Formulaic language and L2 syntactic development

A multiparadigmatic approach

Thomas Antony Hammond

A thesis presented for the degree of

Doctor of Philosophy

Department of English Language and Linguistics

University of Sheffield

July 2024

Declaration

I, Thomas Hammond, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Please note that some of the results and findings have been published in the two articles below, which are both available via the institutional repository, White Rose Research Online (WRRO) without embargo:

Hammond, T.A. and Gil, K.H (2023) "Fixed wh-expressions in classroom second language acquisition: Databases of rules or schemas"? *Journal of the European Second Language Association*, *7*(1), 75-92. https://doi.org/10.22599/jesla.100

Hammond, T.A. and Gil, K.H (2023) "Formulaic chunks, derivational complexity and L1 transfer at the initial state of L2 French". In Paris Gappmayr and Jackson Kellog (eds.) *Proceedings of the 47th annual Boston University Conference on Language Development* (pp. 342-355). Somerville, MA: Cascadilla Press.

Acknowledgements

The biggest acknowledgement and thanks goes to my supervisor, Kook-Hee Gil, without whom this thesis would not have been possible. Her knowledge, expertise and professionalism within the discipline and academia in general have been invaluable for my own understanding of the subject and overall career development. I am extremely grateful to have had the opportunity to work closely with her over the last four years, both as a student and as a friend.

Ideas from this thesis have been presented/published in various outlets over the last four years. I am grateful to all conference audience members, anonymous reviewers and editors for their comments and criticisms. All remaining errors are my own.

The generous funding support from WRoCAH also made this project feasible. Further thanks should go to members of staff and students at the School of English who have helped me along the way in various aspects. In particular, Bea González Fernández, Gabriel Ozón and Yi Liu.

Finally, a thanks to my family and friends, who are the constants that have kept me sane throughout this whole process.

Abstract

This thesis presents a novel, multiparadigmatic approach to the role of formulaic language in the development of second language (L2) syntax. It adopts theoretical frameworks and concepts from generative and usage-based linguistics to analyse two longitudinal corpora documenting interlanguage development from the initial state through transitional state grammars. These datasets constitute L1 Spanish/Catalan learners of L2 English, and L1 English learners of L2 French. Similar formulaic expressions (FEs) are identified in both corpora at the initial state that learners' are able to produce fluently upon contextual cues in absence of related L2 syntactic competence. I track learners' use of these FEs and evidence of related syntactic competence across the data collection periods, as measured by knowledge of underlying computational properties (generative) and schematic patterns (usage-based).

Through adopting a multiparadigmatic approach, similar developmental trends are observed in both corpora that would have otherwise been missed if relying on one framework alone. Outside of the FEs, learners first demonstrate knowledge of lexical categories only, before knowledge of L2 functional categories emerges later. This supports a Weak Continuity view of the initial state. Importantly, the use of FEs seems to be influential on the rate in which learners progress through this trajectory. Correlations are found between a more frequent use of these expressions at the initial state and a greater knowledge of related underlying computational properties longitudinally. Traceback analysis also reveals that the FEs have instantiated utterance schema extraction and generalisation across similar functional structures.

I discuss the interplay of these concepts by drawing on processing models of SLA, suggesting that schematic learning could provide learners with more syntactic and morphological distributional evidence needed for L2 syntactic activation levels to better compete with the existing L1 ones during processing.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The present study investigates the role of formulaic language (FL) on the development of second language (L2) syntax. It analyses longitudinal oral production data of classroom L2 learners taken from two learner corpora; The Barcelona English Language Corpus (BELC) and the French Progression Corpus (FPC). The study offers a unique approach to this topic by drawing on generative and usage-based models of grammar and approaches to second language acquisition (SLA). This chapter provides an overview of the main rationale of my thesis. It begins with a brief introduction to the general research background and research objectives. It then gives an overview of the study and a preview of its main findings. It concludes by outlining the organisation of the thesis.

1.1 Formulaic language in second language acquisition

Formulaic language is an umbrella term used to encompass various 'multiword' language phenomena that seem to be fixed in form, often non-literal in meaning and closely tied to communicative-pragmatic contexts (Van Lancker-Sidtis and Rallon, 2004). Relevant to the present study are conventional expressions, a sub-category of FL which are functional phrases preferred by native speakers in certain commu-

nicative contexts (Bardovi-Harlig, 2019). For example, *what is your name* and *how old are you* would be the English conventional expressions used by a native speaker to ask the name and age of another person, rather than *what are you called, what name do you go by* or *what is your age, what age are you.*

Conventional expressions like these above are clearly useful for the beginner L2 learner, as they can permit immediate interaction in the language-learning classroom/target-language community (Jeremias, 1982; Schmitt, 2010). Indeed, studies examining the interlanguage of L2 speakers at the initial state have found evidence that conventional expressions such as these are produced in advance of respective L2 abilities. Beginner learners have been shown to produce syntactically complex sequences such as *I don't know* and *do you like* fluently, whilst demonstrating no evidence for knowledge of their internal parts elsewhere (Hanania and Gradman, 1977; Schmidt, 1983; Myles et al., 1999; Bardovi-Harlig, 2009).

This discrepancy has led to the popular notion that highly frequent, prototypical phrases are 'entrenched' in beginner learners' minds. This means that, just as they are introduced as one unit in the second language classroom, these expressions are stored and retrieved holistically by learners as opposed to generated in a word-by-word fashion (Nattinger and DeCarrico, 1992; Wray, 2002; Ellis et al., 2015). The short term communicative benefits and perceived processing advantages that come with the holistic retrieval of prototypical, functional FL, has led researchers to enquire about the role that these expressions play in the SLA process. This concept has become a considerable subject of debate in the field of Applied Linguistics, which is divided between two main opposing fields of SLA. Sections 1.2 and 1.3 briefly outline usage-based and generative models of SLA, highlighting the role that FL is perceived to play in each of these frameworks.

1.2 Usage-based approaches

For usage-based linguistics (UBL), linguistic knowledge is a reflex of speakers' domain general cognitive abilities, defined specifically as 'a structured inventory of conventional units' (Langacker, 2000, p. 8). These models reject a sharp distinction between language knowledge and language use. Language learning is initially input-driven and exemplar-based (Roehr-Brackin, 2014), with the emergence of linguistic structure occurring 'through reinforcement of the commonality inherent in multiple experiences' (Langacker, 2000, p. 4). These models therefore reject a sharp distinction between language knowledge and language use, with knowledge of constructions/units emerging from the memories of utterances in usage events and the abstraction of regularities within them. Therefore, as well as input, interaction (in the sense of Long 1996) and output (in the sense of Swain 1985) are essential ingredients of language acquisition. It also follows that highly frequent and prototypical FL that is tied to specific communicative functions and therefore facilitates interaction and output, plays a central role in the development of L2 syntax. UBL proposes a bottom up learning trajectory in which learners move from formulaic phrases (entirely fixed conventional units) to lexically-specific utterance schemas (semi-fixed units) to more abstract schematic patterns (general grammatical category sequencing) (Ellis, 1996, 2012; Eskildsen, 2009; Ellis et al., 2015). These stages are exemplified below with the highly prototypical and conventional expression *what is your name*.

(1) a. *what is your name* [formulaic phrase]

b. [what is] + NOUN PHRASE [lexically-specific utterance schema]
eg. *[what is] the time, [what is] it, [what is] the plan*

c. [WH + COPULA] + X [abstract schematic pattern]
eg. *[what are] you doing, [when is] he coming, [why are] you sad*

A fundamental notion within UBL is that learners move from the formulaic to the

schematic (Eskildsen, 2015, 2020), which is achieved through the abstraction and generalisation of prototypical FL derived from their input (Roehr-Brackin, 2014). Researchers working within the UBL framework adopt what is known as 'traceback' methodology (Lieven et al., 2003, 2009) to explore the relationship between routinisation and creativity in SLA. In these studies, researchers analyse longitudinal data to find evidence that target language use has been instantiated by formulaic exemplars and/or utterance schemas. A number of recent studies investigating the development of various L2s have had success in this regard, demonstrating how learners' novel language utterances can be traced back to prototypical formulaic expressions that exemplify their schematic patterns (Eskildsen, 2015, 2020; Lesonen et al., 2020; Horbowicz and Nordanger, 2021).

1.3 Generative approaches

In generative frameworks, language is modular, and linguistic knowledge is conceptualised as an innate computational system (Chomsky, 1995). Syntax is formalised as 'Merge', which via the operation 'Select', takes items from the lexicon and forms composed elements through recursive computational procedures (Rizzi, 2009). These procedures, namely, computational properties, are driven by features on functional categories and result in a variety of overt surface forms. Merge and Select are universal syntactic operations, a part of Universal Grammar (UG), taken to be an innate endowment of human beings (Collins and Stabler, 2016). Whereas usage-based frameworks are completely input-driven, generative models are based on the 'Poverty of the Stimulus' (POS), which argues that the input is insufficient to explain the eventual linguistic knowledge a speaker comes to acquire (Chomsky, 1975, 1980). This is also known as the 'Logical Problem of Language Acquisition', which generative researchers extend to SLA based on the fact that L2 learners demonstrate knowledge which goes well beyond the L2 input they have may been exposed to (White, 1990).

Therefore, generative SLA is largely concerned with the interplay between UG, knowledge that comes from the L1 and knowledge that comes from exposure to the target language (i.e. the L2 input) (Rothman and Slabakova, 2018). There are competing theories within the paradigm as to how these aspects interact. For example, some models claim full transfer from the first language (L1) at the initial stages of SLA (known as the Strong Continuity Hypothesis) (Poeppel and Wexler, 1993; Schwartz and Sprouse, 1996), whereas others claim that only lexical categories are transferred from the L1 (e.g. NP, VP) and L2 phrase structure develops incrementally thereafter (known as the Weak Continuity Hypothesis) (Vainikka and Young-Scholten, 1998). In either case, the POS argument has meant that traditionally, there has been less emphasis on how properties of the input influence L2 development. Instead, generative SLA has tended to focus on discovering L2 knowledge that cannot have been acquired from the input, and how the L1 and/or UG can be drawn upon to explain this (White, 2012; VanPattern and Rothman, 2015; Rothman et al., 2019). Therefore, researchers within this framework favour the use of experimental designs whereby L2 learners judge and interpret sentences that they may never have encountered before (Juffs and Fang, 2022), rather than the analysis of learner data to discover how input is manipulated. Consequently, frequent and prototypical FL derived from learners' input has played no significant role in generative theories of L2 development. Rather, it has been viewed as a peripheral phenomena separate to the computational system that generates rule-governed, compositional language (Krashen and Scarcella, 1978; Bohn, 1986; Carroll, 2010; Bardovi-Harlig and Stringer, 2017).

1.4 Research Objectives

The above sections highlight the main differences between generative and usage-based models of SLA, which fundamentally come down to how much L2 knowledge can be acquired solely from the input and domain-general cognitive abilities. The strength of generative approaches is the explanatory power that the model offers

through linking surface structures together based on shared underlying computational properties/syntactic categories. This allows for the identification of common L2 developmental sequences based on their emergence in production data, irrespective of the target language. Generative studies fall short when it comes to understanding what learners actually *do* with the language they are exposed to. Usage-based frameworks adopting traceback methodology offer a much better account of this concept, demonstrating how learners can manipulate prototypical formulaic expressions from their L2 input to facilitate creative language use. However, traceback studies are limited in that they often cannot account for common L2 developmental phenomena that occurs outside of item-based, schematic learning.

The present study takes inspiration from recent calls to 'bridge the gap' between these traditionally opposed frameworks, given that SLA as a sub-field would benefit from a multiplicity of approaches (VanPattern and Rothman, 2015; Rastelli and Gil, 2018; Rothman and Slabakova, 2018). It aims to build on these strengths and weaknesses of both approaches to investigate the role of FL in SLA. I believe this area of enquiry is an ideal test ground for the integration of these two frameworks, which could also help for a better understanding of the interplay between L2 input, usage and the development of L2 knowledge more generally. Based on this, the following research objectives can be distinguished.

(2) a. Are formulaic expressions evident in learners' production data at the initial state, and do these influence the development of L2 syntax thereafter?

b. Can this development be captured by schematic learning strategies, or do we find the development of underlying syntactic knowledge more generally? That is, can novel L2 utterances be traced back to formulaic exemplars based on related utterance schemas, or related computational properties?

c. Is there evidence of a common developmental trajectory across learners,

regardless of the use of formulaic expressions and schematic learning strategies?

1.5 Overview of study

To investigate these research objectives, the study analyses two longitudinal learner corpora that document learners oral production data from the initial state and for a significant period after this. One is the Barcelona English Learner Corpus (BELC) (Muñoz, 2006), a collection of spoken transcripts from 9 Spanish/Catalan bilingual classroom learners of English. Naturalistic data was collected from students on four separate occasions over a period of 7 years (age 10, 12, 16 and 17). The other is the French Progression Corpus (FPC) (Mitchell and Martin, 1997; Myles et al., 1998, 1999), a collection of spoken transcripts from 24 classroom English learners of French. Naturalistic data was collected from students on three rounds over a period of two years. The corpora can be seen as complementary datasets; the BELC provides a more general picture of learners development over a longer period of time, whereas the FPC provides a more concentrated picture of L2 development at initial stages of learning.

The study first identifies prototypical formulaic language (FL) typical of the classroom learners' L2 input in both corpora. This is based on representative textbook analysis, previous published work on the learners' linguistic environment, and, more importantly, the analysis of learners' production data. The study then investigates the nature of this FL in comparison to learners' longitudinal interlanguage development. That is, it examines the computational (generative) and schematic (usage-based) properties of the identified FL, and tracks learners' corresponding knowledge of these properties *outside* of the FL over the data collection periods. This allows for the observation of any developmental trajectory that is common across all learners independent of FL use. It also enables the discovery of any individual differences in L2 computational development based on learners' differing use

of FL (generative) and/or the extent to which any of the learners' novel utterances can be accounted for based on schematic learning strategies instantiated by FL use (usage-based).

1.5.1 Preview of findings

In both corpora, similar highly prototypical fixed expressions (FEs) are identified in learners' production data at the initial state. These can be seen below in (3) and (4) respectively.

(3) Fixed expressions in the BELC

a. *what's/is your name?*

b. *how old are you?*

c. *where do you live?*

d. *where are you from?*

(4) Fixed expressions in the FPC

a. *comment t'appelles tu?* (what is your name?)

b. *quel âge as-tu?* (how old are you?)

c. *où habites- tu?* (where do you live?)

d. *quel est le date de ton anniversaire?* (when is your birthday?)

e. *tu as un animal?* (do you have a pet?)

f. *tu as des frères ou des soures?* (do you have any brothers or sisters?)

Under generative frameworks, the FEs are syntactically complex, involving derivations such as wh-movement and subject-verb inversion. Interestingly, in English, wh-movement and subject-verb inversion are obligatory in wh-question formation, but in French, these operations are optional. The wh-word can remain in-situ, and the subject and auxiliary verb uninverted, as in (5) below.

(5) Tu t' appelle comment?
you call yourself how

'what is your name'

Interlanguage analysis reveals that, for all learners in both corpora at the initial rounds of data collection, the FEs are produced in advance of respective L2 competence. That is, there is little evidence for L2 knowledge of the expressions' complex syntactic derivations outside of their use. At these stages, learners are largely seen to rely on the L1 and lexical categories of the L2, such as (6) and (7) below.

(6) BELC: Learner 5; Age 10

a. FE: *what's your name*

b. *no sé que dir-te* [CATALAN]

'I don't know what to say to you'

c. *it's dog eat*

d. *the mum it's*

e. *I study*

(7) FPC: Learner 28; Round 1

a. FE: *où habites-tu; quel âge as-tu, quelle est la date de ton anniversaire*

b. **les yeux?*

the eyes

'what colour are your eyes?' (intended meaning)

Tracking evidence for knowledge of these computational properties outside of the FEs across the data collection period, at the initial rounds, both sets of learners demonstrate knowledge of L2 lexical categories only, before evidence of L2 functional categories (Tense and Complementiser) appears subsequently at the later ages. This developmental trajectory fits that of a Weak Continuity (Vainikka and Young-Scholten, 1998). Interestingly, within the BELC, correlations are found between learners' more frequent use of the FEs at the younger ages, and a better knowledge of their underlying syntactic properties at the later ages. This results in a significant individual difference between learners' L2 knowledge of functional

categories at the later ages and their differing uses of the FEs at the early ages.

Taking a closer look at the longitudinal datasets, a usage-based traceback analysis also reveals evidence of schematic learning. Several of learners' L2 questions of the wh- and yes/no kind in both corpora can be traced back to utterance schemas of the FEs that proceed them ontogenetically in their production data. Schematic learning is particularly prominent in the FPC, where we also find evidence for lexically-specific schemas being used independently of the identified FEs, such as those in (8). Conversely, in the BELC, more abstract schemas are identified, such as that in (9).

(8) Lexically-specific schemas in the FPC

a. *où est* + X (where is + X)
eg. *[où est] l'homme, [où est] la garçon, [où est] la fille*
where is the man, where is the boy, where is the girl

b. *qu'est ce que* + X (what is it that + X)
eg. *[qu'est ce qu]'il a faire, [qu'est ce qu]'elle porte, [qu'est ce qu]'on peut manger*
what is it that he does, what is it that she wears, what is it that you can eat

(9) More abstract, categorical schema in the BELC
[WH + COPULA] + X
[what is] your job, [what are] you studying, [why are] you doing this work, [how are] you

This pattern is predicted by usage-based models, as the later rounds of the BELC are indicative of learners' L2 knowledge 6/7 years after the initial state, whereas the FPC is a window into learners' L2 productions within the first two years of instruction.

The findings identify a great amount of homogeneity for learners of the BELC and FPC. In both corpora, outside of formulaic expressions/utterance schemas, the L2 syntactic developmental trajectory aligns with a Weak Continuity. Here, the initial state consists of memorised, prototypical formulaic wh-questions and lexical categories only. Knowledge of L2 functional categories develops subsequently at the later rounds of data collection, most so for those learners who show an earlier and more frequent use of the identified FEs. Within this development, there is also evidence of schematic learning, some of which is instantiated by the FEs. The discussion draws on processing models of SLA to offer a theoretical explanation as to what this interaction of schematic learning and incremental development of L2 knowledge might look like. Overall, the study highlights the importance of formulaic material in the acquisition of L2 syntax in classroom settings. Perhaps most significantly, it provides a methodological demonstration of how the gap between generative and usage-based approaches to SLA can be bridged to investigate the interplay between input, usage and L2 syntactic knowledge.

1.6 Organisation of the study

Chapter 2 provides a review of the literature on formulaic language, focusing in particular on how this concept is relevant for L2 learners and SLA. It then gives an overview of usage-based and generative frameworks, highlighting their approaches to SLA and how the concept of formulaic language fits within these. Chapter 3 outlines how concepts from both models are implemented within the current study, advocating a multi-paradigm approach to FL in SLA based on the strengths of both frameworks. It then describes the kind of dataset that is required for a systematic analysis of this concept before narrowing down focus to the research objectives. Chapter 4 documents the methodology, giving some background on the learner corpora used for analysis and justifying their choice in light of the research objectives. Chapters 5 and 6 present the analysis and results of the BELC, and Chapters 7 and

8 present those of the FPC. Chapter 9 gives the discussion, bringing the findings from both analyses together. Chapter 10 finally concludes.

Chapter 2

Literature Review

The aim of this chapter is to give an overview of some of the key literature necessary for an understanding of the scope of this thesis and its research aims. Section 2.1 begins by defining 'formulaic language' (FL), narrowing down to how this concept is relevant for L2 learners. Section 2.2 then moves to the role of FL in SLA, presenting how this is central to usage-based models SLA (2.2.1) and less so for generative SLA (2.2.2). The Chapter ends by setting up the discussion in Chapter 3, which argues for a multiparadigmatic approach to this concept, narrowing down to the research objectives.

2.1 Formulaic language: Definitions, categorisation and L2 learners

The most cited definition of FL comes from Wray's (2002) seminal monograph;

> 'a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar' (p. 9).

As can be seen from the definition above, FL carries with it the assumption of being holistically stored and retrieved. This implies a processing advantage to speakers

as opposed to non-formulaic, compositional language. Although Wray (2002) notes over 40 terms in the literature used interchangeably to refer to FL (eg. fixed expressions, lexicalised phrases, lexical chunks etc.), categorically, it is possible to divide the phenomenon up into three broad classifications: idioms (incl. similes and proverbs), corpus derived sequences (incl. collocations, 'lexical bundles') and conventional expressions. That is not to say that these categories are mutually exclusive; it is expected, for example, that many conventional expressions in English would have a certain amount of salience in English-based corpora.

In an attempt to most systematically review the concept of FL with the words available, I will consider some typical characteristics and identificational criteria of each of these broad categories separately below. In these sections, I also include a brief review of past studies who have tested L2 learners' knowledge these various categories, which challenge the notion that all kinds of FL presents processing advantages to all speakers who use it. I finish by discussing previous studies who have instead derived FL directly from L2 production data, which sets up the discussion in Section 2.2 regarding the potential role it plays in the SLA process. I begin with idioms.

2.1.1 Idioms

A key trait of idioms is that many allow for two distinct interpretations (Conklin and Schmitt, 2012). For example, *kick the bucket* and *break the ice* carry both figurative and literal meanings; the former being 'to die' and the latter being 'to do or say something to relieve the tension of a situation'. The figurative meaning of idioms are non-compositional, in that they cannot be derived through the sum of the idiom's parts. You cannot, for example, derive the meaning 'die' from the combination of the words *kick* + *the* + *bucket*, if you were not aware of this idiomatic meaning/use beforehand.

Idioms can sometimes carry idiosyncratic phonological representations. For example, some idioms containing negation only carry figurative meaning when the negative is contracted. Aijmer (1996) (pgs. 14-15) notes that *don't rock the boat* is favoured in its figurative sense to mean 'do not say or do something that will upset people' over *do not rock the boat*, which gives for a literal meaning only. Idioms can also contain idiosyncratic phrase structural properties. For example, there is often a discrepancy between certain idiom's syntactic and lexical representations. Lyons (1969) notes the example *take NP to task* ('to criticise'), where the 'open slot' NP is an obligatory complement of the verb that must be filled for the idiom to be used grammatically, yet the lexical content of the NP is not given in the lexical entry of the idiom. Such 'open slots' can also be restricted in arbitrary ways; for example, the NP in *take NP for a ride* in the figurative sense ('to deceive') must be human, despite this not being a selection property of the verb *take* itself (Chomsky, 1996).

Some idioms go beyond showing idiosyncratic phrase structure properties and can even be described as syntactically ill formed, such as *by and large* and *so far so good*. This is not to say that such idioms lack grammatical structure, but rather that the structures they possess are not made possible by knowledge of the conventional rules of the grammar and how these are generally applied (Fillmore et al., 1988). They can also show greater or less degrees of syntactic flexibility under movement. Jackendoff (2002) notes that *you must draw the line somewhere* can passivise (*the line must be drawn somewhere*), whereas, *kick the bucket* cannot (*the bucket was kicked* loses figurative meaning and carries only the literal one.).

The idiosyncratic nature of idioms as exemplified above has led to the popular notion that they must be stored, processed and accessed holistically in the lexicon, representing 'single choices', similar to individual words (Backus, 2003). There is still no consensus as to what this lexical entry would actually look like (Keller, 2020), despite various models being put forth (eg. Cutting and Bock 1997; Sprenger et al.

2006; Jackendoff 2002). However, there are a number of studies who have set out to test this concept with advanced L2 learners.

For example, Conklin and Schmitt (2008) compared the reading speed of an idiom's literal and figurative meanings by highly proficient L2 speakers and native speakers. They embedded phrases such as *everything but the kitchen sink* used in both figurative and literal contexts in short story passages. The authors confirmed that the native speaker participants were aware of the figurative meanings beforehand, but did not do so for the L2 speakers. The task was for participants to read the passages one line at a time by pushing a button to reveal each new line. The results showed that both native speakers and L2 speakers read the idioms in their figurative context quicker than in the literal one, leading the authors to conclude that idiomatic phrases are processed more efficiently than non-idiomatic ones.

However, other studies have reported different results. Underwood et al. (2004) used an eye-movement paradigm comparing the fixation counts and fixation durations for the final word of an idiomatic phrase compared to a literal one. For example, the idiomatic phrase would read *honesty is the best policy* and the non-idiomatic counterpart would contain the word *policy* in a phrase such as *it seems that his policy of....* As in Conklin and Schmitt (2008), knowledge of the idioms' figurative meanings were checked with the native speakers and not the L2 speakers, who were all university students studying in the UK. The results showed that native speakers demonstrated shorter and fewer fixations on the word in its idiomatic context, but no such differences were found with the L2 speakers.

Similarly, Siyanova-Chanturia et al. (2011) monitored the eye movements of native and non-native university students as they read a series of stories which contained fragments present in an idiom used figuratively *at the end of the day* (ultimately) and literally *at the end of the day* (after the day has finished) or a com-

pletely novel phrase *at the end of the game*. Both native speakers and L2 learners' knowledge of the idioms' figurative meaning was confirmed before the experiment began. The series of stories were presented as passages on a monitor screen, and the participants' eye movements were tracked as they read the fragments, for analysis of first-pass reading time, total reading time and number of fixations. The results showed that native speakers read the figurative passages the quickest, whilst the L2 speakers read these the slowest.

Results from past experimentation therefore suggest that an idiom's figurative meaning can be processed quicker than its literal one by native speakers, but not always for L2 speakers, even at high proficiency levels. This is likely down to the increased exposure that native speakers have with the idioms' figurative meanings, compared to that of the L2 learners. Section 2.1.2 now discusses corpus-derived sequences, another category of formulaic language.

2.1.2 Corpus-derived sequences

Corpus-handling tools contain functions that allow for the automatic retrieval of words that occur most frequently to the right and/or left of a search term, often presented as a Key Word In Context (KWIC) view. This allows for the quick extraction and identification of collocations in a given language. Collocations can be roughly defined as 'usage-determined lexical combinations that are characterised by restricted co-occurrence of elements' (Paquot and Granger, 2012, p. 13). A distinction can be made in this regard between 'restricted' and 'free' collocations. The former are a more preferential selection of word sequences, so much so that the combinations can be seen as partly arbitrary (Kuiper et al., 2007). For example, *catch the bus* is what a native English speaker says when they use public transport, instead of *trapping* or *capturing* the bus, despite *catch, trap* and *capture* being equally arbitrary synonyms in this context. Free collocations instead are more flexible and less arbitrary; the most commonly investigated types are frequently occurring [V +

NP complement] combinations such as *drive a car, read the newspaper, ride a bike etc* (Paquot and Granger, 2012).

Collocations can also display idiosyncratic properties. Kuiper et al. (2007) note that [ADJ + N] combinations such as *heavy smokers* and *artesian wells* have semantically unspecialised heads, yet the adjectives that modify them are specialised, in that they have meanings different than their normal use. Some collocations also seem to have positions open to a restricted number of lexical items which can function in the same position. For example, a *bad mood* is equivalent to a *bad temper*, thus knowing this collocation is knowing that both nouns can be used interchangeably (Kuiper et al., 2007).

As well as collocations, what is usually termed the 'N-gram' function allows the researcher to retrieve from a corpus all repeated strings of words of any given length. These sequences are often termed 'lexical bundles', which are 'recurrent expressions, regardless of their idiomaticity, and regardless of their structural status' (Biber et al., 1999, p. 90). Some examples taken from the British National Corpus (BNC) include *as far as, as discussed in, to the extent in which*. Lexical bundles are products of frequency approaches to formulaicity that are often adopted by researchers, although care should be taken not to rely on frequency counts alone when identifying corpus-derived sequences (Ellis, 2012; Myles and Cordier, 2017). This is because some words have such a high frequency in a corpus that they show as collocates of a search term without having any relation to them at all. For example, the most frequent noun in the BNC is *time*, meaning it is likely to occur close to many search terms solely by chance (Lindquist, 2009).

To combat this 'frequency effect', corpus tools also allow for identification of statistically significant co-occurrences of words, which calculate the frequency of words near the search term in relation to their total frequency in the corpus. One of the

most used is the Mutual Information (MI) statistic, which allows for identification of strongly associated combinations of words, essentially, a measure of 'collocational strength' (Lorenz, 1999). The main difference between such statistical tests and KWIC view/N-gram functions is that the focus is on the association of combination of words rather than their overall frequency, which usually means they are very salient for native speakers (Simpson-Vlack and Ellis, 2010). This also means, however, that prominence is often given to highly rare combinations of words. For example, the top collocate of *educational* in the BNC based on MI score alone is *non-broadcast*, despite this collocation being relatively infrequent in raw numbers (Lindquist, 2009). Due to the limitations of using either frequency-based measures or association tests alone, researchers tend to favour using both when looking to identify corpus-based sequences (see Ellis 2012, for example).

As with idioms, there has been a tendency to assume that corpus-derived sequences present processing advantages to speakers due to their holistic, fixed nature. Schmitt et al. (2004) used recurring clusters derived from corpus analysis to test this concept with L2 learners studying at UK universities. They tested 25 sequences varying between 2-6 words including *you know*, *in the same way as* and *I was going to*. An oral dictation task was employed, where these target sequences were embedded into a recorded story and played out-loud to the participants. Learners then had to repeat 'bursts of words' that they had heard previously, most of which contained one of the target sequences. Results showed that only 4/25 recurring clusters (*as a matter of fact, in the middle of the, you know, on and off*) were produced by half of the L2 learners, with the majority of learners falling into the category of 'partially incorrect', 'disfluent' or 'did not produce' categories. The learners' abilities to recall sequences also showed no sensitivity to the sequences' frequency in the corpora from which they were extracted. The authors therefore conclude their study with a word of caution; 'although there is an unspoken assumption that corpus data is psycholinguistically valid...it is unwise to take recurring clusters in a corpus as evidence

that those clusters are also stored as formulaic sequences in the mind' (Schmitt and Carter, 2004, p. 147). Section 2.1.3 now moves to discuss the final category of FL reviewed here- conventional expressions.

2.1.3 Conventional expressions

Central to conventional expressions is their pragmatic links. Their use requires speakers to have knowledge about and performance consistent with the social norms in specific situations in a given society (Coulmas, 1981; Felix-Brasdefer and Hasler-Barker, 2015). Therefore, their identification relies on examining preferred expressions used by native speakers (NS) in certain contexts. To do this, researchers can carry out field observations of spontaneous conversation in a specific speech community, and record expressions which are most used in particular scenarios. Conventionality is established by setting a cut score above which there is no competing alternative phrase. Traditionally, this is operationalised as when a phrase is used for more than 50% of all responses to a particular scenario (Bardovi-Harlig, 2009; Bardovi-Harlig et al., 2010; Edmonds, 2014). For example, *sorry I'm late* is taken to be the English conventional expression used in the scenario in which an employee fails to arrive to work on time, as this is the phrase that is overwhelmingly favoured by native-speakers in this particular context (Bardovi-Harlig and Stringer, 2017).

Bardovi-Harlig (2019) states that use of conventional expressions requires both sociopragmatic and pragmalinguistic knowledge working together. Sociopragmatics can be seen as the 'specific local conditions on language use' (Leech, 1983, p. 83) and pragmalinguistics as the knowledge of the linguistic repertoire necessary to carry out the sociopragmatics of a specific language and culture. Knowledge of conventional expressions is therefore part of pragmalinguistic competence, and knowledge of their use and the context in which they occur is part of sociopragmatic competence (Bardovi-Harlig, 2019). As conventional expressions are closely associated to specific pragmatic contexts, a mastery of such is advantageous for a learner's ability

to succeed in a second language (Fillmore, 1979), making them popular targets for achieving 'native-like' fluency in many EFL textbooks/syllabuses.

Conventional expressions also show idiosyncratic properties related to their functional uses. As social situations often have conventional expressions attached to them, many of these phrases are very contextually restricted in use, differing somewhat from free combinations of words used outside of these scenarios. Kuiper et al. (2007) use the example of a flight crew announcing *this is your captain speaking*, arguing that 'every small-scale ritual tends to be accompanied by formulae' (p. 317). Outside of this context, this expression is hardly ever used, but this cannot be substituted for any other expression within this context. Paquot and Granger (2012) also note that functions of conventional expressions can differ depending on native/L2 speaker use. An example is the study by Aijmer (2009), who compared the function of *I don't know* in both a native speaker (English) corpus and the Swedish component of the LINDSEI corpus. They found that for the L2 learners, *I don't know* functions as a speech management signal, whereas native speakers use it mainly to avoid answering questions directly.

Bardovi-Harlig and Stringer (2017) tested a variety of L2 learners ranging from low-intermediate to low-advanced levels of proficiency who were attempting to produce the same target conventional expression in the same communicative context. The conventional expressions were selected by a data driven process, including field observations of spontaneous conversations in the community where the study was based and operationalisation with native-speaker use greater than 50%. Elicited expressions were then tested on a control group of native speakers in two pilot studies, as a reassurance of their conventionality. Examples include *I'm sorry I'm late, I'm just looking* and *I really appreciate it*. The data collection procedure consisted of a timed oral Discourse Completion Task (DCT) and a written aural recognition/self-assessment. The oral DCT required the learners to listen and produce a time-

pressured response to 32 created scenarios over individual headsets whilst simultaneously reading the scenarios on a screen. The written aural task was self-paced, where learners heard the 22 conventional expressions and chose from options which best represented their knowledge of such.

Results showed that rather than fluent production of the conventional expressions, learners actually produced interlanguage variations of these which developed across proficiency levels. Specifically, this involved 'the gradual acquisition of a lexical core of the expression that is not fully grammatically specified and is filled in by the learner's interlanguage grammar' (Bardovi-Harlig and Stringer, 2017, p. 79). For example, the target expression *sorry I'm late* exhibited ungrammatical interlanguage variations around the lexical core of *sorry* and *late*, such as **sorry for late, *I'm sorry to late, *I am sorry about late*. The authors conclude that rather than being stored and produced holistically, L2 productions of the elicited conventional expressions seem to be influenced by learners' interlanguage grammars.

2.1.4 Formulaic language in L2 production data

The above sections outline how various categories of FL have been defined and operationalised over the last 20 years. They also demonstrate that many studies testing L2 learners' knowledge of these phenomena present evidence against the notion that FL is processed holistically. Indeed, there have been a growing number of dissenting voices in this regard. Some researchers take issue with the implicit claim that all FL is stored and retrieved whole as per Wray's (2002) 'umbrella' definition, particularly when this comes to L2 learners (Bardovi-Harlig, 2009; Siyanova-Chanturia, 2015; Myles and Cordier, 2017). In other words, what can be defined as FL in a given language (such as idioms, corpus-derived sequences or conventional expressions) cannot be prematurely assumed to have any psycholinguistic salience for any individual L2 learner. Instead, FL that is likely stored and retrieved holistically for L2 learners should be identified through the analysis of their production data.

