110	117	104
Non-adult-like	22	100	82	129
Total	60 – (8.5%)	210 – (29.9%)	199 – (28.3%)	233 – (33.1%)

Table 5 shows that, although the subject complement function is the most favored function (33.1%), it is closely followed by the object (29.9%) and adjunct (28.3%) functions. The subject function is the less favored function (8.5%), which can account for the fact that it is more difficult to produce since its formation differs from the typical formation of wh- questions.

Table 6 deals with the wh- elements included in the non-adult-like wh- questions produced by the 32 participants collapsed. More specifically, table 5 considers all the non-adult-like cases to illustrate which wh- form is more problematic.

Table 6. Non adult-like instances per wh- type – all participants									
	What	Why	When	Where	How	Who	Whom	Whose	Which
Non-adult-like cases	139	36	7	94	23	30	0	4	9
% per wh- type	44.9%	36.3%	41.1%	64.8%	47.9%	50.8%	0%	50%	52.9%

Table 6 shows that the most problematic wh- questions in terms of the wh- element involved are *where* constructions (64.8%). Although the fact that they are constructed as most wh- questions could suggest that they should not be more problematic than other types of wh- questions (e.g. *when*'s 41%), this could be explained by the fact that *where* constructions are one of the first ones that are produced as well as one of the wh- forms more frequently used (see table 3 above). This means that they start being uttered when the MLU is rather low (e.g. Rosie's 'where bike?' at 2;0.13 in file <rosie04>, with an MLU rate of 1.444), which indicates that the child's grammar has yet to be developed. *Who*, *whom*, *whose* and *which* constructions also deserve attention because they appear to be problematic in relation to other constructions (e.g. *which*'s 52.9% versus *why*'s 36.3%). The higher percentage of non-adult-like productions can be explained by the fact that these constructions are more complex than the rest: *which* (52.9%) and *whose* (50%) constructions have the additional intricacy that pied-piping entails, and *who* (50.8%) wh- questions are constructed in an uncommon manner, as explained in section

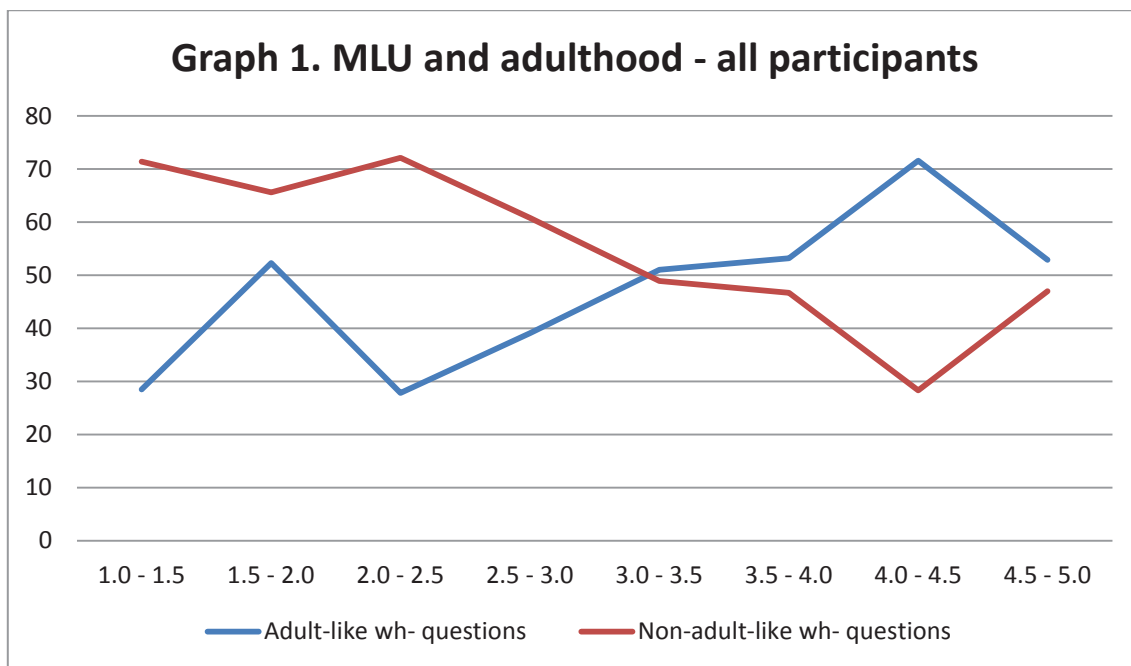
2.4. This could, therefore, account for the higher rate of non-adult-like productions in *who*, *whose* and *which* constructions. In addition to this, and to account for the lack of *whom* cases, it is necessary to point out that *who* and *whom* constructions are interchangeable except for the subject function cases (as seen in section 2.4). The least problematic construction is that involving *why* (36.3%) which is one of the less frequent wh- forms (as in table 4 above).

