MASTER IN ADVANCED ENGLISH STUDIES:

LANGUAGES AND CULTURES IN CONTACT

Departamento de Filología Inglesa

2015-2016

Universidad de Valladolid

Final Master Thesis

The Impact of Biological Gender on the Acquisition of *Wh*-Movement in a Bilingual Context

Cristina Arranz Gozalo

VALLADOLID 2016

The work presented in this MA thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text. The work in this thesis has not been submitted, either in whole or in part, for a degree at this or any other university.

This thesis is submitted in partial fulfillment of the requirements
for the degree of

Master in Advanced English Studies:
Languages and Cultures in Contact

to
Universidad de Valladolid

by

Cristina Arranz Gozalo

September 2016

Student's signature_____

Approved

Dr. Raquel Fernández Fuertes

Supervisor's signature_____

ABSTRACT

This dissertation deals with the impact of biological gender on the acquisition of *wh*-movement in English questions in a simultaneous bilingual context. It provides an empirical analysis of 2L1 English-French child data from CHILDES in an attempt to study the acquisition of this syntactic phenomenon and to compare the English 2L1 production to those of English L1 and L2 speakers. The data are analyzed in terms of the preference and complexity of *wh*-type and function of the *wh*-element as well as the complexity regarding pied-piping, and the adulthood of the productions. The analysis of the data suggests that biological gender does not have an impact on the acquisition of *wh*-movement in English questions.

KEYWORDS: 2L1 acquisition, *wh*-movement, question formation, biological gender, corpus study, CHILDES

RESUMEN

Este trabajo trata sobre el efecto del sexo biológico en la adquisición bilingüe del movimiento *wh* en la formación de preguntas del inglés, y en él se lleva a cabo un análisis empírico de datos de niños bilingües hablantes de inglés y francés procedentes de CHILDES con objeto de estudiar la adquisición de este fenómeno sintáctico y comparar su producción con la de hablantes de inglés como primera y segunda lengua. Los datos son analizados en base a la preferencia y complejidad de la forma y función del elemento *wh*, la complejidad asociada al fenómeno denominado *pied-piping* y la adecuación a la gramática adulta. Este análisis de datos apunta a que el sexo biológico no tiene un efecto sobre la adquisición del movimiento *wh* en las preguntas del inglés.

PALABRAS CLAVE: Adquisición bilingüe, movimiento *wh*, formación de preguntas, sexo biológico, estudio de corpus, CHILDES

Table of contents

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. However, it has not been until the last few decades when interest has been placed upon the impact of biological gender on the acquisition and development of linguistic skills, becoming not only a focus of attention among linguists, but also an object of controversy, since the extent of its influence as a factor involved in processes such as language acquisition has not been agreed on yet.

In an attempt to explore the role of biological gender on the bilingual acquisition of *wh*-movement in English, this dissertation presents an empirical study which strives to answer questions concerning i) the acquisition of this syntactic phenomenon by simultaneous – that is to say, 2L1– English-French bilingual children, comparing it to the acquisition process of children learning English as a second language (L2), and monolingual children with English as a first language (L1); and ii) the impact of biological gender on the English 2L1 acquisition of *wh*-movement, in comparison to that of English L2 and L1 contexts.

The data analyzed provide information about the constructions preferred by children in terms of the following issues:

- The form and function of the *wh*-element involved.
- The biological gender of participants.
- 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 present dissertation is divided into 5 different parts. The second section includes information regarding the 3 issues that constitute the background of the present study: both theoretical and empirical information about language acquisition in relation to i) defining properties of monolingual and bilingual acquisition, ii) research on language acquisition and biological gender, and iii) theoretical accounts on *wh*-movement. In the third section, the main research questions and hypotheses derived from the literature review are presented as the elements which guide the empirical analysis carried out in the present dissertation. The empirical study is presented in the fourth section, including

information about the corpora selected and the participants, the software and programs used to analyze the data, as well as the data classification and the corresponding data analysis. In the last section, the main conclusions reached after analyzing the data are laid out with a reference to the previous research questions initially posed and seeking for the a confirmation to the initial hypotheses.

2. Literature review

2.1 Language acquisition defining properties: monolingual and bilingual acquisition

2.1.1 *Monolingual acquisition*

Six different stages were identified in the language acquisition process by L1 English speakers (Yule, 2014: 171-180). The first one, the babbling stage, goes from 6 to 11 months of age; the one-word-stage comes next, and it starts at around 12 months and continues until 18 months; following this second stage is the two-word stage, which takes place at 18–20 months of age; the next step, the telegraphic speech stage extends from 21 to 24 months. During this stage, strings of words begin to appear in children's production (e.g. this shoe all wet (Yule, 2014: 174)). The fifth and last stage, the multiple word stage, usually begins when the child is 2 or 3 years old, although time differences may appear among children. In this final stage, further morphological, syntactic and semantic developments take place. These developments and, therefore, the multiple-word stage, last until the child is 5 years old or even older than that, until the child finally reaches the adult grammar in all linguistic domains.

Questions and negative sentences are 2 types of constructions acquired within the multiple-word stage. Since the aim of the present dissertation is to study question formation, the next paragraphs only provide information about interrogative sentences.

The production of interrogative (and negative) sentences can be divided into 3 different stages, as proposed by Yule (2014), which occur during the multiple speech stage and are related to syntactic development.

- First stage: children first start producing interrogative sentences by: i) adding a *wh*-word at the beginning of a structure or a sentence, as seen in (1) and (2); ii) raising their intonation at the end of the sentence, as shown in (3) and (4), where no *wh*-form is present but the question mark at the end suggests that the child uttered these with raising intonation.

    (1) **Where** kitty?
    (2) **Where** horse go?
    (3) Doggie?
    (4) Sit chair?

- Second stage: during this stage, a wider range of *wh*-forms is used to form questions, as illustrated in (5) and (6). It is also noticeable that the raising intonation is still used.

    (5) **What** book name?
    (6) **Why** you smiling?

- Third stage: subject-verb inversion and *do* insertion appear for the first time during this stage, as seen in examples (7), (9) and (10). 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 (8) 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 (9) and (10).

    (7) Can I go?
    (8) **How** that opened?
    (9) **What** did you do?
    (10) Will you help me?

Therefore, 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, through development, not only do children produce more complex questions, but also the rate of adult-like utterances is higher and higher.

This development, however, does not occur in isolation. In fact, it takes place simultaneously along with other morphological, syntactic and semantic developments. Thus, we can conclude that language, and more specifically, the grammatical properties of languages are acquired simultaneously at different fronts by children acquiring their first language.

2.1.2 *Bilingual acquisition*

The linguistic development of bilingual children has also been a subject of interest for linguists. Dealing with structural aspects that are different in the 2 languages has been found to be the best option to study how bilingual children acquire their 2 language systems (Barnes, 2006). Although the fact that children can differentiate grammatical systems "as soon as language-specific word order properties and inflectional morphology emerge" (Meisel, 2001: 17) has been agreed on, several theoretical hypotheses strive to account for what happens before that. These possibilities, summarized by De Houwer and Meisel (1996), are described as follows:

- Fusion Hypothesis: a new system combining elements from 2 or more systems is created by the child.
- Differentiation Hypothesis: as soon as they have grammatical knowledge, children differentiate between the 2 systems.
- Interdependent Development Hypothesis: 1 of the languages is used as a developmental guide for the other.
- Autonomous Development Hypothesis: each language system is acquired by bilingual children in the same manner a language is acquired by a monolingual child.

When studying the syntactic acquisition of bilingual children, Paradis and Genesee (1996) developed a hypothesis dealing with interdependent development when 2 L1s interact. The 3 manifestations associated to their own view of the interdependent hypothesis are the following: transfer, acceleration, and delay. The first one takes place when a grammatical property is incorporated from 1 language into another. The second one, acceleration, accounts for the emergence of a grammatical property earlier in bilinguals than it would be the norm in monolingual acquisition. And the third one, delay, deals with the late emergence of a grammatical property in bilinguals' production compared to that of monolinguals' because of the burden that acquiring 2 L1s entails.

2.2 Biological gender and language acquisition

A large number of recent studies sustain that girls develop linguistically earlier than boys. This early development have been seen in early communicative gestures (Özçalışkan and Goldin-Meadow, 2010), early vocabulary growth (Huttenlocher et al.,

1991; Bauer et al., 2002), morphosyntactic growth (Hadley et al., 2011), vocabulary size and syntactic complexity (Ramer, 1976; Fenson et al., 1994; Galsworthy et al., 2000; Lutchmaya et al., 2002; Van Hulle et al., 2004; Berglund et al., 2005; Kern, 2007; Westerlund and Lagerberg, 2008; Bouchard et al., 2009; Lovas, 2011; Simonsen et al., 2014), and consistency in vocabulary production (Barbu et al., 2013). Not exclusive to a single language, these early differences are not to be found systematically in all language skills. That being said, it is also important to mention that girls' advantage is likely to be small, gender differences therefore accounting for a small part of the variance (Fenson et al., 1994; Galsworthy et al., 2000; Berglund et al., 2005; Reilly et al., 2007).

In order to study whether boys produce 3 different types of constructions (argument + argument, predicate + argument, and predicate + predicate) in gesture and speech combinations later than girls, Özçalışkan and Goldin-Meadow (2010) collect, transcribe and analyze data from 40 American children whose age range from 14 to 34 months. Their research shows that sex differences begin to appear in gestures when children start communicating, being girls the first ones to convey sentential meanings by combining gestures with words. In particular, girls started producing argument + argument, and argument + predicate constructions earlier than boys in speech + speech and speech + gesture combinations. However, girls appeared to be behind boys when starting to produce predicate + predicate constructions in both speech + speech combinations and gesture + speech combinations.

Hyde and Linn (1988) carried out a meta-analysis of studies reporting statistics on gender differences in verbal ability in order to examine i) if gender differences in verbal ability existed; and ii) if these differences extended to various measurements of verbal ability, including vocabulary, analogies, and essay writing. The results obtained from their analysis suggest that, although females rank higher than males in measures of speech production, no substantial evidence for gender differences can be found in several types of verbal ability (e.g. vocabulary, reading comprehension and analogies).