Myles and Cordier (2017) propose that such FL will show certain characteristics. One is phonological coherence, as utterance fluency is a reflection of cognitive fluency, and is therefore primary evidence that a string of words is being processed holistically. If, for example, a sequence of words is pronounced haltingly with pauses and/or repairs, it is more likely that this has been put together online rather than processed as a whole (Myles and Cordier, 2017). Another is unity, which can be either semantic or functional. This can include, for example, conventional expressions that are taught and therefore learnt holistically (eg. to give one's name *my name is*), denote time (eg. *last year*) or introduce one's opinion (eg. *I think that*) (Myles and Cordier, 2017). The final criterion the authors propose is the frequency of an expression in a given learner's production data. That is, the more frequently a learner is seen to use an expression, the more likely it is that this sequence is formulaic and holistically retrieved/produced (Myles and Cordier, 2017).

For Bardovi-Harlig (2009), the main indication of an expression's formulaicity, particularly for beginner/initial-state L2 learners, is that it displays a level of fluency and syntactic development that is not found elsewhere in the interlanguage of the learner. A number of SLA studies examining learners' naturalistic L2 production data present evidence of this. Hanania and Gradman (1977), for example, found that their adult subject Fatmah's English output consisted mainly of conventional expressions commonly used in social contexts with children, such as *thank you, I can't* and *do you like*. Fatmah was able to use these expressions appropriately in specific pragmatic contexts, but was unable to use the same words or structures in new combinations outside of these contexts. Similarly, Schmidt (1983) followed the L2 behaviour of an adult learner, Wes, who was able to regularly produce conventional expressions such as *do you have time* and *are you busy*, whilst demonstrating a complete lack of subject-verb inversion in his interlanguage structures outside of these strings. This was also found to be the case by Myles et al. (1998, 1999),

who followed the progression of English classroom adolescent learners of French. At the first rounds of data collection, learners were shown to produce complex wh-questions such as *quel âge as tu* (how old are you) who otherwise lacked wh-fronting and subject-verb inversion in their interlanguages elsewhere. Instead, these same learners were shown to produce ungrammatical utterances in similar functional environments, such as **il âge frère* (*he age brother).

Such studies highlight that formulaicity in SLA is more to do with a sequence's 'degree of entrenchment' in a learner's mind, rather than how the string can be categorically/externally defined (i.e. as an idiom, corpus-derived sequence or conventional expression etc.) (Ellis et al., 2015, pg. 377). FL that is heavily entrenched in beginner learner production data often over-represents their respective L2 abilities (Myles, 2004; Wray and Fitzpatrick, 2008) and can offer short-term communicative benefits in the language-learning classroom/target-language community (Jeremias, 1982; Schmitt, 2010). The kind of phrases that are likely to be highly entrenched for learners, and hence present processing advantages, are those that are high in frequency, functionality and prototypicality in their L2 input (Ellis, 2012; Ellis et al., 2015). These can often be predicted based on learners' linguistic environment (Krashen and Scarcella, 1978). For classroom learners, for example, good candidates are conventional expressions that are closely tied to specific communicative contexts and are taught to permit interaction at initial stages of learning (Towell, 2012), such as *what is your name* or *when is your birthday.* It can be presumed that these expressions, which are introduced holistically in the classroom, are retained this way by learners (Nattinger and DeCarrico, 1992) and provide a pathway to basic routine conversation in the L2.

Given the prominence of this kind of FL in initial state interlanguage production data and its presumed functional/processing advantages for learners, the question of FL's role in the SLA process more generally has been a popular subject of enquiry

in Applied Linguistics. There is considerable debate surrounding the influence, if any, of FL on learners' acquisition of L2 syntax, which depends in large on which theoretical framework is adopted. Section 2.2 below now discusses this debate. It first presents both usage-based and generative frameworks of L2 acquisition before outlining the role that FL is perceived to play in each of these models.

2.2 The role of formulaic language in second language acquisition

2.2.1 Usage-based models of SLA

For usage-based linguistics (UBL), constructions are the basic linguistic unit (Horbowicz and Nordanger, 2021). They propose a lexicon in which 'abstract grammatical patterns and the lexical instantiations of those patterns are jointly included, and which may consist of many different levels of schematic abstraction' (Tummers et al., 2005, pp. 228-229). Linguistic knowledge is therefore conceptualised as 'the cognitive organisation of one's experience with language' (Bybee and Eddington, 2006, p. 711). For UBL, language learning is initially exemplar based, and knowledge emerges from the memories of utterances in usage events (via interaction and output) and the abstraction of regularities within them (Langacker, 1987; Ellis, 2005; Roehr-Brackin, 2014).

The emergence of such knowledge is enabled by general cognitive mechanisms such as 'association, categorization, schematisation and entrenchment' (Lesonen et al., 2020, p. 527). Association refers to when learners relate phonological material with certain meanings/functions. For example, beginner learners of English are hypothesised to initially map the string of sounds in *what is your name* to the function [ask name]. In categorisation, learners can compare novel utterances to

those they have already encountered (Langacker, 2013). Hence, when comparing *what is your name* with *what is the time*, the learner may deduce that both these expressions are used to elicit information from someone, and that *your name* and *the time* are requested entities and belong to the same group of words. Schematisation then occurs when a learner has encountered a number of similar L2 expressions and extracts commonalities within these (Langacker, 2013). In being exposed to a number of wh-questions such as *what is your name* and *what is the time*, gradually the learner could generalise that [*what is*] + NP is a pattern that can be used to elicit information from a person. Finally, entrenchment refers to these memory traces being repeatedly activated and established as a unit (Langacker, 2013; Lesonen et al., 2020). That is, with more frequent use and exposure, schematic patterns can be more easily accessed and activated, as they become more automatic.

In light of this, FL that is high in frequency, functionality and prototypicality plays a significant role in SLA for UBL. Based on a fundamental notion of Emergent models that 'grammar is what results when formulas are re-arranged, or dismantled and re-assembled, in different ways' (Hopper, 1987, p. 145), this framework argues that a learner's long-term knowledge of lexical sequences in FL can be what serves as the database for their language acquisition (Ellis, 1996, 2002a,b). Formulaic phrases are 'databases' in the sense that they are model examples of form-meaning correspondence patterns, which act as frames for whose frequent and early production can encourage pattern extraction and generalisation towards creative L2 use. Specifically, the proposed learning trajectory for SLA is from formulaic phrase to utterance schema (known also as semi-fixed or slot-and-frame pattern) to fully productive schematic pattern (Tomasello, 2003; Ellis, 2012; Ellis et al., 2015). For example, through frequent exposure and usage of the prototypical conventional expression in (1-a), learners can first derive the lexically-specific utterance schema in (1-b) before moving to the more schematic wh-question pattern in (1-c).

(1) a. *where do you live*

b. [where do you] + VERB

eg. [where do you] go, [where do you] pay, [where do you] play

c. [WH + AUX DO] + X

eg. [when does] he get here, [what do] you like to eat

As exemplified above, learners are hypothesised to gradually move from the formulaic to the schematic (Roehr-Brackin, 2014; Eskildsen, 2015, 2020). Schematicity can be viewed as a scale, referring to the level of specificity and abstractness of a given construction (Horbowicz and Nordanger, 2021). Thus, the construction in (1-c) allows several wh-words (eg. *when, what*) and VPs (eg. *he get here, you like to eat*), and is hence more schematic than the fully lexically-specific utterance schema in (1-b), which is more fixed. That is, lexically-specific schemas maintain some of the lexical items from the model formulaic phrase (eg. *where do you live)*, whereas categorically specific ones display the more general grammatical category sequencing.

As UBL frameworks perceive fluidity among linguistic patterns and the abstraction of any generalities within recurring, prototypical exemplars (Eskildsen, 2020), any utterance schema for which a formulaic phrase exemplifies is derivable from its abstract schematic construction. For example, the categories verb (V), noun (N), preposition (P) and 'transitive frame' (TF) in an English verb-object-locative (VOL) construction are all potentially derivable and accessible through the formulaic phrase *put [V] it [N] on [P] the table [P].* This phrase is 'formulaic' by virtue of containing the most frequent, prototypical and generic exemplar in each slot of this particular verb-argument construction (Ellis and Ferreira-Junior, 2009a,b).

A significant component of L2 learning in UBL is therefore the abstraction and generalisation of FL, which can be understood as the gradual expansion of utterance schema use in interaction and output (Roehr-Brackin, 2014). Importantly, FL which is identified as having initiated schematic development must precede all other

instantiations ontogenetically in learner data. That is, learners must be shown to produce the proposed FL in advance of any other manifestation of related schematic patterns. Using the same example as in (1), to argue that the conventional expression *where do you live* has seeded the utterance schemas [where do you] + X or [WH + AUX DO] + X for a particular learner, *where do you live* must appear in their production data before all other utterances which embody these schematic frames.

This constitutes what is often referred to as 'traceback methodology' (Lieven et al., 2003, 2009), a model adopted from child L1 acquisitional studies. This methodology relies on the analysis of longitudinal production data, which allows researchers to explore the relationship between creativity and routinisation in SLA. Essentially, these studies aim to discover evidence that target language use has been instantiated by formulaic exemplars and/or utterance schemas. Some key L2 studies who have adopted this methodology to support the usage-based learning trajectory are now reviewed below.

One of the earliest examples is Wong-Fillmore (1976), who traced the English development of five Spanish-speaking children from the onset of L2 acquisition over the period of one school year. The author noted that many of the children were exposed to typical formulaic question forms that were frequently used in classroom and playroom elicitation activities. For example, learners first learned formulaic phrases such as (2-a) that they broke down into frames (2-b) before they were shown to extract elements and use these in propositional/creative structures (which were often ungrammatical), as in (2-c).

(2) a. *what is it, I don't wanna do these*

b. what is X/X = NP, I don't wanna X/X = VP, NP
what is the paper, what is this colour, I don't wanna play with that one thing

c. *this is for what, you want what, she don't likes me, my mother don't*

want no papers at home

The authors took this as evidence that expressions such as these were memorised and then gradually broken down into their compositional parts to extract rules governing their L2, before creative language use ensued.

Eskildsen and Cadierno (2007) followed an adult L1 Mexican learner of English over a period of five years. By tracking type-token frequencies, they observed that *do*-negation learning was based on one instantiation of the conventional expression *I don't know*. Over the course of acquisition, the learner gradually expanded upon this phrase, using its internal parts with other verbs and pronouns as per the inventory in Table 2.1 below.

I don't ***know*** → *I don't think so* → I don't agree → I don't know exactly…			
↓	**Take**	need	much about
↓	remember		**have** NP
do you know (x)?			↓
yes I do / no I don't			↓
↓			↓
↓			↓
do *you* ***have*** x?	→ you don't **take** a shower? → we don't **have** NP		
You ***have*** x?	**know**?	you didn't see that?	
↓	go?		
↓	come to x?		
↓			
How *do you say x*? → **how** do you write → **how did** you pronounce that?			
	spell		↓
			→ we **did** many things

Table 2.1: Overview of *do* negation inventory (Eskildsen and Cadierno, 2007, p. 11)

The authors concluded that the learner's emerging *do*-negation system was therefore based on the abstraction of regularities from *I don't know* to other constructions. Building on this study, Eskildsen (2008) analysed the same adult learner (Carlos) to track their development of the auxiliary verb *can*. Rather than one dominant formulaic phrase being responsible for initiating schematic development in this regard, the author found that this was instead based on a number of interrelated, locally recurring specific patterns. This goes against the notion that learning has to begin

with reliance on specific, item-based constructions, as was found with *I don't know* in Eskildsen and Cadierno (2007). The early emergence of these patterns can be seen from the examples in Table 2.2 below, where chronology is represented both horizontally and vertically. That is, *I can write* was the first and *you can do* was the last manifestations of these patterns to appear over a two year period (Eskildsen, 2008).

	Write	Spell	Show	See	Help	Take	Call	Sit	Say	Ø	Do
I can Verb, 2001	2										
Can you Verb, 2001	2	1	1	1	1						
Can I Verb, 2002				3		3	2				
You can Verb, 2002								1	3	1	5

Table 2.2: Emergence of *can* patterns (Eskilsden 2008, pg. 341)

Roehr-Brackin (2014) followed an L1 adult English learner of German over 3 years in a learner-tutor classroom-interaction setting. They tracked the long-term development of two German Perfekt constructions, *gehen* ('go', 'walk') and *fahren* ('go by vehicle'). For *fahren*, they identified a developmental sequence that began with formulaic constructions before moving to gradually more abstract utterance schemas based around the past participle *gefahren*. Some identified formulaic, item-based constructions can be seen in (3) and (4), and an example of an abstract schema related to these in (5) below.

(3) Wir sind nach Hause gefahren
we are at home went
'We went/drove back home'

(4) Wir sind in die Stadt gefahren
we are in the town went
'We went/drove into town'

(5) [SUBJ + AUX / AUX + SUB (*zurück/weiter/los) nach* + LOCATION/*nach Hause* + GEFAHREN]

a. eg. *Wir sind mit dem Bus von Flughafen nach in der Nähe vom Hotel gefahren*

'We took the bus from the airport to near the hotel'

b. eg. *Dann sind wir zurück ins Dorf gefahren*
'Then we went/drove back into the village'

Regarding Perfekt constructions with *gehen* (go), the author noted a slightly different pattern. Here, there was evidence that more schematic constructions were being used from the start, similar to what was found for *can* development in Eskildsen (2008). For example, the abstract schema in (6) was posited to be responsible for a number of utterances at initial stages of the observed period, including the examples in (6-a) and (6-b).

(6) [ADVERBIAL + AUX + SUBJ (ADVERBIAL) + PREPOSITION + LOCATION + PARTICLE *gegangen*]

a. eg. *Später sind wir in der Nähe vom Hotel spazierengegangen*
'Later we went for a walk near the hotel'

b. eg. *Dann sind wir wieder zurück zum Auto gegangen*
'Then we went back to the car'

Roehr-Brackin (2014) suggests that with *gehen*, the learner could have been drawing on explicit, top-down processes to override the predicted learning trajectory of formulaic item-based construction to utterance schema.

Eskildsen (2015) used the MAELC corpus to track English question formation development of two Mexican Spanish-speaking learners over five recording periods. One of these was Carlos (as discussed in the previous studies) and the other was Valerio, who had been in the United States for 9 months before joining an ESL programme in Portland Community College. Valerio was tracked from periods between 2003 and 2005, and Carlos from periods between 2001 and 2005. For both learners, Eskildsen (2015) focused on particular question constructions and identified elements of schematic learning. For example, two constructions that were focused

on were wh-questions with the copula *is* and with the auxiliary *do*.

For both learners, the development of WH + COPULA questions seem to be instantiated by conventional expressions such as *where are you from, how are you* and lexically-specific schemas such as [*what are you* + VERB*ing*], [*what is* + X] and [*where is* + X]. Over time, these schemas were used to fulfill tasks in similar functional environments; for example, to ask the whereabouts of other items/classmates (eg. *where is Julio, where is your bike, where is your glasses*) or enquire about things in general (eg. *what is your hobby, what is your favourite holiday, what is the movie*). A similar observation was found for WH + AUX DO questions, where both learners are shown to initially rely on formulaic question forms (eg. *what do you have/say, where do you live*) which develop as new verbs are recruited to similar related utterances (eg. *where do you live/went, what do/did you/he + V*). They conclude by stating that, rather than building structure through a bottom-up process as has been traditionally assumed for L2 question formation (Cazden et al., 1975), learners seem to assemble schematic constructions on the basis of recurring lexically-specific exemplars (Eskildsen, 2015).

Lesonen et al. (2020) traced the development of two verbal constructions (*haluta* 'want' and *tykätä 'like'*) in four adult Finnish L2 learners from various L1 backgrounds. Free response data was collected weekly over a period of 9 months and included both written and spoken components. The authors found that for some learners, *haluta* constructions were initially mostly formulaic, and development of these generally followed the trajectory as posited below in Table 2.3 (from left to right).

Example	*Haluan matkustaa Saksaan* 'I want to travel to Germany'	*Haluan matkustaa* + NP 'I want to travel + NP'	*Haluan* + NFC 'I want + NFC'	HALUTA + NFC WANT + NFC
Type of construction	Lexically specific, formulaic expression	Mostly formulaic expression	Semi-schematic, semi-abstract pattern	Fully schematic, abstract pattern
Fixedness of the construction	Fully fixed	Partially variable: construction has one open slot	Semi-variable: construction has more open slots	Highly variable
Degree of productivity	Not productive			Highly productive

Table 2.3: Continuum from lexically-specific to to productive constructions (Lesonen et al. 2020, pg. 536)

However, the authors also found evidence of more productive semi-schematic phrases early in the data collection period, similar to what was found in Eskildsen (2008) and Roehr-Brackin (2014) with the development of *can* and *gehen*. For example, learners' initial constructions with *tykkään* were much more productive. At the start of the data collection period, all learners are shown to use a number of schemas related to this verb, as presented below in Table 2.4. Here, NP stands for 'noun phrase' and NFC stands for 'non-finite clause', which includes several verbal complements.

Schemas	Example utterances observed in the data
[***tykkään*** + NP] [I like + NP]	***tykkään*** *kahvi* 'I like coffee'
[***tykkään*** + NFC] [I like + NFC]	***tykkään*** *katsoa tvtä* 'I like to watch TV'
[***tykkäätkö*** + NP] [Do you like + NP]	***tykkäätkö*** *pitsasta* 'Do you like pizza?'
[***tykkäätkö*** + NFC] [Do you like + NFC]	***tykkäätkö*** *matkukustaa*? 'Do you like to travel?'

Table 2.4: Identified schemas with *tykään* at initial stages

Lesonen et al. (2020) suggest that one reason for observing productive schemas such as those above at initial stages, rather than fixed formulaic exemplars, is that

adult learners may already have an established L1 schematic system from which they can use as templates for similar L2 expressions. Another suggestion is that this could be down to instruction, as the authors noted that when *tykätä* constructions were taught, emphasis was placed on the communicative functions of its different parts (Lesonen et al., 2020). It is indeed the case that in instructed/classroom contexts, explicit instruction likely plays a part in a learner's deconstruction of model formulaic exemplars, particularly when the focus is on their grammatical properties/functions.

However, the general pattern that was observed overall is that all learners developed more abstract schemas across the period of observation, as predicted by the usage-based learning trajectory. At the final stages, all learners were producing various utterances that could only be captured by more abstract schemas [$TYKÄTÄ$ + NP/NFC] and [$HALUTA$ + NP/NFC]. The authors therefore conclude by stating that some learners start with specific, formulaic expressions whilst others initially use more productive schemas that show a greater number of variable instantiations. Despite these different paths, over time all learners in the study moved from lexically-specific to more abstract schematic patterns (Lesonen et al., 2020).

Finally, Horbowicz and Nordanger (2021) investigated the schematisation process in the development of L2 Norwegian epistemic verb-argument constructions by four adult learners of various linguistic backgrounds/L1s. Recorded conversations on various topics took place between each learner and their former teacher over the course of 17 weeks. The authors tracked the development of the epistemic verbs *tro* ('think/believe') and *vite* ('know') specifically, presenting evidence for a continuum of schematisation as outlined in Table 2.5 below.

Type of construction	Lexically specific; fully fixed	Semi-schematic; semi-fixed; one or more open slot	Highly schematic, productive	Schematic epistemic construction
Examples from the data	*Jeg tror det er morsom*	(i) *Jeg tror det er* + ADJ/NP	SUBJ + *tro* + NP/FC	SUBJ + EPISTEMIC VERB + argument
	'I think it is funny'	'I think it is' + ADJ/NP (ii) *Jeg tror* + FC 'I think' + FC	'think' + NP/FC	

Table 2.5: Continuum of schematisation for epistemic verb-argument constructions (Horbowicz and Nordanger, 2021, p. 447)

For example, for all learners at initial stages, when the subject slot is realised in an argument structure for the verb *tro*, this is filled primarily with the first person pronoun *jeg* ('I'). This gives for the lexically-specific schema [*jeg tror* + X] ('I think + X'), which for some of the learners seems to have been instantiated by multi-word expressions such as *jeg tror at det er* ('I think that it is') and *jeg tror det er morsom* ('I think it is funny'). Reliance on these multi-word expressions is also evident through their appearance in utterances that deviate from the target language norm. For example, use of *jeg tror* in the utterance below (7) violates a norm in Norwegian that adverbs in a subclause should appear in a preverbal position (Horbowicz and Nordanger, 2021).

(7) jeg tro-r det er ikke så stor samling av folk
I think-PRS it be-PRS not such big gathering of people
'I don't think it is such a big gathering of people'

Later on in the data collection period, these schemas are used more productively. For example, *jeg tror det er morsom* (I think it is funny) seems to be functioning as an acquisitional seed for the lexically-specific schema [*jeg tror det er veldig + ADJ*] (I think it is very + ADJ), as exemplified in (8).

(8) jeg tror det er veldig viktig
I think-PRS it be-PRS very important
'I think that it is very important'

Finally, evidence for more schematic learning comes from utterances such as those in (9) and (10), whereby learners are using more abstract schemas exemplifying the general grammatical category sequencing (ie., SUBJ + EPISTEMIC VERB + argument).

(9) jeg aldri to-dde at jeg ville føle meg på denne måt-en
I never think-PRET that I will-PRET feel me on this way-DEF
'I never thought that I would feel this way'

(10) jeg trodde at det var sånn i: alle land-e
I think-PRET that it be-PRET such in all country-PL
'I thought it was like that in all countries'

Horbowicz and Nordanger (2021) conclude that as well as providing learners with useful communicative functions, the formulaic question forms identified in the data seem to provide a stepping-stone for the development of more complex syntax.

In summary, longitudinal UBL research uses traceback methodology to support the usage-based learning trajectory, where FL seems to play a significant role in instantiating schematic development. These studies demonstrate that there is variation between learners and/or the L2 structures being investigated in how exactly this trajectory manifests. Although the learning sequence mostly starts with holistic, rote learned formulaic expressions, there are some instances where learners begin with a broader range of more semi-fixed, productive utterance schemas, which could be a result of instruction or L1 influence. In either case, all studies report that learning proceeds from the formulaic and lexically-specific to the more abstract and schematic. Section 2.2.2 now outlines generative models of SLA, and discusses the far less central role that FL plays within this framework.

2.2.2 Generative models of SLA

Rather than a reflex of the domain-general cognitive systems, generative grammar conceptualises linguistic knowledge as deriving from an innate, domain-specific com-

putational system. Whilst usage-based theories hold that linguistic competence emerges from the schematic analysis of input, generative theories are based on the 'Poverty of the Stimulus' (POS), which argues that the input significantly *under-determines* linguistic competence (Chomsky, 1975). This refers to the fact that 'there is a vast qualitative difference between the impoverished and unstructured environment, on the one hand, and the highly specific and intricate structures that uniformly develop, on the other' (Chomsky, 1980, p. 24). Speakers demonstrate knowledge of things like grammaticality, ungrammaticality and ambiguity, which go well beyond the utterances that they may have been exposed to (White, 1990). This entails that speakers develop linguistic knowledge that cannot be acquired solely from an interaction between the available input and domain-general cognitive systems (Rothman et al., 2019).

Generative studies present evidence for such knowledge through experiments where speakers judge and interpret sentences that they may never have encountered before, correctly rejecting those that are communicatively plausible, but ungrammatical (Juffs and Fang, 2022). Evidence also comes from errors and overgeneralisations that speakers do *not* make in production, which would be far more frequent if not constrained by an internal, domain-specific linguistic knowledge (White and Juffs, 1998; Slabakova, 2016). This is because the input does not contain any examples of ungrammatical utterances, what is often referred to as a 'lack of negative evidence' (Schwartz 1999). For generative researchers, the POS presents the 'Logical Problem of Language Acquisition', and has long been posited for both L1 (Chomsky, 1975, 1980) and L2 development (White, 1989, 1990) alike.

In an attempt to address this 'logical problem' and account for the discrepancy between the impoverished input and acquired linguistic knowledge, generativists posit that humans must be equipped with a Universal Grammar (UG), an innate linguistic knowledge pertaining to all possible human languages (Newmayer, 2008).

The framework used to describe such knowledge, in its most current incarnation, is known as 'Minimalism' (eg. Chomsky 1995). It distinguishes between the lexicon, a finite inventory of elements stored in memory, and syntax, a procedure which combines elements from the lexicon together to form larger, more complex units.

The lexicon contains two major systems of lexical items, the contentive lexicon and the functional lexicon (Rizzi, 2009). The contentive lexicon consists of elements which have substantive lexical-semantic descriptive content characterising events, arguments and qualities etc (i.e. nouns, verbs, adjectives). These are often referred to as 'lexical categories'. The functional lexicon contains lexical elements (and morphemes) with a more abstract semantic content which essentially serve to mark grammatical properties such as tense or modality (determiners, complementisers, auxiliaries etc) (Radford, 2009). These 'functional categories' include functional 'heads' that contain feature specifications related to individual languages (Rizzi, 2009). Features are units of grammar that reflect variation across languages (Rothman and Slabakova, 2018), and are responsible for defining the configurational structure in which the lexical categories are inserted.

Syntax is formalised as the computational property Merge (Chomsky, 2002), which via the operation Select, takes two syntactic (SYN) elements, A and B to form a composed element [A,B] (or [C]). This operation is recursive, in that it can reapply indefinitely to its own output (Rizzi, 2009), and hence structure building is binary branching and hierarchically determined. This means that A and B can be two elements taken directly from the lexicon, or be independent complex entities already created by previous applications of Merge (Collins and Stabler, 2016). For example, Merge combines individual lexical items such as *[I]* + *[eat]* to generate the larger syntactic object *[[I] [eat]]*, which can also combine with previously generated larger syntactic objects such as *[every afternoon]* to generate the more complex item *[[I eat] [every afternoon]]*. Operations such as these are known as 'External Merge'.

In contrast, there is also a process labelled 'Internal Merge', which instead takes one element from within another. This computation is responsible for various sub-operations, particularly those involving dependency relations, also known as 'movement' (Chomsky, 1981, 1986). For example, clause-initial wh-words in interrogative structures such as those in (11) are presumed to have moved from their original complement positions from which their meaning is licensed. I use 't' as shorthand for 'trace', to indicate the original position of the wh-words.

(11) a. *what*$_k$ *did you see* t_k

b. *where*$_k$ *did you go* t_k

c. *who*$_k$ *did you meet* t_k

Once Select has introduced elements (i.e. lexical items, syntactic objects) into the derivation and Merge has combined them into a composed syntactic element, the operation Transfer maps these elements onto both the semantic (SEM) and phonetic (PHON) components (also known as Logical Form and Phonetic Form respectively) (Radford, 2009). Interface representations of sound (PHON) and meaning (SEM) are then interpretable to auditory-articulatory (speech) systems and the conceptual-intentional (thought) systems. A diagram representing this model is shown in Figure 2.1.

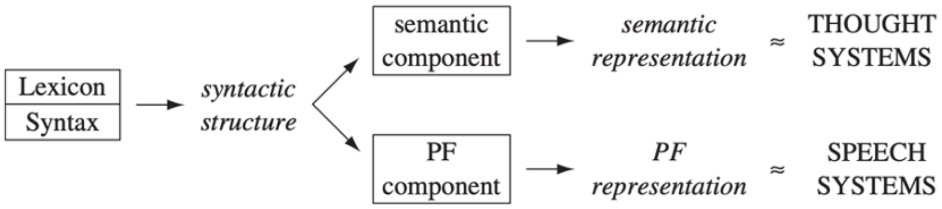

Figure 2.1: The Minimalist Framework (Radford, 2009, p. 14)

Select, Merge and Transfer are universal syntactic operations, and SYN, PHON and SEM are universal sets of features, which together constitute UG (Collins and Stabler, 2016), taken to be an innate endowment of human beings. These mecha-

nisms provided by UG are what determine a speaker's linguistic knowledge, that is, they are the abstract mental representations that are required for the use and realisation of syntax (Towell, 2012). The study of generative SLA is therefore concerned with describing the development of the implicit knowledge of these L2 representations in the minds of the learner (Rothman et al., 2019), which has to consider the interplay between UG, knowledge that comes from the L1 and knowledge that comes from exposure to the target language (i.e. the L2 input) (Rothman and Slabakova, 2018).

There are competing theories as to how these factors interact. For example, some researchers advocate a Strong Continuity Hypothesis, which claims that the initial state L2 grammars consist of the L1 final state (Poeppel and Wexler, 1993; Schwartz and Sprouse, 1996; Grüter, 2006). Under this view, functional categories (such as (T)ense and (C)omplementiser) are present in L2 grammars from the onset of acquisition. In support of this position is that evidence for knowledge of L2 functional categories is often present in beginner learners' production data, for example, wh-questions and/or yes-no questions (Prévost, 2009). Furthermore, where this evidence is absent, it could equally be the case that learners know more than what is present in their production data, which typically reflects their L2 performance as opposed to their underlying competence.

Therefore, the lack of overt evidence in production data, for some researchers, does not entail a lack of knowledge of such (Grondin and White, 1996; White, 2003; Grüter, 2006; Lozano, 2021). The argument here is that all functional categories and feature specifications required for fully grammatical derivations are present from the outset, but that these are just not mapped onto the right morphological/phonological material yet. Lardiere (1998b,a), for example, examined naturalistic longitudinal production data of an end-state Chinese learner of English. She found that the learner seemed to have acquired all features associated with functional category T

(such as verb raising), but had persistent problems with inflectional morphology (eg. agreement morphemes). This was interpreted as being an imperfect acquisition of the mapping between syntactic and morphological development, which occur independently.

On the other hand, some researchers support a Weak Continuity (Hawkins, 2001; Bhatt and Hancin-Bhatt, 2002), one variant of which is known as 'Minimal Trees' (Vainikka, 1994; Vainikka and Young-Scholten, 1998). This position claims that the nature of learners' production data can be indicative of their respective L2 knowledge. Little or no evidence for inflection/auxiliaries/subject-verb inversion in initial state L2 grammars is therefore taken as evidence that functional categories T and C are not yet present. Instead, learners are hypothesised to move from lexical to functional categories incrementally, that is, from NP/VP to TP to CP, where each 'stage' constitutes the most robust grammar for a given speaker (Vainikka and Young-Scholten, 1998).

For example, a learner would be classified as being at the VP stage if their interlanguage *predominantly* consists of a VP-based grammar. That is, there would be a clear lack of evidence for those phenomena associated with functional category T, such as an omission of auxiliaries and/or verbal inflection. At the same stage, learners may also show some evidence of features associated with the next stage (TP), but these rarer instances should be seen as the more developed stage 'competing' with the current one (Vainikka and Young-Scholten, 1998). In order to establish that a learner has moved to the next stage, learners must show consistent productive use of overt material that constitutes evidence for this. In some cases, researchers have attempted to quantify this consistency; for example, the use of a relevant property in 60% of their utterances or more (Vainikka, 1994).

In contrast to the Strong Continuity, which claims that learner production data

can underrepresent their respective L2 knowledge, some proponents of the Weak Continuity posit that this can equally *over*represent this knowledge. Formulaic language plays a big part in this claim. Section 2.1.4 showed that many studies examining learners' production data at the initial state have found early use of complex conventional expressions; for example wh-questions (eg. *what is your name*), and do-negation (*I don't know*). However, because these expressions are produced in advance of respective L2 competence, and are prototypical in nature, they are more likely to be formulaic (Hanania and Gradman, 1977; Schmidt, 1983; Myles et al., 1998, 1999). Generative models conceptualise FL of this kind as a separate system to newly generated language, based on psycholinguistic and neurolinguistic evidence that these phenomena are subserved by different parts of the brain (Makuuchi et al., 2009; Bridges and Van Lancker-Sidtis, 2013; Van Lancker-Sidtis et al., 2015). This means that FL does not constitute reliable evidence for knowledge of L2 functional categories, as these phrases are hypothesised to be retrieved and produced holistically instead of generated by the interlanguage grammar. Therefore, the significance attributed to learners' production data plays a big part in the Strong vs Weak continuity debate (Hawkins, 2001), and relies on the identification of potential FL in L2 initial-state grammars.

As generative models view use of FL as a separate entity to the development of creative language use, the consensus is that L2 knowledge develops independently of FL analysis (regardless of the Strong vs Weak continuity debate). Krashen and Scarcella (1978) conclude from their review of the L2 literature up to that point that most results support the position that the creative construction process develops independently alongside the production of FL. They state that the fact FL appears to be 'immune' to the interlanguage grammar at first implies that these are part of a system that is separate from the process which generates rule-governed, compositional language (p. 286). Similarly, Bohn (1986) followed the naturalistic development of a child learner of English over a six month period. The author noted

that, at early stages, the child used modal auxiliaries in fixed conventional expressions such as *would you like* as a short term production tactic only. He concluded that these constructions were not used to acquire auxiliary verbs, and that no such learning strategy was detected throughout their development. This suggests that the child's production of these expressions were a product of a different system. Granger (1998) agreed with this concept, branding FL as a 'quick and advanced production tactic', as 'there does not seem to be a direct line from prefabs to creative language' (p. 157), inferring that these are necessarily part of a separate system.

However, this stance largely reflects the deficit of L2 longitudinal generative studies which compare learners' use of FL with their corresponding interlanguage development. This is because the central focus of generative SLA has primarily been on discovering the nature of UG, and the extent to which/how learners have access to this. That is, discovering what learners tacitly know without having been learned/exposed to in the input, and what role the L1 could possibly play in this process (Hawkins, 2008; White, 2012; Carroll and George, 2018). Such questions require carefully designed and executed experimental tasks from selected learner demographics that are made to elicit specific language phenomena, rather than the analysis of L2 production data over a period of time. As generative SLA is based around this 'Logical problem of Language Acquisition', there has been far less of a focus on how properties of the input and/or language use may influence the development of L2 knowledge. This is particularly true for SLA in taught- classroom settings (although see Marsden et al. 2018 on how input through instruction influences the development of polarity item *any*).

The above discussion has highlighted the differences between usage-based and generative models of SLA, particularly when it comes to the role that FL is perceived to play in L2 development. Chapter 3 now discusses the advantages of each framework, both methodological and theoretical, and highlights recent calls to bridge the

gap between their mutual exclusivity. It then outlines the kind of data that is required to systematically investigate the interplay between FL use and corresponding grammatical development, focusing on longitudinal datasets that document learners' L2 productions from the initial state and a significant amount of time thereafter. The Chapter concludes by bringing this information together to outline the current study's Research Objectives, which are borne through adopting a novel, multiparadigmatic approach to the concept of FL in SLA. In doing so, it hopes to offer a more comprehensive insight into this phenomenon than previous similar studies who have relied on one framework alone.