The complexity aspect in terms of pied-piping is dealt with in table 7, which accounts for the wh- questions that featured pied-piping produced by the 32 children collapsed, distinguishing between adult-like and non-adult-like cases.

	Compulsory	Optional
Adult-like	23 - (54.7%)	0
Non-adult-like	19 - (45.2%)	0
Total	42 - (100%)	0

Table 7 shows that, although compulsory pied-piping is favored in all the participants' productions, the rate of wh- questions which feature pied-piping in relation to the total number of wh- questions produced is extremely low (42 out of 702, 5.9%). This can account for the fact that those structures featuring pied-piping are more complex than the rest, which would also explain the small difference between the rate of adult-like (54.7%) and non-adult-like productions (45.2%). It could also be the case that these structures are not only more complex but possibly less frequent overall. What is interesting about the data in table 7 is also the fact that pied-piping is only used in compulsory contexts, which again points to its grammatical complexity.

Another aspect this grammatical analysis is based on is the MLU. In this respect, graph 1 deals with the mean length of utterance (MLU) in relation to the adulthood of the production of all the 32 participants in this corpus. This graph aims to analyze whether the MLU is a determining factor in the production of adult-like and non-adult-like cases.



Graph 1 shows that there is a clear supremacy of non-adult-like productions when the MLU rate is at its lowest point (28.5% of adult-like cases versus 71.4% of non-adult like cases when the MLU is between 1.0 and 1.5). Nevertheless, this primacy is slowly curtailed (39.1% of adult-like cases versus 60.8% of non-adult-like cases when the MLU is between 2.5 and 3.0) and finally inverted (71.6% of adult-like cases versus 28.3% of non-adult-like cases when the MLU is between 4.0 and 4.5) .

As stated in the introductory paragraph of this section, a selection of 6 children was made out of the 32 participants involved in Wells' study. The chosen children were those who produced the highest number of wh- questions (see table 2 above). The data belonging to every one of these 6 children are presented below in table 8, and the analysis of their data includes the same properties that have been used in the collapsed data: the different types and functions of wh- constituents and the percentage of each wh- element in relation to the total number of utterances produced, the number of adult-like and non-adult-like utterances and the complexity of the wh- questions in terms of pied-piping. In addition to this, the MLU has been analyzed in relation to the adulthood of the wh- productions.

Table 8 offers an overview of the data selection, including the total number and percentage of wh- questions produced by each child, as well as the total number and percentage of both the adult-like and non-adult like cases found.

	Age range	Wh- questions	Adult-like cases	Non-adult-like cases
Jonathan	1;6.5 - 4;7.14	92 - (100%)	73 - (79.3%)	19 - (20.6%)
Ellen	1;5.26 - 4;9.22	50 - (100%)	16 - (32%)	34 - (68%)
Gavin	1;6.21 - 4;9.18	49 - (100%)	23 - (46.9%)	26 - (53%)
Jack	1;5.26 - 4;9.1	44 - (100%)	20 - (45.4%)	24 - (54.5%)
Iris	1;6.0 - 4;8.4	41 - (100%)	12 - (29.2%)	29 - (70.7%)
Geoffrey	1;6.0 - 3;3.9	40 - (100%)	31 - (77.5%)	9 - (22.5%)
Total		316 - (100%)	175 - (55.3%)	141 - (44.6%)

In table 8, it can be seen that, although in some particular cases the percentage of non-adult-like cases is higher than the adult-like ones (e.g. Iris's 70% of non-adult-like production versus her 29.2% of adult-like cases), the total percentages suggest that the general tendency for all the children analyzed in this selection is to produce more adult-like productions than non-adult ones. As already mentioned in the analysis of the 32 children above in this section (table 4), this suggests that the adult grammar is being implemented. However, as it also happens in the case of the 32 children, in this selection the same tendency appears: the adult-like cases outnumber the non-adult-like ones but the difference is not actually meaningful. In addition to this, the fact that the data belonging to older children (Ellen's age range goes from 1;5.26 to 4;9.22, and Jack's age range goes from 1;5.26 to 4;9.1) do not show a higher adult-like production than those belonging to younger children (Geoffrey's age range goes from 1;6.0 to 3;3.9) reflects that the year and a half difference does not actually play a role. In the same way, the highest percentage of non-adult-like cases does not belong to the production of the child with the smallest age range.