2.3 *Wh*-movement in language acquisition

The field of linguistics, and more specifically the generative grammar tradition was revolutionized a few decades ago by Chomsky's theories on movement and Universal Grammar. This author hypothesized in a collection of 4 articles, *The Minimalist*

Program (Chomsky, 1995), about the existence of a syntactic domain called phase and the capability of a constituent to move out of a phase (assuming that 2 phases named Verbal Phrase –VP– and Complementizer Phrase –CP– form a sentence) as long as it moves to the left part of the phase first. In his work, this is referred to as "phase impenetrability condition" (Chomsky, 1995). The publication of Chomsky's works on this syntactic phenomenon led a high number of linguists and grammarians to take interest in *wh*-movement.

Wh-movement can be understood 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). The properties and motivations of this syntactic movement are the same in the 2 types of structures in which it takes place: relative clauses and questions. However, these 2 constructions are different due to their specific grammatical properties. Relative clauses, as shown in (11), are dependent finite or non-finite clauses. Introduced by a relative pronoun, they follow a noun functioning as antecedent, (the person, in (11)), which they modify. On the contrary, questions are interrogative sentences which can be divided into: i) interrogative root sentences (or yes/no questions), as in (13), and ii) constituent questions or *wh*-questions, as in (12) and the subordinate CP in (13) (Haegeman and Guéron, 1999: 170). The present dissertation deals with the second type. It is also important to mention that *wh*-constituents can be found in direct and indirect questions, as in examples (12) and (13) respectively.

(11) I know the person [**whom** Thelma will meet after lunch].
(12) **What** does he do?
(13) Do you know [**what** he does]?

A relative clause introduced by the relative pronoun *whom* is shown in example (11). This *wh*-element moves from the direct object position of the verb *meet* to the specifier position of the CP level in the subordinate relative clause. It can be therefore determined that this is a declarative sentence, in which information is given about the object of *know*, i.e. *the person*. The function of this relative clause is that of post-modifier of *the person*, and including it, it forms the direct object of *know*. Example (12) shows a direct question introduced by the *wh*-pronoun *what*. In this case, the *wh*-element, which functions as the direct object of the verb *do*, also moves to the specifier position of the CP level. An indirect *wh*-interrogative sentence is shown in example (13), with the *wh*-phrase *what* functioning as the direct object of the verb *does* in the subordinate clause.

The *wh*-questions in (12) and (13) involve a content answer, therefore differing from yes/no questions.

A series of movement theory concepts provided by Haegeman and Guéron (1999: 172) are considered in the grammatical representation of *wh*-movement in direct and indirect questions, as shown in examples (11), (12) and (13) above. These theoretical concepts are explained as follows and illustrated in (14) to (16) below:

- Moved element: the element which undergoes the movement operation, which, in this particular case, is the *wh*-element (i.e. *whom* in (11) and *what* in (12) and (13)).
- 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*.

The grammatical representations shown in (14), (15) and (16) correspond to the *wh*-movement operations in the examples (11), (12) and (13), taking the concepts explained above into consideration:

(14) I know the person [**whom**$_i$ Thelma will meet t_i after lunch].
(15) **What**$_i$ does he do t_i?
(16) Do you know **[what**$_i$ he does t_i]?

As shown in examples (14), (15), and (16), *wh*-movement can take place at 2 different levels: at a main clause, as in (15); and at a dependent clause, as in examples (14) and (16). In these 3 examples, it can be seen that the *wh*-phrases *whom* and *what* have landed in the specifier position of the CP they belong to (their respective landing sites) after moving out from their extraction positions (as objects of their respective verbs). A trace can be found in the extraction site, which is co-indexed with the moved constituent, *whom* and *what*.

A number of characteristics are shared by relative clauses and *wh*-questions, including *wh*-forms. Besides, the *wh*-constituents which can be found in both types of constructions are subject to the same constraints (Haegeman and Guéron, 1999: 176). Therefore, the presentation in the subsequent sections is focused on the target structure for the present dissertation, *wh*-movement in questions, but the same applies to *wh*-movement in relatives.

2.3.1 *Types of wh-words and their grammatical functions*

Huddleston (1984) provides a typology of 9 different interrogative *wh*-words used to produce questions. In addition, this author offers a classification of *wh*-words in terms of the function they play in the sentence. Taking his own categorization of *wh*-types as a point of departure, he distinguishes, among others, 4 functions of *wh*-elements: i) subject, ii) object, iii) subject complement, and iv) adjunct. The distribution of functions discussed in this section of the present dissertation corresponds to the functions found in child data.

- *Who*, *whom*, and *whose*: although these 3 forms differ in their case specification (*who* is nominative, *whom* is accusative, dative or ablative and *whose* is genitive), they are in fact different representations of the same pronoun *who*. The first 2 *wh*-words, *who* and *whom*, are usually fused in *who*, as shown in example (17). Furthermore, *who* constructions incorporate the [+ human] feature, in opposition to *what* constructions. *Whom* can appear with or without a following head noun, as in (19a) *versus* (19b).

  (17) a. **Who** are you talking to?
  b. **Whom** are you talking to?
  (18) **Whose** entry won the prize?
  (19) a. **Whose** team won?
  b. **Whose** won?

  With respect to their grammatical function, *who* usually functions as a subject, as in 20, or as the complement of a preposition, as in (21). *Whose* appears as a determiner in a Determiner Phrase –DP–, which can function as direct object, as in (22), or as subject complement.

  (20) **Who** is she?
  (21) **Whom** were they talking to?
  (22) **Whose** wallet did they steal?

- *What*: this *wh*-word can appear either as a pronoun, as in (23), as a determiner in a DP (as in 24), or as a complement to the possessive clitic, as in (25). *What* accounts for the [-human] feature when appearing as a pronoun, as in (26), in contrast with *who* constructions.

  (23) **What** caused that?
  (24) **What** book are you reading?
  (25) **What** schoolchild's imagination could fail to be simulated by such a challenge?
  (26) **What** happened?

  When appearing as a pronoun, *what* can either take the function of subject or direct object, as in (26) and (27). Nevertheless, examples (28), (29) and (30) show *what* as a determiner being part of a DP functioning as subject complement, direct object and adjunct respectively.

  (27) **What** did you buy?
  (28) **What** time is it?
  (29) **What** movie are we watching?
  (30) **What** time are they coming?

- *Which*: it can also be a pronoun, as in (31) or a determiner, as in (32) and it is neutral as to the [+/- human] contrast, as in (33).

  (31) **Which** do you want?
  (32) **Which** candidates do you support?
  (33) **Which** of the versions shall we use?

  Regarding its grammatical function, it can appear as a pronoun or as a determiner. Moreover, it can function as subject complement and direct object or as part of a DP which can, in turn, function as a subject complement, as in (34) or as a direct object, as in (35).

  (34) **Which** one is John?
  (35) **Which** one do you want?

- *When*, *where*, *why*, and *how*: *when* and *where* can occur as adverbs, as in (36) or pronouns as complements of a preposition, as in (37). Moreover, *why* appears as an adverb, as it can be seen in (38). With respect to the *wh*-word *how*, it can occur either as a degree adverb modifying an adjective, as in (39) or as an adjective, as in (40), where it functions as a subject complement.

(36) **When** did he arrive?
(37) **Where** does he come from?
(38) **Why** did she leave?
(39) **How** big is it?
(40) **How** are you?

The main function of *when*, *where* and *why* is that of an adjunct, as in (41) and (42). Likewise, *how* can function as adjunct or as subject complement, as in (43) and (44).

(41) **Why** are you going to the park?
(42) **Where** are you?
(43) **How** did you make the cake?
(44) **How** are you?

Equally important as the differentiation of direct questions in terms of the form and function of the *wh*-element is the fact that subject and object questions are not constructed in the same manner most *wh*-questions are formed, as shown in (45) and (46).

(45) [CP **who**$_i$ [IP t_i annoyed him]?
(46) [CP **who**$_i$ [IP did she marry to t_i]?

Example (45) shows a subject *wh*-question with the *wh*-element *who*, which has moved from the subject position in the specifier of IP to the specifier of CP, a movement that is not overtly marked in the sentence level. Although the same *wh*-word (*who*) is used in example (46), in this case it is used to form an object question. In this example, *who* has moved from the object position in IP to the specifier CP, a movement that is overtly marked in syntax, as the representation in (46) shows.

Taking Chomsky's argumentation that "the language learner assumes that there is syntactic movement only where there is overt evidence for it" (1986: 50) as a point of departure, Trotta (2004) carries out an analysis on interrogative sentences, exclamative sentences, as well as free and bound relative sentences with the aim to deal with vacuous subject movement (VSM). That is to say, over *wh*-movement not taking place for *wh*-subjects.

	Dependent	Interdependent
(47) interrogative:	I don't know **who** was present.	**Who** was present?
(48) exclamative:	You won't believe **what** strange people were on the tram today.	**What** strange people were on the tram today!
(49) free relative:	He threw **what** was left of the cheesecake at the dog.	--
(50) bound relative:	The author **who** wrote the novel Fight Club has a long last name.	--

Trotta (1999: 4)

This author concludes that Chomsky's assumption implies that language learners are unaware of other analogous types of movement operations which, according to Trotta, "would lead them to a different (tacit) treatment of the structure in question" (2004: 15). Moreover, he states that the existence of a moved *wh*-subject is not contradicted by any empirical data or theoretical foundation. Following Trotta's argument, it can also be added that subjects are never originally in the CP level. They are located in the IP level instead. This implies that a *wh*-subject moves from the IP to the CP level, too, even if this movement is not overt.

2.3.2 *Levels of complexity*

When studying *wh*-movement, the level of complexity also needs to be taken into consideration. Not only does this apply to the function played by the *wh*-element (subject *versus* other function, as presented above), but it also relates to the effects of movement. To be more specific, these effects are associated to the phenomenon of pied-piping. In a few words, pied-piping can be understood as the movement of other elements along with the *wh*-element itself. That is to say, the *wh*-word drags other constituents along with it. Pied-piping can occur when the *wh*-element is either part of a prepositional phrase, as in (51a), or a modifier within a SP, as in (52a). While the contrast between examples (51a) and (51b) illustrates optional pied-piping, the ungrammaticality of (52b) shows that pied-piping is compulsory when the *wh*-word occurs as a modifier within a DP.

(51) a. **To whom**$_i$ did you speak t_i?
b. **Whom**$_i$ did you speak to t_i?

(52) a. **Which book**$_i$ are you reading t_i?
b. ***Which**$_i$ are you reading t_i book?