Chapter 3

The current study

Chapter 3 narrows the literature review down to the focus of the current study. Section 3.1 discusses the strengths of both usage-based and generative paradigms, outlining how the present study builds on these to provide a more unified approach to investigate the interplay between learners' use of FL and corresponding grammatical development. Section 3.2 then outlines the kind of dataset that is needed to investigate this phenomenon systematically, before Section 3.3 concludes with some research objectives.

3.1 The implementation of different paradigms

The discussion in Chapter 2 defined and operationalised the concept of formulaic language (FL), and outlined the role this is perceived to play in both usage-based and generative accounts of SLA. The fundamental difference between these frameworks is the extent to which L2 knowledge can be acquired solely from an interaction between the available input and domain-general cognitive systems (Rothman et al., 2019). As usage-based proposals are based on input manipulation, the interaction with and production of highly frequent and prototypical FL is seen as a catalyst for L2 acquisition. Through this usage, learners can analyse and subsequently extract patterns from exemplary surface forms through general cognitive mechanisms such as association, categorization, schematisation and entrenchment.

Conversely, in the generative tradition, the nature of UG and its interaction with other language properties and the L1 is what is used to explain the eventual outcome of SLA. This entails that L2 grammars have their own time course of development, despite what input manipulation and general cognitive systems can achieve. Hence, in this framework, FL is seen as a peripheral phenomenon and plays far less of a role in the development of L2 knowledge.

Both paradigms present different strengths when it comes to the description and analysis of SLA. The main strength of generative SLA is the centrality of linguistic theory within the framework, which provides a model to make sense of linguistic properties that have little obvious surface connection. This is because it is able to connect superficial surface phenomena to the same underlying syntactic category/feature, which allows researchers to uncover trends in acquistion based on underlying properties that are inherently related to one another (Rothman and Slabakova, 2018). For example, functional category Tense is responsible for surface phenomena such as inflectional morphology, verb-raising, auxiliary verbs and negation. It is only through examining all of these phenomena together that we can have an understanding of when/how learners' knowledge of T(ense) manifests. It also allows us to identify persistent difficulties related to this functional category that are common to learners of various demographics, and give systematic suggestions for why this might be. Furthermore, it enables generative researchers to uncover common interlanguage developmental stages, based on underlyingly related surface phenomena that emerge at the same time in development across various learner demographics- regardless the target language or evidence in the input. This is the basis for the Strong vs. Weak Continuity debate, as presented in Chapter 2.

On the other hand, the main advantage of usage-based approaches is that they can provide insight into how learners manipulate the input to facilitate L2 devel-

opment, particularly with regard to lexical learning. Their use of corpus-linguistics and traceback methodology with longitudinal datasets (as reviewed in Chapter 2) are a valuable tool to reveal what learners actually *do* with the language they are exposed to. These methodologies have success in relating some novel surface structures back to phrases that have clearly been deconstructed and subsequently generalised. They also highlight the importance of frequent, prototypical, formulaic phrases in the early stages of SLA, both as initial gateways to conventional L2 communication and as model exemplars of L2 structure thereafter.

Traditionally, there has been a marked divide between researchers working within either framework, leading the field of SLA to become somewhat polarised. Zyzik (2009) notes that, for some, the differences between each framework are 'irreconcilable' (see for example, Tomasello and Abbot-Smith 2002). This friction was clear in the past from open peer commentaries with titles such as 'Universal Grammar is Dead' (Tomasello, 2009) and 'Grammar is Grammar and Usage is Usage' (Newmayer, 2008).

However, there has been recent recognition that this gap needs to be bridged. It is clear that SLA as a sub-field would benefit from a multiplicity of approaches, given that one single approach cannot adequately address all aspects of this phenomena (VanPattern and Rothman, 2015; Rothman and Slabakova, 2018). In particular, the success of usage-based traceback methodologies in uncovering patterns of input manipulation has lead to a recent shift within generative SLA to focus more on how properties of the input can influence L2 development (eg.,Lidz and Gagliardi 2015; Yang and Montrul 2017; Hicks and Dominguez 2020), a question that was always recognised but not fully emphasised in this framework (although see Carroll 2001).

In fact, it has even been claimed that the area of mutual exclusivity between the two frameworks is not as wide as previously thought. Rothman and Slabakova

(2018) state that:

> 'The strict divide between the so-called sides of cognitive approaches to SLA is more a matter of tradition and mutual misunderstanding than tangible. Much work can be done at the crossroads of where data are neutral' (pg. 436)

The role of the input in SLA presents one such 'crossroad'. Rastelli and Gil (2018) state that the increased interest in input properties within generative SLA means that 'it is time to engage in an upfront manner with the hypothesis that statistical manipulation of the input and cognitive learning principles may impact on how both the peripheral and the core properties of the L2 grammar are learned' (pg. 254). Some generative scholars believe that usage-based notions and methodologies can prove useful in this regard. Rothman and Slabakova (2018) admit that 'other sub-fields of SLA have much to offer generative SLA in terms of methodology and beyond' (pg. 436), and Rastelli and Gil (2018) claim that in order to gain a better understanding of the influence of input properties on SLA, 'generative SLA should encompass concepts, methodologies and techniques of statistical and cognitive investigation mostly developed outside the generative field' (pg. 249).

This study takes inspiration from these recent calls for a more 'unified' approach to SLA, particularly with regard to how one salient property of the input, namely formulaic language, can influence L2 development. It aims to build on the strengths offered by both frameworks to provide a more systematic and comprehensive study of the role of FL in interlanguage development. In this way, it is hoped that developmental trends can be uncovered that would otherwise be missed if relying on one paradigm alone. Therefore, this study adopts a generative framework of grammar, where surface structures are presumed to be manifestations of underlying computational properties. This is because of the explanatory power that this framework offers in terms of linking surface structure phenomena together and identifying developmental sequences based on their emergence in L2 production data. It then

also applies a usage-based traceback methodology to the dataset, recognising that formulaic expressions and related schematic learning clearly play a large part in bootstrapping learners into creative L2 use. Ultimately, it aims to bring these analyses together to uncover and subsequently explain any developmental trends that emerge from the longitudinal datasets. Section 3.2 now discusses more about the kind of datasets that are required to carry out such an analysis.

3.2 Capturing the interplay of FL and grammatical development: The initial state and longitudinal data

In order to investigate the interplay of FL and corresponding grammatical acquisition (under whichever theoretical framework), it is important to analyse data that captures the nature of learners' L2 initial state and individual development thereafter. This is because, as reviewed in Section 2.1.4, highly entrenched FL that is most likely to be influencing syntactic development is characteristically present at the initial state, identifiable as being syntactically 'advanced' in comparison with other L2 utterances. These are often prototypical and highly frequent expressions derived from their early L2 input, which learners hold onto as 'phrasal teddy-bears' (Ellis, 2012). Longitudinal data across a significant period of time is then required to track how/if these phrases feed into the grammatical system of individual learners under analysis (Doughty and Long, 2003; Myles, 2005, 2015; Verspoor et al., 2021; Lozano, 2021). It is important that longitudinal datasets span over a significant period of time in order to identify clear developmental stages, and that the data is dense enough to give a representative picture of learners' L2 knowledge at each of these stages. Specifically, in order to best capture the nature of learners' L2 knowledge at various stages of learning, it is generally agreed that spontaneous, oral production data is preferable (Ellis, 2002c). Written output tasks are more susceptible to the conscious application of a pedagogical, metalinguistic

or idiosyncratic rule, and therefore cannot serve as evidence that the learner has accessed/acquired/internalised the underlying properties behind this rule (Paradis, 2004). The real-time pressure of the communicative situation, however, does not allow such a recourse to the metalinguistic system to the same extent, which gives for a better window into learners' implicit grammatical knowledge (Myles, 2015).

There are other desirable approaches to measuring the development of L2 knowledge; for example the analysis of longitudinal L2 comprehension data (e.g. White 2003; Grüter 2006) or repeated controlled experimental tasks over a period of time (e.g. Juffs and Fang 2022). However, such datasets are extremely rare (if not, absent) and difficult to construct in practice, meaning that any studies who adopt these approaches tend to triangulate their data along with oral production data (Tracy-Ventura and Myles, 2015). It is also difficult to identify potentially entrenched FL at the initial state without examining learners' oral production data, since an inherent characteristic of FL at this stage is learners' fluent production of such in advance of other evidence for associated L2 knowledge. Oral language in general (for L1 and L2 speakers) is far denser in FL than written language (Ellis et al., 2009), and it is well understood that the greater the working-memory demands of the processing task, the greater the need to rely on FL (Kuiper, 1996). If it is the case that FL is seeding grammatical acquisition, it is far more likely that this process will be identifiable in learners' spoken rather than written/comprehension data.

Despite consensus that the analysis of L2 longitudinal production data presents an ideal window into the nature of L2 knowledge development, this kind of dataset is notoriously difficult to obtain. As a result, this methodology has been overlooked and underused in previous studies of SLA- particularly those in the generative tradition. Difficulties largely pertain to the associated costs and time-consuming processes involved in having access to the same learners for a prolonged period of time (Myles and Cordier, 2017; Granger, 2021). As a result, previous studies investi-

gating any stage of longitudinal interlanguage development have traditionally relied on the analysis of production data from ad-hoc case studies (e.g. Haznedar and Schwartz 1997; Lardiere 1998b,a, 2008) or cross-sectional designs (e.g. Vainikka and Young-Scholten 1998; Bhatt and Hancin-Bhatt 2002). For example, Lardiere's (2008) longitudinal analysis of the L2 end-state relied on four recordings and 25 emails from one learner only (Patty), whilst Bhatt and Hancin-Bhatt (2002) derived developmental patterns of L2 English based on production data from different groups of learners in a public school across Grades 6-12. Likewise, many longitudinal usage-based studies examining the formulaic to creative continuum (as reviewed in Section 2.2.1) relied on production data from one or two learners only over a short period of time (e.g. Eskildsen 2008, 2015; Horbowicz and Nordanger 2021; Lesonen et al. 2020). Whilst from an SLA perspective generalisability is often not the goal as 'it is sufficient to know that a particular phenomenon has occurred' (Gass, 2013, p. 35), previous datasets such as these are potentially limited and lack extrapolability to the wider learner population (Lozano, 2021).

In order to systematically investigate the role of FL in the development of L2 grammatical knowledge, larger and better constructed databases are therefore required; in particular, those that document L2 productions from a number of learners at the initial state to a significant amount of time after this. Tracking a larger number of learners over a longer period of time means that, through the comparison of individual trajectories, common developmental patterns/trends can also be observed. Dense, naturalistic production data from more open-ended tasks is particularly favourable, where all possible contexts for the production of linguistically related surface phenomena are present so to best establish that a learner is within a certain developmental stage (Rankin, 2009; Myles, 2015). In light of these requirements, this study analyses data from two spoken learner corpora, which both document learners' L2 productions at the initial state and a significant period of time after this. Like most learner corpora, these datasets are large electronic collections

of L2 production data emanating from '(near) natural situations where learners can use their own wording rather than being prompted to produce a specific linguistic feature' (Granger, 2021). They can be seen as an ideal domain to observe naturalistic L2 output which comes from more open ended types of contextualised production tasks assigned to L2 learners (Le Bruyn and Paquot, 2021). Before introducing these two longitudinal learner corpora in the Methodology (Chapter 4), Section 3.3 below first outlines the research objectives that the analysis of these datasets aims to address.

3.3 Research Objectives

Based on the multiparadigmatic analysis of longitudinal learner production data as outlined above, the following research objectives can be distinguished:

(1) a. Are formulaic expressions evident in learners' production data at the initial state, and do these influence the development of L2 syntax thereafter?

b. Can this development be captured by schematic learning strategies, or do we find the development of underlying syntactic knowledge more generally? That is, can novel L2 utterances be traced back to formulaic exemplars based on related utterance schemas, or related computational properties?

c. Is there evidence of a common developmental trajectory across learners, regardless of the use of formulaic expressions and schematic learning strategies?

Chapter 4 now documents the methodology and introduces the two longitudinal corpora used for analysis, and reinforces why a combination of these datasets is the ideal test ground to address the research objectives as stated above.

Chapter 4

Methodology

This Chapter presents the Methodology, in terms of the data used for analysis. It presents the two longitudinal learner corpora that are chosen for analysis in light of the requirements outlined in the previous chapter. These are the Barcelona English Language Corpus (BELC) (Section 4.1) and the French Progression Corpus (FPC) (Section 4.2). These sections focus on their learner demographics and various task types involved in their construction. The Chapter finalises briefly by highlighting the strengths of combining an analysis of both corpora to address the research objectives (Section 4.3).

4.1 The Barcelona English Language Corpus (BELC)

The Barcelona English Language Corpus (BELC) is made up of spoken transcripts from longitudinal Spanish/Catalan bilingual learners of English as Foreign Language (EFL). The BELC is a corpus built by the Barcelona Age Factor (BAF) project (Muñoz, 2006) to examine the effects of age on the acquisition of English as a Foreign Language. The research was conducted with students from state schools in Catalonia (Spain), which shared a similar socioeconomic background and similar teaching methodologies. Catalonia is a bilingual community with a majority language, Spanish, and a minority language, Catalan, which is the language of instruction in the state school system in Catalonia. The age of onset to the L2 for all

students was 9 years old, when they were first exposed to English in limited taught sessions in their state school classroom. Recordings began from age 10, where learners can still be classified as beginners and therefore representative of the L2 initial state.

Spoken data was collected from the students on four separate occasions, which can be seen in Table 4.1 below.

	age	hours of instruction
early years	10	200
	12	416
later years	16	726
	17	826

Table 4.1: The four rounds of data collection and corresponding hours of classroom English instruction (accumulative)

Pupils with only school exposure to English fulfilled the conditions for comparison in the data. For example, it was not the case that any of these pupils had more hours of instruction via extracurricular exposure or retaking a course grade. Out of all the participants, 55 of these could be classified as longitudinal learners, that is, those who took part in various spoken tasks at more than one age. Out of these 55, there were only a handful of learners who participated at every age (10, 12, 16, 17), and only some who participated over 3 ages. As the focus of the present study relies on the close analysis of language features over the course of learners' development, it was key that all longitudinal participants I chose to focus on had participated across at least 3 different age groups, in an attempt to best identify clear stages in their interlanguage development. This resulted in 9 learners available for analysis.

The data was collected via several spoken tasks, which were created to elicit naturalistic production data. These consisted of an interview, narrative and role-play. The interviews are semi-guided, beginning with a series of questions about the learner's family, daily life and hobbies. This also constituted a warming-up phase

that helped students feel more at ease. The interviews were made to elicit as many responses as possible from the learners, and the interviewers would accept learner-initiated topics in order to create a natural and interactive situation. The learners were also given the opportunity to pose questions to the interviewer at the end of the interview.

The narrative task was elicited from a series of six pictures which the learners could freely look at before and during their telling of the story to the interviewer. All learners participated in the same narrative task which involved two main protagonists, a boy and a girl, who are getting ready for a picnic; a secondary character, their mother; and a character that disappears and later reappears- a dog that gets into the food basket and eats all the children's sandwiches. The role-play task was performed in randomly chosen pairs, where one of the students was given the role of the parent and the other the child, which they would swap after completing an interaction. The learner acting as the child was required to ask permission to have a party at home, and both students were asked to negotiate arrangements i.e. setting, time, activities etc. The interviewer would give the initial instructions and when needed also elicited talk by reminding learners of topics for discussion or led the task to its completion by asking about the outcome of the negotiation.

Transcripts of the BELC are accessible via slabank.talkbank, which separates the production files into the three separate task types. However, I do not distinguish between task types in the following sections when discussing BELC learner production data. I instead combined all tasks for individual learners (by copying and pasting the transcripts into an empty Word document), resulting in large individual learner transcripts that contained spoken data from every task they participated in, at each different age. This made it clear to see each individual learner's utterances across each age/proficiency level, which enabled me to examine developmental trajectories and conduct a traceback analysis.

4.2 The French Progression Corpus (FPC)

The second dataset chosen for analysis are the semi-naturalistic spoken transcripts from English adolescent classroom learners of L2 French, taken from the French Progression Corpus (Myles et al., 1998, 1999). One of the main objectives of creating the FPC was for the authors to explore the role of prefabrication (i.e. use of formulaic expressions) and its relationship with creative language use. This means that the corpus is an ideal test ground for the present study which seeks to address a similar concept. The corpus was constructed within the context of a three-year project, 'Progression in Foreign Language Learning', and recordings took place in secondary schools in Southern England.

The project tracked 60 pupils in two secondary schools during six terms of classroom French (from their second term of French language study to their seventh term, inclusive), and individual pupils' progress in spoken French was monitored through a programme of interviews and problem solving activities once per school term (Myles et al., 1999). All learners were 11 or 12 at the first round of data collection (in Term 2 of 'Year 7' at school) and had started their first formal study of National Curriculum French one term before. They can therefore be classified as beginner learners and representative of the L2 initial state. By the end of the data collection period (in Term 1 of 'Year 9' at school), learners were either 13 or 14.

The data collection period therefore spans across two years, and amounts to around 200 hours of production data across 6 school terms of instruction[1]. Audio recordings were transcribed and produced as individual learner files, which are openly accessible online via FLLOC (French Learner Language Oral Corpora).

[1]The exact number of instructional hours that this amounts to is not specified in the corpus metadata or indeed any publications describing the building of such.

Myles and colleagues make clear that the secondary school classroom was the prime source of French language experience for all learners, with exceptions being the possibility of family holidays abroad and school exchange trips, which were not controlled for (Myles et al., 1998, 1999; Myles, 2004). The authors also recorded the French as a foreign language lessons attended by the learners of the corpus on a fortnightly basis, which allowed for an insight into the language learning classroom/linguistic context.

The present study analyses a subset ($n = 24$) of the FPC from various tasks across Rounds 1 (Year 7 Term 2 Ages 11-12), 5 (Year 8 Term 3 Ages 12-13) and 6 (Year 9 Term 1 Ages 13-14), as highlighted in Table 4.2 below.

Round 1 Year 7 Term 2 Ages 11-12	Round 2 Year 7 Term 3 Ages 11-12	Round 3 Year 8 Term 1 Ages 12-13	Round 4 Year 8 Term 2 Ages 12-13	Round 5 Year 8 Term 3 Ages 12-13	Round 6 Year 9 Term 1 Ages 13-14
A: One to one with investigator (photos stimulus)	A: One to one with investigator (photos stimulus)	L: Loch Ness narration	A: One to one with investigator (photos stimulus)	I: Interrogative elicitation task (picture description)	L: Loch Ness narration
D: Pair task (conversation and role play)	D: Pair task (conversation and role play)	R: Pair task (picture description)	G: Information gap pair task	D: Pair task (conversation and role play)	G: Information gap pair task
R: Pair task (picture description)					

Table 4.2: The three rounds of data collection and corresponding spoken tasks chosen for analysis

Reasons for choosing these specific tasks and rounds of data collection are as follows. As with the BELC, learners' varied in how many rounds of data collection they participated in, which limited the specific tasks I could analyse in order to achieve a representative longitudinal comparison across learners. In their studies, Myles and colleagues analysed learners' production data across a larger range of rounds, but their subsets were much smaller; 8 learners in Myles et al. (1998), and 16 in Myles et al. (1999). As the subset of the BELC is relatively small ($n = 9$), I opted for analysing a larger sample of learners over less rounds, in an attempt to complement the small sample of learners in the BELC analysis.

Task D at Round 1 (Pair Task conversation and role play) allowed pairs of learners to exchange personal information, asking and answering questions and complet-

ing a written information card about their partner. Task R (Picture description pair task) at this round was also done in pairs and encouraged learners to ask questions, as one learner had to reproduce a picture of a cartoon person that was described to them by the other learner, which was given on a range of different cue cards. For Task I at Round 5 (Interrogative elicitation task picture description), learners worked on a one-to-one basis with an interviewer. The interviewer had a picture of a French townscape scene, including four characters doing different activities, and the learner had the same picture with the characters missing. The learner had to ask questions about what each character was doing to then add these to the townscape, before comparing their final pictures to the original ones of the interviewer.

Task L (Loch Ness Narration) again involved learners working individually with the interviewer, who told learners a story with picture cards that learners then had to re-tell afterwards with limited prompting and support. Finally, Task G (Information gap pair tasks) saw learners back in pairs, where they had to exchange information about their own likes and dislikes to make arrangements to go out in an imaginary town (Belleville), and to decide whether to take a friend. Learners were provided with different cue sheets, which included diary information and possible Belleville activities.

4.3 Using the two learner corpora to investigate the present research objectives

As mentioned in Section 3.2, traditionally, longitudinal SLA studies feature much smaller samples of learners due to the difficulties in acquiring data in this fashion. The analysis of a larger number of learners in both the BELC and FPC can therefore be seen as offering a unique longitudinal perspective, in that it provides the opportunity to track each learner individually as well as observe common developmental trajectories through the comparison of these individuals.

Furthermore, the BELC and FPC can be seen as complimentary longitudinal datasets when it comes to investigating the current research objectives. The BELC gives a general picture of interlanguage development over a substantially prolonged period of time (7 years), whereas the FPC is a more concentrated window into learners' development within the first two years of classroom instruction. This will allow for a comprehensive analysis of how FL may be manipulated at different intervals throughout the SLA process, and any effect this may have on learners' creative language use at each developmental stage. In addition, the comparison of learners' acquiring two different target languages (English and French) will allow for interesting observations regarding the kind of prototypical FL that is used in either context. As both groups are classroom learners, it could be that the kinds of functional, prototypical phrases taught in their L2 input are similar. Related to this, English and French display some syntactic variation. If learners are using similar formulaic expressions, it will therefore be interesting to see if these are manipulated differently in both languages, as a result of the their differing syntactic features. Finally, comparing the acquisition of two different L2s will perhaps allow for the discovery of any common patterns of syntactic development in these datasets- whether based on FE analysis or not.

What is also important for the present study is that both the BELC and FPC were created to examine age and classroom instruction as variables on L2 acquisition. This means that both corpora began tracking learners from the initial state and were constructed to minimise all other variables that may influence the rate of L2 development. The BELC specifies, for example, that learners in the sample could not have spent any time abroad in an English-speaking country or have attended out-of-school English classes. As learners grew older, these conditions could not always be met, and so the number of potential longitudinal participants gradually decreased (Celaya, 2019). Similarly, no learners of the FPC took part in extracurric-

ular French lessons outside of their state-school classroom. Controlling for external factors such as these means that both sets of learners are extremely comparable, and that any results from the analysis are likely to be representative of the 'classroom learner' demographic more generally. Additionally, it means that the learners' linguistic environment is highly predictable, which will help to identify prototypical FL in their classroom input.

The two chapters that follow now document the analysis of the two corpora as presented above. Chapter 5 begins with the BELC.

Chapter 5

Analysis: The Barcelona English Language Corpus

Chapter 5 presents the analysis of the BELC. Sections 5.1 - 5.3 show how formulaic expressions (FEs) were identified, and Section 5.4 documents their production by all learners across the data collection period. Section 5.5 outlines the assumed syntactic derivation and related computational properties of the FEs under mainstream generative models. It also details how these computational properties can manifest in production data (5.5.6) and how I measured learners' knowledge of such (5.5.7). Section 5.6 then analyses the FEs as abstract schematic constructions under usage-based models, and presents how learners could potentially derive related structures by using these abstract templates. Finally, I document how I have tailored the analysis to best suit the two longitudinal datasets, focusing specifically on the normality of data (5.7.2) and significance (p values) levels (5.7.3).

5.1 Learners' linguistic environment

Although the learners of the BELC did not attend the *same* state school, they all attended state schools in Catalonia, which share similar teaching materials and methodologies. The state-school language learning classroom was also the only en-

vironment that learners were exposed to English.[1] This entails that learners' principle L2 input was that of EFL teaching materials and teaching instructors inside the state-school classroom.

Wray (2008) believes that a given sequence of language is formulaic and hence subject to holistic retrieval when there is a greater than chance-level probability that the speaker has encountered the sequence before. Additionally, in foreign language classes, learners are taught holistic sequences to fit typical communicative purposes (Nattinger and DeCarrico, 1992; Lewis, 1993; Myles and Cordier, 2017). It is well known, in particular, that *textbooks* constitute the main and sometimes only source of input for language practice both in and outside the classroom in EFL contexts (Richards, 2005), and that 'much of the English language teaching to L2 speakers that occurs throughout the world today is conducted through the medium of textbooks' (Menkabu and Harwood, 2014, p. 145).

We can therefore predict that a significant amount of formulaic L2 input learners were exposed to is derivable from the EFL textbook used in their class. It has generally been agreed that 'textbook analyses of particular constructions compared with corpus data is a good point of departure for future, more rigorous studies of input in instructed L2 settings' (Zyzik, 2009, p. 56). Whilst information regarding the exact textbooks used in each of the learners' state schools is absent from the metadata of the corpus, it is possible to identify textbooks which would be representative of those teaching materials used in this particular learning context.

[1]By this, I mean by as far as could be controlled by the Age Factor Project, i.e. no time spent abroad in an English speaking country, no extra curricular classes in English etc. Whether or not learners were practising English independently at home/watching English television programmes is something that is harder to monitor, and therefore more unpredictable.

5.2 Identifying 'representative' EFL textbooks for analysis

Given the variety and number of EFL textbooks is high, any 'representative' sample is by necessity a convenience sample. A distinction can be made between *local* and *global* textbooks. The former are 'designed in and for a situated context, with a given audience in mind at a national or regional level' (Damien, 2018, p. 12) and the latter are 'a coursebook which is not written for learners from a particular culture or country but intended for use by any class of learners in the specified level and age group anywhere in the world' (Tomlinson, 1998, p. 15). Despite global textbooks being more commonly found in international EFL contexts, to achieve the best representativeness of textbook language presented to the BELC learners, it makes sense to analyse both global and local textbooks that are commonly used in Spain. The inclusion of local textbooks in the sample frame is particularly important when it is considered that these learners are students of Spanish state secondary schools, hence their English teaching environment is local rather than international.

Criado and Sanchez (2009) identify seven representative local EFL textbooks of different educational levels and modalities in Spain. They divide these up into secondary, upper secondary, teenager and adult textbooks. All EFL textbooks identified in the study refer to the first year in each of the educational levels selected, which is important since the present study aims to extract potential formulaic phrases that are presented *early* to the BELC learners and hence more likely to act as acquisitional seeds throughout their course of L2 development. Criado and Sanchez (2009) divide compulsory educational levels in Spain into Compulsory Secondary Education (Enseñanza Secundaria Obligatoria- 11-15 year old students) and Upper Secondary Education (Bachillerato/Baccalaureat- 16-18 year old students). The authors also identified representative textbooks for teenagers who attend private language schools in Spain, as well as those of the Official Schools of Languages (Escuelas Oficiales de

Idiomas), who are above the age of 16 and can be classified as adult learners. As the BELC participants are EFL learners at local state-schools, the representative textbooks of compulsory secondary education for children aged 11-18 were chosen as the sample local textbooks for analysis. This also aligns with the ages of learners when they participated in the data collection tasks, which were between 10-17 years old. The two 'local' (Spanish) EFL textbooks identified for analysis are therefore the following:

(1) *Challenge for ESO 1.* By Charlotte Addison and Pamela Field. Burlington Books Espana, 2006.

(2) *Bachillerato Made Easy 1.* By Adela Fidalgo, Alberto Fontanillo, Immaculada Mayorga and Sarah Dague. Santillana Richmond Pubishing, 2001.

The selection of representative global EFL textbooks was a relatively simpler process. Following Burton (2019), the main selectional criteria for this sample was that both titles should be published by international publishing houses and should be commercially successful. Burton (2019) identifies the top 5 global EFL textbooks in this regard. From these, I chose to use the Elementary series of *New Headway* and *New English File*, the former of which is now in its fourth edition and has sold over 100 million copies (Oxford Annual Report of the Delegates of the University Press 2010/2011: 7). The full titles can be seen below:

(3) *New Headway Elementary 4th Edition.* By Soars and Soars; Soars, Soars and Hancock. Oxford University Press, 2009-2015.

(4) *New English File 2nd Edition.* By Oxenden et al; Oxenden and Latham-Koenig; Oxenden, Latham-Koenig and Seligson. Oxford University Press, 2005-2010.

Once the textbooks were chosen for analysis, the next stage was to extract all speaking exercises from all four representative textbooks, and identify all phrases

that were presented to learners in each of these tasks. It was spoken tasks from the first half of the textbooks only that were chosen for analysis. This was done for a number of reasons. Firstly, these are the stages which constitute the learners' earliest L2 input/exposure, and FEs that are presented early and frequently to the learner are most likely to play a role in their language development (Wray and Fitzpatrick, 2008; Ellis, 2012; Ellis et al., 2015). It can be presumed, therefore, that these preliminary sections are more likely to contain functional phrases that are subject to memorisation and internalisation. This is also in line with psycholinguistic studies who posit the first-in-last-out principle, whereby formulaic language that is acquired early in the acquisition process has the propensity to remain the longest in a learner's mind (Lindholm and Wray, 2011). Secondly, it is widely accepted that a representative picture of the nature of EFL teaching materials can be obtained through the analysis of a very small sample. Littlejohn (2011) states that the analysis of 10-15 percent of instructional materials suffices to gain an accurate understanding of their nature, and McDonough et al. (2013) suggest analysing two or three chapters from a textbook is enough to obtain a comprehensive analysis of the textbook.

I chose to extract phrases used in the spoken tasks only, as these are the instances in which learners are required to actually *use* the L2 via the spoken mode, whether these be through various role-play scenarios or simple repetition/drilling exercises. It can be presumed that these phrases which learners are required to make use of repeatedly in the classroom are those that are more likely to be entrenched in their mind/subject to memorisation (e.g., Swain 1985). Further, as the BELC consists of these learners' *spoken* production data, it is likely that any expressions they choose to use during elicited production tasks will be those that they commonly encounter during speaking tasks in the classroom. Once all phrases from the spoken tasks were collected, I identified the four most frequent expressions included in all spoken tasks across all four textbooks. I then considered the extent to which these expressions

could potentially be formulaic, on account of their prototypicality, functionality and high frequency in learners' classroom input.

5.3 Formulaic expressions identified for analysis

The four most frequent expressions extracted from speaking exercises presented holistically across the first half of all four representative textbooks can be seen below.

(5) Four most frequent expressions extracted from representative textbook analysis

a. *what's/is your name?*

b. *how old are you?*

c. *where do you live?*

d. *where are you from?*

These expressions are the only ones that are repeated consistently and continually across the spoken tasks of the sample EFL textbooks, therefore we can assume that learners have been frequently exposed to them in their classroom input. The expressions are clearly conventional in nature, where conventionality refers to a native speaker's overwhelmingly preferred way of expressing a particular concept in a particular social context (Coulmas, 1981; Bardovi-Harlig and Stringer, 2017; Bardovi-Harlig, 2019). 'Concept' here relates to the functions that these expressions are fulfilling in specific discourse/pragmatic scenarios. We could posit, for example, that they are a native speaker's preferred choice for realising the concepts [ask name] in 1, [ask age] in 2 [ask living destination] in 3 and [ask place of birth/origin] in 4.[2]

We can establish the conventionality of the expressions by running them through a referential native speaker corpus to compare their frequency against other possible

[2]See Myles (2004) for a similar proposal of how classroom EFL learners map conventional formulas to functional concepts.

grammatical alternatives in the same discourse/pragmatic contexts. This was done for the above phrases via the British National Corpus (BNC). The results are conclusive to these expressions' conventionality, which can be seen below. I have also included the 'echo' question form of the identified expressions (where the wh-word is in-situ) for comparison.

(6) Concept/function: [ask name]

a. *what is your name?* [0.48 instances per million words]
b. *what are you called?* [no matches]
c. *what do they call you?* [no matches]
d. *your name is what?* [no matches]

(7) Concept/function: [ask age]

a. *how old are you?* [1.17 instances per million words]
b. *what is your age?* [no matches]
c. *you are how old?* [no matches]

(8) Concept/function: [ask living destination]

a. *where do you live?* [0.42 instances per million words]
b. *where are you living?* [0.04 instances per million words]
c. *where is your home?* [0.01 instances per million words]
d. *where is your house?* [0.01 instances per million words]
e. *you live where?* [no matches]

(9) Concept/function: [ask place of birth/origin]

a. *where are you from?* [0.35 instances per million words]
b. *where do you come from?* [0.29 instances per million words]
c. *where do you originate from?* [0.1 instances per million words]
d. *where is your place of birth?* [no matches]
e. *you are from where?* [no matches]

As these expressions are the most *frequently* used expressions by *native speakers* in specific communicative contexts, it can be said that the phrases pass both the 'frequency'[3] as well as the 'prototypical' criteria, irrespective of their raw frequency tokens in the representative EFL textbooks. All expressions are also high in functionality, as they are closely tied to specific discourse/pragmatic contexts. Their salience in EFL textbooks is therefore unsurprising, given that a mastery of such at early stages of acquisition would be beneficial to learners in conventional communicative L2 contexts (Fillmore, 1979). Furthermore, as these expressions are all wh-questions, they are syntactically complex, displaying, for example, wh-fronting and subject-verb inversion. There is therefore potential for these expressions, at beginner stages, to act as model exemplars of these syntactic operations for learners who are exposed to/use them. To summarise, under all definitional/categorical accounts, these fixed wh-expressions (henceforth FEs^{wh}) are prime candidates for acquisitional seeds. Section 5.4 now investigates the nature of learners' production of these expressions across the data collection period.

5.4 Learner productions of the FEs^{wh} across the data collection period

As I had collated the BELC transcripts as individual learner files, (see Section 4.1), this allowed for the identification of the FEs^{wh} in learners' transcripts across the data collection period. Table 5.1 shows that the extracted FEs^{wh} are produced at each age by every learner under analysis. Where learners did not participate in the spoken elicitation tasks at a particular age (as the criteria for selection in Chapter 4 specified learners had to participate across at least three out of the four rounds

[3]Frequency here refers to how the identified conventional expressions are more frequent compared to their other grammatical alternatives. When taken as raw figures in isolation, some researchers would interpret these counts as low. Biber et al. (2004), for example, posit that expressions must meet or exceed a range of 10-40 occurrences per million words in a native speaker corpus. These kinds of figures are associated with approaches who use frequency *alone* to determine conventionality, rather than as a criterion used along with native-norms, predictability of exposure and prototypicality in use.

of data collection) this is indicated by *no transcript*. Where learners participated in spoken elicitation tasks but did not produce the FEs^{wh}, this is indicated by a dash -. Note also that the FEs^{wh} appear in the table as an exact representation of the production by a particular learner. For example, *what is your name* implies a fluent production of the expression, whereas one such as *hm < whatare > [//] what [//] what is your name* indicates both a false start (by means of < *and* >) and repetition of *what* (by means of [//]).

learner	age 10	age 12	age 16	age 17
2	-	-	*Hm <what are> [\ \] what [\ \] what is your name how old are you*	*what's your name*
5	*what's your name*	-	*what's your name*	*what's your name*
7	*what's your name*	*what's your name how old are you where do you live*	-	*what's your name first of all how old are you (x2) well and where do you live*
13	-	-	*what's your name where do you live what's her name*	*what's your name and where do you live how old are you*
18	-	*what's your name how old are you *what do you live*	-	-
27	*no transcript*	-	*how old are you where do you live*	*how old are you where do you live what is your name*
38	*no transcript*	*what's your name *where you live*	*what's your name*	*where do you live now*
42	*no transcript*	-	*what is your name where are you from*	*what is your name *where is you from*
47	*how old are you what's your name*	*no transcript*	*how old are you where do you live*	*how old are you (x2) what's your name where are you from*

Table 5.1: Productions of the extracted FEs^{wh} across the BELC

Overall production frequencies can be compared below.