Table 9 shows the numbers and percentages of each wh- type used to account for the preference of wh- type in the wh- questions produced by these 6 children.

	What	Why	When	Where	How	Who	Whom	Whose	Which
Jonathan	48 - (52.1%)	20 (21.7%)	1 (1%)	11 (11.9%)	6 (6.5%)	3 (3.2%)	0	0	3 (3.2%)
Ellen	14 (28%)	5 (10%)	5 (10%)	16 (32%)	3 (6%)	5 (10%)	0	2 (4%)	0 (0%)
Gavin	35 (71.4%)	2 (4%)	1 (1%)	4 (8.1%)	3 (6.1%)	4 (8.1%)	0	0	0 (0%)
Jack	28 (63.6%)	4 (9%)	2 (4.5%)	7 (18.1%)	1 (2.2%)	2 (4.5%)	0	0	0 (0%)
Iris	8 (19.5%)	9 (21.9%)	0	6 (14.6%)	8 (19.5%)	5 (12.1%)	0	0	5 (12.1%)
Geoffrey	15 (37.5%)	7 (17.5%)	3 (7.5%)	6 (15%)	5 (12.5%)	4 (10%)	0	0	0
Total	148 (46.8%)	47 (14.8%)	12 (3.7%)	50 (15.8%)	26 (8.2%)	23 (7.2%)	0	2 (0.6%)	8 (2.5%)

Table 9 shows that, despite particular cases (e.g. Ellen’s 32% of *where* constructions versus her 28 % of *what* constructions), the general wh- type constructions of choice are *what* constructions (46.8%), which shows the same tendency as in the overall data distribution (table 6 above). The extremely low percentages of *whose* and *which* constructions in comparison with the second most preferred wh- type (*why*’s 14.8%) can be explained by the level of complexity that constructions that feature pied-piping entail, as already mentioned above in this section.

Table 10 deals with the wh-types used in all the wh- questions produced by the 6 chosen participants in relation to adulthood, aiming to give an answer to which wh- elements can be considered more and less problematic.

		Jonathan	Ellen	Gavin	Jack	Iris	Geoffrey	Total
What	Adult-like	39 (81.2%)	5 (35.7%)	17 (48.5%)	12 (42.8%)	3 (37.5%)	12 (80%)	88 (59.4%)
	Non-adult-like	9 (18.7%)	9 (64.2%)	18 (51.4%)	16 (57.1%)	5 (62.5%)	3 (20%)	60 (40.5%)
Why	Adult-like	17 (85%)	3 (60%)	2 (100%)	3 (75%)	3 (33.3%)	5 (71.4%)	33 (70.2%)
	Non-adult-like	3 (15%)	2 (40%)	0	1 (25%)	6 (66.6%)	2 (28.5%)	14 (29.7%)
When	Adult-like	1 (100%)	1 (20%)	0	2 (100%)	0	3 (100%)	7 (58.3%)
	Non-adult-like	0	4 (80%)	1 (100%)	0	0	0	5 (41.6%)
Where	Adult-like	9 (81.8%)	2 (12.5%)	1 (25%)	1 (14.2%)	1 (16.6%)	3 (50%)	17 (34%)
	Non-adult-like	2 (18.1%)	14 (87.5%)	3 (75%)	6 (85.7%)	5 (83.3%)	3 (50%)	33 (66%)
How	Adult-like	4 (66.6%)	0 (0%)	3 (100%)	1 (100%)	0	4 (80%)	12 (46.1%)
	Non-adult-like	2 (33.3%)	3 (100%)	0	0	8 (100%)	1 (20%)	14 (53.8%)
Who	Adult-like	2 (66.6%)	3 (60%)	0	1 (50%)	1 (20%)	4 (100%)	11 (47.8%)
	Non-adult-like	1 (33.3%)	2 (40%)	4 (100%)	1 (50%)	4 (80%)	0	12 (52.1%)
Whom	Adult-like	0	0	0	0	0	0	0
	Non-adult-like	0	0	0	0	0	0	0
Whose	Adult-like	0	2 (100%)	0	0	0	0	2 (100%)
	Non-adult-like	0	0	0	0	0	0	0
Which	Adult-like	1 (33.3%)	0	0	0	4 (80%)	0	5 (62.5%)
	Non-adult-like	2 (66.6%)	0	0	0	1 (20%)	0	3 (37.5%)