2.3.3 *Motivation for movement*

According to Carnie (2007: 362), the fact that *wh*-phrases move to the specifier position of the CP to be near the [+WH] feature present in the head level of the CP (i.e. in C) explains the motivation for movement. In other words, they move to check this feature. Interrogative sentences containing a *wh*-element therefore have a [+WH] feature, which triggers the movement by attracting any *wh*-constituent within the sentence. Interrogative sentences also have a [+Q] feature, which is located in C as well. However, other types of structures (e.g. relative clauses) have a [-Q] feature. As already explained, *wh*-movement takes place when the *wh*-phrase moves to be near the [+WH] feature, regardless of whether the sentence is a question ([+Q]) or a non-question ([-Q], e.g. in relative clauses, in negative clauses, in statements). Therefore, interrogative *wh*-sentences have a double feature in the CP level: [+Q] (as questions *versus* statements) and [+WH] (as *wh*-questions *versus* yes/no questions).

2.3.4 *Wh-movement and language acquisition*

Many linguists have taken interest in language acquisition, resulting in the creation of extensive work on this particular topic also within the generative tradition. A summary of two studies which deal with the acquisition of *wh*-movement is presented below.

Although several empirical studies have studied question formation by analyzing data from L1 speakers, this dissertation focuses on 2 particular studies: i) Gavruseva and Thornton's (1999) analysis of the acquisition of questions with possessor phrases; and ii) Van Valin's (1998) study on the acquisition of interrogative sentences. These two studies have been selected due to the fact that they deal with *wh*-elements and *wh*-sentences and provide insightful information regarding the acquisition of *wh*-movement.

Gavruseva and Thornton (1999) conduct an empirical analysis with 12 L1 English participants whose ages range from 4;5 to 6;0 in order to explore the acquisition of questions with *wh*-possessor phrases, as seen in (49).

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

The example in (53a) 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 in the genitive case (*whose*) and thus differs from the non-possessor counterpart (*who*), as explained in section 2.3.1 above. The lack of pied-piping in (53b) and (53c), and the incorrect case marking in the *wh*-word in (53c) lead to non-adult-like productions.

The data obtained were classified in terms of the three possessive structures illustrated in the example (53) above. This appears in table 1.

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

Children	Extrac-tion of *who*	Extraction of *whose*	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: Gavruseva and Thornton's (1999: 166) table 3.

Table 1 shows that only 1 child (Mary – 4;11) has a 100% rate of correctness. On the contrary, Matt and Mandy (6 year-old children) have a 100% rate of non-adult-like production as they produce both types of non-adult-like questions with *wh*-possessor phrases: lack of pied-piping (*whose* extraction) and lack of genitive marking (*who* extraction). The other 9 participants produce both adult-like and non-adult-like questions with *wh*-possessor phrases, which suggests that English-speaking children go through a developmental stage at which they alternate adult-like and non-adult-like constructions.

The results obtained from the analysis of the data they elicited from the 2 experiments carried out in the form of a guessing game led Gavruseva and Thornton to the conclusion that *whose* questions are more difficult than *who* questions, given their complexity associated to both morphological marking and pied-piping of the possessed noun. This conclusion is based on the fact that, as seen in table 1, half of the *whose*

structures produced by the 12 participants were adult-like (55%), as in (49a). On the contrary, most of the non-adult-like cases corresponded to the movement of *who* (38%), as in (53c).

The aim of Van Valin's (1998) empirical study was to provide 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 have never been exposed to before. Moreover, this author's study is based on the analysis carried out by Stromswold (1995) on the early production of *wh*-questions by 12 children in the CHILDES database.

The data analyzed by Stromswold show that, with respect to *who wh*-questions, both subject and object questions appear around the same time. Regarding subject and object *wh*-questions, 6 children produced object *wh*-questions first, 4 produced subject *wh*-questions first, and 1 produced both of them simultaneously. Regarding *what* and *which wh*-questions, the production of this group of children favored object *wh*-questions first. While 7 of these children produced object *what wh*-questions first, 4 produced both types at the same time, and 1 child produced subject *what* questions first. Concerning *which wh*-questions, 5 children produced object questions first, and 1 child produced both types at the same time.

The analysis of the data leads Stromswold to conclude that i) object *wh*-questions begin to be produced earlier than subject *wh*-questions; and ii) object questions start being produced earlier in complex sentences, too.

After conducting his analysis, Van Valin concluded that, despite appearing to be simpler than other types of *wh*-questions (since there is no subject-auxiliary inversion and they only require to replace the subject with a *wh*-element to be formed), subject *wh*-questions are not the first type of *wh*-questions that children comprehend and produce. He also added 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: 642). 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, without necessarily having been ever exposed to it before.

Two different conclusions can be drawn from the studies above:

i) There is a developmental path in the acquisition of *wh*-questions by L1 English children.
ii) Syntactic complexity can account for the appearance of non-adult-like constructions during the acquisition process of *wh*-questions.

3. Research questions and hypotheses

The present dissertation deals with the acquisition of *wh*-movement in English questions, and the effect of biological gender in the process of acquisition of this syntactic phenomenon in an English bilingual context. In order to do so, an analysis of data from simultaneous bilingual children, that is, children exposed to two L1s from birth (2L1) recorded in a naturalistic setting is provided. The data belonging to this group of 2L1 participants are compared to data elicited from a group of English monolingual children (L1), and a group of children learning English as a second language (L2).

Taking as a point of departure the different studies presented in section 2 dealing with both the grammatical properties of *wh*-movement and the acquisition of this structure, the analysis presented below strives to answer the following questions:

1. Do males and females acquire *wh*-movement in the same way? That is to say, do they acquire it at the same time? And, is gender a key factor to be considered when accounting for the L1, 2L1 and L2 acquisition of English *wh*-movement?
2. Is the English production of 2L1 bilinguals similar to the production of L1 monolingual children and children learning English as an L2 in terms of i) the order of acquisition of the *wh*-constituent, ii) the grammatical function of the *wh*-element and iii) the grammatical complexity?
3. Is the English production of the 2L1 bilingual children affected by interdependence and, if so, which specific effect is seen (transfer,

acceleration or delay)? Or, on the contrary, is there no bilingual effect reflected in their production?

Based on the research carried out by previous authors as discussed before (section 2.2.2), the initial hypotheses of the present dissertation are the following: i) *wh*-movement and the different properties it entails (order of acquisition and grammatical function of the *wh*-element, as well as their grammatical complexity) would be gradually acquired; ii) there would be a delay in the acquisition of *wh*-movement by 2L1 speakers, due to the burden that acquiring 2 different linguistic systems at the same time implies; and iii) no substantial differences across genders would appear in the children's production in terms of preference for a specific *wh*-type and function, as well as in terms of the rate of correctness of their production of *wh*-questions.

4. Empirical analysis on *wh*-movement and question formation

With the aim to provide an answer to the research questions presented above and to test the hypotheses previously described, an empirical analysis has been carried out. Four different stages form this analysis, which include the following: a description of the corpora used to select data, information about the participants whose data form these corpora, the data classification criteria, and the data analysis.

4.1 Data selection

4.1.1 Selected corpora

Three different corpora have been selected in order to study the impact of biological gender on the acquisition of *wh*-movement and question formation in an English bilingual context. These 3 corpora have been extracted from TalkBank, and more specifically, from CHILDES (MacWhinney, 2000). TalkBank offers several different L1 and L2 browsable and downloadable databases. CHILDES is the child component of TalkBank and so it deals with child language. The first 2 corpora selected for this analysis, the Genesee and the Paradis corpora, have been extracted from the Biling section. These 2 corpora include transcriptions of the recorded conversational interactions of, on the one hand, 2L1 English-French bilinguals in Canada, and, on the other hand, children learning English as an L2, respectively. Moreover, the third corpus used in this analysis has been extracted from the Eng UK-MOR section. This third selected corpus is the Wells corpus, and it includes transcriptions of the recorded

conversational interactions of British monolingual children. The Paradis and the Wells corpora are used as control groups in order to compare the production of the 2L1 group with those of the L1 and L2 groups.

Not only were these 3 corpora selected because they featured English-speaking L1, L2 and 2L1 bilingual 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 2.2.1.1). As previously explained, and within the multiple-word stage, *wh*-elements begin to appear in the first stage of this process. Nevertheless, a different kind of syntactic movement (a head movement in the form of subject-auxiliary inversion) is first produced alongside *wh*-movement (Cook and Newson, 2007) during the third stage of the process of question formation suggested by Yule (2014). In other words, it is not until this third stage when the sentences produced by children can be classified as adult-like. Hence why the combination of the selected corpora is optimal to study the acquisition of *wh*-movement in bilingual speakers, and also to compare it with the acquisition of the same phenomenon in the production of L1 and L2 English: they cover a period in which the first questions emerge (between 1 year and a half and 2 years) and go until the age of 5 – and beyond – when adult-like production is expected.

A more detailed account of each of the corpora selected appears next, which includes information taken from the database manual section available in CHILDES.

4.1.1.1 *The Genesee corpus*

This corpus, compiled by Fred Genesee, is divided into 3 sections containing data in i) English, ii) French, and iii) both languages. The English data section, which includes the data under analysis in the present dissertation, is formed by 17 files containing data from 5 children. The age of the 2 male and 3 female children ranges from 1 year and 10 months to 3 years and 7 months approximately.

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

Table 2. The Genesee corpus: general information and *wh*- production				
Age range	**# of participants**	**# of utterances**	**# of *wh*-sentences**	**# of *wh*-questions**
1;10.20 to 3;7.09	5	2,484 (100%)	74 (2.9%)	43 (1.7%)

Table 2 shows that out of the total number of utterances produced by these 5 children (2,482), 1.7% of these are *wh*-questions (43), which, as previously mentioned, are the ones the present dissertation focuses on.

The data compiled in the Genesee corpus belong to children living in Canada. More specifically, all the participants and their families lived in Montreal or the surrounding communities. It is important to mention that the linguistic community of Montreal is formed of English-French bilingual speakers who use both languages on a daily basis. Moreover, children do not only receive input in these 2 languages from their primary care-takers, but also from the media and through contact with other children and adults in this bilingual community.

Regarding the context in which the data were elicited, the 5 participants lived in houses where both languages were used, and each parent generally used 1 of the 2 languages to interact with their children. It was in the children's homes (usually in the living-room, playroom, or kitchen) where the data were elicited by having an assistant or graduate student record the parents interacting and talking with their children. However, said assistants were to be ignored as much as possible by the parents and their children during the sessions.

Although each session lasted longer, only 20 to 30 minutes of each session with each child were transcribed. The first 5 minutes of each session were discarded, since it was the amount of time the children required to get used to the taping equipment. However, if the children did not produce 100 intelligible utterances during that period of time, the recording continued until that number of intelligible utterances was reached.