(10) Learner productions of the FEs^{wh} across the corpus

a. *what is your name?* [15 tokens]

b. *how old are you?* [12 tokens]

c. *where do you live?* [10 tokens]

d. *where are you from?* [3 tokens]

Learners' production of the FEs^{wh} across the data collection period supports the categorisation of these expressions as functional and prototypical. Although these may seem like low frequency counts out of context, for low-proficiency levels during naturalistic L2 production tasks, this level of conformity and consistency can be considered quite high. It also confirms the representativeness of the textbooks chosen for analysis for this particular learner demographic and classroom context. The majority of these productions are during the latter stages of the interview task, where learners are given the opportunity to pose questions to the interviewer. It is unsurprising that learners fall back on the FEs^{wh} in such a scenario, given their high prototypicality in 'greeting' scenarios and assuming that they have had a sufficient amount of practice with these phrases via frequent exposure in their EFL teaching materials/tasks.

Further, nearly all of learners' FE^{wh} productions are fluent. This suggests that learners have had sufficient practice with and exposure to these expressions. The only erroneous productions are Learner 18 at age 12- **what do you live?*, Learner 38 at age 12- **where you live?* and Learner 42 at age 17- **where is you from?*. There is only one instance of repair/repetition; Learner 2 at age 16- *hm < whatare > [//] what [//] what is your name?*. There are also only four instances of the expressions being produced in utterances containing additional material, which can be seen below.

(11) Learner 4 age 17: *what is your name [first of all]?* (FE^{wh} with idiomatic adverbial phrase)

(12) Learner 4 age 17: *[well and] where do you live?* (FE^{wh} with clause initial discourse marker/conjunction)

(13) Learner 38 age 17: *where do you live [now]?* (FE^{wh} with external adverbial modifier)

(14) Learner 13 age 17: *[and] where do you live?* (FE^{wh} with conjunction)

I have included Learner 13's production of *what's her name* as a variant of *what's your name*, since this differs by the replacement of the possessive pronoun only. Note that no other learner was shown to produce grammatical and/or lexical variations such as these, and instead stuck to the FEs^{wh} in their 'base' form as presented in the EFL textbooks.

To summarise, the fluent and consistent nature of learners' productions of the FEs^{wh} as documented above is indicative of the expressions' formulaicity to these learners. To investigate this concept further, I now outline the expressions' underlying syntactic properties under a generative framework (Section 5.5) and abstract schematic constructions under a usage-based model (Section 5.6). This is to set up the corresponding analysis of learners' knowledge of these properties alongside use of the FEs^{wh} across the data collection period, the results of which are presented in Chapter 6.

5.5 The assumed syntactic derivation and related computational properties of the FEs^{wh} under generative frameworks

This section outlines the assumed syntactic derivation of the FE^{wh} following a mainstream Minimalist framework (e.g. Chomsky 1995) as adopted by Radford (2009). Since the FEs^{wh} are all wh-expressions, they all share the same four computational derivations, driven by features on functional categories T(ense) and C(omplementiser). These are wh-movement, T-C movement (subject-verb inversion), V-T movement (v-raising) and A-movement. *Where do you live* also involves

do-support. The expressions' syntactic structures will be displayed graphically as binary branching tree diagrams, as these best represent information about hierarchical structures (i.e. their structural relations) (Yang, 1999). I will present the derivation of each FE^{wh} separately, before bringing these together to examine all possible surface structures which are manifestations of the FEs^{wh}' shared computational properties. I will then discuss how these surface structures can be analysed as evidence for knowledge of these computational properties and their associated functional categories in L2 production data. This is to understand the nature of learners' L2 syntactic knowledge alongside their FE^{wh} use at each stage of data collection.

5.5.1 *What is your name?*

Figure 5.1 below shows the syntactic derivation of the FE^{wh} *what is your name.*

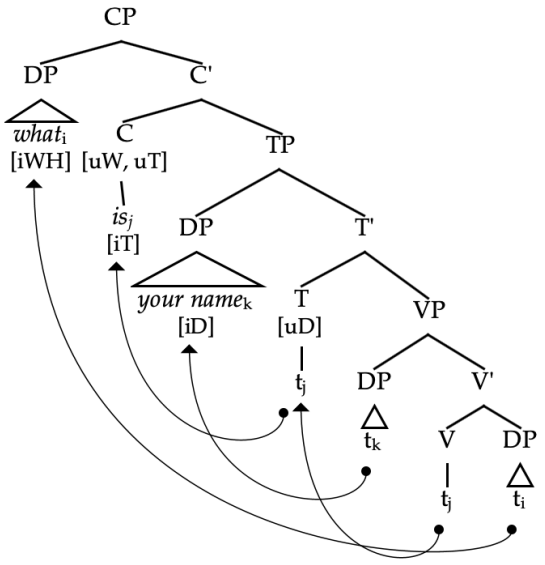

Figure 5.1: *What is your name?* assumed syntactic structure

The assumed derivation would be as follows. *What* is a determiner phrase (DP) which carries an interpretable wh-feature [iWH] by virtue of being a wh-word/expression [4], which is 'in-situ' as a complement of the head V *is*. This DP

[4] A wh-expression is an expression containing an interrogative word beginning with wh- i.e. *what, which, who* etc, but also encompasses *how* where its behaviour is syntactically similar (Radford et al., 1999).

is merged with V (*is*) to form the VP (verb phrase) *is what*, which has the DP *your name* as an internal subject in its specifier position in accordance with the VP-Internal Subject Hypothesis (VPISH) [5] (Radford, 2009). As *your name* is a nominal subject in specifier VP, it carries a [+interpretable] D feature (Alexiadou and Anagnostopoulou, 1998).

The whole VP is then merged with a finite T (tense) constituent which triggers raising of the V *is* to T [6], an instance of internal merge sometimes referred to as V-T movement (Radford, 2004; Sportiche et al., 2014) or Verb Raising (O'Grady, 2011). Now *is* has been moved to T and is therefore finite, it carries an interpretable tense feature [iT]. T also contains the [EPP] (Extended Projection Principle) with a [-interpretable] D feature, which requires it to have a nominal syntactic subject which it agrees with in person/number as its specifier. To check the EPP property, this [-interpretable] D feature attracts the [+interpretable] D feature of the DP *your name* to the specifier position within the TP, a process known as A-movement.

The TP *your name is what* is then merged with the null interrogative complementiser C, which carries an uninterpretable interrogative feature [uWH] requiring it to have an interrogative specifier (i.e. a specifier which carries an interpretable interrogative freature [iWH]) in accordance with the Interrogative Condition[7] (Radford, 2004). The [uWH] on C thus attracts the [iWH] on the DP *what* to move from its VP complement position, a process known as wh-movement. Since the FE^{wh} *what is your name* is a main-clause/root question, C additionally carries an

[5]The VP-Internal Subject Hypothesis claims that non-expletive subjects originate internally within the VP containing the relevant verb (Radford, 2009).

[6]Here I am assuming, following Radford (2009) and Sportiche et al. (2014), that *be* in its use as progressive auxiliary or copula verb originates in a position lower than T and consequently undergoes raising in finite structures. Evidence for such can be seen from contrasting utterances such as *she may not be well*, where non-finite copula *be* follows the negative particle *not* and hence occupies the head V position in the VP, compared with finite copula *be* where *is* precedes the negative particle and occupies the head T position of TP in examples such as *she is not well*. However, because surface structure evidence for V-T movement is so limited in English, I choose not to investigate learners' knowledge of such in the sections that follow.

[7]The Interrogative Condition states that a clause is interpreted as a non echoic question if it is a CP with an interrogative specifier.

uninterpretable tense feature [uT] which attracts the interpretable tense feature [iT] of *is* to move from T to C, attaching to a null affixal interrogative complementiser (Radford, 2009). This process is commonly referred to as T-C movement, or subject/verb inversion (Radford, 2009; Sportiche et al., 2014).[8] As can be seen from Figure 5.1, I have used the notation 't' as shorthand for 'trace' with corresponding integers to indicate that a moved element leaves behind a (full, not pronominal) null copy of itself in the position out of which it moves, i.e. its extraction site, following Chomsky's copy theory of movement (1995).

To summarise, the syntactic derivation of *what is your name* involves various suboperations of Internal Merge, namely V-T movement, A-movement, T-C movement and wh-movement, which are driven by features on functional categories T and C.

5.5.2 *How old are you?*

Figure 5.2 below shows the assumed syntactic structure of *how old are you?*.

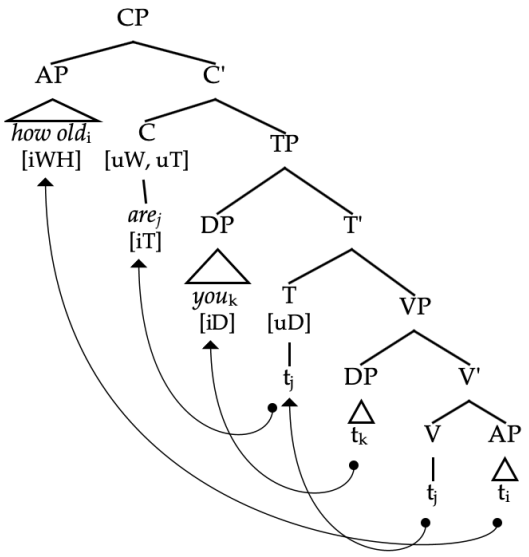

Figure 5.2: *How old are you?* assumed syntactic structure

The assumed syntactic derivation of *how old are you?* would be somewhat

[8] Wh-movement and T-C movement can also be analysed as strategies for checking the EPP property of the [uWH] and [uT] features on C (Prévost, 2009).

identical to that of *what is your name?*, involving the computational properties V-T movement, A- movement, T-C movement and wh-movement. One thing to draw attention to is the fact that in this structure, the whole AP (adjective phrase) is moved (via wh-movement) to the specifier of C, rather than the wh-word in isolation *how*. Assuming the AP *how old* has the internal structure of Figure 5.3 below, movement of *how* alone would result in the ungrammatical surface structure manifestation **how are you old?*[9].

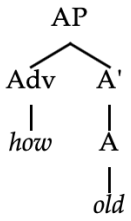

Figure 5.3: *How old* assumed syntactic structure

This process is commonly referred to as 'pied piping' (Ross, 1967) or 'convergence' (Radford, 2004) and refers to when subordinate material in the c-command domain of the wh-word is dragged along with it in wh-movement. This is by way of satisfying both the Chain Uniformity Principle (Chomsky, 1995) and the Attract Smallest Condition (Radford, 2009), whereby C attracts the smallest possible interrogative constituent which is uniform with regard to its phrase structure status.

5.5.3 *Where do you live?*

Figure 5.4 below shows the assumed syntactic structure of *where do you live?*.

[9]I mean ungrammatical here for the intended meaning being [ask age].

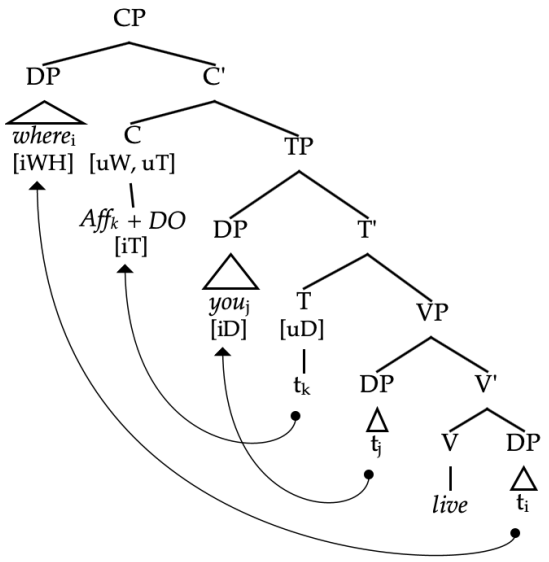

Figure 5.4: *Where do you live?* assumed syntactic structure

The assumed derivation of *where do you live?* does not differ greatly from those of *what is your name?* and *how old are you?* as seen above. What differs is use of the dummy auxiliary *do*, which I analyse as follows. Since C contains an uninterpretable tense feature [uT] by virtue of *where do you live?* being a main-clause/root question, this attracts whatever is contained within T to adjoin to it via T-C movement. In this case, it is an abstract affix on T with an interpretable tense feature [iT] which is attracted to C. As this moved affix on C is stranded/unhosted (i.e. it is not attached to a verbal host as its V feature requires), it is instead spelled out as an appropriately inflected form of the dummy auxiliary *do*, a process commonly referred to as do-support/do-insertion[10] (Radford, 2004; Sportiche et al., 2014).

5.5.4 *Where are you from?*

Figure 5.5 below shows the assumed syntactic structure of *where are you from?*.

[10]Note importantly that, although this analysis takes do-support as applying at the spell out stage of derivation, this is still taken to be an instance of T-C movement as the null (abstract) affix on T has to move to C in order to be spelled out as an appropriately inflected form of *do*. This same analysis of T-C movement of an abstract affix is also assumed for the derivation of yes/no questions such as *do you like sport?*.

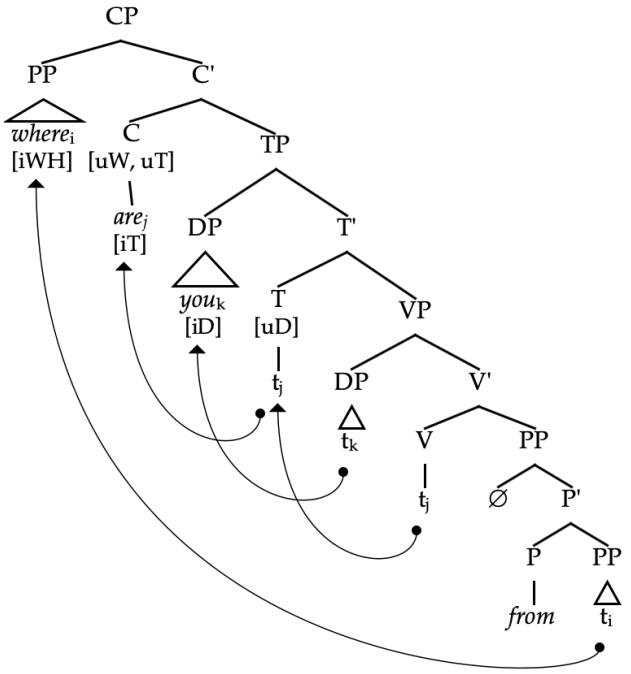

Figure 5.5: *Where are you from?* assumed syntactic structure

The assumed derivation of *where are you from?* mirrors that of *what is your name?* and *how old are you?* (minus the pied piping of subordinate material in the latter), thus involving the computational properties V-T movement, A-movement, T-C movement and wh-movement.

5.5.5 Shared computational properties of the FEs^{wh}

As the FEs^{wh} are all root interrogatives, their syntactic derivations are similar, involving the computational properties A-movement, V-T movement, T-C movement, wh-movement and do-support. These are driven by features on functional categories T and C. It can be said that their syntactic makeup is therefore complex, when it is considered that syntactic complexity is often measured in terms of respective Merge operations and in particular movement (Jakubowicz, 2011; Durrleman et al., 2016). These computational properties are not exclusive to the derivation of root interrogatives, rather, they have the potential to manifest via a large range of surface structures (to be outlined and elaborated upon in Section 5.5.6 which follows).

Therefore, when measuring learners' corresponding knowledge of the FEs^{wh,} associated syntactic properties and functional categories, it is necessary to capture all surface structure manifestations as corresponding evidence for this knowledge. The magnitude of such a task is well attested to (Lardiere, 2008), and requires the careful analysis of interlanguage production data by the same learner over a given period of time. As stated in a previous footnote, I do not measure learners' knowledge of V-T movement in this study, as surface structure evidence is extremely limited in English and would therefore prove too unreliable as an indication of knowledge of this property. The following sections therefore concentrate on the surface structure manifestations and measuring learners' corresponding knowledge of the computational properties A-movement, T-C movement, wh-movement and do-support.

5.5.6 Identifying the manifestation of related computational properties in learners' interlanguages

Wh-movement is a computational property assumed to be involved not only in root interrogatives but also exclamative clauses (15-a), relative clauses (15-b) and interrogative complement clauses (15-c).

(15) a. [CP*what rubbish*$_{\mathrm{i}}$[C[TP*he*$_{\mathrm{k}}$[T*would*[VP_t_{k}[V*talk*_t_{i}]]]]]

b. [CP[C[TP*it*$_{\mathrm{i}}$[T*is*$_{\mathrm{k}}$[VPt_{i}[Vt_{k} *something*[CP*which*$_{\mathrm{j}}$[C[TP*you*$_{\mathrm{y}}$[T*can*[VPt_{y}[V*do*_t_{j}]]]]]]]]]]]

c. [CP[C[TPI_{i}[T*wonder*$_{\mathrm{k}}$[VPt_{i}[Vt_{k}[CP*how much money*$_{\mathrm{j}}$[C[TP*they*$_{\mathrm{y}}$[T*have*$_{\mathrm{l}}$[VPt_{y}[Vt_{l}_t_{j}]]

The main difference between these alternative instances of wh-movement is that C is presumed to only carry an uninterpretable interrogative feature [uW], not an uninterpretable tense [uT] feature as in root interrogative structures. Hence, wh-movement occurs in the absence of T-C movement.

Similarly, do-support in English is required not just for question formation but also some structures of negation, as can be seen from a typical example in Figure 5.6. Here, I adopt the separate Negation Phrase (NEGP) projection analysis of English

negation, where the negative particle *not* originates in its specifier position[11], rather than as an adverb in the specifier position within the VP. This is based on the notion that only an argument of a verb can occupy the specifier position within the VP (i.e. its subject), and *not* in a negative sentence such as that in Figure 5.6 is not an argument selected by the verb. Note also that in some of the following examples I will exclude the CP projection from the tree structures for sake of notational simplification, although the assumption is that all grammatical finite and complete clauses are CPs headed by either an overt or null category on C (Radford, 2004).

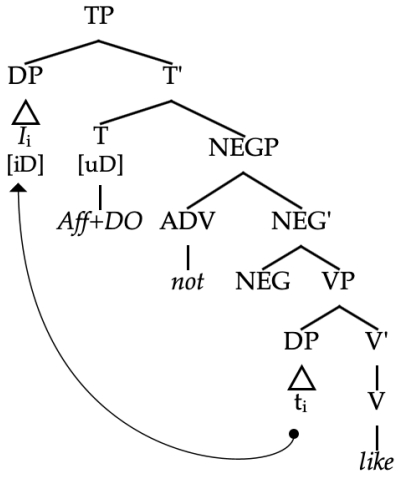

Figure 5.6: *I do not like*: assumed syntactic structure of negative clause containing do- support

One issue in SLA is the reliability of surface structure evidence in determining whether a certain computational property has actually manifested/ the associated functional category is being projected. For example, A-movement is assumed in the full derivation of simple finite clauses with copula *be* such as *I am happy*, by virtue of the finite form *am* participating in a TP (tense phrase) projection, as can be seen below in Figure 5.7.

[11]This is following Radford (2009), although it should be pointed out that many researchers posit an alternative analysis where *not* is taken to be the head NEG constituent of NEGP.

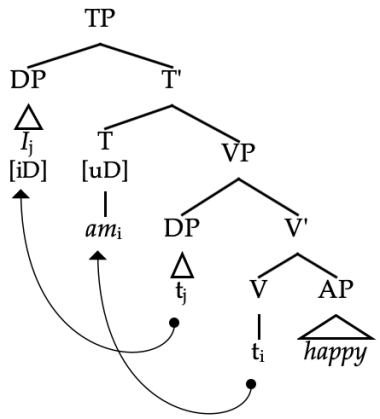

Figure 5.7: *I am happy*: assumed syntactic structure of a simple finite clause containing A-movement and V-T movement

However, when a learner produces an utterance such as this in isolation, it is ambiguous as to whether the subject pronoun has moved to a higher clausal projection. No inflectional morphology associated with T is required in this environment, as 'be' is a suppletive form. It could be the case that this learner has a VP- based grammar only in the L2, and is yet to acquire TP, meaning that the subject pronoun remains VP internal as in Figure 5.8.

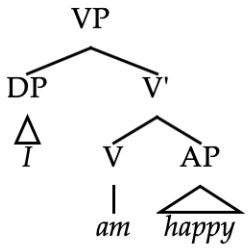

Figure 5.8: *I am happy*: VP projection

It could also be the case that the first-person conjugation *I am* has been memorised by way of being irregular and frequent in use, in which sense it would be non-finite. This is common for second language learners of English who are often encouraged to learn the conjugations of *have* and *be* in a 'formulaic-like' fashion, through consistent repetition of *I am, you are, he/she is, we are* etc (Samian and Tavakoli, 2012). In this case, there would be no VP or TP projection, and hence the utterance would simply be a fixed, lexicalised chunk. Indeed, it is common

amongst previous interlanguage studies investigating L2ers' knowledge of functional category T to count agreement/tense/person marking on lexical verbs only, and discount common, irregular and suppletive forms (Lardiere, 1998b,a; Vainikka and Young-Scholten, 1998).

To combat this ambiguity, it is necessary to identify more reliable surface features associated with A-movement as evidence that a learner has acquired knowledge of this computational property. Since A-movement is triggered by D feature checking on T requiring it to project a subject as its specifier, better evidence for this property would be those surface structures that more reliably indicate a tense phrase (TP) projection. These in English include the use of overt subjects with either; corresponding finite verbal inflection (TNS, NUM, AGR), modal verbs/auxiliaries, the dummy auxiliary *do*, negation and 'infinitival *to*'. In the latter structure, infinitival *to*, it is assumed that *to* is a (non-finite) infinitival tense particle occupying the head T position of the TP. Support for such an analysis is the similar function of *to* in this environment to that of modal auxiliaries such as *will* in the examples below, which demonstrate future time-reference.

(16) a. *They are expecting the game [to] be won tomorrow*

b. *They are expecting that the game [will] be won tomorrow*

Other evidence for A-movement would be in surface structures that imply a TP projection by means of showing overt evidence for a CP projection, as head C necessarily takes a TP as its complement. This entails that those surface structures showing overt subjects with wh-movement and T-C movement, for example, can also be presumed to involve A-movement.

The above demonstrates that a variety of surface structures can be used as more overt and therefore reliable evidence of a computational property's manifestation. It is these more reliable surface structures that I will consider when measuring learn-

ers' knowledge of the FEs^{wh}' associated syntactic properties, rather than assuming their manifestation in ambiguous surface structures that lack overt evidence. Table 5.2 below summarises the FEs^{wh}' computational properties and corresponding reliable surface structures that I take as evidence for their manifestation in learners' production data.

computational property	***wh*- movement**	**T-C movement**	***do*- support**	**A- movement**
surface structure evidence	- wh- words occupying a clause-initial position in root interrogatives - exclamative clauses - relative clauses - interrogative complement clauses/ embedded wh-clauses	- inversion of the subject and (auxiliary) verb	- via negation - via question formation	overt subjects used with structures that imply a TP projection, including: - corresponding finite verbal inflection [TNS, NUM, AGR] in declaratives (not including *is/are*) - modal/auxiliary verbs (including dummy *do*) - *wh*-movement - T-C movement - `infinitival to'

Table 5.2: FEs^{wh}' computational properties and reliable surface structures that are taken to evidence their manifestation

The next section now documents how I measured learners' knowledge of these properties, which takes into account all L1, translanguaging, accurate L2 and inaccurate L2 utterances in learners' trascripts.

5.5.7 Measuring learners' knowledge of the computational properties

When measuring learners' knowledge of the computational properties, an attempt has been made to be as conservative as possible, so to reduce the risk of unrepresentative and misleading results. This conservatism has already been reflected in the restriction of surface forms that are being accepted as overt evidence for the properties as presented in the previous section. For example, the choice to eliminate overt subjects used with uninflected verb forms and suppletive uses of *be* and *have* as instances of A-movement (see Section 5.5.6).

As learners of the BELC are taking part in L2 spoken tasks whereby the target language is English, it is important to note that *all* of these tasks are contexts

which require the L2. However, there are four possible ways for learners to realise an utterance in these contexts; in the L1, via translanguaging[12], *accurately* in the L2 or *inaccurately* in the L2. Essential is that all of these realisations are considered in order to achieve a more accurate understanding of a learner's L2 knowledge. I will use do-support as an example. Say that in a learner's transcript at age 16, there were 9 contexts that required do-support in English, and the learner realised these as below:

i	*I don't go to school*
ii	*he not like it*
iii	*no sé* [I don't know]
iv	*no sé* [I don't know]
v	*no want eat tonight*
vi	*want go there tonight?*
vii	*he doesn't gustar la comida* [he doesn't like the meal]
viii	*te gusta la musica?* [do you like music?]
ix	*do you speak English?*

As all of the instances above if uttered in English require do-support, the total required contexts for the computational property do-support for this learner at this age would be 9. Out of these 9 contexts where do-support should manifest in English, 3 are realised in the L1 (iii, iv and viii), 1 via translanguaging (vii) and 5 are attempted in the L2 (i, ii, v, vi and ix). Out of these 5 L2 attempts, only 2 of these utterances are accurate (i.e. *grammatical*) (i and ix). All 9 utterances constitute this learner's realisation of do-support at this particular age. This realisation is represented in Table 5.3, which distinguishes between the respective computational properties' possible surface structure manifestations as identified in Section 5.5. In the case of do-support, this would be split between negation and question formation.

[12]Note that I adopt the term 'translanguaging' rather than code-switching. This is because, for the sample EFL classroom learners (especially at ages 10 and 12), use of the L1 is likely a 'fallback' strategy used to communicate meaning in absence of L2 knowledge, rather than being a constrained alternation occurring at specific points in communicative episodes (Przymus, 2023). That said, many instances of translanguaging in learners' transcripts could be classed as instances of 'code-switching', if looked at purely objectively/out of context.

Example Learner (age 16)								
DO Support								
	Negation				Question Formation			
Total	L1	Translang.	L2 Att.	L2 Acc.	L1	Translang.	L2 Att.	L2 Acc.
9	2	1	3	1	1	0	2	1

Table 5.3: Token count: The realisation of do-support (example learner)

It is important to note that, when adding the contexts together in Table 5.3 above, the resulting total is 11, not 9, despite there only being 9 total contexts at this age. This is because of the crossover when distinguishing between all attempted L2 utterances and *accurate* (grammatical) L2 utterances, as the accurate L2 utterances are a number out of the attempted L2 ones. In the case of our example learner, this can be understood as follows. Out of 3 L2 attempts of negation where do-support is required, only 1 of these is grammatical, and out of 2 L2 attempts of question formation where do-support is required, only 1 of these is grammatical. The total number of contexts can therefore be understood as being divided between L1, translanguaging and L2 realisations, where the latter can be further subdivided into those that are accurate/grammatical[13]. The importance of such a distinction becomes clear when relative percentages are considered, as now discussed below.

After all contexts that require an L2 computational property have been organised into L1, translanguaging and L2 realisations, each of these is taken as a relative percentage out of the total contexts. These relative percentages reflect the computational properties as whole entities, and do not distinguish between their different surface structure manifestations. That is, relative percentages of the example learner's realisation of do-support in all required contexts include both negation and question formation structures, as shown in Table 5.4 below.

Example Learner (age 16)				
DO Support				
Total	L1	Translang.	L2 Att.	L2 Acc.
9	3 (33%)	1 (11%)	5 (56%)	2 (22%)

Table 5.4: Relative percentages: The realisation of do-support (example learner)

[13]Note that the corresponding 'ungrammatical' utterances are not given as a separate figure in the table, as this is percentage is inferred from the 'grammatical/accurate' ones.

Note, importantly, that the relative percentage of accurate/grammatical L2 tokens is a percentage out of the total number of contexts, rather than out of the learners' L2 attempts. This is again to be as conservative/cautious as possible when it comes to measuring a learner's knowledge of an L2 computational property. If, for instance, the example learner's L2 accuracy percentage rate was taken instead out of their L2 attempts only, it would read much higher- 2 out of 5 instances, hence 40%. However, to say this learner has a 40% L2 accuracy rate of do-support is somewhat misleading, when it is considered that in 44% of contexts (i.e. 4 out of 9 total contexts) where do-support is required in the L2, the learner reverts to realising the utterance via their L1/translanguaging. This, taken together with their inaccurate L2 realisations, constitutes 79% of contexts in which the learner fails to realise do-support accurately where required in the L2.

A more extreme example can better highlight how only considering learners' L2 productions could lead to an over-representation of their L2 syntactic knowledge. Say, for instance, that another learner has a total of 9 contexts where do-support is required in the L2, and in 8 of these, they use the L1. The other realisation is an accurate/grammatical L2 realisation. If grammatical L2 percentage rates were taken instead out of L2 attempts only, this learner would be perceived to have a 100% L2 accuracy rate of do-support. Hence it could be inferred that this learner has fully acquired this particular computational property. If this is the case, why then would the learner have to revert to their L1 when realising utterances in *all* other contexts which require do-support in the L2? As mentioned previously, the context in which the production data of these learners was collected was one which required (and indeed was *testing*) their L2 use. Therefore, all L1/translanguaging utterances are likely instances of *reliance*, that is, instances where the learners are unable to realise an utterance in the L2 and fall back on their L1 to express the same concept.

Clearly, whether or not L1/translanguaging utterances can solely be attributed to the learner's lack of L2 knowledge is hard to prove. They could also be the result of a lack of vocabulary knowledge, or other translanguaging/code-switching phenomena. However, when the context in which the production data was collected is considered, along with the conservative selection of various potential surface structures that are counted together as instances of a computational properties' manifestation, I believe the implication is strengthened. Furthermore, when a learner produces an L1 utterance in a context in which an L2 computational property is required, it is possible to check other interlanguage productions for use of the same lexical items. Continuing with do-support as an example, consider a learner who produces *no va al parque* [he doesn't go to the park] in the L1, but also produces the L2 lexical items *he*, *go* and *park* outside of contexts which require do-support at the same age. This would suggest that it is lack of L2 do-support knowledge that forces the learner to revert to their L1 in this context, rather than a deficit of L2 vocabulary knowledge.

Taking into account all learner productions when measuring their L2 knowledge is also adhering to a fundamental principle of corpus linguistics, whereby all data must be considered in an analysis rather than a favourable subset (McEnery and Hardie, 2012). However, learner corpus research, in particular, often comes under scrutiny with regards to its efficiency in representing learners' competence as opposed to their performance. Elicitation/experimental tasks have traditionally been used within generative SLA (see Chapter 2), as these are said to tap into a learner's competence more directly than corpora do. This is because the production of a specific form in a corpus may not necessarily reflect the learner's competence, and as such, the absence of a form does not entail a learners' lack of knowledge (Lozano, 2021). This concept holds for learners' naturalistic production data more generally, as when categories fail to show up in production data, one should be wary of concluding that they are altogether absent (Grondin and White, 1996).

It is indeed the case that transcripts of the BELC may not be a complete representation of each learner's individual L2 competence, as they are, after all, a 'sample' of their language use. In turn, the complete lack of an L2 computational property in a learner transcript of the BELC does not imply a learners' lack of knowledge of such. Imagine that in a learner's transcript, there are 5 'question formation structure' contexts which require wh-movement in English, and this learner realises all of these utterances in the L1. If it was that L2 accuracy rates were counted as a percentage out of L2 attempts *only*, and L1 utterances were not taken into consideration in the analysis, an immediate criticism to positing this learner's L2 wh-movement knowledge as zero could be that 'the learner simply didn't want/was never in a position to ask questions'. However, as all L1 productions are taken into consideration, we know that this learner has had the opportunity/ did want to ask questions to the interviewer, as we see this was the case in 5 separate contexts. Inclusion of L1 utterances in the analysis therefore not only helps give a more accurate representation of a learners' respective L2 syntactic competence, but helps to better overcome limitations associated with the analysis of naturalistic instead of experimental production data.

Another point to emphasise is that the computational properties under investigation are being investigated as separate, independent procedures. In an utterance which requires more than one computational property under investigation, it is therefore possible for learners to show an accurate use of a one of these computational properties but still produce an ungrammatical L2 utterance. This investigation also only concerns these computational properties as realised in the L2, not the L1. It follows then that it is possible for learners to realise a computational property accurately in accordance with their L1 feature specifications which results in an ungrammatical utterance in the L2 (English).

Take, for example, a learner who produces **where go you tomorrow.* This ut-

terance is ungrammatical in English as it lacks do-support, and instead shows inversion of a lexical verb with the subject pronoun. In terms of a computational property analysis, do-support here would be taken as an attempted but inaccurate/ungrammatical L2 count. This is because this computation is required in English in this context- and the learner has attempted the utterance in English- yet do-support is absent. Similarly, T-C movement would be taken as an attempted but inaccurate L2 count. Whilst Spanish lexical verbs can raise to T and thereafter undergo T-C movement, English lexical verbs (like *go*) do not permit V-raising, and are therefore unable to reach the head T. This means lexical verbs in English are not licensed for T-C movement- hence the ungrammatically of **where go you tomorrow?*. However, because the wh-word is fronted, this would be taken as an accurate L2 count of wh-movement. Similarly, by virtue of the wh-word being fronted, this implies a CP projection and therefore a TP complement, which thus implies A-movement of the subject DP *you* to the specifier of TP. This would be taken as an accurate L2 count of A-movement. Hence the ungrammatical utterance **where go you tomorrow* would give accurate counts for wh-movement and A-movement, but attempted and inaccurate ones for do-support and T-C movement under the analysis assumed in the present study.

Diverging from the generative analysis presented above, Section 5.6 now outlines how usage-based models would analyse the FEs^{wh} as abstract schematic constructions, and presents how the deconstruction of such could lead learners to an acquisition of similar functional creativity in the L2.