The results shown in table 10 suggest that the less problematic wh- questions are *why* constructions (70.2% of adult-like cases versus 29.7% of non-adult-like cases). Although the difference between adult-like and non-adult-like cases is not as high as in the case of *why* constructions, wh- questions with *what* and *when* cannot be considered to be especially problematic. However, *where*, *how* and *who* constructions show a higher percentage of non-adult-like cases in relation to the adult-like cases. As mentioned above, this could be explained by the fact that *who* questions are constructed in a different way than most wh- questions. Moreover, the fact that some wh- elements begin to appear before others (when the MLU is still low) can account for the higher

percentage of non-adult-like cases of *where* and *how* constructions. In addition to this, it is relevant to mention that, although the percentage of adult-like *whose* and *which* questions is higher than the non-adult-like ones, the number of wh- questions with these two wh- elements is extremely low in comparison with other wh- constituents, which could suggest that these two types of wh- questions are more complex than others, and therefore less frequent overall. These results confirm the general tendency already discussed above when the results from the 32 selected children were analyzed (see table 6).

Table 11 deals with the preferred function of wh- element in the wh- questions produced by the 6 selected children, including the total number of wh- questions belonging to each function, as well the percentage of each grammatical function.

Table 11. Per wh- function – all 6 participants				
	Subject	Object	Adjunct	Subject Complement
Jonathan	4 - (4.3%)	26 - (28.2%)	34 - (36.9%)	28 - (30.4%)
Ellen	5 - (10%)	12 - (24%)	15 - (30%)	18 - (36%)
Gavin	4 - (8.1%)	30 - (61.2%)	4 - (8.1%)	11 - (22.4%)
Jack	2 - (4.5%)	15 - (34%)	9 - (20.4%)	18 - (40.9%)
Iris	5 - (12.1%)	7 - (17%)	21 - (51.2%)	8 - (19.5%)
Geoffrey	4 - (10%)	12 - (30%)	14 - (35%)	10 - (25%)
Total	24 - (7.5%)	102 - (32.2%)	97 - (30.6%)	93 - (29.4%)

Table 11 shows that the wh- function of choice overall is the object function (32.2%), closely followed by the adjunct (30.6%) and subject complement (29.4%) functions. The less preferred grammatical function is the subject function (7.5%), which can be explained by the fact that, as already mentioned before, *who* questions are not constructed in the same way most wh- questions are formed, which implies an additional complexity. These results, however, show a slightly different tendency in terms of preference when compared to the results belonging to the 32 children (see table 5). Table 5 shows that the subject complement function is the most favored function (33.1%), and the second most favored function is the object function (29.9%). Nevertheless, the results in table 11 suggest that the supremacy of the subject complement function is not relevant, as it can be seen that, in the case of the 6 selected children, the object function surpasses the subject complement function.

Table 12 deals with the functions of the *wh*- elements in the *wh*- questions produced by the 6 selected children in relation to adulthood, with the aim to state which functions are less and more problematic.

	Subject		Object		Adjunct		Subject Complement	
	Adult-like	Non-adult-like	Adult-like	Non-adult-like	Adult-like	Non-adult-like	Adult-like	Non-adult-like
Jonathan	3 (75%)	1 (25%)	19 (73%)	7 (26.9%)	29 (85.2%)	5 (14.7%)	22 (78.5%)	6 (21.4%)
Ellen	3 (60%)	2 (40%)	4 (33.3%)	8 (66.6%)	5 (33.3%)	10 (66.6%)	4 (22.2%)	14 (77.7%)
Gavin	0	4 (100%)	15 (50%)	15 (50%)	3 (75%)	1 (25%)	5 (45.4%)	6 (54.5%)
Jack	1 (50%)	1 (50%)	7 (46.6%)	8 (53.3%)	5 (55.5%)	4 (44.4%)	7 (38.8%)	11 (61.1%)
Iris	1 (20%)	4 (80%)	4 (57.1%)	3 (42.8%)	6 (28.5%)	15 (71.4%)	1 (12.5%)	7 (87.5%)
Geoffrey	4 (100%)	0	9 (75%)	3 (25%)	12 (85.7%)	2 (14.2%)	6 (60%)	4 (40%)
Total	12 (50%)	12 (50%)	58 (56.8%)	44 (43.1%)	60 (61.8%)	37 (38.1%)	45 (48.3%)	48 (51.6%)