The criteria used in order to transcribe the utterances in the transcriptions included the addressee (e.g. parent, dog toy) and the language of the utterance. Regarding the language, the utterances were sorted out into 5 categories: i) French only, ii) English only, iii) mixed, iv) neutral, and v) unintelligible. Utterances which contained both English and French were classified as mixed utterances (e.g. '*ça go pas là' = this$_{FR}$ go$_{EN}$ not$_{FR}$ there$_{FR}$'*). Likewise, utterances that could belong to either language were

considered to be neutral (e.g. animal sounds such as '*meow'*; the word '*okay'*, and proper names). Nevertheless, if a neutral word appeared as part of a longer utterance in 1 single language, the entire utterance was to be categorized as being in that language. Finally, 1 or 2 bilingual assistants – native-speakers of the primary language of the session – reviewed all the transcripts.

4.1.1.2 *The Paradis corpus*

Created by Johanne Paradis in 2005, the Paradis corpus is formed by 106 files containing data which belong to 25 children learning English as an L2. The data included in this longitudinal corpus have been sorted out into 5 rounds, which extend to a 2-year period of time. However, only data belonging to 19 participants were compiled in all the 5 rounds.

The main features of the Paradis corpus are summarized in table 3.

Table 3. The Paradis corpus: general information and *wh*- production

Age range	# of participants	# of utterances	# of *wh*-sentences	# of *wh*-questions
4;02.12 to 8;9.13	25	52,358 (100%)	2,071 (3.9%)	466 (0.8%)

Table 3 shows that these 25 participants (17 males and 8 females) produced a total number of 52,358 utterances. Moreover, it can be seen that 3.9% of these utterances (2,071) are *wh*-sentences, and also that 0.8% of them (469) are *wh*-questions which are the focus of analysis in the present study.

The participants whose data are included in the Paradis corpus are children from newcomer (immigrant and refugee) families to Canada. In addition, all the participants are sequential bilinguals, since they started learning English as a second language once their first language had been acquired. Despite the fact that some of these children were born in Canada, they were functionally monolingual until they first attended an English language preschool or school.

For 5 rounds, data were collected every 6 months and the participants were 5 years and 6 months old on average when the study started. Furthermore, they had been exposed to English as a second language for about 10 months in a preschool or school program. In order to collect the data, the children were video-recorded speaking with a student research assistant in their homes. These taping sessions lasted for 45 minutes

approximately. With the aim to start a conversation or to introduce new topics, questions were asked to the participants by the assistants. However, questions were only used if the children did not introduce their own topics or if the conversation did not move forward. Thus, not all children answered questions in each round of data collection.

4.1.1.3 *The Wells corpus*

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.

A summary of the main features of the Wells corpus appears in table 4.

Table 4. The Wells corpus: general information and *wh*- production

Age range	# of participants	# of utterances	# of *wh*-sentences	# of *wh*-questions
1;6 to 5;0	32	33,050 (100%)	1,783 (5.3%)	702 (2.1%)

In table 4, it is shown that out of the total number of utterances produced by these 17 males and 15 females (33,050), 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 a project carried out in 1973 ("The Bristol Language Development Study: language development in preschool children") in which children were recorded in a lifelike setting 10 times at 3-monthly periods. Moreover, all the samples are 90 seconds long, after which the recorder used to tape the conversations automatically stopped recording. 24 examples at 20 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.

Instead of having a researcher present during the recording sessions, a microphone was attached to the child's clothes (which allowed for movement and freedom around the house) when the child was getting dressed up in the morning. In addition, a radio receiver, a tape recorder and a programmed timing mechanism were also used for the taping sessions, which were separated by irregular periods of time so as to avoid: the child's awareness of being recorded, and the parents planning any activity beforehand. Regarding the context of the data, it was retrieved thanks to the parents, since they were

able to provide details about the location, participants and context of the conversation after listening to the sample recorded the very same day.

4.1.2 Participants

These 3 different corpora available in CHILDES, the Genesee corpus, the Paradis corpus, and the Wells corpus, have been selected in order to carry out an analysis of the data extracted from them. As already mentioned in section 4.1.1, these 3 corpora contain data belonging to a group of 2L1 English-French bilingual children, a group of children learning English as an L2, and a group of English monolingual children respectively.

Since the present study focuses on the effect of biological gender, 4 participants (2 males and 2 females) have been selected from each corpus. The data belonging to the 2L1 group will be compared, on the one hand, with those belonging to the English monolinguals, the Wells corpus; and, on the other hand, with those belonging to a group of children learning English as an L2.

Table 5 provides an overview of the participants selected, including their name, gender, age range, MLU range, number of utterances, and number of *wh*-questions.[1]

[1] 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).

Table 5. Information about all the participants					
The Genesee corpus: 2L1 participants' information					
Name	**Gender**	**Age Range**	**MLU range**	**# of utterances**	**# of *wh*-questions**
Gene	M	1;10.26 – 3;7.09	1.976 – 2.972	631	15 – (2.3%)
Jessica	F	1;10.20 – 1;11.23	1.173 – 1.432	606	0
Joelle	F	2;4.06 – 2;4.15	1.381 – 1.485	591	19 – (3.2%)
Olivier	M	1;11.15 – 2;10.29	1.875 – 2.621	656	9 – (1.3%)
Sub-total		1;10.20 – 3;7.09	1.173 – 2.972	2,484	43 – (1.7%)
The Paradis corpus: L2 participants' information					
Name	**Gender**	**Age range**	**MLU range**	**# of utterances**	**# of *wh*-questions**
CNDX	F	6;08.22 – 8;3.15	3.275 – 4.800	1,788	13 – (0.7%)
DVDC	M	6;03.23 – 8;4.15	3.615 – 4.603	2,120	18 – (0.8%)
JNNH	F	5;10.26 – 7;11.21	3.451 – 4.289	1,673	13 – (0.7%)
SMNS	M	5;06.23 – 7;6.29	2.764 – 3.933	2,719	22 – (0.8%)
Sub-total		5;06.23 – 8;4.15	2.451 – 4.800	8,300	66 – (0.7%)
The Wells corpus: L1 participants' information					
Name	**Gender**	**Age range**	**MLU range**	**# of utterances**	**# of *wh*-questions**
Jonathan	M	1;6.5 – 4;7.14	1.307 – 4.485	2,109	92 – (4.3%)
Abigail	F	1;5.28 – 4;8	1.417 – 4.827	1,072	29 – (2.7%)
Debbie	F	1;6.9 – 1;11.25	1.241 – 4.098	1,208	35 – (2.8%)
Benjamin	M	1;5.21 – 5;0.24	1.376 – 4.418	1,361	18 – (1.3%)
Sub-total		1;5.21 – 5;0.24	1.307 – 4.827	5,750	174 – (3%)
Total		1;5.21 – 8;4.15	1.173 – 4.800	16,534	283 – (1.7%)

Table 5 includes information about the 12 participants selected from the 3 corpora used to carry out this analysis, 6 males and 6 females. This information shows that the total number of *wh*-questions that are involved in the analysis is 283 and that this corresponds to 1.7% of these children´s overall production. However, more importantly, it also demonstrates that these differences across groups are important. In this regard, as shown in table 5, the total initial MLU rate belonging to the L2 group of participants is considerably higher than that of the children belonging to the L1 and 2L1 groups due to the fact that the data compiled in the Paradis corpus were elicited when the participants were significantly older (from 5;06.23 to 8;4.15) than those children in the other 2 groups (from 1;10.20 to 3;7.09 in the case of the Genesee corpus; and from 1;5.21 to 5;0.24 in the case of the Wells corpus).

Considering the fact that it is impossible to perform an MLU match between the participants of the three corpora (as the total MLU rate ranges from 1.173 to 2.972 in the case of the Genesee corpus; and from 2.451 to 4.800 in the case of the Paradis corpus), the data belonging to the control groups have been divided into 2 different

stages, with 2 different aims: i) to compare the production of the English monolingual participants in stage 1 with that of the 2L1 children and ii) to compare the production of the English monolingual participants in stage 2 with that of the children learning English as an L2. Therefore, this double comparison using the L1 group 2 stages makes groups much more parallel in terms of linguistic development.

The MLU match carried out in order to analyze the data is presented below in table 6.

TABLE 6. Information about participants and MLU match			
Stage 1 of the analysis: MLU range from 1 to 3			
2L1 group		**L1 control group**	
Participant	**MLU range**	**Participant**	**MLU range**
Gene	1.976 – 2.972	Jonathan	1.307 – 1.690
Jessica	1.173 – 1.432	Abigail	1.417 – 2.884
Joelle	1.381 – 1.485	Debbie	1.241 – 2.972
Olivier	1.875 – 2.621	Benjamin	1.376 – 2.879
Total	1.173 – 2.972	**Total**	1.241 – 2.972
Stage 2 of the analysis: MLU range from 3 to 5			
L2 group		**L1 control group**	
Participant	**MLU range**	**Participant**	**MLU range**
CNDX	3.275 – 4.800	Jonathan	3 – 4.485
DVDC	3.615 – 4.603	Abigail	3.202 – 4.827
JNNH	3.451 – 4.289	Debbie	3.014 – 4.098
SMNS	3.254 – 3.933	Benjamin	3.255 – 4.418
Total	3.254 – 4.800	**Total**	3 – 4.827

As seen in table 6, the MLU rate of the 2L1 and the L1 participants in section 1 of the analysis ranges from 1.173 to 2.972, and from 1.241 to 2.972 respectively. Likewise, the MLU rate of the L2 and L1 participants in section 2 of the analysis ranges from 3.254 to 4.800, and from 3 to 4.827 respectively. Therefore, this double comparison using the L1 group two stages makes groups much more parallel in terms of linguistic development.

4.2 Data classification criteria

The data analyzed have been classified taking into consideration several different factors, which are described below.

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 2 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 (54) and (55); and *wh*-movement in relative clauses was not considered either, as in (56)); and ii) not have been the product of the imitation of an adult's previous utterance, or an already made sentence, as in (57) or unproductive language, such as songs.

(54) **How**? – Smns (7;6)
(55) **What** Mummie? – Abigail (4;8)
(56) Look **what** I've done – Jonathan (3;5)
(57) When I say "**who**'s your friend"? – Debbie (1;11)

The criteria used in order to classify the data from the 12 participants include the following variables: general inclusion criteria, *wh*-word types, *wh*-word function, pied-piping, adult-like form, and MLU rate.