5.6 The FEs^{wh} as abstract schematic constructions

Rather than a syntactic derivation based on functional categories, features and computational properties, the level of ultimate abstractness for UBL consists of schematic knowledge of symbolic units (Lieven et al., 2003; Eskildsen, 2009, 2020),

that is, the storage of lexical items as a range of fully schematic constructions. Following Eskildsen (2015), the FEs^{wh} would represent the fully schematic constructions below.

(17) a. *what's/is your name?* [WH + COPULA + PossDET + NOUN]
b. *how old are you?* [WH + ADJ + COPULA + PRN]
c. *where do you live?* [WH + AUX DO + PRN + VERB]
d. *where are you from?* [WH + COPULA + PRN + PREP]

UBL models posit an acquisition of fully schematic constructions and/or utterance schemas through the analysis and subsequent generalisation of prototypical, fixed expressions that exemplify these constructions. Due to their salience, prototypicality and formulaicity for all learners under analysis, the FEs^{wh} are good candidates for acquisitional seeds in this proposed developmental sequence. Adopting this learning strategy, for example, learners could gradually move from the FE^{wh} [*what is your name*] to a lexically-specific utterance schema (a fixed part and open slot) [what is + PossDET + NOUN], to the fully schematic construction [WH + COPULA + PossDET + NOUN].

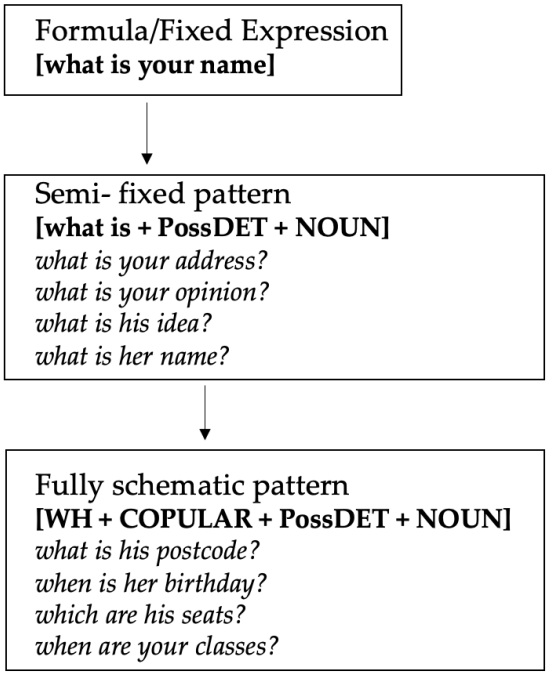

Figure 5.9: A usage based developmental trajectory of the schematic construction WH + COPULA + PossSUBJ + NOUN derived from the formulaic exemplar *what's your name*

Equally, as past studies on L2 interrogative development have suggested, learners can derive more general 'wh-question' utterance schemas from the FEs^{wh}, rather than acquiring their fully schematic patterns (see Chapter 2). Utterance schemas based on fixed wh-questions in English traditionally comprise the [WH + VERB] element, based on evidence that a child's earliest wh-questions produced with an auxiliary and/or copula can be explained with reference to formulaic patterns that begin with a limited range of these schemas (Fletcher, 1985; Rowland and Pine, 2003; Eskildsen, 2015). Fletcher's (1985) longitudinal study of Sophie, for example, demonstrated that at the age of three she was only shown to produce wh-questions with successful subject-auxiliary inversion with the combinations *what are*, *how do* and *why do* (pp. 105-107). Utterance schemas based on the [WH + VERB] are also deemed to be more likely because the range of possible wh-words and auxiliaries is more limited than the range of potential subjects and verb phrases (Rowland and Pine, 2000).

Based on the FEs^{wh}, this would give for the following utterance schemas, which have the potential to be lexically (18) and/or categorically (19) specific[14].

(18) a. *what's/is your name?* [what is/'s] + X
b. *how old are you?* [how old are] + X
c. *where do you live?* [where do] + X
d. *where are you from?* [where are] + X

(19) a. *what's/is your name?* [WH + COPULA] + X
b. *how old are you?* [WH + ADJ + COPULA] + X
c. *where do you live?* [WH + AUX DO] + X
d. *where are you from?* [WH + COPULA] + X

However, as any utterance schema is potentially extractable from formulaic exemplars, learners could also extract the FEs^{wh}' [VERB + SUBJ] utterance schemas and omit the wh-element to derive yes/no questions. These lexically (20) and categorically specific (21) yes/no question utterance schemas are given below.

(20) a. *what's/is your name?* [is your] + X
b. *how old are you?* [are you] + X
c. *where are you from?* [are you] + X
d. *where do you live?* [do you] + X

(21) a. *what's/is your name?* [COPULA + PossDP] + X
b. *how old are you?* [COPULA + PRN] + X
c. *where are you from?* [COPULA + PRN] + X
d. *where do you live?* [AUX DO + PRN] + X

Learners' knowledge of the FEs^{wh}' abstract schematic properties will be measured by how many L2 interrogatives (yes/no and wh-) they produce that embody any

[14]Note that I adopt the term 'lexically' specific following Eskildsen (2015) to refer to those schemas that are more concrete and hence retain the same particular words (i.e. *what*, *do* etc.) of the FEs^{wh}. 'Categorically specific' utterance schemas instead display the same abstract word categories (i.e. wh-word, auxiliary verb etc.) as the FEs^{wh}, but with different words.

variation of the FEs^{wh}' utterance schemas (lexically and/or categorically specific) and fully schematic patterns across the data collection period. To examine whether learners' L2 questions share an utterance schema/fully schematic pattern of a previously used FE^{wh} in their production data, a traceback methodology is adopted. I created individual learner tables documenting their FE^{wh} productions and all L2 questions across the four rounds of data collection. Underneath each FE^{wh} and L2 question, I specify their lexically (i) and categorically (ii) specific utterance schemas, as well as their fully schematic patterns (iii). I then underlined instances where those of an L2 question matched those of a previously used FE^{wh}. Learner 13's L2 wh-questions can be seen in Table 5.5 as an example.

#	age 10: wh-Q	age 12: wh-Q	age 16: FE	age 16: wh- Q	age 17: FE	age 17: wh-Q
13	-	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]	*what's her name* i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN]	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN]	*where do you go the last weekend* i. [where do] + X ii. [WH+AUX DO]+ X iii. [WH+AUX DO + PRN + VERB] **what language talk you in the house* i. [what language talk] + X ii. [WH+N+V] + X iii. [WH+NOUN+VERB+PRN+Prep+Det+NOUN]

Table 5.5: Learner 13

Table 5.5 shows that one L2 wh-question in Learner 13's transcripts shares the same wh-question utterance schema and fully schematic pattern of a previously produced FE^{wh}. This is *where do you go the last weekend* produced at age 17 after using *where do you live* one year previously at age 16. These question forms share the same lexically specific utterance schema ([where do] + X) and fully schematic pattern ([WH + AUX DO + PRN + VERB]). For all learners, the same procedure is also adopted for yes/no questions.

Before presenting the results of the BELC analysis, I now discuss the use of parametric statistical tests in the present study. I first document what tests are appropriate for investigation of the research objectives, before discussing how nor-

mality and significance levels will be operationalised for the small SLA dataset under analysis.

5.7 The use of statistics in the present study

5.7.1 Statistical tests adopted

Parametric statistical tests will be applied to aid the investigation of Research Objectives 1 and 2, which are repeated below.

a. Are formulaic expressions evident in learners' production data at the initial state, and do these influence the development of L2 syntax thereafter?

b. Can this development be captured by schematic learning strategies, or do we find the development of underlying syntactic knowledge more generally? That is, can novel L2 utterances be traced back to formulaic exemplars based on related utterance schemas, or related computational properties?

In order to capture any relationship between learners' use of FEs and corresponding knowledge of related underlying syntactic properties under a generative framework, relative L2 accuracy rates of these properties as raw numbers will be compared to the frequency of FEs produced by learners, and the age in which they do this. This is because, as reviewed in Chapter 2, FEs that are used early and frequently in the learning process are those that are most likely to have an impact on later syntactic development (Ellis, 2002b, 2012; Ellis et al., 2015). Relative L2 accuracy rates will be classified as continuous dependent variables, and number of FEs and age of FE production will be continuous independent variables. Correlation analyses are used to investigate the relationship between two continuous variables (Pallant, 2010). Therefore, I will adopt this procedure to compare both learners' frequency and age of FE production with knowledge of related L2 syntactic properties sepa-

rately, using SPSS Version 28.0 software[15]. In order to best interpret any results from the analysis, correlations will be presented both visually in the form of scatterplots and numerically in the form of Pearson product-moment coefficients (r). The measurement of effect sizes for r is traditionally based on the criteria set forth by Cohen (1988), who proposes $r = .10$ to .29 as a small effect, $r = .30$ to .49 as a medium effect and $r = .50$ - 1.0 as a large effect. Following Larson-Hall (2016), I will adopt the more specific guidelines for effect sizes as proposed for second language acquisition by Plonsky and Oswald (2014). Here, $r = .2$ is a small effect, $r = .4$ is a medium effect and $r = .6$ is a large effect. If differences can be observed between groups of learners in terms of frequency and age of FE usage and L2 accuracy rates at the end of the data collection period, an independent samples T-test will be used to verify if this difference is significant or not. The correlation and t-test analyses will allow for a general insight on the relationship between FE use and later L2 syntactic knowledge, which will allow for further qualitative exploration of the data in light of these results.

For the usage-based analysis, the investigation of FE use and related utterance schemas is necessarily more qualitative, as the traceback methodology will involve a manual analysis of the data comparing learners' novel L2 question forms with any FEs produced previously in their data. The number of question forms that can be traced back to FEs will be presented and discussed as relative percentages out of the total number of question forms observed across the data, but no parametric statistical tests will be applied here. Section 5.7.2 now discusses the normality of the BELC data to further justify the use of the parametrical statistical tests outlined above.

[15]IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp

5.7.2 Small datasets and 'normality' of the data

With small-data sets, it is questionable as to whether the data is 'normally' distributed or not, which has consequences for the use of parametric statistical tests. It is commonplace for SLA research designs to use small sample sizes, meaning that the statistical power of a test of a normal distribution may be low. It has even been claimed that real data sets in Applied Linguistics are likely hardly ever normally distributed (Larson-Hall and Herrington, 2009), and it is impossible to tell whether small samples represent an exactly normal distribution or not (Larson-Hall, 2016). Nevertheless, studies in SLA continue to use parametric statistics which assume a normal distribution, without carrying out preliminary tests of normality (Larson-Hall and Plonsky, 2015). Or, they simply report results of their experiments without saying whether the data have satisfied the assumptions underlying parametric statistical tests (Larson-Hall and Plonsky, 2015).

There are ways to assess the normality of a given dataset, both graphically and numerically. There is an overall consensus for the use of formal, numerical tests of normality for small datasets, as 'at the level of 10 or 15 samples per group, it is extremely unlikely that even samples generated from normal distribution will follow the smooth line of a normal distribution' (Larson-Hall, 2016, p. 105). Another advantage with using numerical tests is that they are hypotheses tested, and therefore often perceived to be more precise than graphical representations such as histograms and boxplots. There is, however, a disagreement as to which particular numerical test works best for small data samples. Larson-Hall (2016) recommends that for sample sizes under 30, the best test of normality is seeing whether the data are skewed or not, as a skewness level over 1 indicates a significant departure from normality (Porte, 2002; Pallant, 2010). Other researchers recommend the Shapiro-Wilk statistical test for smaller sample sizes (Field et al., 2012; Ricci, 2005), although Turner (2014) claims that this isn't very rigorous when used with samples of 30 participants or less. Dörnyei (2007) instead recommends that the Kolmogorov-Smirnov

statistics works best when testing for normality of small samples, yet this has also been claimed to not be sensitive enough (Staudte and Sheather, 1990; Wilcox, 2003).

Whilst there is a disagreement between what numerical tests of normality work best for small data sets, imperative (and indeed consensual in the field of SLA) is that normality *is* tested for in the first instance. This is because one risks obtaining statistically inaccurate results when applying parametric statistics to non-normal data, as the model distribution is not reflected in the actual data at all (Larson-Hall and Mizumoto, 2020). For this reason, I carried out all of the above numerical tests for normality with the learners' L2 computational accuracy percentage rates at later ages (16&17), to justify the use of parametric tests to investigate their relationship with early/frequent FE^{wh} use. I carried these out with the distribution of learners' L2 accuracy of each computational property individually, and then with these combined as a mean computational accuracy percentage.

	Kolmogorov-Smirnova			**Shapiro-Wilk**			**Skewness**	
Operation	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	St. Error
WH	0.19	9	.200*	0.891	9	0.203	0.28	0.717
T-C	0.174	9	.200*	0.953	9	0.728	0.502	0.717
DO	0.144	9	.200*	0.938	9	0.56	0.322	0.717
A	0.235	9	0.164	0.911	9	0.322	-0.212	0.717
Mean	0.258	9	0.85	0.872	9	0.192	0.291	0.717

* This is a lower bound of the true significance.

Figure 5.10: BELC production data: Tests of normality

Normality is indeed indicated with skewness levels under 1 and Kolmogorov-Smirnov/ Sapiro-Wilk p (significant) values equal to or greater than 0.05. Note, however, that passing these tests of normality does not entail that any results obtained from parametric analysis can be immediately generalised to the wider L2 learner population; the sample size is far too small and restricted to make such a claim. It does however justify the use of parametric statistical tests on the dataset at hand, and indicates that any relationships found between early and frequent FE^{wh} use and corresponding computational development are likely true for the group of

learners under analysis.

5.7.3 Small datasets and significance levels

It is important here to also address the issue of statistical significance when analysing small data samples. Alpha (p) values are traditionally set at <.05 (2-tailed), but it is likely that measuring the relationship between two variables in the current data set (n= 9) will deliver insignificant results in this capacity, by consequence of its small sample size. Indeed, Larson-Hall (2016) states that 'the p-value is not necessary to understand anything about a correlation; if the sample size is large enough, any correlation will be 'statistically significant' and if the sample size is quite small, almost any correlation will be 'statistically insignificant' (p. 211). The author continues by saying that in SLA, researchers should place importance on the effect size of the r value, without worrying about whether the p-value is lower than 0.05 or not.

To combat the low power of small sample studies with participants of 20 participants or less, some researchers propose adjusting the alpha level (Pallant, 2010). Stevens (1996) for example, suggests that when small group sizes are involved it is necessary to adjust the alpha level to .10 or .15 to compensate. With this in mind, I will consider any results derived from statistical tests with alpha levels of <.10 and <.15 as significant. Furthermore, recent developments in the application of statistics in SLA specifically argue for a correction of alpha levels based on sample size, and even a move away from traditional p values altogether. Specifically for learner corpus research, Paquot and Plonsky (2017) recommend the following when using statistics:

'...conduct fewer tests of statistical significance and correct for the alpha level... be skeptical of p values... calculate, report and interpret descriptive statistics, including effect sizes and confidence intervals' (p. 87.)

The use of confidence intervals (CIs) is indeed one of the main changes in the

'new statistics' era (Larson-Hall, 2016; Larson-Hall and Mizumoto, 2020), along with the use of bootstrapping, which are being used to replace assumptions of significance traditionally derived by p values. CIs indicate how much confidence we can have in an effect by giving a range of plausible values where we could expect to find the true relationship between two variables, with 95% confidence. It is said to be an improvement on p values, as the CI can tell us more about the nature of the size effect (i.e. the range in which the true effect is located), rather than just whether there was an effect or not. If the CI goes through 0, this is said to be similar to the p-value being above 0.05, and the conclusion is that there was not an effect for a relationship between the variables tested. If the CI instead does not go through 0, this is similar to the p-value being below 0.05, and hence we can assume there is an effect for a relationship among variables (Larson-Hall, 2016). CIs can be used in tandem with bootstrapping, a procedure which generates an empirical distribution of data from random re-samples of the original data sample. CIs can therefore be calculated from a new 'bootstrapped' empirical distribution data.

With all this in mind, I will consider any results derived from correlation and t-test analyses with alpha levels of $<.10$ and $<.15$ as significant, and run these in tandem with bootstrapped confidence intervals. Chapter 6 now presents the results of the BELC analysis.

Chapter 6

Results: The Barcelona English Language Corpus

This Chapter reports the results of the BELC analysis. It first analyses learners' use of the FEs^{wh} and their corresponding L2 knowledge of associated syntactic properties outside of these expressions across the data collection period (generative). It then adopts a traceback analysis, comparing the abstract schematic patterns of learners' L2 interrogatives (both wh- and yes/no) to those of preceding FEs^{wh} in their production data (usage-based).

6.1 Learners' use of FEs^{wh} and corresponding knowledge of associated computational properties

6.1.1 Learners' initial use of the FEs^{wh} and corresponding knowledge of associated computational properties

Chapter 2 stated that formulaic language is identifiable in L2 production data predominantly by a level of fluency and syntactic development that is not found in the interlanguage of the learner (Bardovi-Harlig, 2009). Recognised, then, is the need to compare candidate formulas (our FEs^{wh}) to the generative competence of individual

learners at a given point of interlanguage development (Myles, 2004). Typically, in beginner learner production data, there is a discrepancy between complex chunks uttered fluently and simple utterances generated online by the same learner (Weinert, 1995; Myles et al., 1998, 1999; Myles, 2004; Myles and Cordier, 2017). Now that the FEs^{wh} have been identified as candidate formulas in the BELC, it is necessary to compare learners' *initial* productions of these expressions with other interlanguage realisations at the same age, to see if there is a discrepancy between the two. That is, the comparison of individual learners' FE^{wh} productions in bold from Table 6.1 below with their other interlanguage realisations at that same stage of data collection.

learner	age 10	age 12	age 16	age 17
2	-	-	**Hm <what are> [\ \] what [\ \]** **what is your name** **how old are you**	*what's your name*
5	**what's your name**	-	*what's your name*	*what's your name*
7	**what's your name**	*what's your name* *how old are you* *where do you live*	-	*what's your name first of all* *how old are you (x2)* *well and where do you live*
13	-	-	**what's your name** **where do you live** **what's her name**	*what's your name* *and where do you live* *how old are you*
18	-	**what's your name** **how old are you** ***what do you live**	-	-
27	*NT*	-	**how old are you** **where do you live**	*how old are you* *where do you live* *what is your name*
38	*NT*	**what's your name** ***where you live**	*what's your name*	*where do you live now*
42	*NT*	-	**what is your name** **where are you from**	*what is your name* **where is you from*
47	**how old are you** **what's your name**	*NT*	*how old are you* *where do you live*	*how old are you (x2)* *what's your name* *where are you from*

Table 6.1: Learners' first productions of the extracted FEs^{wh} across the BELC (bold)

Adopting this analysis will determine whether the FEs^{wh} in the first instance are likely products of holistic retrieval or online generation. Sections 6.1.1.1 - 6.1.1.4

below demonstrate this adopted procedure with four out of the nine learners as representative examples; Learners 5, 7, 18 and 42. Following this, Section 6.1.2 then compares learners' knowledge of the FEs^{wh}' computational properties across the four rounds of data collection. In doing so, I can observe the developmental trajectory of individual learners' interlanguages alongside their FE^{wh} use.

6.1.1.1 Learner 5

Learner 5 produces *what's your name* for the first time fluently at age 10. Table 6.2 below shows the percentage rates of this learner's realisations of the respective L2 computational properties in all required contexts outside of the FE^{wh} at this age.

Learner 5 Age 10/10;9					
FwhE: *what's your name*					
all contexts which require L2 computational properties					
Movement	**total**	L1	codeswitch	english att.	(english acc.)
WH	**2**	2 (100%)	0 (0%)	0 (0%)	0 (0%)
T-C	**2**	2 (100%)	0 (0%)	0 (0%)	0 (0%)
A	**7**	6 (86%)	0 (0%)	1 (14%)	0 (0%)

Table 6.2: Learner 5 age 10: Realisation of related computational properties in all required contexts

It can be seen by the English accuracy rates (0%) that this learner fails to realise any of the FEs^{wh}' related properties accurately in English at this stage of data collection. Wh-movement is required in a root interrogative clause (1-a) and an interrogative complement clause (1-b), which are both realised in the L1.

(1) a. *donde vivo?* [SPANISH]
where do I live?

b. *hm (.) no se que dir-te* [CATALAN]
I don't know what to tell you

However, their production of wh words/phrases in isolation varies between the L1 and L2. Examples can be seen below.

(2) a. *what sisters?*

b. *quantas habitacions?* [CATALAN]

how many rooms?

This suggests that it is wh-movement as a computation that is lacking in the learner's interlanguage at this stage, rather than English knowledge of wh-phrases altogether. T-C movement is also realised in the L1 where required, both in the root interrogative above (1-a) and the one below in (3)

(3) *mi ho pots repetir?* [CATALAN]

can you repeat it to me?

A-movement is equally as unsuccessful in English. In 86% of required contexts this is realised in the L1, examples of which include the utterances in (1-a) and (1-b) above. The other 14% of contexts constitute the one instance where the learner attempts a verb with a third person subject in English. However, this is realised inaccurately, as seen below.

(4) **it's dog eat* (when describing a picture of a dog eating)

The only accurate English productions with overt subjects and verbs outside of the FE^{wh} are utterances with the verb *be* (5) and uninflected verb forms with overt first person subjects (6-a) (6-b). These are assumed to be VP projections only, as they show no overt evidence for a TP projection and corresponding A-movement.

(5) *it's a mum and girl*

(6) a. *I study*

b. *girl and boy see*

There is also evidence for the verb *be* being used with overt subjects holistically, through the overextension of the contraction *it's* in ungrammatical environments

such as (7) and (8) below.

(7) **the mum it's*

(8) **is it's is*

All other English productions by Learner 5 at this age are limited to subjectless (first person) verb + noun collocations (9), single lexical items (10) and coordinated lexical items (11).

(9) a. _ *see television*

b. _ *play music*

(10) a. *yes*

b. *home*

c. *big*

d. *dog*

(11) a. *mum and girl*

b. *girl and boy*

c. *map and dog*

It is hard to construe that such simple utterances are products of the same computational system that is responsible for the generation of the fluent production of *what is your name?*. Hence, it is likely that the FE^{wh} is a product of holistic retrieval at this stage of data collection.

6.1.1.2 Learner 27

Learner 27 produces the FEs^{wh} *how old are you* and *where do you live* for the first time at the age of 16. Table 6.3 shows their realisation of these expressions' related computational properties at this age across all other interlanguage productions.

Learner 27 Age 16/16;9					
FwhE: *how old are you, where do you live*					
all contexts which require L2 computational properties					
Movement	**total**	L1	codeswitch	english att.	(english acc.)
WH	**11**	11 (100%)	0 (0%)	0 (0%)	0 (0%)
T-C	**4**	4 (100%)	0 (0%)	0 (0%)	0 (0%)
A	**43**	41 (95%)	0 (0%)	2 (5%)	1 (2%)
DO-support	**26**	26 (100%)	0 (0%)	0 (0%)	0 (0%)

Table 6.3: Learner 27 age 16: Realisation of related computational properties in all required contexts

Table 6.3 shows that the expressions' related syntactic properties in required contexts are realised overwhelmingly in the L1 (all are 100% in the L1 apart from A-movement at 95%). Wh-movement is required in root interrogative clauses (12-a) including those that require pied piping of subordinate material (12-b) and interrogative complement clauses (12-c). These are realised in the L1, as the examples below show.

(12) a. *que hice la semana pasada?* [SPANISH]
what did I do last week?

b. *a que hora vengo aqui?* [SPANISH]
what time did I come here?

c. *ahora no me acuerdo como se dice* [SPANISH]
now I don't remember how to say

T-C movement and do-support are realised similarly, examples of which include those above in (12-a) and (12-b). As well as root interrogatives, where do-support is required in utterances with negation, these are also realised in the L1. This can be seen from the clause in (12-c) and other interrogative complement clauses such as that below in (13).

(13) *no sé explicarlo* [SPANISH]
I don't know how to explain it

In contexts which require A-movement, this learner only attempts 5% of these in English. Only 1 utterance is grammatical, as seen in the progressive aspect structure below.

(14) *we are eating*

The two ungrammatical English attempts include the absence of overt subjects, auxiliaries and verbal inflection in the progressive and perfective structures below. These appear during the narration task requiring the learner to describe a picture showing the actions of a dog.

(15) a. *_ _ *look*_ *(.) look*_ *in the bag* [he is looking in the bag]

b. **the dog (.) _ eaten* [the dog has eaten]

The only grammatical L2 utterances are overt subjects used with those verbs that require no inflection (16), although even in these contexts the learner's use of overt subjects is somewhat inconsistent (17).

(16) a. *I study*

b. *two children (.) go out*

(17) a. *_ *go to the house* [I go to the house]

b. *_ *go to the cinema* [I go to the cinema]

c. *_ *go out with my friends* [I go out with my friends]

Such instances above are likely lexical projections, when it is considered that the majority of this learner's L2 productions are single lexical items only, such as *five, the map, mother, dog* etc. Therefore, based on all other available interlanguage productions at this stage of data collection, it seems that the FEs^{wh} have been committed to memory and produced holistically rather than generated online.

6.1.1.3 Learner 18

Learner 18 produces *what's your name?*, *how old are you* and **what do you live?* for the first time at age 12. Table 6.4 below shows the realisation of these expressions' related computational properties in all other interlanguage utterances at this particular age.

Learner 18 Age 12/12;9					
FwhE: *what's your name, how old are you, *what do you live*					
all contexts which require L2 computational properties					
Movement	**total**	L1	translanguage	L2 attempt	L2 accurate
WH	**1**	0 (0%)	0 (0%)	1 (100%)	0 (0%)
T-C	**3**	0 (0%)	0 (0%)	3 (100%)	3 (100%)
A	**8**	0 (0%)	0 (0%)	8 (100%)	3 (38%)
DO-support	**4**	1 (100%)	0 (0%)	3 (75%)	3 (75%)
V-T	**0**	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 6.4: Learner 18 age 12: Realisation of related computational properties in all required contexts

In the one context which requires wh-movement in the L2, the learner attempts this in the L2. However, this utterance is ungrammatical as the learner fails to 'pied-pipe' the subordinate material in the quantifier phrase (QP) *what sports*, as seen below.

(18) **what do you like sports?* [what sports do you like?]

This is perhaps evidence to suggest that the fluent production of *how old are you* is a result of holistic retrieval and production, given that, if generated online, *how old are you* requires movement of the whole AP *how old* to spec CP. Conversely, Table 6.4 shows that the realisation of T-C movement, A-movement and do-support seems fairly successful, with 100%, 38% and 75% L2 accuracy rates respectively. However, all of these accurate instances are counts of these derivations manifesting in the same phrase three times- *do you*- in the utterances below.

(19) a. **what do you like sports?*

b. *do you.. do you have any pets?*

c. *do you like pets?*

These utterances constitute the only contexts in which T-C movement is required at this age. Therefore, one interpretation is that this learner has fully acquired this computational property in the L2 at this stage of proficiency. However, this learner's manifestation of A-movement is less successful; the three utterances in (19) constitute the 38% of accurate L2 utterances where A-movement is required. In all other A-movement contexts, this learner's L2 productions are ungrammatical. This is via double subjects (20-a), non-corresponding AGR inflection/missing auxiliary verbs (20-b), missing subjects (20-c) and the preposition *for* (20-d), in structures which require and are clearly lacking a TP projection.

(20) a. **the sister and brother I go to the mountain* [the sister and brother go to the mountain]

b. **the dog are hungry and eating the sandwich* [the dog is hungry and is eating the sandwich]

c. **_ go to the home* [I go home]

d. **I go to the park for to play football* [I go to the park to play football]

All accurate utterances involving do-support are also those in (19). Where do-support is required outside of these productions, the learner instead reverts to the L1, as can be seen below.

(21) *no sé* [SPANISH]
I don't know

All other accurate L2 productions outside of the utterances in (19) are those with overt first/second person subjects with uninflected verbs, as in the examples below.

(22) a. *I drink coca-cola*

b. *I go to my house*

Given the lack of accurate evidence for T-C movement, A-movement and do-support outside of *do you*, another plausible interpretation is that *do you* is an unanalysed chunk and hence formulaic, similar to this learner's productions of the FEs^{wh} under investigation.

6.1.1.4 Learner 42

Learner 42 produces the FEs^{wh} *what is your name?* and *where are you from?* for the first time at the age of 16. Table 6.5 shows their realisations of the expressions' related computational properties in all other required contexts at this age.

Learner 42 Age 16/16;9					
FwhE: *what is your name, where are you from*					
all contexts which require L2 computational properties					
Movement	**total**	L1	codeswitch	english att.	(english acc.)
WH	**2**	2 (100%)	0 (0%)	0 (0%)	0 (0%)
T-C	**4**	2 (50%)	0 (0%)	2 (50%)	1 (25%)
A	**28**	18 (64%)	2 (7%)	8 (29%)	1 (4%)

Table 6.5: Learner 42 age 16: Realisation of related computational properties in all required contexts

In the two contexts which require wh-movement, the learner realises these in the L1, as can be seen below (note that INV stands for 'interviewer' and PAR stands for 'participant' i.e. the learner).

(23) INV: *what did you do last weekend?*
PAR: *el que vaig fer la setmana passada?* [CATALAN]
what did I do last week? [TRANSLATION]
INV: *weekend*

(24) INV: *but what will you do?*
PAR: *xxx.* INV: *no (laugh) que faras?*
what will you do? [TRANSLATION]
PAR: *que fare?* [CATALAN]
what will I do? [TRANSLATION]

The wider discourse context around these L1 utterances shows that the learner is echoing the questions of the interviewer, which are initially given to them in English and also involve T-C movement by way of being root interrogatives. This is perhaps an indication that this learner's lack of L2 syntactic knowledge not only limits their production of root interrogatives in the L2, but also their comprehension of such. Lack of T-C movement knowledge in the L2 is also implied through the learner's utterance in (25) below.

(25) INV: *what time did you arrive this morning?*
PAR: *repeat please*
INV: *what time did you arrive to the school this morning?*

The exchange in (25) further indicates this learner's difficulty in the comprehension of L2 root interrogatives, and also emphasises their lack of L2 T-C movement knowledge. They are seemingly forced to rely on lexical items only (V *repeat* + ADV *please*) rather than combining these in a corresponding yes/no interrogative structure involving T-C movement *can/could you repeat that please?*. This could also highlight a deficit in this learner's vocabulary knowledge. It is clear, for example, from the exchange in (23) that they misunderstood the lexical item *weekend* for *week*, and similarly, it could be their lack of knowledge of modal/auxiliary verbs that result in their inability to understand/form corresponding questions in the L2 in (24) and (25). This angle is given further support when their productions of wh-words in isolation or as part of noun/quantifier phrases are taken into consideration. All of these instances are L1 realisations, some examples of which are given below.

(26) a. *que* (x3) [CATALAN]
what?
b. *como?* [CATALAN]
what/how?
c. *que mes?* [CATALAN]

what else?

Clearly, this learner's inability to realise questions in the L2 is not solely a consequence of restricted L2 syntactic knowledge, but also seems to be down to their limited L2 vocabulary at this stage of proficiency. The only other context which requires T-C movement is actually an accurate L2 realisation which can be seen below.

(27) **do you have_?*

However, any implications that this demonstrates T-C movement knowledge in the L2 should be treat with caution. The learner produces this utterance in the final stages of the interview task when asked to pose questions to the interviewer, and fails to realise an overt complement to the V *have*, which actually renders the utterance in its totality ungrammatical in English. It is more likely that this production is one of holistic retrieval, where the learner has committed the 'question-frame' *do you have* to memory and is able to produce it upon functional/contextual cues (i.e. the interviewer requiring them to ask a question). Further support for this utterance being lexical in nature is its fluent production in comparison with all other realisations of do-support in required contexts (manifested across both question formation and negation structures), which are shown in Table 6.6 below (note the 1 accurate instance is *do you have*).

Learner 42 Age 16/16;9						
all contexts which require L2 computational properties						
Movement	**total**	L1		codeswitch	english att.	(english acc.)
DO-support	**9**	8 (89%)		0 (0%)	1 (11%)	1 (11%)

Table 6.6: Learner 42 age 16: Realisation of do-support in all required contexts

Do you have also constitutes the only accurate L2 realisation of A-movement (see Figure 6.5). In other required contexts, this computation is realised in the L1 or ungrammatically in the L2, as per the examples below.

(28) a. *_ *eating and* _ *see the television* (I am eating and watching the television)

b. **childrens is a mountain* (the children are looking at a mountain)

c. **the childrens (.) bye* (the children are saying goodbye)

Therefore, when examining Learner 42's interlanguage productions outside of the FEs^{wh} and *do you have*, it can be inferred with some confidence that these are products of holistic retrieval and production. The discrepancy between their fluent production and other interlanguage utterances suggests a deficit in not only syntactic but lexical L2 knowledge at this stage of proficiency.

6.1.2 Learners' L2 computational knowledge across the data collection period

Section 6.1.1 above used four learners of the BELC as representative examples to suggest that the FEs^{wh} are produced initially in advance of respective L2 syntactic competence. The interlanguage analysis at these ages reveals very little evidence for knowledge of the FEs^{wh}' associated computational properties *outside* of the expressions. Instead, the overwhelming majority of their other L2 utterances are ungrammatical and/or of a much lower syntactic complexity, and learners still rely heavily on the L1. This is further support of the FEs^{wh}' psycholingusitic salience and formulaicity for all learners under analysis.

Now that this concept has been established, I move to compare learners' knowledge of associated computational properties across the four rounds of data collection outside of the FEs^{wh}. The graph in Figure 6.1 below shows the mean L2 accuracy rates of associated computational properties at each age of data collection for all 9 learners under analysis. I have also included L2 accuracy rates of what are ambiguous VP/TP projections, that is, overt subjects used with uninflected lexical

verbs and *be/have*, such as *I like football, I am ten years old* (henceforth OS in VP). This is to visualise how L2 accuracy of these clauses differs from accuracy of clauses that are taken to be more solid evidence of TP projections and hence involve A-movement, such as *I went to school, he can play very well*. Note also that I have chosen to exclude Learner 18's repeated use of *do you* at age 12 as reliable evidence for knowledge of associated computational properties at that particular age. This is due to its formulaic nature and overextension in use as well as its overall discrepancy in syntactic complexity compared to other interlanguage productions at this age, as discussed in the previous section.