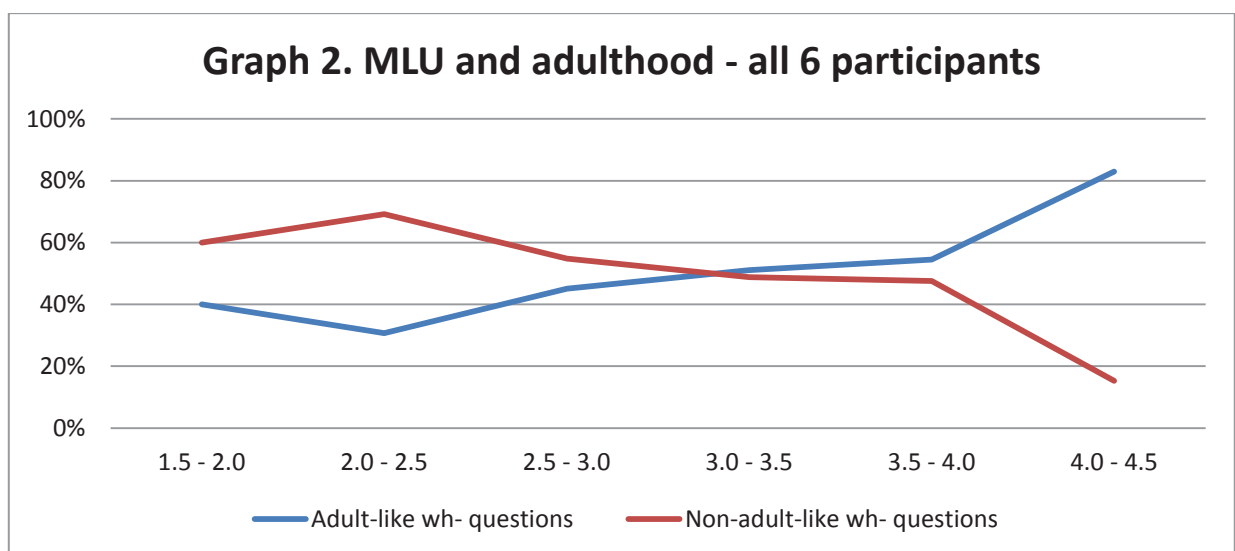
The results in table 12 suggest that the less problematic function is the adjunct function (61.8% of adult-like cases versus 38% of non-adult like cases). The second less problematic function is the object function, although the difference between adult-like and non-adult like cases is significantly smaller (56.8% of adult-like productions versus 43.1% of non-adult-like productions). On the contrary, the most problematic grammatical function is the subject complement function (51.6% of non-adult-like cases versus 48.3% of adult-like cases), closely followed by the subject function (50% for both adult-like and non-adult like productions). This can be explained by the fact that both of these grammatical functions feature *who* constructions (e.g. Iris' 'who got hair cut?' in file <iris07>; and Betty's 'who are you?' in file <betty21>). As previously argued, *who* constructions are problematic because they are not constructed as most *wh*-questions are. When compared to the analysis of the data belonging to the 32 children, it can be seen that they follow the same tendency: the most problematic function is the subject complement function, and the less problematic one is the adjunct function. That being said, it is relevant to mention that, the subject function is slightly more problematic for these 6 children. Hence why the general tendency suffers a slight change in this respect.

Table 13 deals with the aspect of complexity considering the wh- questions that feature pied-piping produced by the selected 6 children. Moreover, it addresses pied-piping in relation to adulthood.