Regarding *wh*-types, 9 *wh*-word forms were isolated: *what*, as in (58), *why*, as in (59), *when*, as in (60), *where*, as in (61), *how*, as in (62), *who*, as in (63), *whose*, as in (64) and *which*, as in (65). The only *wh*-element which does not appear in the data is *whom,* as in (66).

(54) **What** title is it*?* – Jnnh (6;5)
(55) Mom, **why** you play this? – Gen (3;0)
(56) **When** was my birthday? – Dvdc (8;4)
(57) **Where** do black olives comes from? – Cndx (8;3)
(58) **How** do you take this thing off? – Dvdc (6;10)
(59) **Who** is it? – Olivier (2;10)
(60) **Whose** is that? – Debbie (3;11)
(61) **Which** one does I mean? – Dvdc (8;3)
(62) **Whom** did you talk to?

With respect to *wh*-word function, 6 categories were distinguished: subject, as in (67), object, as in (68), subject complement, as in (69) and adjunct, as in (70). Other functions mentioned in section 2.1.1 above do not appear in the data analyzed.

(63) **Who** ate it? – Smns (7;6)
(64) Puppy, **what** are you doing over there? – Dvdc (6;3)
(65) **What** was that? – Dvdc (8;3)
(66) **Why** it's always fall off? – Cndx (6;8)

Concerning pied-piping, a division was made in terms of it being compulsory, as in (71) or optional, as in (72).

(67) **Which** one is that other meanies guy's name? – Dvdc (8;3)
(68) **What** grade is she in? – Smns (7;0)

The adulthood of the utterances is another property that was taken into account when classifying the data, dividing the selected data into adult-like examples, as in (73) and non-adult-like examples, as in (74).

(69) **What** do you mean? – Jnnh1 (5;10)
(70) **What** she likes to talk about? – Jnnh5 (7;11)

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 (75), subject missing, as in (76), lack of subject-verb agreement, as in (77), as well as wrong verbal tense and wrong verbal tense form, as in (78).

(71) **What** you say thank you for? – Jonathan (2;9)
(72) **Where** is? – Jonathan (2;9)
(73) Well **who** is your friends? – Debbie (1;11)
(74) **What** did happened to the third little pig? – Olivier (2;10)

4.3 Data Analysis

This section is divided into 2 different parts: the first part deals with the computerized programs used to analyze the data, and, taking this as a point of departure, 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.

4.3.1 Automatic searching: the CLAN programs

In order to study the data from the 3 selected corpora, 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.

The MLU program calculates the mean length of utterance (MLU), that is to say, the ratio of morphemes or words per utterances. A typical MLU output appears in (79).

```
(79)
mlu  +t*CHI @
From file <c:\PARADIS\PARADIS\round5\yssf5.cha>
MLU for Speaker: *CHI:
        Number of: utterances = 485, morphemes = 1597
        Ratio of morphemes over utterances = 3.293
        Standard deviation = 2.143
```

The MLU output in (79) includes the following information about a chosen speaker (Yssf, in this case) within a certain file or files (file <yyssf5>, in this case which appears as @ in the syntax line <mlu +t*CHI @>): the number of utterances (485), the number of words/morphemes (1597), the ratio of morphemes over utterances (i.e. the actual MLU value: 3.2) and the standard deviation (2.1).

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 or files. A typical FREQ output is shown in (76).

```
(80)
freq+t*CHI +s"where"  @
From file <c:\PARADIS\PARADIS\round5\yssf5.cha>
Speaker: *CHI:
  23 where
------------------------------
   1 Total number of different item types used
  23 Total number of items (tokens)
```

The FREQ output in (80), provides the number of times the word *where* has been uttered by Yssf in the file <yssf5> and this amounts to 23 occurrences.

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 (81).

```
(81)
kwal  +t*CHI  +s"what" -w2+w2  @
From file <c:\PARADIS\PARADIS\round5\yssf5.cha> line 810. Keyword: where
*CHI:     no!
*EXP:    no?
*CHI:    what is it mama?
*CHI:    what?
*CHI:    what is what? xx.
```

In (81) KWAL was used to provide the context in which *what* has been produced by Yssf. In file <yssf5>, 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 (81) KWAL shows the 2 utterances before (-w2) and 2 utterances following (+w2) the target utterance. If more context is needed, the corresponding commands in the syntax line can be changed in order to show more context.

The search carried out with these automatic programs provided the following information regarding each analyzed corpus: the total number of *wh*-words used that fitted the inclusion criteria; 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 used to establish the 2 L1 group stages. This information is the basis for the grammatical analysis shown next.

4.3.2 Grammatical analysis

The data belonging to the 12 participants from the 3 selected corpora (see section 4.1.2) are presented in this section. As previously mentioned, these data are analyzed in terms of *wh*-movement in question formation and accounting for 4 factors: the type of *wh*-element used, the function of the *wh*-constituent, the complexity in terms of pied-piping, and the adulthood of the examples. These data belonging to the 12 participants are shown in tables 7, 8, 11, 12, 15 and 16.

Moreover, and with the aim to shed light as to whether biological gender has an impact on the acquisition of *wh*-movement in a bilingual context (and to compare it to English in L1 and L2 contexts), the data belonging to the 12 participants have also been classified in terms of preference associated to biological gender. This data classification is shown in tables 9, 10, 13, 14, 17 and 18.

Tables 7 and 8 provide information regarding the performance of the 2L1, L2 and L1 participants. Table 7 focuses on the preference of *wh*-type comparing 2L1 and L1 children; while table 8 is intended to study the same issue comparing L2 and L1 children.

Table 7. *Wh*-type – 2L1 and L1		
	2L1 – all 4 participants	**L1 – all 4 participants**
What	12 – (27.9%)	12 – (42.8%)
Why	3 – (6.9%)	0
When	0	0
Where	27 – (62.7%)	13 – (46.4%)
How	0	1 – (3.5%)
Who	1 – (2.3%)	0
Whom	0	0
Whose	0	0
Which	0	2 – (7.1%)
Total	43 – (100%)	28 – (100%)

Table 7 shows that, when comparing the production of *wh*-questions of the 2L1 and L1 groups (whose MLU rate ranges from 1 to 3), the preferred *wh*-element by both groups is *where* (62.7% in the case of the 2L1 group and 46.4% in the case of the monolingual group). Likewise, the second preferred *wh*-phrase in both groups of participants is *what*. However, it is important to mention that, in the case of the L1 group, *where* (46.4%) is closely followed by *what* (42.8%). This does not happen in the 2L1 group, since there is a higher difference in terms of percentage between *where* and *what* (62.7% *versus* 27.9%). Moreover, table 7 also illustrates that *wh*-elements such as *when*, *how*, *who*, *whom*, *whose* and *which* are less frequent in the production of both groups of participants. The results in this table therefore suggest that these *wh*-phrases start being produced later on and that, therefore, are less frequent in these children's output.

Table 8. *Wh*-type – L2 and L1		
	L2 – all 4 participants	**L1 – all 4 participants**
What	43 – (65.1%)	74 – (50.6%)
Why	2 – (3%)	34 – (23.2%)
When	4 – (6%)	1 – (0.6%)
Where	10 – (15.1%)	16 – (10.9%)
How	4 – (6%)	8 – (5.4%)
Who	1 – (1.5%)	9 – (6.1%)
Whom	0	0
Whose	0	1 – (0.6%)
Which	2 – (3%)	3 – (2%)
Total	66 – (100%)	146 – (100%)

In terms of preference of *wh*-type in the case of the L2 and L1 groups of participants (whose MLU rate ranges from 3 to 5), table 8 shows that the preferred *wh*-form for both groups is *what* (65.1% in the case of the L2 group and 50.6% in the case of the L1

group). Nevertheless, the second preferred option is different for the 2 groups: while *where* (15.1%) is the next most frequent *wh*-phrase in the production of the L2 group, *why* (23%) is the second *wh*-word of choice for the L1 group. It is also important to mention that, despite the fact that 8 out of 9 *wh*-forms are uttered in these 2 groups of participants, their production of *wh*-questions is clearly dominated the 3 *wh*-types mentioned in this paragraph (*what*, *where*, and *why*).

The results shown in tables 8 and 9 are in line with Gavruserva and Thornton's (1999) conclusions discussed in section 2.1.4 with regards to *whose* constructions being complex constructions. In fact, 1 instance of *whose* appears in all the data analyzed.

Tables 9 and 10, deal with the preference of the 3 groups of participants in terms of *wh*-constituent, and the data appear divided in terms of biological gender and *wh*-type.

Table 9. *Wh*-type – 2L1 and L1: M *versus* F participants				
	2L1 group		**L1 group**	
	M	**F**	**M**	**F**
What	12 – (100%)	0	0	12 – (100%)
Why	3 – (100%)	0	0	0
When	0	0	0	0
Where	8 – (29.6%)	19 – (70.3%)	2 – (15.3%)	11 – (84.6%)
How	0	0	0	1 – (100%)
Who	1 – (100%)	0	0	0
Whom	0	0	0	0
Whose	0	0	0	0
Which	0	0	0	2 – (100%)
Total	24 – (55.8%)	19 – (44.1%)	2 – (7.1%)	26 – (92.8%)

As shown in table 9, male participants produce an overall higher number of *wh*-types than females (55.8% *versus* 44.1% respectively) in the case of the 2L1 group. Although this difference is not very significant, the fact that females only produce 1 *wh*-form (*where* – 70.3%) contrasts with the male production, in which 4 *wh*-types are identified. On the contrary, in the case of the L1 group, female participants produce an overall higher number of *wh*-forms than males (92.8% *versus* 7.1%). Moreover, the production of the female participants in this group is dominated by 4 different *wh*-types, whereas in the case of the males, only 1 (*where*) is identified.