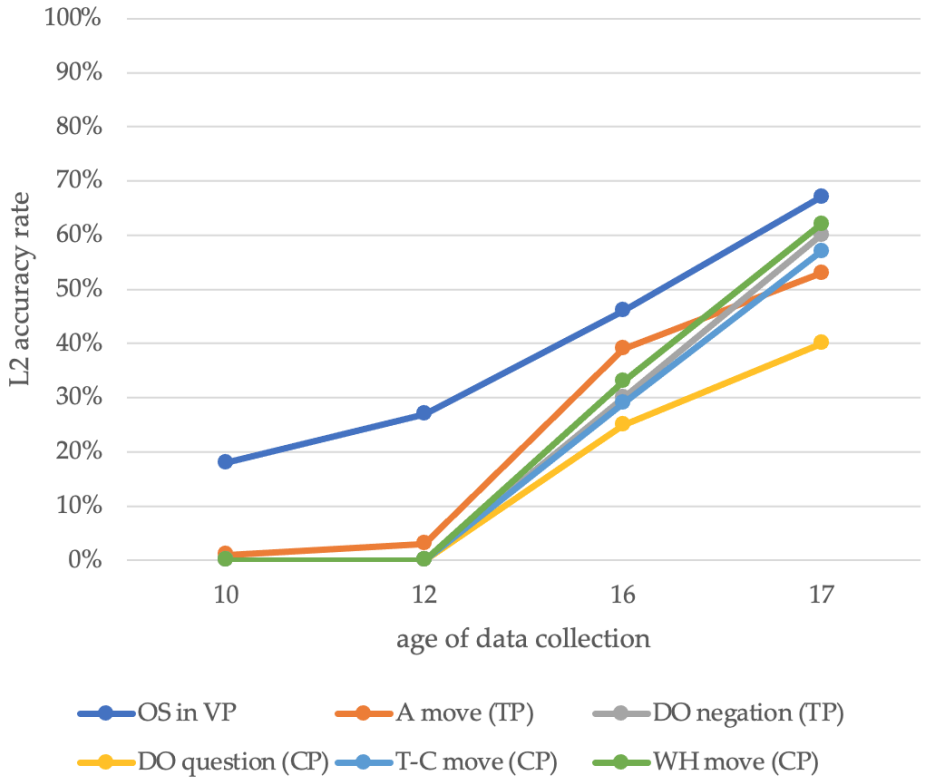

Figure 6.1: Learners' mean L2 accuracy of the computational properties' manifestations over the four rounds of data collection

The graph in 6.1 shows that at ages 10 and 12, learners have the highest L2 accuracy rates with ambiguous VP/TP utterances, which I argue are instances of VP-internal overt subjects (OS in VP). At these early ages, L2 accuracy of the computational properties remains close to 0%. A paired-samples t-test reports a significant increase in L2 computational accuracy between these early ages (10 &

12) (M= 0.55%, SD= 0.73%) and the later ages (16 & 17) (M= 43.66%, SD= 28.75%); t (8) = 4.54, p <0.05 (two tailed). The first evidence for knowledge of the computational properties wh-movement, T-C movement and do-support appear for the first time at age 16. Tables 6.7 and 6.8 below give some example utterances from Learner 38 and Learner 42's transcripts across the data collection period, to exemplify the nature of this developmental trajectory.

age 10	age 12	age 16	age 17
	FE: what's your name? *where you live?	**FE: what's your name?**	**FE: where do you live?**
no transcript	*they age, I am thirteen years old, I eat, foot, chair, a bed, they ate, I play football, they play football, *the dog see the the sandwiches, *her dog hm eat the food, you prepare my birthday?, it's my birthday, house in my house?*	*I wanna go to university, I don't know, I like to do in my freetime, I can't sorry, can I do the party? can you do something for eat?, *why are doing this kind?, *what you wanna say?, I have to study so I do the homework, I went to the cinema*	*I need to sleep, I've been at this school half of my life and I love this house, I'll finish at [\](.) at half past twelve, they don't speak Catalan, I think that we have five rooms, why are you doing this work?, I will go to the pub with my girlfriend or something similar*

Table 6.7: Learner 38: Example utterances across the four rounds of data collection

age 10	age 12	age 16	age 17
	FE: -	**FE: what is your name? where are you from?**	**FE: what is your name? *where is you from?**
no transcript	*half past one, homework, *_listen the music. *_like the computer, nine o'clock, *_speak English?, seventeen, one, two children, eat, girl, the basket. the sandwiches. *_eaten the eat [\ \] eaten the dog. girl and boy.*	*_repeat please?. *_listen to music, *_eating and _see the television, I study, *_is the childrens and mother and dog, *childrens is a mountain, window table, childrens and a dog and a mother, *_meeting my friends, *_going to the bar*	*the childrens see a and a dog, mother say goodbye, the children climb the mountain, I had dinner hm at night and hm bars, how are you?, *the dog is hm basket, what?*

Table 6.8: Learner 42: Example utterances across the four rounds of data collection

The development from the tables above can be seen as representative of all learners' utterances at each age of data collection, and they show how at early ages, as well as some accurate VP internal overt subject clauses (eg. *I play football*), learners' other L2 utterances are either ungrammatical (eg. **her dog eat the food, _speak English*) or restricted to single lexical items only at this stage (eg. *homework, girl*

and boy). Evidence of L2 computational knowledge (and more generally, functional categories T and C) appears at the later ages (e.g. *can I do the party, I had dinner hm at night*). This developmental trajectory fits a Weak Continuity view of the L2 initial state (Vainikka, 1994; Vainikka and Young-Scholten, 1998; Hawkins, 2001; Bhatt and Hancin-Bhatt, 2002).

Despite all learners' similar developmental trajectory in the L2 across the data collection period, we still find intra-learner variation within L2 accuracy rates for the computational properties at the later ages (note the high standard deviation for mean L2 accuracy rates at the later ages $SD = 28.75\%$). It is clear from Tables 6.7 and 6.8 above, for example, that Learner 38's interlanguage is more developed at ages 16 and 17 than that of Learner 42. Learner 38 produces a variety of utterances which demonstrate knowledge of the computational properties associated with functional categories T and C (incl. *can I do it, why are you doing this work, I think that we have five rooms*), whereas Learner 42's interlanguage at these ages- whilst clearly more developed from the earlier ages- shows far less evidence for knowledge of these properties and often makes errors where these should manifest (eg. **the childrens see a and a dog, *the dog is hm basket what?*). To quantify this difference across the corpus, I calculated each learner's L2 computational accuracy at the early ages (10 &12) and later ages (16 & 17) as a mean percentage, as exemplified in the table below with Learner 7.

Learner 7					
age	WH	T-C	DO	A	**mean**
10	(0/1) 0%	(0/1) 0%	n/a	(0/2) 0%	0%
12	(0/4) 0%	(0/1) 0%	(0/9) 0%	(1/26) 4%	1%
mean	0%	0%	0%	2%	**1%**
Learner 7					
age	WH	T-C	DO	A	**mean**
16	(1/3) 33%	(0/4) 0%	(1/9) 11%	(6/19) 32%	19%
17	(3/6) 50%	(2/4) 50%	(2/7) 29%	(14/41) 34%	41%
mean	42%	25%	20%	33%	**30%**

Table 6.9: Learner 7: L2 computational accuracy at early ages (10&12) and later ages (16&17)

This allowed for a comparison of all learners' L2 computational accuracy rates at early ages versus later ages of data collection, which can be seen in Table 6.10 below.

Learner	L2 accuracy (%)	
	ages 10 & 12	ages 16 & 17
2	1%	12%
5	0%	72%
7	1%	30%
13	0%	37%
18	1%	80%
27	0%	1%
38	2%	77%
42	0%	24%
47	0%	81%

Table 6.10: Learners' L2 computational accuracy in all required contexts at early vs later ages

Table 6.10 shows that all learners have a similar lack of L2 computational accuracy at the initial stages, but show variation in these accuracy rates at the later ages. Sections 5.4 and 6.1.1 also made clear that learners differ in how many of the FEs^{wh} they produce across the data collection period, and the first age when they are shown to do this. Table 6.11 is repeated from Section 5.4 below to illustrate this, where we see that 5 learners produce an FE^{wh} at the early ages (10 & 12) and 4 fail to do so until age 16.

learner	age 10	age 12	age 16	age 17
2	-	-	*Hm <what are> [\ \] what [\ \]* *what is your name* *how old are you*	*what's your name*
5	*what's your name*	-	*what's your name*	*what's your name*
7	*what's your name*	*what's your name* *how old are you* *where do you live*	-	*what's your name first of all* *how old are you (x2)* *well and where do you live*
13	-	-	*what's your name* *where do you live* *what's her name*	*what's your name* *and where do you live* *how old are you*
18	-	*what's your name* *how old are you* **what do you live*	-	-
27	*no transcript*	-	*how old are you* *where do you live*	*how old are you* *where do you live* *what is your name*
38	*no transcript*	*what's your name* **where you live*	*what's your name*	*where do you live now*
42	*no transcript*	-	*what is your name* *where are you from*	*what is your name* **where is you from*
47	*how old are you* *what's your name*	*no transcript*	*how old are you* *where do you live*	*how old are you (x2)* *what's your name* *where are you from*

Table 6.11: Productions of the extracted FEs^{wh} across the BELC

Tables 6.7 and 6.8 showed how Learner 38 demonstrates a higher knowledge of the FEs^{wh}' computational properties at the later ages than Learner 42, and they also show that Learner 38 produces two FEs^{wh} at age 12 (*what's your name, *where you live*). Conversely, Learner 42 fails to produce any until age 16 (*what is your name, where are you from*). To investigate this trend further, I divided those 5 learners that were shown to produce FEs^{wh} at the early ages (Learners 5, 7, 18, 38, 47) from those that produced them for the first time at the later ages (Learners 2, 13, 27, 42) and compared their mean L2 computational accuracy rates at the later ages.

An independent samples t-test reports a significant difference in L2 computational accuracy rates for those learners who produced an FE^{wh} at the early ages (Early FE^{wh}) (M = 61.2%, SD = 24.8%) and those who failed to do so until the

later ages (Later FE^{wh}) (M = 22.3%, SD = 14.2%; t (7) = 2.95, p = <0.05). The four graphs below can be used to illustrate how this difference manifests over the data collection period. The blue line shows the mean L2 computational property accuracy rates of the Early FE^{wh} learners, and the orange line shows the mean rates of the Later FE^{wh} learners. The grey bars show the mean average of all learners.

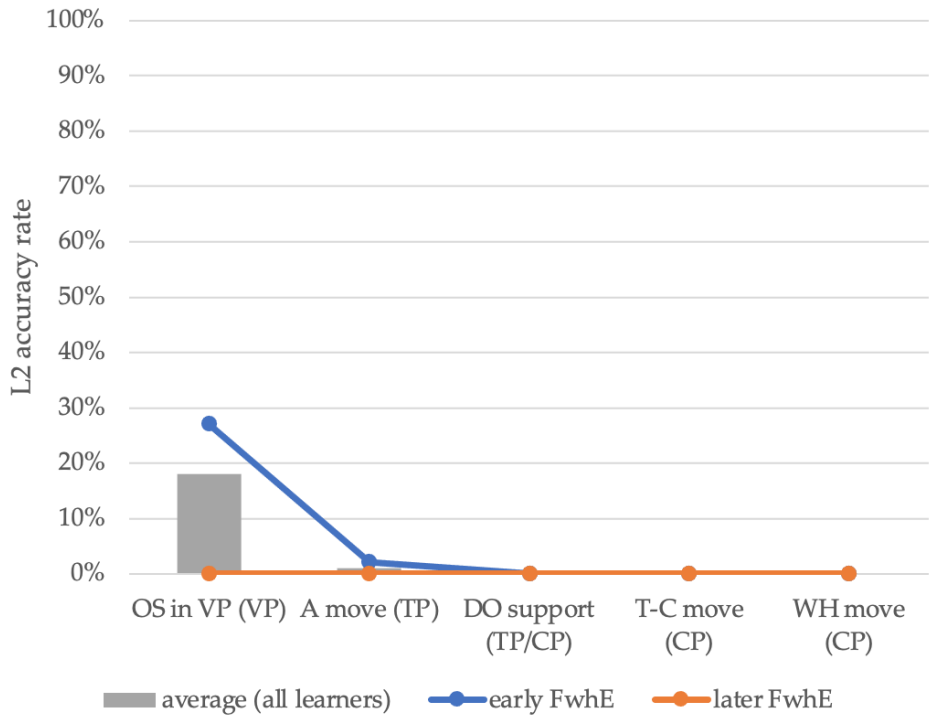

Figure 6.2: Learners' L2 computational accuracy age 10: Early vs later FE^{wh} use

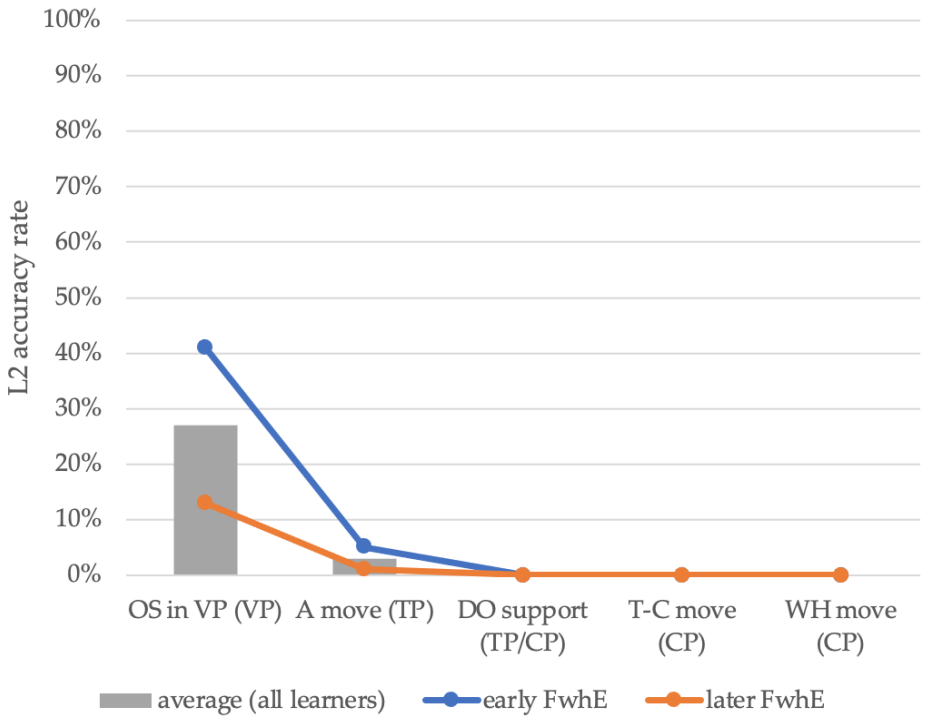

Figure 6.3: Learners' L2 computational accuracy age 12: Early vs later FE^{wh} use

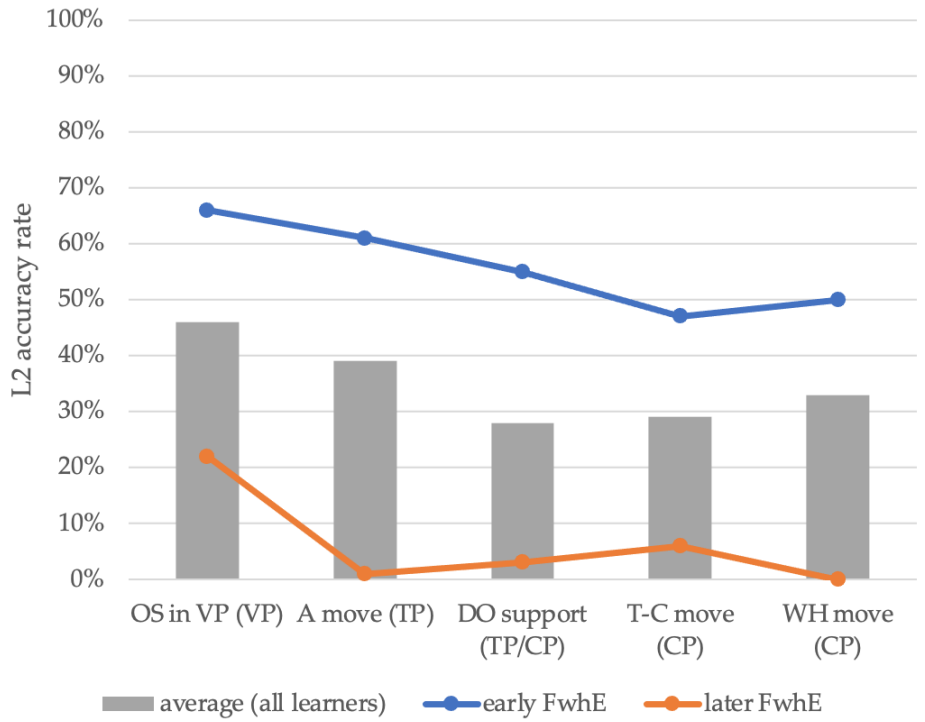

Figure 6.4: Learners' L2 computational accuracy age 16: Early vs later FE^{wh} use

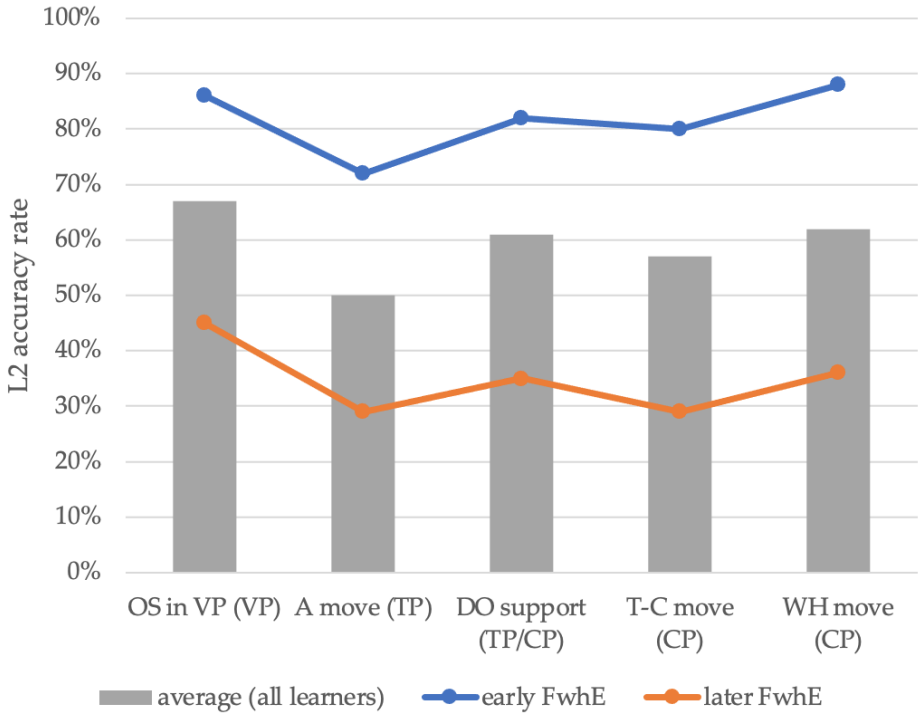

Figure 6.5: Learners' L2 computational accuracy age 17: Early vs later FE^{wh} use

The graphs above show how all learners pass through a similar developmental trajectory, which builds incrementally from L2 accuracy of VP to TP to CP projections. This is compatible with a Weak Continuity, specifically a Minimal Trees hypothesis (Vainikka and Young-Scholten, 1998), which states that the L2 initial state consists of lexical categories only, whilst functional ones develop in succession. However, those learners with earlier FE^{wh} use seem to be 'bootstrapped' into this trajectory, showing higher L2 accuracy rates of the FEs^{wh}' associated computational properties at each stage. Sections 6.1.3 and 6.1.4 now investigate this relationship more closely. They analyse the effect of learners' age of first FE^{wh} production and number of FE^{wh} productions at the early ages (10 & 12) as independent variables on their later L2 knowledge of the expressions' associated syntactic computations. Table 6.12 shows these variables alongside learners' L2 accuracy rates as raw numbers and relative percentages, and the Sections below present statistical analyses of these figures to discover developmental trends within this data.

Learner			L2 accuracy of computational property at later ages as mean averages (16 & 17)													
#	1st FE prod.	# of FEs	*wh*-movement			T-C movement			*do*- support			A-movement			Mean (total)
	age	ages 10 & 12	age 16	age 17	**mean**	age 16	age 17	**mean**	age 16	age 17	**mean**	age 16	age 17	**mean**	
2	16	0	(0/1) 0%	(0/1) 0%	**0%**	(0/1) 0%	(0/1) 0%	**0%**	(0/6) 0%	(4/8) 50%	**25%**	(1/20) 5%	(4/10) 40%	**23%**	**12%**
5	10	1				(0/1) 0%	(1/1) 100%	**50%**	(7/8) 88%	(1/1) 100%	**94%**	(12/18) 67%	(6/8) 75%	**71%**	**55%**
7	10	4	(1/3) 33%	(3/6) 50%	**42%**	(0/4) 0%	(2/4) 50%	**25%**	(1/9) 11%	(2/7) 29%	**20%**	(6/19) 32%	(14/41) 34%	**33%**	**30%**
13	16	0	(0/1) 0%	(2/2) 100%	**50%**	(0/1) 0%	(2/4) 50%	**25%**	(0/1) 0%	(2/3) 67%	**34%**	(3/10) 30%	(8/16) 50%	**40%**	**37%**
18	12	3	(4/5) 80%	(2/2) 100%	**90%**	(3/3) 100%	(5/5) 100%	**100%**	(1/7) 14%	(4/4) 100%	**57%**	(14/21) 67%	(23/29) 79%	**73%**	**80%**
27	16	0	(0/11) 0%	(0/3) 0%	**0%**	(0/4) 0%	(0/3) 0%	**0%**	(0/26) 0%	(0/8) 0%	**0%**	(1/43) 2%	(1/24) 4%	**3%**	**1%**
38	12	2	(2/2) 100%	(1/1) 100%	**100%**	(5/8) 63%	(1/2) 50%	**57%**	(6/8) 75%	(4/5) 80%	**78%**	(36/50) 72%	(20/27) 74%	**73%**	**77%**
42	16	0	(0/2) 0%	(3/7) 43%	**22%**	(1/4) 25%	(2/3) 67%	**46%**	(1/9) 11%	(4/19) 21%	**16%**	(1/28) 4%	(7/35) 20%	**12%**	**24%**
47	10	2	(2/4) 50%	(3/3) 100%	**75%**	(5/7) 71%	(6/6) 100%	**86%**	(6/11) 55%	(7/7) 100%	**78%**	(36/52) 69%	(29/29) 100%	**85%**	**81%**

Table 6.12: Learners age of first FE^{wh} production & frequency of FE^{wh} production at the early ages with their L2 accuracy of related computational properties at the later ages

6.1.3 Age of first FE^{wh} production and corresponding L2 knowledge of associated computational properties

The scatterplot in Figure 6.6 below shows learners' L2 computational accuracy rate (as a mean average between all properties under investigation) alongside the first age they are shown to produce an FE^{wh}.

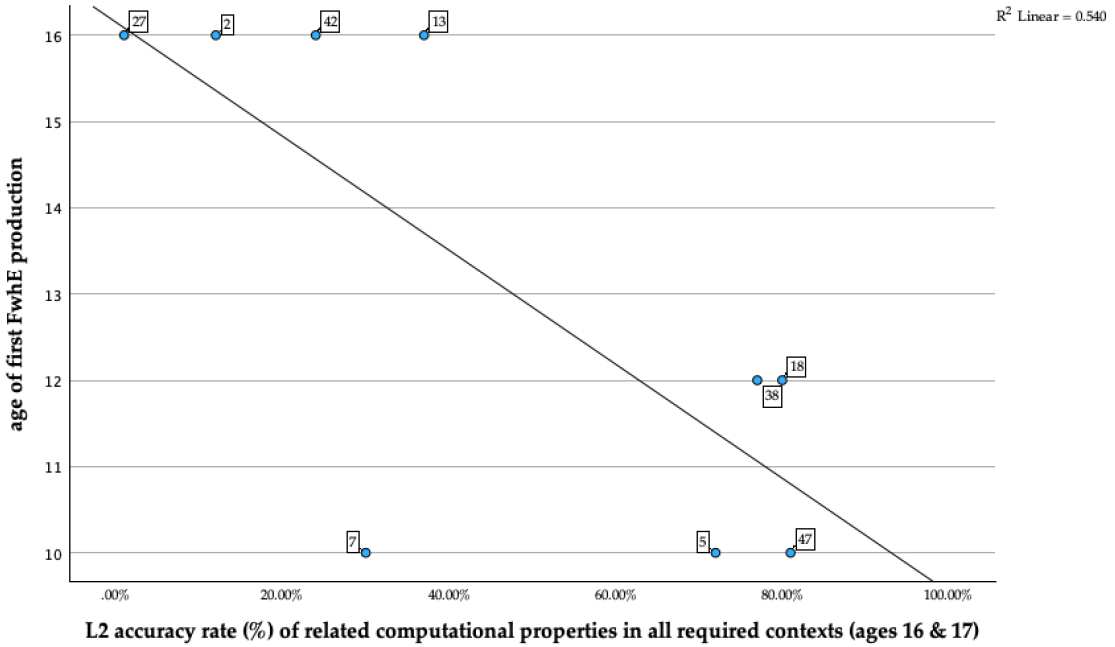

Figure 6.6: Learners' age of first FE^{wh} production and corresponding L2 computational accuracy at later ages as a mean rate (16&17)

Figure 6.6 above shows a relatively steep regression line with a negative slope, reading an 'R Squared Linear value' of 0.540 which, for such a small sample size, indicates an amount of linearity between the variables. This indicates that the younger the age of learners' first FE^{wh} production, the higher the L2 accuracy of their related computational properties at the later ages. I then ran Pearson product-moment correlation coefficients between these variables for each associated computational property individually as well as these combined as a mean average (as seen in the scatterplot). I include their corresponding p values and bootstrapped BCa 95% confidence intervals, which can be seen below in Table 6.13.

L2 accuracy (ages 16 & 17)	**age of first FE^{wh} production**	sig. (2-tailed)	bootstrap (BCa) 95% Confidence Interval
Mean	-.735	*.024	**-.959/-.284
wh- movement	-.683	*.062	**-.954/-.294
T-C movement	-.593	*.093	**-.903/-.028
do- support	-.705	*.034	**-.973/-.210
A- movement	-.779	*.013	**-.985/-.320

*significant at the adjusted p <.15 for small sample sizes in SLA
** CI effect for a relationship among variables

Table 6.13: Pearson product-moment correlation coefficient between age of first FE^{wh} production and L2 accuracy of computational properties

Table 6.13 above shows rather large negative effect sizes, which indicate that a younger age of first FE^{wh} production correlates with a higher L2 accuracy of associated computational properties at the later ages.

Note that some learners (27, 38 and 42) did not participate in the spoken tasks at age 10, and therefore have no production data available for analysis at this age[1]. This means that it is impossible to tell whether or not these learners would have produced a FE^{wh} at age 10, which could impact the correlations found above. However, this issue is somewhat reduced when it is considered that two of these learners (27 and 42) failed to produce an FE^{wh} at the age of 12 when they were recorded for the first time. We can infer that since this is the case, it is unlikely that they would have been able to produce a FE^{wh} during the same set of tasks, two years prior. Learner 38, on the other hand, did produce an FE^{wh} at age 12, so it could be reasonably suggested that they would have produced one at age 10. However, this would only strengthen the correlation, as Learner 38 is one of the learners who shows the highest L2 accuracy rates at the later ages. The fact that some learners missed the first round of data collection also supports the divide of the BELC into early (10 & 12) and later (16 & 17) ages, to ensure that a representative picture of both FE^{wh} productions and L2 computational accuracy can be given at these

[1]The metadata of the BELC does not indicate a reason for this.

comparative stages for every learner.

Section 6.1.4 now compares learners' L2 computational accuracy at the later ages with their number of FE^{wh} productions at the early ages (10&12) as an independent variable.

6.1.4 Number of early FE^{wh} productions and corresponding L2 knowledge of associated computational properties

'Number of FE^{wh} production' refers to the number of FEs^{wh} learners produce at the early ages (10 & 12)[2]. Figure 6.7 shows these frequencies alongside learners' mean L2 accuracy rates represented as a scatterplot.

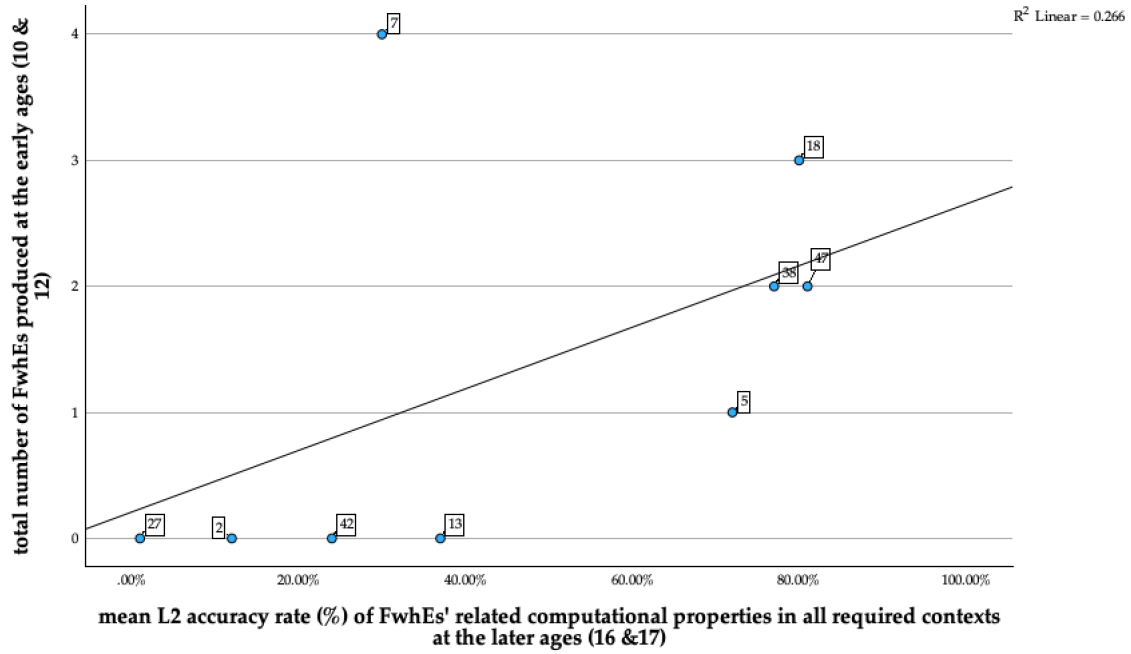

Figure 6.7: Scatterplot comparing learners' number of FE^{wh} productions at the early ages (10&12) and corresponding L2 computational accuracy at later ages (16&17)

Figure 6.7 shows a relatively steep regression line with a positive slope, reading an 'R Squared Linear value of 0.26 which, for such a small sample size, indicates an amount of linearity between a higher number of FE^{wh} productions at the early ages

[2]For example, Learner 7 produces *what's your name* at age 10 and *what's your name, how old are you* and *where do you live* at age 12. This Learner's total number of FE^{wh} productions at the early ages would therefore be 4.

and higher corresponding L2 computational accuracy at the later ages. To further investigate this relationship, Pearson product-moment correlation coefficients were ran between the variables for each associated computational property individually as well as these rates combined as a mean average (as in the scatterplot), which can be seen in Table 6.14.

L2 accuracy (ages 16 & 17)	**number of FE^{wh} productions (ages 10 & 12)**	sig. (2-tailed)	bootstrap (BCa) 95% Confidence Interval
Mean	.516	*.150	**.001/.967
wh- movement	.615	*.104	**.143/.968
T-C movement	.520	*.148	**.001/.966
do- support	.305	.425	-.199/.969
A- movement	.515	*.150	-.0.95/.945

*significant at the adjusted p <.15 for small sample sizes in SLA
** CI effect for a relationship among variables

Table 6.14: Pearson product-moment correlation coefficient between number of FE^{wh} productions at the early ages and corresponding L2 accuracy of computational properties at the later ages

Table 6.14 shows that a higher number of FE^{wh} productions at the early ages correlates with higher corresponding L2 computational accuracy rates at the later ages. Under the criteria of effect sizes and significance adopted in the present study, these effect sizes are strong and significant for all syntactic properties apart from do-support, which fails to reach significance.

6.1.5 Summary

Sections 6.1.3 and 6.1.4 show that learners' later accuracy of each L2 computational property is correlated with a younger age of first FE^{wh} production and a higher number of FE^{wh} productions at the early ages. However, as is widely recognised, correlation does not equal causation (Larson-Hall, 2016). When it is considered, however, that measuring the correlation between these variables was borne through testing a developmental hypothesis, the strength of causation certainly increases. The effect of the FEs^{wh} in this capacity is also supported by the homogeneity of

the learners under analysis, and the fact that external factors deemed likely to effect a learner's L2 performance were heavily controlled throughout the data collection period (see Section 4.1). Maintaining the homogeneity of learners' linguistic environment in this way strengthens the claim that individual differences in their development (at least in part) could be attributed to their differing FE^{wh} use.

Furthermore, the relationship between FE^{wh} use and L2 accuracy of associated computational properties does seem to be developmental. To be sure, we find a clear linearity between learners' differing use of these expressions at the early stages of data collection and differing L2 accuracy rates at the later stages. Conversely, if we count learners' individual FE^{wh} productions across the *entire* data collection period (across ages 10, 12, 16 and 17), and then compare these differing frequencies with their L2 computational accuracy rates at the later ages, we find no relationship. Instead, when analysing these variables, Figure 6.8 below shows a scatterplot with a relatively flat regression line. Table 6.15 also shows that overall frequency of FE^{wh} production across the data collection period does not correlate with higher L2 accuracy of any associated computational property individually or these as a mean average.