	Adult-like cases	Non-adult-like cases	Total
Jonathan	3 - (50%)	3 - (50%)	6 - (100%)
Ellen	1 - (50%)	1 - (50%)	2 - (100%)
Gavin	1 - (50%)	1 - (50%)	2 - (100%)
Jack	0	0	0
Iris	4 - (57.1%)	3 - (42.8%)	7 - (100%)
Geoffrey	0	1 - (100%)	1 - (100%)
Total	9 - (50%)	9 - (50%)	18 - (100%)

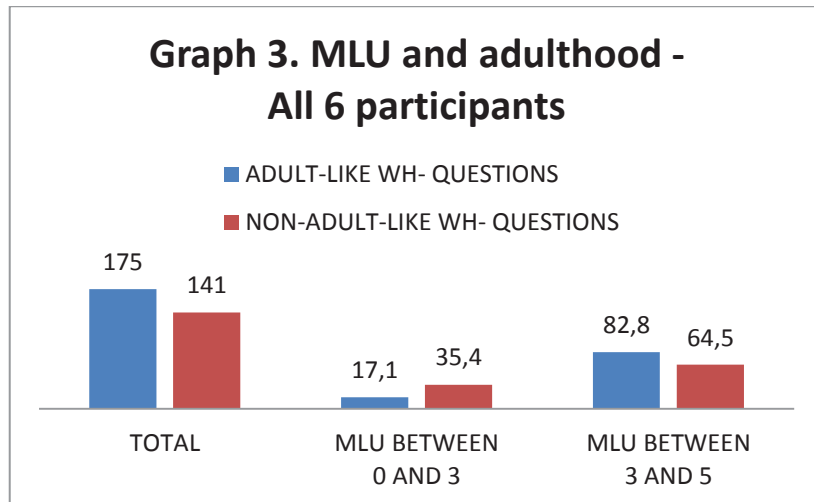
As in table 7 above, no cases of optional pied-piping were found in the data. The results in table 13 suggest that: pied-piping is not frequent at all in the data selected; moreover, these results show that, except for one particular case (Iris' 57.1% of adult-like cases versus her 42.8% of non-adult-like cases), overall adult-like and non-adult-like productions share the same percentage, which leads to believe that wh- questions featuring pied-piping are complex and problematic for these children, and that they are not frequent at all, something that is corroborated by the low number of cases found, too.

Graph 2 deals with another aspect this grammatical analysis is based on: the MLU. More specifically, this graph shows the relation between the MLU value and the rate of adult-like and non-adult-like cases found in the production of wh- questions of the 6 selected children.



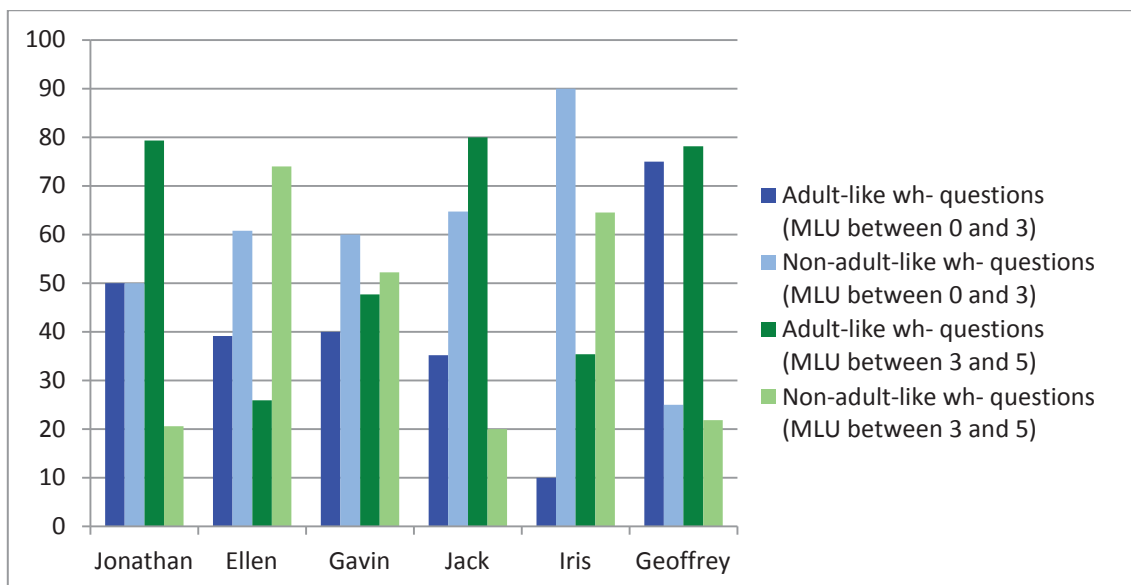
Graph 2 shows that there is a direct relation between the increase of the MLU rate and the rate of adult-like cases. This means that, as children's linguistic development progresses (and the MLU rate consequently rises), there is a change in their production in terms of adulthood: from a higher percentage of non-adult-like productions, when MLU values are between 1.5 and 3, to a higher percentage of adult-like productions, from MLU 3 upwards. More specifically, in the case of the 6 selected children, when the overall MLU is between 1.5 and 2.0 (the initial stage), there is a 40% of adult-like cases versus a 60% of non-adult like cases. This pattern continues (although being slightly altered) until the MLU rate is between 3.0 and 3.5, when there is a 51.10% of adult-like productions versus a 48.80% of non-adult-like productions. This change of pattern in which adult-like forms outnumber non-adult-like ones becomes bigger and bigger as time progresses (54.50% of adult-like cases versus 47.60% of non-adult like cases when the MLU rate is between 3.5 and 4.0; and a 83% of adult-like productions versus a 15.30% of non-adult-like productions when the MLU rate is between 4.0 and 4.5). As already mentioned above in this section, this progressive change in terms of adulthood of the productions suggests that the adult's grammar is being implemented in the case of wh- question formation, hence the constant reduction of non-adult-like cases. These results are in line with those belonging to the 32 children already discussed above (see graph 1).