Table 10. *Wh*-type – L2 and L1: M *versus* F participants				
	L2 group		**L1 group**	
	M	**F**	**M**	**F**
What	24 – (55.8%)	19 – (44.1%)	60 – (81%)	14 – (18.9%)
Why	0	2 – (100%)	24 – (70.5%)	10 – (41.6%)
When	3 – (75%)	1 – (25%)	1 – (100%)	0
Where	7 – (70%)	3 – (30%)	11 – (68.7%)	5 – (31.2%)
How	3 – (75%)	1 – (25%)	6 – (75%)	2 – (25%)
Who	1 – (100%)	0	3 – (33.3%)	6 – (66.6%)
Whom	0	0	0	0
Whose	0	0	0	1 – (100%)
Which	2 – (100%)	0	3 – (100%)	0
Total	40 – (60.6%)	26 – (39.3%)	108 – (73.9%)	38 – (26%)

Table 10 shows that, in both groups of L2 and L1 participants, males produce a higher overall number of *wh*-types (60.6% *versus* 39.3% in the case of the L2 group; and 73.9% *versus* 26% in the case of the L1 group). More specifically, in the case of the L2 group, males produce a higher percentage each *wh*-word than females. However, there is an exemption: the *wh*-form *why*. With regards to the L1 group, male participants also produce a higher percentage of each *wh*-type than females, with the exception of *who* (66.6% *versus* 33.3%), and *whose*. However, these results may not be related to biological gender and have to do with the idiosyncrasy of each child instead.

Tables 11 and 12 provide information regarding the performances of the different groups in terms of *wh*-function type. Table 11 compares 2L1 and L1 children, while table 10 compares L2 and L1 children.

Table 11. *Wh*-type – 2L1 and L1		
	2L1 – all 4 participants	**L1 – all 4 participants**
Subject	6 – (13.9%)	0
Object	4 – (9.3%)	10 – (34.4%)
Adjunct	30 – (69.7%)	14 – (48.2%)
S.C.	3 – (6.9%)	5 – (17.2%)
Total	43 – (100%)	29 – (100%)

Table 11 shows that the preferred *wh*-function type in the groups of 2L1 and L1 children is that of adjunct (69.7% and 48.2% respectively). However, differences between both groups arise when dealing with the next most favored functions. In this sense, the second and third preferred functions in the 2L1 group are those of subject (13.9%) and object (9.3%). On the contrary, the object function (34.4%) is the second

preferred option for the L1 group, and the third most favored function is that of subject complement (17.2%).

Table 12. *Wh*-function – L2 and L1		
	L2 – all 4 participants	**L1 – all 4 participants**
Subject	4 – (6%)	10 – (6.8%)
Object	26 – (39.3%)	43 – (29.6%)
Adjunct	18 – (27.2%)	63 – (43.4%)
S.C.	18 – (27.2%)	29 – (20%)
Total	66 – (100%)	145 – (100%)

Table 12 illustrates that, in the case of the L2 speakers, the preferred grammatical function is that of subject complement (39.3%), while in the case of the L1 group, the preferred functions is that of adjunct (43.4%). The second most preferred functions in the L2 group are those of adjunct and subject complement (27.2% each one of them). However, the second most favored function in the L1 group is that of object (29.6%), which is closely followed by the subject complement function (20%).

The results in tables 11 and 12 lead us to believe that, although the adjunct function is preferred by the three groups of participants, this choice of function becomes less frequent in the production of the older groups of participants. Instead of dominating the participants' production, the adjunct function (although still the most frequent one) shares importance with other *wh*-function types, such as the object function and the subject complement function. This may be explained by the fact that these other 2 functions are acquired later than the adjunct function. Moreover, the subject function being quite infrequent in the participants' production has to do with the fact that subject *wh*-questions are acquired later due to their complexity. This is in line with Van Valin's (1998) empirical analysis (section 2.1.4), which points that children do not learn subject *wh*-questions first.

Tables 13 and 14 deal with the preference of the 3 groups of participants in terms of the grammatical function of the *wh*-constituent, too, but this time dividing the data in terms of biological gender.

Table 13. *Wh*-function – 2L1 and L1: M *versus* F participants				
	2L1 group		L1 group	
	M	F	M	F
Subject	6 – (100%)	0	0	0
Object	4 – (100%)	0	0	10 – (100%)
Adjunct	11 – (36.6%)	19 – (63.3%)	2 – (14.2%)	12 – (85.7%)
S.C.	3 – (100%)	0	1 – (20%)	4 – (80%)
Total	24 – (55.8%)	19 – (44.1%)	3 – (10.3%)	26 – (89.6%)

Table 13 shows that, in the case of the 2L1 group, males produce a higher overall number of *wh*-function types (55.8% *versus* 44.1% respectively). More specifically, males produce a higher percentage of every function with one exception, that of adjunct (females producing 63.3% of the adjunct function *versus* males producing 36.6%). Regarding the L1 group, females produce not only a higher overall number of *wh*-function types (89.6% *versus* 10.3%), but also a higher percentage of every function classified. These results suggest that the preference of production of these *wh*-function types is not conditioned by the biological gender of the participants.

Table 14. *Wh*-function – L2 and L1: M *versus* F participants				
	L2 group		L1 group	
	M	F	M	F
Subject	2 – (50%)	2 – (50%)	4 – (40%)	6 – (60%)
Object	9 – (34.6%)	17 – (65.3%)	29 – (67.4%)	14 – (32.5%)
Adjunct	12 – (66.6%)	6 – (33.3%)	48 – (76.1%)	15 – (23.8%)
S.C.	17 – (94.4%)	1 – (5.5%)	26 – (89.6%)	3 – (10.3%)
Total	40 – (60.6%)	26 – (39.3%)	107 – (73.7%)	38 – (26.2%)

As seen in table 14, males produce a higher overall number of *wh*-function types in the case of the L2 and L1 groups (60.6% and 73.7% respectively). Furthermore, male participants produce a higher percentage of every function with one exception in both groups (the object function being preferred by females in the L2 group (65.3%), and the subject function being favored by females in the case of the L1 group (60%)). Despite the fact that males produce more *wh*-function types and a wider variety of them does not prove that biological gender has a role in it, as has been shown in table 14.

The complexity aspect in terms of pied-piping is dealt with in tables 15 and 16, which account for the *wh*-questions that featured pied-piping produced by the three groups of participants collapsed, distinguishing between adult-like and non-adult-like cases.

Table 15. Pied-piping – 2L1 and L1				
	2L1 group		L1 group	
	Compulsory	Optional	Compulsory	Optional
Adult-like	0	0	1 – (50%)	0
Non-adult-like	1 – (100%)	0	1 – (50%)	0
Total	1 – (100%)	0	2 – (100%)	0

Table 15 shows that, although compulsory pied-piping is favored over optional pied-piping in the production of both groups of participants, the rate of pied-piped *wh*-questions in relation to the total number of *wh*-questions produced by both 2L1 and L1 groups is extremely low (1 and 2 instances respectively). The fact that pied-piping is only used in compulsory contexts, and also the fact that the production of this type of structure is not dominated by adult-like instances points to its grammatical complexity.

Table 16. Pied-piping – L2 and L1				
	L2 group		L1 group	
	Compulsory	Optional	Compulsory	Optional
Adult-like	8 – (88.8%)	1	7 – (63.6%)	6
Non-adult-like	1 – (11.1%)	0	4 – (36.3%)	0
Total	9 – (100%)	1 – (100%)	11 – (100%)	6 – (100%)

Table 16 shows that in the case of the L2 and L1 group of participants, pied-piping is favored in the production of both of these groups of participants. As mentioned in the previous table, the rate of *wh*-questions which feature pied-piping in relation to the total number of *wh*-questions produced is extremely low (10 and 17 instances respectively). Although the production of the L2 and L1 groups is dominated by adult-like instances of pied-piped *wh*-questions, the fact that these structures are favored in compulsory contexts suggests that they are grammatically complex.

In terms of the adult-like production in these 3 different groups of participants, 2 different analyses have been carried out: the first one, analyzing the overall production of all the participants belonging to each group; the second one, comparing the performance of the male and the female participants. These analyses strive to answer the question regarding how *wh*-movement is acquired in a bilingual context (and to compare it with L2 and L1 contexts), and whether or not biological gender has an effect on the acquisition process of *wh*-movement.

Table 15 and table 16 deal with the rate of correctness of the participants from the 3 corpora, including the participants' name, their gender, the number of *wh*-questions produced by them, and the instances of adult-like and non-adult-like cases.

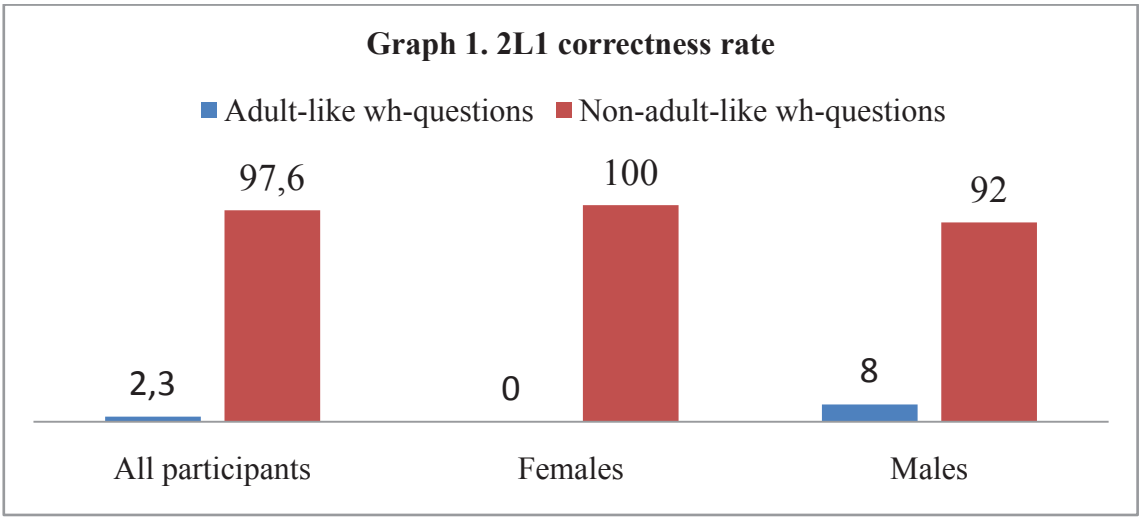
Table 17. Rate correctness of all the participants				
The Genesee corpus: all 4 participants				
	Gender	***Wh*-questions**	**Adult-like**	**Non-adult-like**
Gene	M	15 – (100%)	1 – (6.6%)	14 (– 93.3%)
Jessica	F	0	0	0
Joelle	F	19 – (100%)	0	19 – (100%)
Olivier	M	9 – (100%)	0	9 – (100%)
TOTAL:		43 – (100%)	1 – (2.3%)	42 – (97.6%)
The Wells corpus: all 4 participants (MLU rate from 1 to 3)				
	Gender	***Wh*-questions**	**Adult-like**	**Non-adult-like**
Jonathan	M	0	0	0
Abigail	F	16 – (100%)	8 – (50%)	8 – (50%)
Debbie	F	10 – (100%)	5 – (50%)	5 – (50%)
Benjamin	M	2 – (100%)	1 – (50%)	1 – (50%)
TOTAL:		28 – (100%)	11 – (50%)	11 – (50%)

Table 17 shows that the production of the 2L1 participants is clearly dominated by non-adult-like instances (97.6% of non-adult-like *wh*-questions *versus* 2.3% of adult-like examples). In the case of the L1 children, there are the same amount of adult-like and non-adult-like instances (50% in both of them). The results obtained from the analysis of the data belonging to both groups of 2L1 and L1 speakers can be explained by their grammatical age measured in terms of their MLU rates, since an overturn regarding the adulthood of their production is expected to happen as they develop linguistically.