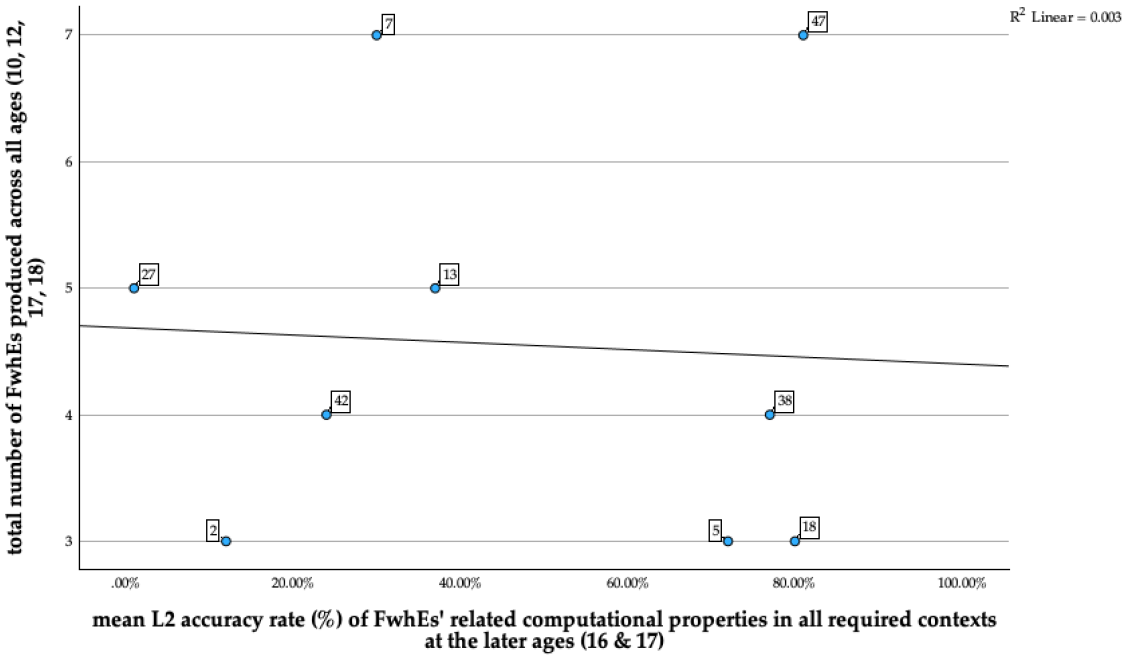

Figure 6.8: Scatterplot comparing learners' *total* frequency of FE^{wh} production across the corpus and corresponding L2 accuracy of computational properties at later ages (16&17)

L2 accuracy (ages 16 & 17)	**total number of FEs^{wh} produced (across all ages)**	sig. (2-tailed)	bootstrap (BCa) 95% Confidence Interval
Mean	-.057	.884	-.767/.598
wh- movement	.062	.883	-.682/.840
T-C movement	-.018	.963	-.817/.680
do- support	-.170	.662	-.746/.509
A- movement	.001	.997	-.727/.646

Table 6.15: Pearson product-moment correlation coefficient between *total* frequency of FE^{wh} productions across the corpus and L2 accuracy of computational properties at the later ages

Therefore, a higher L2 accuracy of associated computational properties seems to correlate specifically to a higher number of FE^{wh} productions at younger ages, rather than a frequent production of the expressions across the entire data collection period. This is suggestive of a more developmental relationship between FE^{wh} use and underlying L2 syntactic knowledge.

Section 6.1.6 now exemplifies this developmental effect by comparing pairs of learners who differ with regards to their FE^{wh} use and later L2 computational accuracy. In each pair, there is one learner who shows later use of the FEs^{wh} (i.e. delays usage until age 16) compared with one learner who shows early use of the FEs^{wh} (i.e. produces one at age 10 and/or 12). I compare their later utterances at ages 16 and 17 for evidence of knowledge of associated L2 computational properties.

6.1.6 Exemplification of correlations

6.1.6.1 Learners 27 and 47

Learners 27 and 47 can be compared to demonstrate the correlation between early and frequent FE^{wh} use and higher L2 syntactic accuracy at the later ages. Learner 27 delays usage of an FE^{wh} until the age of 16, when they produce *how old are you?* and *where do you live?* during the latter stages of the interview task. This learner's other L2 utterances at the later ages are restricted to what we can presume to be VP projections only, as these frequently show a lack of accurate corresponding overt subjects, auxiliary verbs and/or inflectional morphology. Some examples are given below.

(29) a. **I eating* (I am eating)

b. **the mother say goodbye and the child look* (the mother says goodbye and the child looks)

c. **_ return to the house* (they return to the house)

d. **the dog (.) eaten* (the dog has eaten)

The above utterances indicate this learners' lack of TP projections and corresponding A-movement in English. Where the other computational properties under investigation are required, this learner fails to even attempt them in the L2, and resorts to translanguaging/L1 utterances in all of these required contexts.

(30) a. *no estamos de acuerdo en como se llaman* [SPANISH]

we don't agree on what they are called

b. *what es que no se* [TRANSLANGUAGING]
what I don't know

c. *donde vivo?* [SPANISH]
where do I live?

d. *que hice la semana pasada?* [SPANISH]
what did I do last week?

Despite producing *where do you live?* fluently, this learner in the same data collection period resorts to producing *where do I live?* (30-c) in the L1. In the L2, these expressions differ only in the use of a personal pronoun; (*I* in place of *you*). This supports the notion that, at this age, the FEs^{wh} are likely unanalysed and being produced holistically. The utterances in (30-c) and (30-d) above are instances where the learner is echoing the interviewer, who poses these questions to the learner initially in the L2. The fact that the learner has echoed these back in the L1 can be interpreted two ways. One interpretation is that this echoing not only demonstrates a lack of L2 production competence, but perhaps also indicates a lack of comprehension. The other is that the L1 echo actually demonstrates the learner's comprehension (as they echo back the correct meaning in the L1), and this method is being used to clarify their understanding with the interviewer via the production system that they are comfortable with (the L1). If the latter is the case, it raises a question of where this initial comprehension comes from- the learners' phonological memory or their underlying syntactic knowledge? Learner 27 is clearly familiar with the meaning of the L2 phonological form *where do you live?*, as they are shown to produce this in the same interview transcript as a memorised FE^{wh}, so their comprehension can be hypothesised to be from their phonological memory. The other question that the interviewer posed to them, *what did you do last week*, could have been understood by the learner through the recognition of the clause-initial question word *what* and lexical items *last week* in the context of being questioned, in which case they were using material stored in phonological memory to decipher meaning of

similar wh-constructions. The alternative is that their understanding of this phrase came from their underlying L2 syntactic knowledge, which would demonstrate a quicker development of L2 comprehension over L2 production as this learner gives no L2 surface structure evidence to suggest syntactic competence of any of this wh-construction's composition parts.

On the other hand, Learner 47 is shown to produce the FEs^{wh} *how old are you?* and *what's your name?* for the first time at age 10. Their respective L2 computational accuracy at the later ages is the highest of all learners, at 81%. Examples of their L2 utterances at these later ages are given below. These often display a combination of the different computational properties together in complex/subordinate surface structures, indicative of CP projections.

(31) a. *what do you do in your free time?*
b. *do you play some instrument?*
c. *I suppose I will go to the university*
d. *I think that the children are going to the house to have more food and drink*

6.1.6.2 Learners 2 and 18

Learners 2 and 18 can be compared similarly. Learner 2 shows a mean computational accuracy of only 25% at the later ages, and produces a FE^{wh} for the first time at age 16. Similarly to Learner 27, the majority of their L2 utterances are likely lexical in nature, due to their lack of overt subjects with corresponding verbal inflection/and or auxiliary verbs.

(32) a. **_ is (.) is hm hm (.) _ have a bed* (they have a bed)
b. ** _ going to (.) excursion* (they are going on an excursion)
c. **the mother (.) hm read the map* (the mother is reading the map)
d. **_is a big room and hm _ have two beds* (it is a big room and has two

beds)

Utterances such as those above constitute this learners' low L2 accuracy of A-movement, which results in an overall 35% L2 accuracy rate. In those contexts which require manifestations of the other computational properties under analysis, the learner realises most of these in the L1.

(33) a. *no lo entiendo* [SPANISH]
I don't get it
b. *cuanto tiempo pasamos?* [SPANISH]
how much time did we spend?

When this learner is presented with potential required contexts of wh-movement and T-C movement in the form of root interrogatives, they opt instead for producing the L2 wh-word as part of a quantifier phrase in isolation, with the possible subordinate finite material elided.

(34) a. *what foods?* (what food do you want/will you have?)
b. *what drink?* (what drink do you want/will you have?)

Learner 18 on the other hand shows the second highest overall L2 computational accuracy at the later ages, at 80%, and produces three FEs^{wh} at the age of 12. Some of their accurate L2 utterances are given below. Like Learner 47, these include surface structures demonstrating a combination of the computational properties under analysis in CP projections.

(35) a. *what do (.) are you studying?*
b. *...they go to a mountain and (.) when they (.) they go to eat the breakfast the dog (.) they see that the dog eats the...*
c. *do you like your job?*
d. *they are (.) preparing the the breakfast because they go out of house*

6.1.6.3 Learners 42 and 38

Learners 42 and 38 present another large disparity, with mean L2 computational accuracy rates at 24% and 77% respectively. Learner 42 displays some accurate L2 utterances indicating the manifestation of the computational properties under analysis. However, the majority of their utterances in these required contexts are in the L1 and/or ungrammatical in the L2. They produce the FEs^{wh} *what is your name?* and *where are you from?* for the first time at age 16.

(36) a. *and how many people hm are hm going?*

b. **__ repeat please?* (can you repeat please?)

c. *una historia es que no se muy bien* [SPANISH]
a story that I do not know very well

d. *que et sembla si no tens 'menjar'?* [SPANISH]
what do you think if you don't have 'food'?

Learner 38 instead displays a range of utterances which demonstrate a much higher L2 computational accuracy at the later ages, and are shown to produce the FEs^{wh} *what's your name?* and *where you live?* for the first time at the age of 12. Some of these utterances are given below.

(37) a. *can you do something to eat?*

b. *why are you doing this work?*

c. *we don't need it to be happy*

d. *I've been at this school half of my life and I love this house*

Section 6.2 now conducts a usage-based traceback analysis of the BELC. It compares learners' use of of the FEs^{wh} with their later productive knowledge of the expressions' associated schematic constructions.

6.2 Learners' use of the FEs^{wh} and later knowledge of their schematic constructions

The results outlined in Section 6.1 found that learners' more frequent use of the FEs^{wh} at younger ages correlates with a higher L2 accuracy of the expressions' associated computational properties at the later ages. This section now compares learners' use of the FEs^{wh} with their corresponding knowledge of the expressions' schematic constructions, as conceptualised under usage-based linguistic frameworks. I analyse the form of learners' L2 interrogatives (both yes/no and wh-) across the longitudinal data collection period, to see if their utterance schemas match those of an FE^{wh} that precedes these ontogenetically in learners' production data. I begin by looking at learners' wh-questions.

6.2.1 Wh-questions

As seen in Section 5.6, the FEs^{wh} have the potential to represent lexically (38-a) and categorically (38-b) specific 'wh question utterance schemas', as well as fully schematic patterns (38-c). These are repeated and exemplified below with *where do you live*.

(38) *where do you live*

a. lexically specific utterance schema: [where do] + X

b. categorically specific utterance schema: [WH + AUX DO] + X

c. fully schematic pattern: [WH + AUX DO + PRN + VERB]

Tables 6.16 - 6.24 below show learners' FE^{wh} productions and all other L2 wh-interrogatives at each age of data collection. Where FEs^{wh} are not shown for a certain age, this means that the learner did not produce an FE^{wh} at this age. Underneath the FEs^{wh} and wh-questions are their lexically (i) and categorically (ii) specific utterance schemas as well as their fully schematic patterns (iii). I have

underlined instances where these of their wh-questions match those of a previously produced FE^{wh}. 'NT' refers to 'no transcript', meaning that the learner did not participate in that round of data collection, and a dash '-' indicates that learners did participate but were not shown to produce any L2 interrogatives in full wh-question form at this stage.

#	age 10: wh-Q	age 12: wh-Q	age 16: FE	age 16: wh- Q	age 17: FE	age 17: wh-Q
2	-	-	***what is your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA + PRN]	-	***what is your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+ NOUN]	-

Table 6.16: Learner 2

#	age 10: wh-Q	age 10: FE	age 12: wh-Q	age 16: wh-Q	age 16: FE	age 17: wh-Q	age 17: FE
5	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+Poss DET+NOUN]	**what time of eat hemes des comprares?*		***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+Pos sDET+NOUN]	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+Poss DET+NOUN]

Table 6.17: Learner 5

#	age 10: wh-Q	age 10: FE	age 12: wh-Q	age 12: FE	age 16: wh-Q	age 17: wh-Q	age 17: FE
7	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+P ossDET+NOUN]	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+Pos sDET+NOUN] ***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPUL A] + X iii. [WH+ADJ+COPUL A + PRN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]	*why we have these questionnaires* i. [why + we] + X ii. [WH + PRN] + X iii. [WH + PRN + VERB + DET + NOUN]	*what have in your free time buenos* i. [what + have] + X ii. [WH + VERB] + X iii. [WH + VERB + PREP + PossDET + NOUN]	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+ PossDET+NOU N] ***how old are you*** i. [how old are] + X ii. [WH+ADJ+COP ULA] + X iii. [WH+ADJ+COP ULA + PRN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]

Table 6.18: Learner 7

#	age 10: wh-Q	age 12: wh-Q	age 16: FE	age 16: wh- Q	age 17: FE	age 17: wh-Q
13	-	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]	*what's her name* i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN]	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN]	*where do you go the last weekend* i. [where do] + X ii. [WH+AUX DO]+ X iii. [WH+AUX DO + PRN + VERB] **what language talk you in the house* i. [what language talk] + X ii. [WH+N+V] + X iii. [WH+NOUN+VERB+PRN+Prep+Det+NOUN]

Table 6.19: Learner 13

#	age 10: wh-Q	age 12: wh-Q	age 12: FE	age 16: wh-Q	age 17: wh-Q
18	**what is your party?* i. [what is] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PossDET+NOUN]	**what do you like sports?* i. [what do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB+NOUN]	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA + PRN] ****what do you live*** i. [what do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO+PRN +VERB]	*what is your job* i. [what is] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PossDET+NOUN]	*what are you studying?* i. [what are] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PRN+VERB -ing]

Table 6.20: Learner 18

#	age 10	age 12: wh-Q	age 16: FE	age 16: wh-Q	age 17: FE	age 17: wh-Q
27	NT	-	***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA + PRN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]	-	***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA + PRN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB] ***what is your name*** i. [what is] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN]	-

Table 6.21: Learner 27

#	age 10	age 12: wh-Q	age 12: FE	age 16: wh-Q	age 16: FE	age 17: wh-Q	age 17: FE
38	NT	-	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***where you live*** i. [where + you] + X ii. [WH + PRN] + X iii. [WH + PRN + VERB] + X	*why are you doing this kind?* i. [why + are] + X ii. [WH + COPULA] + X iii. [WH+COPULA +PRN+ VERB -ing+DET+NOUN] **what you wanna say?* i. [what + you] + X ii. [WH + PRN] + X iii. [WH + PRN + AUX + VERB]	***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA] +PossDET+NOUN]	*why are you doing this work?* i. [why are] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PRN+VERB-ing+DET+NOUN]	***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]

Table 6.22: Learner 38

#	age 10	age 12: wh-Q	age 16: FE	age 16: wh-Q	age 17: FE	age 17: wh-Q
42	NT	-	***what is your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PossDET+NOUN] ***where are you from*** i. [where are] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PRN+PREP]	-	***what is your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+PossDET+NOUN] ***where is you from*** i. [where are] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PRN+PREP]	*how are you* i. [how are] + X ii. [WH+COPULA] + X iii. WH+COPULA+PRN] *what do you want* i. [what do] + X ii. [WH + AUX DO} + X iii. [WH+AUX DO+PRN+VERB] *how many people are going* i. [how many people are] + X ii. [WH+ADJ+NOUN]+ X iii. [WH+ADJ+NOUN+COPULAR+VERBing]

Table 6.23: Learner 42

#	age: 10 wh-Q	age 10: FE	age 12	age 16: wh-Q	age 16: FE	age 17: wh-Q	age 17: FE
47	-	***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA] + PRN] ***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA +PossDET+NOUN]	NT	**how many frequency do you go to the cinema?* i. [how many frequency do] + X ii. [WH+QUANT+NOUN+AUX DO] + X iii. [WH+QUANT+NOUN+AUX DO+PRN+VERB+PP+DET+NOUN] *which film do you like?* i. [which film do] + X ii. [WH+NOUN+DO] + X iii. [WH+NOUN+DO+PRN+VERB]	***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA] + PRN] ***where do you live*** i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO + PRN +VERB]	*what do you do in your free time?* i. [what do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO+PRN+VERB +PP+PossDET+ADJ+NOUN] *where do you work?* i. [where do] + X ii. [WH+AUX DO] + X iii. [WH+AUX DO+PRN+VERB]	***how old are you*** i. [how old are] + X ii. [WH+ADJ+COPULA] + X iii. [WH+ADJ+COPULA] + PRN] ***what's your name*** i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PossDET+NOUN] ***where are you from*** i. [where are] + X ii. [WH+COPULA] + X iii. [WH+COPULA+PRN+PREP]

Table 6.24: Learner 47

The Tables above show that a total of 20 full clause L2 wh- interrogatives are produced by 7 out of the 9 learners under analysis across the data collection period.

Out of these 20 wh-interrogatives, 17 are preceded by FE^{wh} use. Therefore, these 17 wh-interrogatives are the ones fit for analysis and comparison, as they are the utterances that follow FE^{wh} use ontogenetically in learners' data. Learner 13, for example, produces both the FE^{wh} *what's your name* and the wh-interrogative *what's her name*- which share the same fully schematic pattern- at the same data collection point at age 16. As these utterances are both produced at the same age, and no other FE^{wh} production precedes these at age 10 and 12, *what's her name* cannot be analysed as an instance of FE^{wh} utterance schema extraction, as it does not follow an FE^{wh} ontogenetically in this learner's available production data.[3]

Out of the 17 interrogatives that follow FEs^{wh} in learners' production data, 9 of these embody the same categorically-specific utterance schemas of a previously produced FE^{wh}, which are produced by five of the learners under analysis (Learners 13, 18, 38, 42 and 47). Three out of these five learners are Early FE learners (18, 38, 47), that is, they are those learners that are shown to produce FEs at the early ages (10 & 12). Out of these 9 wh-interrogatives, 3 embody the same lexically-specific utterance schemas and 4 show the same fully schematic patterns of a previously produced FE^{wh}. This accounts for 53% of the total wh-interrogatives that proceed FE^{wh} use in the longitudinal data, and 7/9 (77.7%) of these are produced by the Early FE learners. I now briefly discuss these patterns below.

Learner 13 produces *where do you live* at age 16 and *where do you go the last weekend* at age 17 with the same fully schematic pattern [WH + AUX DO + PRN + VERB]. Learner 18 produces *what's your name* at age 12 and then produces *what is your job* and *what are you studying* at ages 16 and 17. *What is your job* shares the same fully schematic pattern [WH + COPULA + PossDET + NOUN], and *what are you studying* shares the same categorically specific utterance schema [WH + COPULA] + X. Note that at age 10, this learner also overextends [what is] in

[3]This is also why *what's her name* is treated as a variant of the FE^{wh} *what's your name*, as presented in 5.4.

the erroneous **what is your party?* (instead of *where* is your party), which could be further suggestive of a learning sequence based on this [WH + COPULA] utterance schema.

Learner 38 produces *what's your name* at age 12 and *why are you doing this kind* and *why are you doing this work* at ages 16 and 17, which share the same utterance schema [WH + COPULA] + X. They also produce another FE^{wh} erroneously at age 12 **where you live*, and seem to adopt this [WH + PRN] + X utterance schema which leads to another erroneous wh-question at age 16 **what you wanna say*. Learner 42 produces the FEs^{wh} *what is your name* and *where are you from* at age 16 and *how are you* at age 17, which all share the same categorically specific utterance schema [WH + COPULA] + X. Finally, Learner 47 makes productive use of the fully schematic pattern [WH + AUX DO + PRN + VERB], producing *what do you do (in your free time)* and *where do you work* at age 17 after producing the FE^{wh} *where do you live* at age 16. Section 6.2.2 now looks at learners' yes/no questions in the L2.

6.2.2 Yes/No- questions

As well as the wh-question utterance schemas presented above, the FEs^{wh} have the potential to represent lexically (39-a) and categorically (39-b) specific 'yes/no utterance schemas', as repeated and exemplified below with *where do you live*.

(39) *where do you live*

a. lexically specific utterance schema: [do you] + X

b. categorically specific utterance schema: [AUX DO + PRN] + X

Tables 6.25 - 6.33 show learners' FE^{wh} productions and yes/no questions in the L2 across the data collection period. Underneath the FEs^{wh} and yes/no questions are their lexically (i) and categorically (ii) specific utterance schemas. I have underlined

instances where these of their yes/no questions match those of a previously produced FE^{wh}.

learner	age 10: y/n Q	age 12: y/n Q	age 16: FE	age 16: y/n Q	age 17: FE	age 17: y/n Q
2	-	-	***what is your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X	*can you repeat (x5)* i. [can you] + X ii. [ModalAUX + PRN] + X	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X	n/a

Table 6.25: Learner 2

#	age 10: y/n Q	age 10: FE	age 12: y/n Q	age 12: FE	age 16: y/n Q	age 16: FE	age 17: y/n Q	age 17: FE
5	n/a	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X	-	-	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X	*can you leave me the house this weekend?* i. [can you] + X ii. [ModalAUX + PRN] + X	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X

Table 6.26: Learner 5

#	age 10: y/n Q	age 10: FwhE	age 12: y/n Q	age 12: FwhE	age 16: y/n Q	age 17: y/n Q	age 17: FwhE
7	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X	-	*can I do it?* [can I] + X [ModalAUX + PRN] + X *are you sure?* [are you] + X [COPULA + PRN] + X	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X

Table 6.27: Learner 7

learner	age 10: y/n Q	age 12: y/n Q	age 16: FwhE	age 16: y/n Q	age 17: FwhE	age 17: y/n Q
13	-	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X ***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X	*do you like your work* i. [do you] + X ii. [AUX DO + PRN] + X

Table 6.28: Learner 13

#	age 10: y/n Q	age 10: FE	age 12: y/n Q	age 12: FE	age 16: y/n Q	age 17: y/n Q
18	-	-	*do you have any pets?* i. [do you] + X ii. [AUX DO + PRN] + X *do you like pets?* i. [do you] + X ii. [AUX DO + PRN] + X	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X ****what do you live*** i. [do you] + X ii. [AUX DO + PRN] + X	**do you born in Spain?* i. [do you] + X ii. [AUX DO + PRN] + X *do you like your job?* i. [do you] + X ii. [AUX DO + PRN] + X	*are you studying?* i. [are you] + X ii. [COPULA + PRN] + X *do you live in Barcelona?* i. [do you] + X ii. [AUX DO + PRN] + X *do you live on your own?* i. [do you] + X ii. [AUX DO + PRN] + X *do you have any brothers or sisters?* i. [do you] + X ii. [AUX DO + PRN] + X

Table 6.29: Learner 18

learner	age 10: y/n Q	age 12: y/n Q	age 16: FE	age 16: y/n Q	age 17: FE	age 17: y/n Q
27	NT	-	***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X	-	***how old are you*** i. [are you] + X ii. [COPULA + PRN] + X ***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X ***what is your name*** i. [is your] + X ii. [COPULA + PossDET] + X	-

Table 6.30: Learner 27

#	age 10	age 12: y/n Q	age 12: FE	age 16: y/n Q	age 16: FE	age 17: y/n Q	age 17: FE
38	NT	-	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X ****where you live***	*can I make a party?* i. [can you] + X ii. [ModalAUX + PRN] + X *can I do the party?* i. [can you] + X ii. [ModalAUX + PRN] + X **can you do something for eat?* i. [can you] + X ii. [ModalAUX + PRN] + X *can you do something for the party?* i. [can you] + X ii. [ModalAUX + PRN] + X **can we any drink?* i. [can you] + X ii. [ModalAUX + PRN] + X	***what's your name*** i. [is your] + X ii. [COPULA + PossDET] + X	-	***where do you live*** i. [do you] + X ii. [AUX DO + PRN] + X

Table 6.31: Learner 38

learner	age 10: y/n Q	age 12: y/n Q	age 16: FE	age 16: y/n Q	age 17: FE	age 17: y/n Q
42	NT	-	*what is your name* i. [is your] + X ii. [COPULA + PossDET] + X *where are you from* i. [are you] + X ii. [COPULA + PRN] + X	-	*what is your name* i. [is your] + X ii. [COPULA + PossDET] + X **where is you from* i. [*is you] + X ii. [COPULA + PRN] + X	-

Table 6.32: Learner 42

#	age 10: y/n Q	age 10: FE	age 12	age 16: y/n Q	age 16: FE	age 17: y/n Q	age 17: FE
47	-	*how old are you* i. [are you] + X ii. [COPULA + PRN] + X *what's your name* i. [is your] + X ii. [COPULA + PossDET] + X	NT	*do you like?* i. [do you] + X ii. [AUX DO + PRN] + X *can you repeat the question please?* i. [can you] + X ii. [ModalAUX + PRN] + X	*how old are you* i. [are you] + X ii. [COPULA + PRN] + X *where do you live* i. [do you] + X ii. [AUX DO + PRN] + X	*do you have brothers or sisters?* i. [do you] + X ii. [AUX DO + PRN] + X *do you like music?* i. [do you] + X ii. [AUX DO + PRN] + X *do you play some instrument?* i. [do you] + X ii. [AUX DO + PRN] + X	*how old are you* i. [are you] + X ii. [COPULA + PRN] + X *what's your name* i. [is your] + X ii. [COPULA + PossDET] + X *where are you from* i. [are you] + X ii. [COPULA + PRN] + X

Table 6.33: Learner 47

Adopting the same procedure as with wh-interrogatives, the tables show that a total of 23 yes/no questions are produced by 7 out of the 9 learners under analysis across the data collection period. Of these 23 yes/no interrogatives, 21 of them follow an FE^{wh} in learners' data ontogenetically. 11 of these yes/no questions produced by 4 learners (Learner 7, 13, 18 and 47) embody the same categorically specific utterance schemas as a previously produced FE^{wh} (53%). Three out of these four learners are Early FE learners (7, 18, 47), that is, they produce FEs at the early ages (10 & 12). All 11 of these yes/no interrogatives also share the same lexically-specific utterance schemas as the FEs^{wh}. Out of the 11 yes/no questions that match utterance schemas of previously used FEs, 10 of these (90.1%) are produced by learners who show early FE use (at ages 10 & 12). I briefly describe the observed patterns in the paragraph below.

Learner 7 produces *are you sure* at age 17 which contains the same lexically-specific utterance schema [are you] + X as the FE^{wh} *how old are you*, produced

previously at age 12. Learner 18 also makes use of this [are you] + X utterance schema; they produce *are you studying* at age 17 after also producing the FE^{wh} *how old are you* at age 12. Learner 13 makes use of the [do you] + X utterance schema at age 17 with *do you like your work*, after producing *where do you live* one year previously at age 16. Learner 18 also produces the erroneous FE^{wh} **what do you live* at age 12, and goes on to produce five yes/no questions with the [do you] + X utterance schema at ages 16 and 17, including *do you like your job, do you live in Barcelona?* and *do you have any brothers or sisters*. Further evidencing productive use of this utterance schema is their overextension of such in the ungrammatical **do you born in Spain*. Lastly, Learner 47 also makes use of the [do you] + X utterance schema in 3 interrogatives at age 17 (*do you have brothers or sisters, do you like music, do you play some instrument*) after producing *where do you live* at age 16.

6.2.3 Summary

To summarise, 53% of L2 (20/38) interrogatives in the longitudinal data can be traced back to utterance schemas of previously produced FEs^{wh}. These interrogatives are produced by 6 out of the 9 learners under analysis. All of these utterance schemas are categorically specific to preceding FEs^{wh} (20/20), 70% are lexically specific (14/20) and 44% of learners' wh-interrogatives (4/9) share the same fully schematic patterns of preceding FEs^{wh}. The most productive wh- question utterance schema is [WH + COPULA] + X (4/9) and the most productive yes/no question utterance schema is [do you] + X (9/11). No learner's total number of grammatical interrogatives produced in the L2 can be linked back to utterance schemas of previously used FEs^{wh}.

Chapter 7 now presents the analysis of the French Progression Corpus (FPC).

Chapter 7

Analysis: The French Progression Corpus (FPC)

This chapter documents the analysis of the FPC data. The analysis proceeds similarly to that of the BELC. First I identify FEs^{wh} in the corpus and document learners' productions of these across the longitudinal data collection period (Sections 7.1 and 7.2). Then, I outline the assumed syntactic derivation and related computational properties of these expressions under mainstream generative frameworks (Section 7.3). Finally, I present the expressions' abstract schematic constructions and extractable utterance schemas under usage-based models (Section 7.4). However, the FPC differs from the BELC in a number of important ways.

Firstly, it presents a more concentrated picture of learners' development at the initial stages of acquisition, containing recordings of learners at three intervals during their first two years of L2 instruction. Secondly, as will be seen, the FEs^{wh} identified in the FPC are very similar to those identified in the BELC. However, French permits a large range of syntactic variation with respect to question formation structures, as opposed to English which is far more restricted. This means that, whilst the FEs^{wh} in the BELC represent *all* possible question formation structures in the L2 to the learners, this is not the case in the FPC.

Therefore, comparing learners' use of the FEs^{wh} with their other L2 interrogatives *outside* of these expressions will allow to see whether learners reproduce structures/surface forms that mirror those of the FEs^{wh}, or whether they opt for other structures that are not exemplified by the FEs^{wh} but are also permissible in the target language in these environments. Due to this, unlike the BELC's generative analysis which compared knowledge of *all* surface structures related to the FEs^{wh}, computational properties, the generative analysis of the FPC will only compare learners' use of the FEs^{wh} with their L2 *question forms* outside of these across the data collection period. Section 7.1 now explains how the FEs^{wh} were identified in the FPC, before Section 7.2 documents learners' longitudinal productions of these across the corpus.

7.1 Learners' linguistic environment

As mentioned in Chapter 4, like learners of the BELC, the state-school foreign language classroom was the main source of L2 input for the FPC learners. Unlike the investigators who compiled the BELC, Myles and colleagues also recorded the French foreign language lessons attended by the learners under analysis on a biweekly basis, which allowed for a comprehensive insight into the language learning classroom context and learners' linguistic environment. Mitchell and Martin (1997) noted that the teaching style in this classroom was strongly oral and teacher centered, with a significant emphasis on the rehearsal and memorisation of conversational exchanges, often involving question-and-answer sequences.

Adding to this, Myles et al. (1999) describe that these activities typically followed from whole class practice 'in which the teacher modeled and rehearsed the target expressions globally, with little variation or analysis' [p. 55]. It has been well documented in a number of previous works (Myles et al., 1998, 1999; Myles, 2004, 2015) that these exchanges involved the rote learning of memorised formulaic

question forms derived from learners' input, similar to those identified in the BELC. This makes the FPC another ideal test ground for investigating the effect of such memorised FEs on associated L2 development. The identified FEs, and learners' corresponding productions of such, are now outlined in Section 7.2.

7.2 Formulaic expressions identified for analysis and learners' production of these across the corpus

This section now documents learners' productions of the identified FEs across the corpus, before Sections 7.3 and 7.4 analyse their structural form under generative and usage-based frameworks respectively. The fixed expressions used by learners of the FPC, as documented by Myles and colleagues in previous works, can be divided into formulaic wh- (FE^{wh}) and yes/no ($FE^{y/n}$) questions, which can be seen in (1) and (2) below.

(1) Formulaic wh-expressions (FEs^{wh})

comment t'appelles tu?
how call-yourself you

'what is your name?'

quel âge as-tu?
what age have-you

'how old are you?'

où habites-tu?
where live-you

'where do you live?'

quel est le date de ton anniversaire?
when is the date of your birthday

'when is your birthday?'

(2) Formulaic yes/no expressions (FEs$^{y/n}$)

tu as un animal?
you have an animal

'do you have a pet?'

tu as des frères ou des soures?
you have the brothers or the sisters

'do you have any brothers or sisters?'

Note that three of the FEs^{wh} in (1) are the French equivalents to those identified in the BELC. It is unsurprising that these expressions, which are highly conventional and prototypical in nature, constituted a significant part of these learners' early L2 input, given that their foreign language classes favoured role play and memorisation activities to mimic typical conversational exchanges. Tables 7.1 and 7.2 below document each learners' productions of the French FEs across the three rounds of data collection. The tables are given to primarily demonstrate the salience of the FEs at the initial round of data collection for all learners under analysis. All are shown to produce a number of the FEs at this stage ($Sum = 103$, $M = 4.29$, $SD = 1.42$), and the vast majority of these productions are the expressions in their target second-person structural forms as presented to learners holistically in their language classroom input (i.e. as in (1) and (2) above).

learner	Round 1: Year 7, Term 2 (ages 11-12)	Round 5: Year 8, Term 3 (ages 12-13)	Round 6: Year 9, Term 1 (ages 13-14)
1	*comment t'appelles tu | où habites-tu*	**ehm un garçon le café ehm (.) comment t'appelles tu | *a fille le parc comment t'appelles tu |*comment t'appelles tu le monsieur*	-
2	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date le mon anniversaire*	*où habites-tu*	-
3	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date le mon anniversaire | tu as un animal à la maison | tu as des frères et des sœurs*	**elles comment t'appelles tu | *comment elle t'appelles tu | *comment il t'appelles tu*	-
4	*comment t'appelles tu| quel âge as-tu | quelle est la date de ton anniversaire*	-	-
5	*comment t'appelles tu | *il a quel âge | *il habite où | *il a un animal à la maison*	-	-
6	*comment t'appelles tu | où habites-tu | quel âge as-tu | tu as un animal | tu as un frère ou un sœur*	-	-
7	*tu as quel âge | *comment tu appelles | où habites-tu | tu as un animal | tu as un frère ou un sœur*	**comment t'appelle le garçon | *comment t'appelle la fille*	-
9	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as un animal | tu as un frère ou un sœur*	**comment t'appelle t' elle | *comment t'appelle t' il*	-
10	*quel âge as-tu | quelle est la date de ton anniversaire | *il a un animal à la maison*		-
11	*comment t'appelles tu | quel âge as-tu | où habites-tu | tu as un frère ou un sœur*		-
12	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as un animal*	-	*quel âge as tu*
13	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as un animal à la maison*	-	-
14	*quel âge as-tu | *quelle heure est date de ton anniversaire | *comment se s'appellent elles [\ \] t' ils?*		**quel âge as-tu quel âge as Thingy*

Table 7.1: Production of the FEs: Learners 1-14

learner	Round 1: Year 7, Term 2 (ages 11-12)	Round 5: Year 8, Term 3 (ages 12-13)	Round 6: Year 9, Term 1 (ages 13-14)
15	*tu as quel âge | *tu as habites | tu as un animal*	-	-
18	*quel âge as-tu | quelle est la date de ton anniversaire | où habites-tu | *j'ai (.) j'ai un frère ou du sœur |*	-	-
19	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as des frères et des sœurs*	-	**comment t'appelles tu | *quel âge as-tu*
20	*comment t'appelles tu | quel âge as-tu | quelle est la date de ton anniversaire*	**comment t'appelles tu (third person reference) | *comment t'appelles tu euh (.) un fille | *comment t'appelles tu e un garçon | *comment t'appelles tu et l'homme*	-
22	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as des frères et des sœurs*		**quel âge as-tu*
23	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | tu as des frères et des sœurs | tu as un animal à la maison*	**comment t'appelles tu le (.) petit garçon | *comment t'appelles tu le petit fille | *comment t'appelles tu*	-
24	*comment t'appelles tu | où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire | *des frères de tu des sœurs | tu as les animaux*	-	*comment t'appelles tu*
26	*Tu as quel âge | tu as des frères ou des sœurs | *tu un un est animal*	**comment tu t'appelles le (.) garçon*	-
28	*où habites-tu | quel âge as-tu | quelle est la date de ton anniversaire*	-	-
29	*comment tu t'appelles | tu as un animal*	-	-
30	*comment tu t'appelles | tu as quel âge | *quelle la date de ton anniversaire | où habites tu | tu as un animal | tu as des frères ou des sœurs*	-	-

Table 7.2: Production of the FEs: Learners 15-30

At the later rounds 5 and 6 (columns 2 and 3 in the tables), the FEs are produced far less ($Sum = 25$, $M = 1.041$, $SD = 1.20$). Analysing these productions more closely, it can be seen that the majority are over-extensions in similar functional contexts, when learners are attempting to ask the name of a third-person subject by attaching a referential NP to the FEs^{wh} in their second-person form. Two examples are given below.