Graph 3 illustrates the change in the production in terms of adulthood and in relation to the MLU rate already addressed in graph 1 above. Graph 3 accounts for the number of adult-like and non-adult like cases found in the wh- questions produced by the 6 selected children, separating them in terms of the MLU rate.



Graph 3 shows the change already mentioned in graph 2 in terms of adulthood in relation to the MLU rate, accounting for the total number of adult-like and non-adult like cases found. This graph shows that, when the MLU rate is between 0 and 3, there is a higher number of non-adult like cases (17.1% of adult-like wh- questions versus 35,4% of non-adult-like wh- questions), although the difference between adult-like and non-adult-like productions is not extremely high. However, when the MLU is between 3 and 5, there is a change in the production, adult-like cases being more frequent than non-adult-like cases (82.8% of adult-like wh- questions versus 64.5% non-adult-like wh- questions). Again, this suggests that the increase in the MLU rate is directly related to the higher rate of adult-like productions, as already seen in the case of the 32 children (see graph 1).

Graph 4 deals with the production of each one of the 6 selected children, dividing their total production of wh- questions into several different categories: adult-like and non-adult-like wh- questions when the MLU rate is between 0 and 3, as well as adult-like and non-adult-like wh- questions when the MLU rate is between 3 and 5.



The results shown in graph 4 confirm the general overview on this issue shown in graphs 1 and 2. In most cases, the percentage of non-adult-like cases (in lighter color) was higher than the percentage of adult-like cases when the MLU rate is between 0 and 3 (with the exceptions of Geoffrey's 75% of adult-like cases versus his 25% of non-adult-like cases, and Jonathan's 50% of adult-like productions versus his 50% of non-adult-like productions). When the MLU rate is between 3 and 5, there is a change in terms of the adulthood of the productions (e.g. Jonathan's 79.3% of adult-like wh- questions versus his 20.6% of non-adult-like wh- questions, Geoffrey's 78.1% of adult-like productions versus his 21.8% of non-adult-like productions, and Jack's 80% of adult-like productions versus his 20% of non-adult-like productions). This change can also be accounted for in the smaller gap between adult-like cases and non-adult like cases in the production of those children in which the overturn is yet to take place (e.g. Gavin's 47.7% of adult-like cases versus his 52.2% of non-adult-like cases). In fact, this shows what is really relevant about these results: the fact that not every child's linguistic and grammatical development take place at the same pace.

5. Conclusions

The present dissertation has offered an empirical analysis of wh- movement and question formation in order to explore the nature of this grammatical phenomenon in relation to its acquisition. In order to do so, British English child data from the Wells corpus in CHILDES have been selected and analyzed. This analysis has led to a number

of conclusions being reached in relation to the research questions and the hypothesis presented in section 3.

The data analyzed show that, although all types of *wh*- constituents start being used around the same time, there are clear preferences. It can be established that these preferences have to do with the different level of complexity of the different *wh*- questions (the most complex *wh*- structures being the less frequent ones). This complexity can derive from the way in which questions are formed, or from pied-piping. The evidence for this lies in the fact that more complex *wh*- constructions have a higher rate of non-adult productions than others, and are less frequent in production than others.