Table 18. Rate correctness of all the participants				
The Paradis corpus: all 4 participants				
	Gender	***Wh*-questions**	**Adult-like**	**Non-adult-like**
CNDX	F	13	7 (53.8%)	6 (46.1%)
DVDC	M	18	14 (77.7%)	4 (22.2%)
JNNH	F	13	11 (84.6%)	2 (15.3%)
SMNS	M	22	21 (95.4%)	1 (4.5%)
TOTAL:		66 (100%)	53 (80.3%)	13 (19.6%)
The Wells corpus: all 4 participants (MLU rate from 3 to 5)				
	Gender	***Wh*-questions**	**Adult-like**	**Non-adult-like**
Jonathan	M	92 – (100%)	73 – (79.3%)	19 – (20.6%)
Abigail	F	13 – (100%)	9 – (69.2%)	4 – (30.7%)
Debbie	F	25 – (100%)	15 – (60%)	10 – (40%)
Benjamin	M	16 – (100%)	11 – (68.7%)	5 – (31.2%)
TOTAL:		146 – (100%)	108 – (73.9%)	38 – (26%)

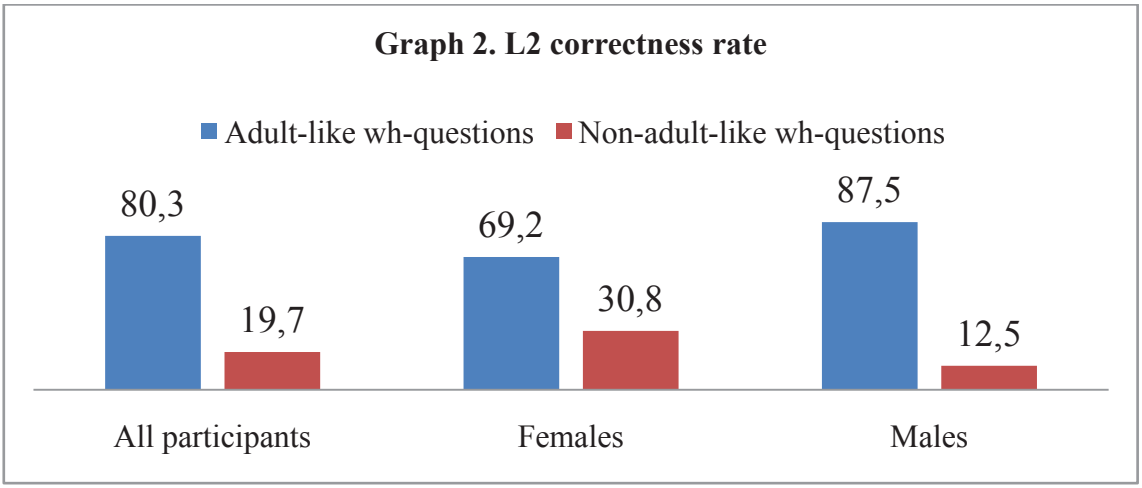
Table 18 shows that the production of both groups of L2 and L1 participants is dominated by adult-like instances (80.3% and 73.9%). By the time the children have reached this grammatical age (measured by their MLU rate ranging from 3 to 5), the overturn regarding the adulthood of their production has already taken place, as it can be seen in the increase of adult-like production shown in this table 18 if compared to that in table 17.

Graph 1 compares the production of adult-like *wh*-questions and non-adult-like *wh*-questions from the group of 2L1 bilinguals so that a comparison can be established between their overall production (as also seen in tables 17 and 18 above) and the production of the male and female participants separated.



As for the group of 2L1 participants, graph 1 shows that the overall production is dominated by non-adult like *wh*-questions (97% of adult-like *wh*-questions *versus* 2.3% of non-adult-like questions). While the female performance shows a 100% of non-adult-like instances; the male production shows that 8% of the *wh*-questions uttered by these children are adult-like, and 92% of them are non-adult-like. These results suggest that the adulthood of these children's production is not related to their biological gender given the low percentage of adult-like production that separates males and females. It is also necessary to point out, too, that, even if the difference is small between the 2 gender groups, the only group that contributes grammatical structures is the male group.

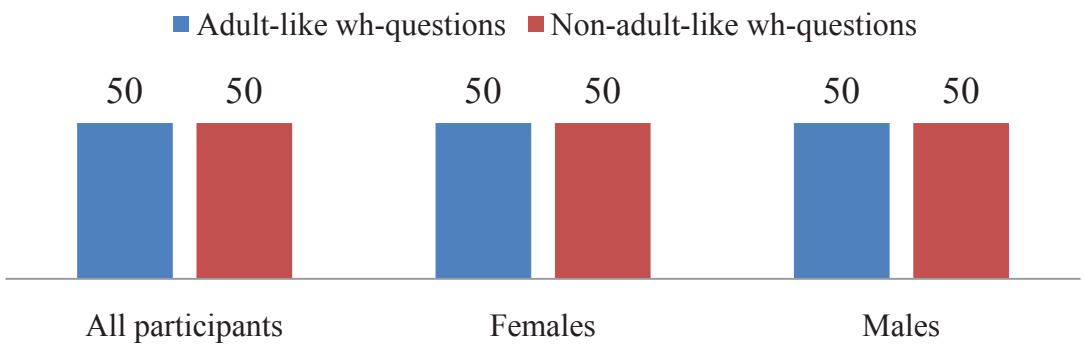
Likewise, graph 2 compares the adult-like and non-adult-like production of the participants belonging to the L2 group (in overall terms, as well as making a comparison between the production of male and female participants).



Regarding graph 2, which deals with the group of children learning English as an L2, it can be seen that 80.3% of the production features adult-like *wh*-questions (in contrast with the 19.7% of non-adult-like instances). In terms of gender, it is also shown that the female production features a 69.2% of adult-like production (in contrast with the 30.8% of non-adult-like instances). As for the male production, graph 2 also shows that 87.5% of the *wh*-questions produced by these participants are adult-like instances; while the other 12.5% features non-adult like *wh*-questions. The production being dominated by adult-like examples can be explained due to the grammatical age of the participants, which is significantly higher than the one of the 2L1 group shown in graph 1 above. Moreover, graph 2 does not provide evidence that biological gender plays a role on the adulthood of the production of these children, since the difference between the adulthood of the production of both genders is not significant.

Graph 3 deals with the production of the L1 group of participants in terms of adulthood (in overall terms and comparing the production of male and female children as well) in stage 1 of the analysis, that is to say, when the MLU rate ranges from 1 to 3.

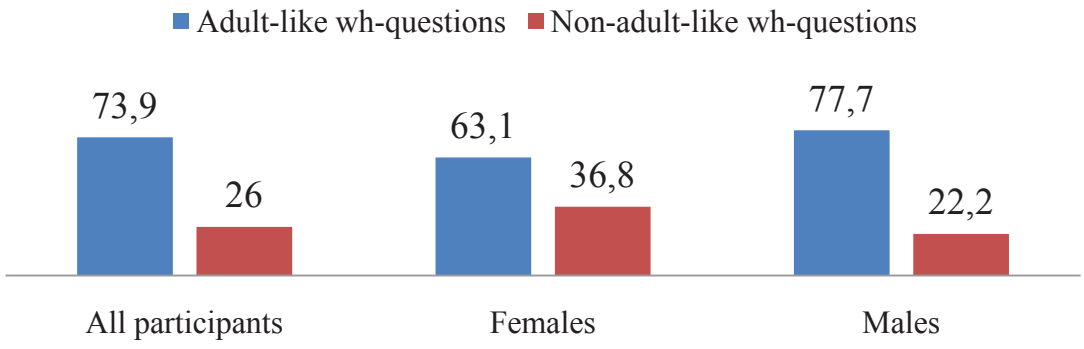
Graph 3. L1 group correctness rate - MLU rate from 1 to 3



Graph 3 shows that 50% of the production of L1 speakers in stage 1 features adult-like *wh*-questions, while 50% of these *wh*-questions are non-adult-like. With regards to gender, it can be seen that the female production of *wh*-questions features 50% of adult-like examples (in contrast with the 50% of non-adult-like instances). The results in this graph suggest that there is a balance between the production of adult-like and non-adult-like instances, which points at the fact that *wh*-questions are still in the process of being acquired. Regarding gender, graph 3 shows that there is no difference in terms of adulthood between the production of males and females (50% of adult-like occurrences *versus* 50% of non-adult-like occurrences in the case of both genders). These results suggest that biological gender does not play a role in the acquisition of *wh*-movement in this stage.

Likewise, graph 4 compares the production of the L1 group of participants in terms of adulthood (in overall terms and comparing the production of male and female children as well) in stage 2 of the analysis, that is to say, when the MLU rate ranges from 3 to 5.

Graph 4. L1 group correctness rate - MLU rate from 3 to 5



Graph 4 shows that 73.9% of the production of L1 speakers in stage 2 features adult-like *wh*-questions, while 26% of these *wh*-questions are non-adult-like. With regards to gender, it can be seen that the female production of *wh*-questions features 63.1% of adult-like examples (in contrast with the 36.8% of non-adult-like instances). Furthermore, the male production of *wh*-questions features 77.7% of adult-like examples (as opposed to the 22.2% of adult-like instances). Moreover, graph 4 also shows that not only is the percentage of adult-like *wh*-questions higher in the case of male production (77.7% *versus* the 63.1% produced by females), but also the percentage of non-adult-like *wh*-questions is lower in the case of males (22.2% *versus* 36.8% in the case of female production). As suggested by the results seen in this graph, the production of the L1 group of participants in stage 2 being dominated by adult-like instances is related to their linguistic development as they grow older (graph 3 *versus* graph 4). Furthermore, the difference between genders in terms of adulthood is not so significant so as to suggest that it has to do with the biological gender of the participants.

Moreover, graphs 1 and 3 show that, the production of the L1 group of participants does not resemble that of the 2L1 bilingual children in terms of adulthood, since non-adult-like *wh*-questions dominate the production of the simultaneous bilinguals, which is not the case for the L1 group (97.6% of non-adult-like occurrences in the case of the 2L1 children *versus* 50% of non-adult-like *wh*-questions in the case of the L1 group). However, if we compare the results obtained from the data belonging to the L2 and L1 groups (as seen in graphs 2 and 4), it can be seen that the production of these 2 groups is dominated by adult-like instances (80.3% in the case of the L2 group and 73.9% in the case of the L1 children). These results also show that the 2L1 production of adult-like *wh*-questions is lower than that of the English monolingual group, and the L1 production of non-adult-like *wh*-questions is higher than that of the L1 group of participants.

Although the production of males and females in the 3 groups of 2L1, L1 and L2 children is not very different, it is important to mention that males behave more adult-like linguistically speaking, as they not only produce more adult-like occurrences, but also less non-adult-like instances (with the only exception of the L1 group in stage 1). The small advantage that boys have over girls in relation to several linguistic skills (as discussed by Özçalışkan and Goldin-Meadow (2010), Huttenlocher et al. (1991),

Hadley et al. (2011), among others) cannot be applied to the acquisition of *wh*-movement, as it has been shown in the 4 graphs presented above.

5. Conclusions

An empirical analysis of *wh*-movement in question formation has been carried out in the present dissertation in order to explore not only the nature of this syntactic phenomenon, but also its acquisition. And more specifically, since one of the main aims of this study is to analyze the impact of biological gender on the acquisition of *wh*-movement, the selected data (which were extracted from the Wells corpus, the Paradis corpus and the Genesee corpus in CHILDES) have been analyzed and classified in terms of the biological gender of the participants considering the preference of *wh*-type and *wh*-function, and the adulthood of the production as well. This analysis has led to a number of conclusions being reached in relation to the research questions and the hypotheses presented in section 3.

Concerning research question 1, the analysis of the data shows that biological gender does not have an impact on the acquisition of *wh*-movement in any of the 3 groups of participants whose data have been analyzed. While it is true that differences do appear between male and female productions in the 3 groups, they are not strong enough to prove that biological gender does in fact play a role on the acquisition of this linguistic phenomenon. This is in line with previous works on the role of gender in acquisition (Hyde and Lynn, 1988). However, it is important to remember that, in fact, males produce more adult-like *wh*-questions and less non-adult-like *wh*-questions than their female peers.

Regarding research question 2, the analysis of the data shows that the production of the 2L1 bilingual group is not similar to that of the monolingual group. These participants do not produce *wh*-elements in the same proportion and adult-like rate, as it has been shown in the tables presented in the previous section of this dissertation.

Moreover, and with respect to research question 3, this analysis reveals that, although there is not a delay effect in the production of the 2L1 group, *wh*-movement questions appear more frequently in the production of monolingual children than in the production

of simultaneous bilingual children. In addition to this, the production of these 2 groups of participants is not similar in terms of adulthood.

Finally, and to account for the hypotheses of this dissertation, it can be established that:

- *Wh*-movement and the different properties it entails (order of acquisition and grammatical function of the *wh*-element, as well as their grammatical complexity) is gradually acquired. Hypothesis 1 has consequently been confirmed.
- A delay in the acquisition of *wh*-movement by simultaneous bilingual speakers is not found in the data analyzed. Thus, hypothesis 2 has been rejected.
- No substantial differences across genders appear in the children's production in terms of preference for a specific *wh*-type and function, as well as in terms of the rate of correctness of their production of *wh*-questions. Hypothesis 3 has therefore been confirmed.

Further research can shed light on the acquisition of *wh*-movement by bilingual speakers and whether or not biological gender plays a role in it. In this sense, and although the present dissertation only focuses on the analysis of the English data elicited from the bilingual group of speakers, it can also be insightful to study the French data, that is, their other L1, with the aim to compare the children's production in their 2 L1s and determine whether the conclusions reached from the analysis of the English data can also be applied to the other language these children speak as an L1. In addition, further analyses on bigger groups of participants would shed more light as to whether biological gender has an impact on the rate correctness of the production of *wh*-questions.

6. Bibliography

- Barbu, Stéphanie, Aurélie Nardy, Jean-Pierre Chevrot, and Jacques Juhel. "Language Evaluation and Use during Early Childhood: Adhesion to Social Norms or Integration of Environmental Regularities?" *Linguistics*. 51.2 (2013): 381-411. Print.
- Barnes, Julia D. *Early Trilingualism: A Focus on Questions*. Clevedon, England: Multilingual Matters, 2006. Print.

- Bauer, Daniel J., Beverly A. Goldfield, and J. Steven Reznick. "Alternative Approaches to Analyzing Individual Differences in the Rate of Early Vocabulary Development." *Applied Psycholinguistics*. 23.3 (2002): 313-335. Print.
- Berglund, Eva, Marten Eriksson, and Monica Westerlund. "Communicative Skills in Relation to Gender, Birth Order, Childcare and Socioeconomic Status in 18-month-old Children." *Scandinavian Journal of Psychology*. 46.6 (2005): 485-491. Print.
- Bouchard, Caroline, Natacha Trudeau, Ann Sutton, Marie-Claude Boudreault, and Joane Deneault. "Gender Differences in Language Development in French Canadian Children between 8 and 30 Months of Age." *Applied Psycholinguistics*. 30.04 (2009): 685-707. Print.
- 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, Takuya Goro, and Rosalin Thornton. "Language Acquisition is Language Change". *Journal of Psycholinguistic Research*. 35.1 (2005): 31-48. Print.
- De Houwer, Annick, and Jürgen Meisel. "Analyzing the relationship between two developing languages in bilingual first language acquisition: methodology, data and findings." *Workshop on Language Contact: Linking Different Levels of Analysis*, Netherlands Institute for Advanced Studies (NIAS). Wassenaar, the Netherlands. 1996. Print.
- Fenson, Larry, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal, Stephen J. Pethick, Michael Tomasello, Carolyn B. Mervis, and Joan Stiles. "Variability in Early Communicative Development." *Monographs of the Society for Research in Child Development*. 59.5 (1994): 1-173. Print.

- Galsworthy, Michael J., Ginette Dionne, Philip S. Dale, and Robert Plomin. "Sex Differences in Early Verbal and Non-verbal Cognitive Development." *Developmental Science*. 3.2 (2000): 206-215. Print.
- Gavruseva, 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. 21 Feb. 2016.
- Hadley, Pamela A., Matthew Rispoli, Colleen Fitzgerald, and Alison Bahnsen. "Predictors of Morphosyntactic Growth in Typically Developing Toddlers: Contributions of Parent Input and Child Sex." *Journal of Speech Language and Hearing Research*. 54.2 (2011): 549-566. Print.
- 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.
- Hulle, Carol A. Van, H. H. Goldsmith, and Kathryn S. Lemery. "Genetic, Environmental, and Gender Effects on Individual Differences in Toddler Expressive Language." *Journal of Speech Language and Hearing Research*. 47.4 (2004): 904-912. Print.
- Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons. "Early Vocabulary Growth: Relation to Language Input and Gender." *Developmental Psychology*. 27.2 (1991): 236-248. Print.
- Hyde, Janet S., and Marcia C. Linn. "Gender Differences in Verbal Ability: A Meta-analysis." *Psychological Bulletin*. 104.1 (1988): 53-69. Print.
- Kern, Sophie. "Lexicon Development in French-speaking Infants." *First Language*. 27.3 (2007): 227-250. Print.
- Lovas, Gretchen S. "Gender and Patterns of Language Development in Mother-toddler and Father-toddler Dyads." *First Language*. 31.1 (2010): 83-108. Print.
- Lutchmaya, Svetlana, Simon Baron-Cohen, and Peter Raggatt. "Foetal Testosterone and Vocabulary Size in 18- and 24-month-old Infants." *Infant Behavior and Development*. 24.4 (2001): 418-424. Print.
- MacWhinney, Brian. CHILDES - *Child Language Data Exchange System*. 2000. Web. 14 Apr. 2016. <http://childes.psy.cmu.edu/>.
- MacWhinney, Brian. *Talkbank*. 2003. Web. 23 Apr. 2016. <https://talkbank.org/>.

- Meisel, Jürgen M. "The Simultaneous Acquisition of Two First Languages." *Trends in Bilingual Acquisition Trends in Language Acquisition Research*. (2001): 11-41. Print.
- Özçalışkan, Şeyda, and Susan Goldin-Meadow. "Sex Differences in Language First Appear in Gesture." *Developmental Science*. 13.5 (2010): 752-760. Print.
- Paradis, Johanne, and Fred Genesee. "Syntactic Acquisition in Bilingual Children." *Studies in Second Language Acquisition.* 18.1 (1996): 1-25. Print.
- Reilly, Sheena, Mark Onslow, Ann Packman, Melissa Wake, Edith L. Bavin, Margot Prior, Patricia Eadie, Eileen Cini, Catherine Bolzonello, and Obioha C. Ukoumunne. "Predicting Stuttering Onset by the Age of 3 Years: A Prospective, Community Cohort Study." *Pediatrics*. 123.1 (2009): 270-277. Print.
- Simonsen, Hanne G., Kristian E. Kristoffersen, Dorthe Bleses, Sonja Wehberg, and Rune N. Jørgensen. "The Norwegian Communicative Development Inventories: Reliability, Main Developmental Trends and Gender Differences." *First Language*. 34.1 (2014): 3-23. Print.
- Stromswold, Karin. "The Acquisition of Subject and Object Wh-Questions." *Language Acquisition*. 4.1-2 (1995): 5-48. Web. 22 Feb. 2016.
- Ramer, Andrya L. H. "Syntactic Styles in Emerging Language." *Journal of Child Language*. 3.1 (1976): 49-62. Print.
- Trotta, Joe. *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. 10 Mar. 2016.
- 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.

- Westerlund, Monica, and Dagmar Lagerberg. "Expressive Vocabulary in 18-month-old Children in Relation to Demographic Factors, Mother and Child Characteristics, Communication Style and Shared Reading." *Child: Care, Health and Development*. 34.2 (2008): 257-266. Print.
- Yule, George. *The Study of Language*. Cambridge: Cambridge University Press, 2014. Print.