This analysis also reveals that, although the preferred grammatical function of the *wh*- element within the interrogative sentence is the subject complement function, this preference is closely followed by the object function. However, what is interesting in this regard is the fact that the subject function is the less preferred one. Once again, this points to the fact that *who* subject constructions are less produced because *wh*- subject questions are more complex and difficult for children to acquire, considering that they are not formed following the regular pattern for question formation.

Grammatical complexity as measured by the presence of pied-piping does play a role in the acquisition process, indeed. *Wh*- questions featuring pied-piping are almost nonexistent across the overall production and are only used in compulsory contexts. This leads to believe that not all *wh*- types and functions are acquired at the same time and in the same way.

Furthermore, language development and complexity play a role in the non-adult-like constructions. There is a higher percentage of non-adult-like cases in the initial stages of acquisition in most children. However, not every child develops at the same pace. With respect to complexity, complex *wh*- forms and functions, as well as structures featuring pied-piping do not always have a higher error rate than their less complex counterparts. However, the difference in the rate between adult-like and non-adult-like cases is not as big as in simpler structures. Moreover, they are less frequent than their less complex counterparts, which can also account for how *wh*- questions are acquired.

Linguistic development along the acquisition process (measured by the MLU rate) is indeed closely related to the usage of wh- movement in root questions. This can be explained, not only by the obvious relation between the increase of the MLU rate and the age growth, but also by the drastic change that the child's production undergoes: there is an overturn regarding the adulthood of the productions, the supremacy of non-adult-like cases being first curtailed. As the child grows up and his/her MLU consequently rises, adult-like cases take over and dominate the child's production in terms of adulthood.

Finally, and to account for the hypothesis of this dissertation (presented in section 3), it can be established that wh- movement and the different properties it entails is gradually acquired, as the correlation between MLU increase and the increase of general rate of correctness implies, regardless of grammatical complexity.

6. Bibliography

- Brown, Roger. *A First Language: the Early Stages*. Cambridge, MA: Harvard University Press, 1973. Print.
- Carnie, Andrew. *Syntax: a Generative Introduction*. Malden, MA: Blackwell Publishers Ltd, 2007. Print.
- Chomsky, Noam. *The Minimalist Program*. Cambridge, MA: The MIT Press, 1995. Print.
- Cook, Vivian, and Mark Newson. *Chomsky's Universal Grammar: An Introduction*. Oxford: Blackwell Publishers Ltd, 2007. Print
- Crain, Stephen, and Takuya Goro, and Rosalin Thornton. "Language Acquisition is Language Change". *Journal of Psycholinguistic Research*. 35.1 (2005): 31-48. Print.
- Gavrusseva, Elena, and Rosalind Thornton. "Possessor Extraction in Child English: A Minimalist Account". *University of Pennsylvania Working Papers in Linguistics*. 6.1. (2014): 155-170. Web. 27 Feb. 2015.
- Haegeman, Liliane, and Jacqueline Guéron. *English Grammar: A Generative Perspective*. Malden, MA: Wiley-Blackwell, 1999. Print.
- Huddleston, Rodney D. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press, 1984. Print.
- MacWhinney, Brian. CHILDES - *Child Language Data Exchange System*. 2000. Web. 16 Apr. 2015. <<http://childes.psy.cmu.edu/>>.
- MacWhinney, Brian. *Talkbank*. 2003. Web. 23 Apr. 2015. <<https://talkbank.org/>>.

- Stromswold, Karin. "The Acquisition of Subject and Object Wh-Questions." *Language Acquisition* 4.1-2. (1995): 5-48. Web. 21 Feb. 2015.
- Trotta, J. *Comments on Vacuous Subject Movement*. Göteborg: Halmstad University, 2004. Print.
- Valin, Robert D. Van, and Robert D. Van Valin Jr. "The acquisition of WH-Questions and the Mechanisms of Language Acquisition". In M. Tomasello, ed., *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1998. [reprinted] *Journal of Child Language* 27.1 (2000): 213-223. Web. 5 Mar. 2015.
- Valin, Robert D. Van, and Randy J. LaPolla. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press, 1997. Print.
- Wells, C. Gordon. *Learning Through Interaction: the Study of Language Development*. Cambridge: Cambridge University Press, 1981. Print.
- Wells, C. Gordon. *Language Development in the Pre-School Years*. Cambridge: Cambridge University Press, 1985. Print.
- Yule, George. *The Study of Language*. Cambridge: Cambridge University Press, 2014. Print.