(3) a. Learner 20: Round 5 (Year 8, Term 3)

*comment t'appelles tu un fille?
what call-yourself you a girl

'what is the girl's name' (intended meaning)

b. Learner 3: Round 5 (Year 8, Term 3)

*elles comment t'appelles tu?
she what call-yourself you

'what is her name' (intended meaning)

Fluent FEs in their base (2nd person) form at the later ages are also marked with an asterisk (*) in the table when it is clear from the context that they were intended for third person reference, such as that of Learner 19 below.

(4) Round 6 (Year 9, Term 1)

Learner 19: *quel âge as tu?*

'how old are you'

Researcher: *um (.) Richard est (.) quatorze ans.*

'Richard is fourteen'

Note that I have included variations of the FE$^{y/n}$ *tu as un animal* which carry slightly different meanings, such as the plural form produced by Learner 24.

(5) Learner 24: Round 1 (year 7, Term 2)

tu as les animaux?

you have the pets

'do you have any pets'

Importantly, it can be seen that the FEs^{wh} are overwhelmingly produced by learners with subject-clitic inversion and a fronted wh-word, and the FEs$^{y/n}$ are all in declarative form, despite the fact that French exhibits a large amount of variation with

respect to the structure of interrogatives (Prévost, 2009). Section 7.3 now outlines this variation. It analyses the syntactic derivation of the FEs under mainstream generative models in the structural forms as produced by learners.

7.3 The assumed syntactic derivation and related computational properties of the FEs under generative frameworks

Wh-interrogatives in French display a large range of structural variation. The wh-word can remain in-situ (13-a) or be fronted (i.e. wh-moved). Wh-movement can occur without T-C movement (subject/clitic - verb inversion) (13-b), with the question marker *est-ce que*[1] (13-c), with clefting (13-d) or with T-C movement (13-e).

(6) Wh-questions in French e.g. *where do you work?*

a. wh in situ + no inversion [wh IN SITU]
vous travaillez où?
you work where

b. wh-movement, no inversion [wh + NO INV]
où vous travaillez?
where you work

c. wh-movement + *est-ce que* [wh + ESK]
où est-ce que vous travaillez?
where [ESK] you work

[1]The structural status of *est-ce que* is disputed in the literature, as to whether it is a fixed question marker or analysed sequence involving the inversion of *ce* (it) and *est* (is) and an embedded clause introduced by the complementiser *que* (that). However, I follow Prévost (2009) in analysing *est-ce que* as a fixed question marker due to the fact that it cannot be inflected for tense **sera ce-que* (will it be that) or modified by adverbs **est-ce que peut-être que* (is it perhaps that).

d. wh-movement with clefting [wh + CLE]
c'est où que vous travaillez?
it is where that you work

e. wh-movement + inversion [wh + INV]
où travaillez-vous?
where work you

Yes/no questions are also grammatical with subject-verb inversion (7-a) or without (7-b).

(7) Yes/no questions in French eg. *do you have a pen?*

a. yes-no inversion [INV]
as-tu un stylo?
have you a pen

b. yes- no without inversion [DEC]
tu as un stylo?
you have a pen

The syntactic derivation for each of the structural variations under generative frameworks are assumed to be as follows. Here, I concentrate specifically on aspects of the derivation that can account for the different interrogative surface forms that are permitted in French. As mentioned when discussing the derivation of the English FEs^{wh} in the BELC, it is assumed that wh-words carry an interpretable wh-feature [iWH] and finite verbs in T carry an interpretable tense feature [iT]. In root interrogatives, C carries an uninterpretable interrogative feature [uWH] and an uninterpretable tense feature [uT] (Chomsky, 1995). Where interrogatives display [wh IN SITU]

(8), it is assumed that neither the [uWH] nor the [uT] features on C carry the EPP property. This is why there is no movement of the wh-word or finite verb to specifier and head positions of CP respectively (Prévost, 2009; Prévost et al., 2010).

(8) *vous travaillez où?* [wh IN SITU]
you work where

Conversely, in those interrogatives which display wh-movement, the [uWH] feature is presumed to carry the EPP property. Where wh-movement occurs without inversion [wh + NO INV] (9), the [uT] feature on C is still lacking the EPP property. Hence, the finite verb remains uninverted in head T.

(9) *où vous travaillez?* [wh NO INV]
where you work

With those interrogatives that show wh-movement + *est ce-que* [wh + ESK] (10), the [uT] feature does carry the EPP property, which is checked via the External Merging of the interrogative particle/question marker [ESK] in C (Prévost, 2009; Prévost et al., 2014).

(10) *où [est ce-que] vous travaillez?* [wh + ESK]
where [is it that] you work

When wh-movement occurs with clefting [wh + CLE] (11), it is assumed that the wh-word is directly merged in the matrix CP projection. Taking root scope by means of Agree with a [uWH] feature in C, it moves an empty operator to the embedded C together with the embedding of the wh-cleft (Jakubowicz, 2011).

(11) *c'est où que vous travaillez?* [wh + CLE]
it is where that you work

Lastly, in those interrogatives displaying wh-movement and T-C movement [wh +

INV] (12) (as in English), both the [uWH] and [uT] features carry the EPP property. Hence, both the wh-word and finite verb are moved to the CP projection.

(12) *où travaillez-vous?* [wh + INV]
where work you

It is assumed that those structures involving more External/Internal Merge operations are more derivationally complex than those involving less operations (Hamann, 2006; Jakubowicz, 2011; Durrleman et al., 2016). Therefore, the French wh-question structures given in (6) above can be seen as ranked in order of syntactic complexity, from least derivationally complex to most derivationally complex. These are repeated below in (13).

(13) Wh-questions in French e.g. *where do you work?*

a. wh in situ + no inversion [wh IN SITU]
vous travaillez où?
you work where

b. wh-movement, no inversion [wh + NO INV]
où vous travaillez?
where you work

c. wh-movement + *est-ce que* [wh + ESK]
où est-ce que vous travaillez?
where [ESK] you work

d. wh-movement with clefting [wh + CLE]
c'est où que vous travaillez?
it is where that you work

e. wh-movement + inversion [wh + INV]
où travaillez-vous?
where work you

As discussed in the analysis of the BELC data, surface structures can be ambiguous as to whether they constitute evidence for the functional categories and associated computational properties that are assumed in their generation. Unlike English, French requires overt morphology on the lexical verb which indicates a relevant T head and can be used as surface evidence to disambiguate between VP and TP projections. However, French wh-interrogatives of the [wh IN SITU] kind which feature subject pronouns used in combination with the verbs *avoir* (have) and *être* (be) (used in a lexical sense) present an ambiguous case, as these conjugations are often rote-learned by memory in a formulaic- like fashion inside the EFL classroom (Samian and Tavakoli, 2012). Hence, it is also possible that these utterances are products of a 'clipping' strategy which combines these sequences with other lexically rehearsed material. Say, for example, a beginner learner of French produces the [wh + IN SITU] question form below in (14).

(14) *tu as quelle voiture?* (you have what car)

This utterance could equally be the learner producing the relevant memorised conjugation of *avoir* [*tu as*] clipped together with an appropriate referential DP [*quelle voiture*] as it could be a full CP projection.

As done with the BELC, it is necessary to examine other L2 question formation strategies from the same learners at the same point in time, to confirm the likelihood of either scenario. Say, for example, that outside of this utterance, the learner expresses L2 interrogatives largely via NPs only (with assumed rising intonation), such as that below in (15).

(15) *le nom?* (the name)

(intended: what is her name?)

This could suggest a lack of sufficient knowledge of L2 functional categories T and C and hence their [wh IN SITU] utterance *tu as quelle voiture* is likely lexical in nature. I will evaluate the likelihood of both scenarios where relevant from the results of the analysis.

Moving now to the derivation of yes/no questions in French, these can show subject-auxiliary/verb inversion [INV][2] via the EPP feature checking of T as shown below in (16).

(16) *As-tu un stylo?* [INV]
have you a pen

They can also be in declarative form [DEC] where it is assumed that the [uWH] feature on C has scope over the sentence which allows it to be interpreted as a question in either case, as in (17).

(17) *Tu as un stylo?* [DEC]
you have a pen

Based on the number of Merge operations involved in each derivation, [INV] yes/no question structures are assumed to be more syntactically complex than [DEC] ones. Like with wh-interrogatives of the [wh IN SITU] kind though, yes/no questions with the verbs *être* (be) and *avoir* (have) (when used lexically) are more ambiguous as to being genuine instances of this computational derivation outlined above or products of a clipping strategy involving rote-learned conjugations. Again, I will consider the likelihood of either scenario as they present themselves in the results based on learners' other strategies for L2 question formation at that particular point of data collection.

[2]The most common kind of inversion in French is also known as 'subject-clitic inversion' (Hulk and Zuckerman, 2000), where the subject clitic follows the finite verb as in (16).

To summarise, French interrogatives permit a large amount of variation with regard to their syntactic derivation, unlike English. There is, however, still some overlap. English wh-questions are of the [wh + INV] kind, as they require both wh-movement and T-C movement (subject-verb inversion). English yes/no questions are of the [INV] kind, as they also require T-C movement. It is necessary to consider this overlap in the results section, when analysing the structures of the FPC learners' L2 interrogatives *outside* of the FEs (i.e. their Novel Interrogatives (NIs)).

Returning now to examine the structures preferred in learners' productions of the identified FEs in the FPC, we see that out of all 24 learners under analysis, only four are shown to produce an identified FE^{wh} in [wh IN SITU] form. These are shown below in (18) - (21). The example in (21) shows that one of these learners also produces another FE^{wh} in [wh + NO INV] form.

(18) Learner 7: age 11-12 [wh IN SITU]

tu as quel âge?
you have what age

'how old are you?'

(19) Learner 15: age 11-12 [wh IN SITU]

tu as quel âge?
you have what age

'how old are you?'

(20) Learner 26: age 11-12 [wh IN SITU]

tu as quel âge?
you have what age

'how old are you'

(21) Learner 30: age 11-12

a. *tu as quel âge* [wh IN SITU]
you have what age
'how old are you'

b. *comment tu t'appelles* [wh + NO INV]
how you call yourself
'what is your name'

All other FEs^{wh} as produced by learners are in the [wh + INV] form, and all FEs$^{y/n}$ are produced in declarative [DEC] rather than inverted [INV] form.

(22) Formulaic wh- expressions (FEs^{wh})

a. *comment t'appelles tu?* (what is your name?) [wh + INV]

b. *quel âge as-tu?* (how old are you?) [wh + INV]

c. *où habites- tu?* (where do you live?) [wh + INV]

d. *quel est le date de ton anniversaire?* (when is your birthday?) [wh + INV]

(23) Formulaic yes/no expressions (FEs$^{y/n}$)

a. *tu as un animal?* (do you have a pet?) [DEC]

b. *tu as des frères ou des soures?* (do you have any brothers or sisters?) [DEC]

This homogeneity across learner productions supports the notion that, as similarly reported in Myles et al. (1998, 1999), learners were presented with the FEs in these structural forms in their EFL classroom as opposed to the other derivational possibilities.

The first part of the FPC results section concerns the generative analysis, which compares learners' FE productions with the structural form of their French inter-

rogatives *outside* of the FEs (i.e. their Novel Interrogatives (NIs)) across the three rounds of data collection. This is to see how their knowledge of the FEs' associated computational properties manifests when these are optional in the target structures. That is, do learners overwhelmingly produce wh-questions of the [wh + INV] type and yes/no questions of the [DEC] type, as exemplified by the FEs, or do they opt for other structures which are, for example, less derivationally complex ([wh + IN SITU]) or more similar to the L1 ([INV])? Before presenting these results, Section 7.4 now departs from the generative analysis of the FEs' syntactic derivation and outlines how the FEs as produced by learners in the FPC would be analysed under usage-based schematic models.

7.4 The FEs as abstract schematic constructions

Following Eskildsen (2015) as adopted for the analysis of the FEs^{wh} in the BELC, the French FEs exemplify the following fully schematic abstract constructions. Note that REF stands for 'reflexive', used to distinguish between reflexive/lexical verbs [REF VERB/VERB] and reflexive/ subject pronouns [REF PRN/SUBJ PRN].

(24) Formulaic wh-expressions (FEs^{wh})

a. *comment t'appelles tu* [WH + REF PRN + REF VERB + SUBJ PRN]

b. *quel âge as tu* [WH + ADJ + VERB + SUBJ PRN]

c. *où habites-tu* [WH + VERB + SUBJ PRN]

d. *quel est le date de ton anniversaire* [WH + COPULA + DET + NOUN + PREP + PossDET NOUN]

(25) Formulaic yes/no- expressions (FEs$^{y/n}$)

a. *tu as un animal* [SUBJ PRN + VERB + DET + NOUN]

b. *tu as des frères ou des soures* [SUBJ PRN + VERB + DET + NOUN + CONJ + DET + NOUN]

As posited for the FEs^{wh} in the BELC, learners have the potential to derive more general 'wh-question' and 'yes/no' question utterance schemas through analysis of the FEs, which can be lexically and/or categorically specific. As the FEs^{wh} are all in [wh + INV] word order form, they can function as wh-interrogative utterance schemas for wh-questions of the [wh + INV] kind only, as demonstrated below in 7.1 with *quel est le date de ton anniversaire*.

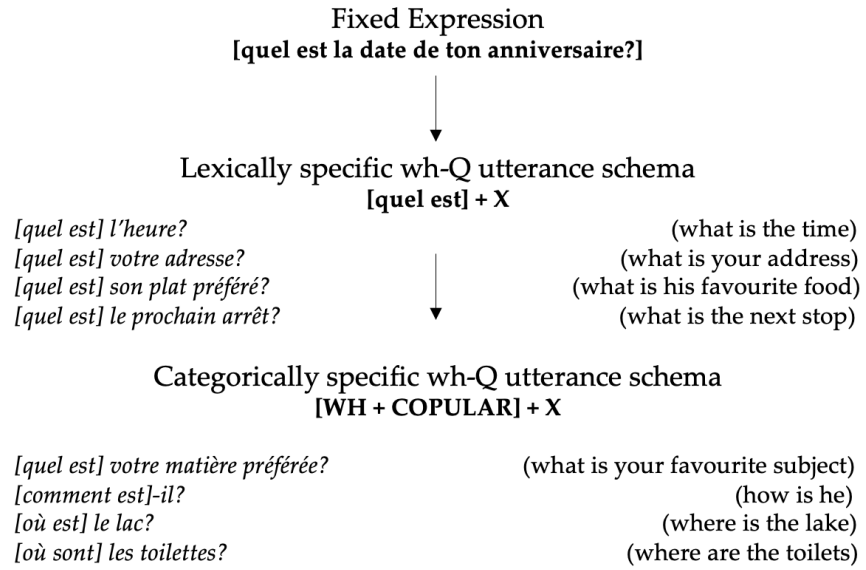

Figure 7.1: Example developmental trajectory of [wh + INV] wh-question forms from FE analysis

The FEs^{wh} can also function as yes/no-interrogative utterance schemas for yes/no questions of the [INV] kind, again demonstrated with *quel est le date de ton anniversaire* shown below.

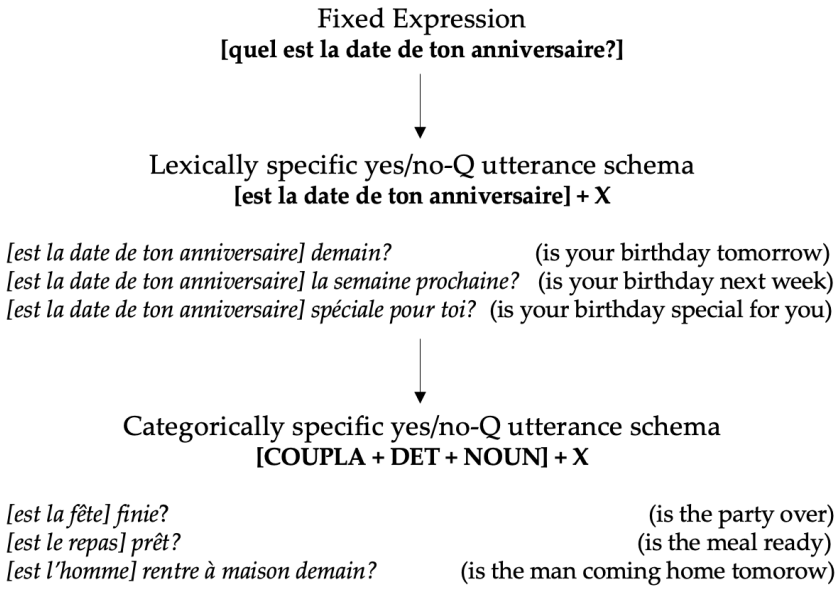

Figure 7.2: Example developmental trajectory of [INV] yes/no-question forms from FE analysis

Examples (26) and (27) below outline each FE^{wh}'s lexically and categorically specific utterance schemas for [wh + INV] questions as extractable from their schematic patterns.

(26) FE^{wh} and wh-Q utterance schemas: lexically specific

a. *comment t'appelles tu*: [comment t'appelle] + X

b. *quel âge as tu*: [quel âge as] + X

c. *où habites-tu?*: [où habites] + X

d. *quel est le date de ton anniversaire*: [quel est] + X

(27) FE^{wh} and wh-Q utterance schemas: categorically specific

a. *comment t'appelles tu*: [WH + REF PRN + REF VERB] + X

b. *quel âge as tu*: [WH + ADJ + VERB] + X

c. *où habites-tu?*: [WH + VERB] + X

d. *quel est le date de ton anniversaire*: [WH + COPULA] + X

Examples (28) - (29) show the same for yes/no [INV] question utterance schemas.

(28) FE^{wh} and yes/no-Q utterance schemas: lexically specific

a. *comment t'appelles tu*: [t'appelle tu] + X

b. *quel âge as tu*: [as tu] + X

c. *où habites-tu?*: [habites-tu] + X

d. *quel est le date de ton anniversaire*: [est le date de ton anniversaire] + X

(29) FE^{wh} and yes/no-Q utterance schemas: categorically specific

a. *comment t'appelles tu*: [REF PRN + REF VERB + SUBJ PRN] + X

b. *quel âge as tu*: [VERB + SUBJ PRN] + X

c. *où habites-tu?*: [VERB + SUBJ PRN] + X

d. *quel est le date de ton anniversaire*: [COPULA + DET + NOUN] + X

I move now to the fixed yes/no expressions ($FEs^{y/n}$). Since these are all in [DEC] word order form, they can function as wh-interrogative utterance schemas for wh-questions of the [wh IN SITU] kind, and yes/no interrogative utterance schemas for yes/no questions of the [DEC] kind, as hypothesised below in 7.3.

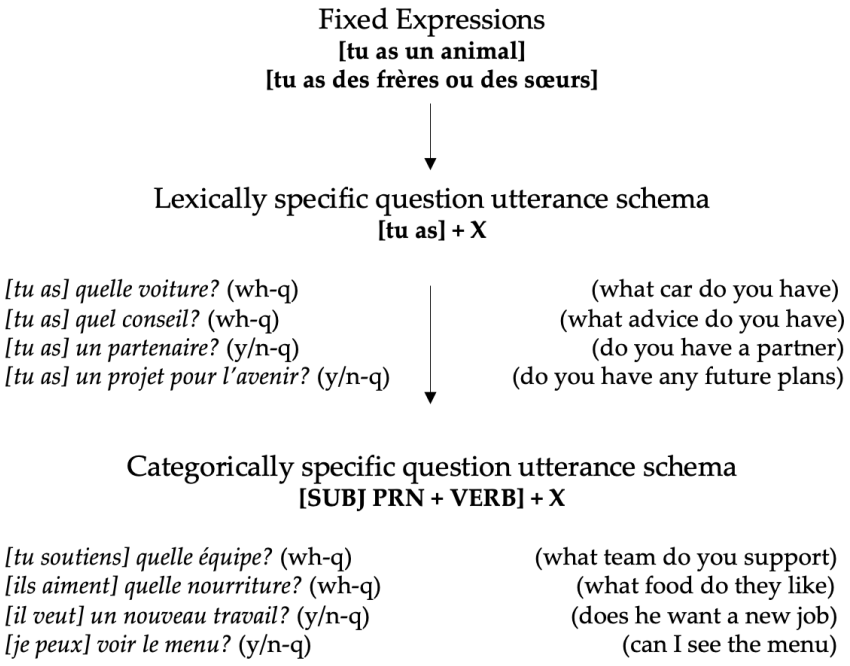

Figure 7.3: Example developmental trajectory of wh- and yes/no- question forms from FE analysis

The FEs$^{y/n}$' lexically and categorically specific utterance schemas for both [wh IN SITU] and [DEC] questions are therefore as stated below.

(30) FE$^{y/n}$ and utterance schemas: lexically specific / categorically specific

a. *tu as un animal*: [tu as] + X / [SUBJ PRN + VERB] + X

b. *tu as des frères ou des soeurs*: [tu as] + X / [SUBJ PRN + VERB] + X

The second part of the FPC results section concerns the usage-based analysis, which examines whether learners' grammatical L2 question forms at the latter stages of data collection (Rounds 5 and 6) can be linked back to utterance schemas of previously produced FEs, as outlined above. To do this, a traceback methodology was adopted across the FPC. Individual learner tables were created to document their FE productions and all L2 questions across the three rounds of data collection. Underneath each FE and L2 question, I specify their lexically (i) and categorically (ii) specific utterance schemas, as well as their fully schematic patterns (iii). I then underlined instances where those of an L2 question matched those of a previously used FE. This is exemplified in Table 7.3 below, with wh-questions produced by Learner 2 of the FPC.

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
2	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** [quel âge as] + X ***quelle est la date le ton anniversaire*** [WH + COPULA] + X	***où habites-tu***	*qu'est ce que fait la fille?* *qu'est ce que fait garçon?* *qu'est ce que la boulangerie?* *qu'est ce que la madame?* *qu'est ce que la monsieur?*	-	*quel âge a Richard?* [quel âge as] + X *où est la café?* [WH + COPULA] +X *qu'est ce que c'est?*

Table 7.3: Learner 2

Table 7.3 shows that two wh-questions are produced at Round 6 of data collection that can be traced back to an FE at Round 1. These are *quel âge a Richard* (how old is Richard), which shares the same lexically-specific schema [quel âge as] + X as the FE *quel âge as-tu* (how old are you). The other is *où est la café* (where

is the café), which shares the same more abstract categorically-specific schema [WH + COPULA] + X as the FE *quel est la date de ton anniversaire* (when is your birthday).

Chapter 8 now presents the results from both the generative and usage-based analyses of the FPC data.

Chapter 8

Results: The French Progression Corpus

This chapter provides the generative and usage-based analysis of the FPC data. It compares learners' productions of the FEs with all other accurate L2 interrogatives *outside* of the FEs across the three rounds of data collection. I refer to these as learners' Novel Interrogatives (NIs). The generative analysis examines how the computational properties within the FEs manifests in learners' NIs. On the other hand, the usage-based analysis examines the utterance schemas of learners' NIs and compares whether these correspond to the utterance schemas of previously used FEs in their production data. I begin with the generative analysis.

8.1 Production of the FEs and learners' L2 interrogative structures outside of these (NIs)

This section now examines the manifestation of learners' NIs across the three rounds of data collection to compare their structural form to those of the FEs. Section 8.1.1 first analyses learners' NIs in Round 1 of the data collection period.

8.1.1 Round 1 (ages 11/12)

At Round 1, Tables 7.1 and 7.2 in Chapter 7 showed that all 24 learners made extensive use of the FEs ($Sum = 103$, $M = 4.29$, $SD = 1.42$). Outside of the FEs at this age, learners often rely on lexical categories to express interrogatives, which are either coordinated (1) or used in isolation (2).

(1) Learner 6 Round 1

*un petit ou un grande?
a short or a tall

'are they short or tall?' (intended meaning)

(2) Learner 28 Round 1

*les yeux?
the eyes

'what colour are your eyes?' (intended meaning)

Another common strategy is linking lexical items with the verb *avoir* (have) in an attempt to form the intended question, as exemplified in (3) and (4).

(3) Learner 2 Round 1

*il age a frère?
the age has brother

'how old is your brother?' (intended meaning)

(4) Learner 4 Round 1

*la couleur a cheveux?
the colour has hair

'what colour is your hair?' (intended meaning)

Interestingly, these strategies found for NIs (i.e. single/coordinated lexical items, *avoir* (have) as a linking verb) are similar to what was found at the initial rounds of data collection for learners of the BELC. It further highlights the FEs' formulaicity for the FPC learners, as there is little evidence for knowledge of similar complex structures outside the use of these expressions.

I now move to examine all learners' NIs that were produced as full/complete utterances, rather than single/co-ordinated/linked lexical items as presented above. Table 8.1 below shows the manifestation of learners' NIs of the wh-question kind at Round 1. Raw numbers and relative percentages out of all structural possibilities (as presented in Chapter 7) are given.

wh- interrogatives						
derivation	wh IN SITU	wh + NO INV	wh + ESK	wh + CLE	wh+ INV	**TOTAL WH Q**
number	12	1	0	0	1	14
%	**86%**	**7%**	**0%**	**0**	**7%**	

Table 8.1: Learners' grammatical manifestations of wh- interrogatives outside of the FEs^{wh} (NIs) at Round 1

Table 8.1 shows that outside of the FEs^{wh}, only 14 grammatical wh-interrogatives were produced in full utterance form by 7 out of the 24 learners under analysis. Given the large number of FEs^{wh} identified in all learners' transcripts at this stage (which were all produced in the complex [wh + INV] form), this highlights the discrepancy in syntactic complexity between the FEs^{wh} and learners other NIs. Out of these NIs, 93% lack inversion (13/14) (i.e. the [wh IN SITU] and [wh + NO INV] structures) and are mostly of the [wh + IN SITU] kind (12/14). Two examples are below in (5) and (6).

(5) Learner 1 Round 1 [wh IN SITU]

il a les cheveux quelle couleur?
he has the hair what colour

'what colour is his hair?'

(6) Learner 6 Round 1 [wh IN SITU]

ça s'écrit comment?
that writes-itself how

'how do you spell it?'

Note also that some of these instances of [wh + IN SITU] are learners asking the name or age of a third person referent after producing the FEs^{wh} *comment t'appelles tu* and *quel âge as-tu* in second person [wh + INV] form, as shown in (7) and (8).

(7) Learner 12 Round 1 [wh IN SITU]

elle s'appelle comment?
she calls-herself how

'what is her name?'

(8) Learner 5 Round 1 [wh IN SITU]

il a quel âge?
he has what age

'how old is he?'

There are only two instances where the wh-word is fronted (i.e. wh-moved) at this round of data collection, one without corresponding inversion (9) and one with it (10).

(9) Learner 9 Round 1 [wh + NO INV]

comment ça s'écrit?
how that writes-itself

'how do you spell it?'

(10) Learner 28 Round 1 [wh + INV]

comment s'appellent- ils?
how call-themselves they

'what are their names?'

Important to note, also, is that within the NIs of the [wh IN SITU] kind, 83% (10/12) of these feature personal pronouns in combination with the verb *avoir* (have), such as the examples below.

(11) Learner 1 Round 1 [wh IN SITU]

il a les cheveux quelle couleur?
he has the hair what colour

'what colour is his hair?'

(12) Learner 5 Round 1 [wh IN SITU]

il a les yeux comment?
he has the eyes how

'how are his eyes?'

As outlined in Chapter 7, it is possible that utterances such as these are lexical in nature, and that learners are reproducing rote learned [pronoun + verb] combinations along with the wh-word/lexical object in isolation. For example, the utterance below in (13) could be a full CP projection where the [uWH] and the [uT] features on C lack the EPP property, or could equally be produced via memorising the conjugation [*tu as*] (you have) and combining this with the object [*les cheveux*] (hair) and wh-word [*comment*] (how).

(13) Learner 12 Round 1 [wh IN SITU]

tu as les cheveux comment
you have the hair how

'what style is your hair?'

It is likely utterances such as these are lexical in nature at Round 1 of data collection, given the other observed strategies for question formation at this stage and general lack of evidence for knowledge of complex derivations in the L2 outside of the FEs.

Moving now to NIs of the yes/no question kind, Table 8.2 shows the distribution of learners' realisations as raw numbers and relative percentages.

yes/ no interrogatives			
derivation	INV	DEC	TOTAL Y/N
number	0	20	20
%	**0%**	**100%**	

Table 8.2: Learners' grammatical manifestations of yes/no interrogatives outside of the FEs$^{y/n}$ (NIs) at Round 1

Table 8.2 shows that outside of the FEs$^{y/n}$, only 20 other grammatical yes/no questions are produced in full structures by 10 out of the 24 learners under analysis. Like the FEs$^{y/n}$, these are all in [DEC] rather than [INV] form. Some examples can be seen below.

(14) Learner 19 Round 1 [DEC]

tu aimes la animal?
you like the animal

'do you like your pet?'

(15) Learner 20 Round 1 [DEC]

il a le cheveux longs?
he has the hair long

'has he got long hair?'

(16) Learner 5 Round 1 [DEC]

il est petit ou grand?
he is small or big

'is he short or tall?'

As found with the NIs of the wh-question kind, the majority of yes/no question NIs (18/20 = 90%) contain subject pronouns in combination with the verbs *avoir* (have) and *être* (be). This could imply that these utterances are lexical in nature, produced in part via a rote memorisation strategy as discussed above for the [wh IN SITU] structures. When comparing these yes/no questions with the high proportion of *avoir* [wh IN SITU] wh-questions, along with the other observed lexical/ungrammatical question formation strategies, it is likely that learners are relying on lexical rather than functional categories at this initial stage of data collection.

To summarise, through comparison of learners' fluent productions of the FEs and all other question formation strategies at Round 1, we can strongly infer that learners are producing the FEs holistically, rather than generating these online. Outside of the FEs at Round 1 of data collection, one plausible interpretation is that learners are seen to rely heavily on lexical categories to express interrogatives, both wh- and yes/no. This, as was found with the learners of the BELC, is perhaps indicative of an incremental development of L2 phrase structure. At the initial state, learners seem to lack sufficient knowledge of functional categories T and C, forcing them to rely on lexical means of expression (NPs, rote-learned conjugations etc.). This would be further support for a Weak Continuity view of the initial state (Vainikka, 1994; Vainikka and Young-Scholten, 1998; Hawkins, 2001; Bhatt and Hancin-Bhatt, 2002).

Alternatively, if the identified grammatical interrogatives are taken to be genuine instances of computational derivation, we see that learners overwhelmingly opt for

the least complex syntactic derivations; [wh + IN SITU] structures for wh-questions (86%) (which lack T-C movement and wh-movement) and [DEC] structures for yes/no questions (100%) (which also lack T-C movement). Although this would imply that some learners are projecting functional categories T and C at these initial stages, these categories are likely still underdeveloped, as [wh IN SITU] and [DEC] structures are the least derivationally complex options which are available in the target language. Section 8.1.2 now adopts the same procedure for learners' NIs at the end of the data collection period at Rounds 5 and 6, to examine how these manifestations have developed over a two year period of classroom instruction and interaction with the FEs.

8.1.2 Rounds 5 and 6 (ages 12/13)

Two years later at Rounds 5 and 6, the FEs^{wh} are produced far less than in Round 1 ($Sum = 25$, $M = 1.041$, $SD = 1.20$) and by only 13 out of the 24 learners under analysis. These productions are almost always overextensions, as exemplified in (17) below.

(17) Learner 20 Round 5

*comment t'appelles tu un garçon?
how call-yourself you the boy

'what is his name?' (intended meaning)

We see no productions of the FEs$^{y/n}$ at this stage. Table 8.3 below show the manifestation of learners' NIs of the wh-question kind.

wh- interrogatives						
derivation	wh IN SITU	wh + NO INV	wh + ESK	wh + CLE	wh+ INV	TOTAL WH
number	12	7	32	0	51	102
%	**12%**	**7%**	**31%**	**0%**	**50%**	

Table 8.3: Learners' grammatical manifestations of wh- interrogatives outside of the FEs^{wh} (NIs) at Rounds 5 and 6

At Rounds 5 and 6, 14 out of the 24 learners now produce a total of 102 grammatical wh-questions outside of the FEs^{wh}. This, along with their lower reliance on the FEs^{wh}, suggests a slight overall increase in L2 competence from the first round of data collection. There are still, however, some instances where learners rely on single/co-ordinated NPs to express interrogatives, as was found in Round 1. Examples of such are below in (18) and (19).

(18) Learner 4 Round 5

nom la madame?
name the woman

'what is the woman's name?' (intended meaning)

(19) Learner 15 Round 5

grande ou petit?
big or small

'are they tall or short?' (intended meaning)

Table 8.3 also shows that the majority of NIs at these latter stages of data collection are now in [wh + INV] form, some examples of which are below.

(20) Learner 2 Round 5 [wh + INV]

quel âge a Richard?
what age has Richard

'how old is Richard?'

(21) Learner 7 Round 6 [wh + INV]

où habite le garçon?
where lives the boy

'where does the boy live?'

(22) Learner 24 Round 5 [wh + INV]

où est le manger en Belleville?
where is the restaurant in Belleville

'where is the restaurant in Belleville?'

(23) Learner 23 Round 5 [wh + INV]

a quelle heure est diner?
at what time is dinner

'what time is dinner?'

It is worth noting that 62% of these [wh + INV] NIs (32/51) are with *où est* (where is), some examples of which can be seen below.

(24) Learner 5 Round 5 [wh + INV]

où est la fille?
where is the girl

'where is the girl?'

(25) Learner 20 Round 5 [wh + INV]

où est le homme?
where is the man

'where is the man?'

(26) Learner 6 Round 6 [wh + INV]

où est le dejeuner
where is the breakfast

'where is the breakfast?'

Past studies of both L1 and L2 acquisition of French have found that learners' ini-

tial [wh + INV] interrogatives feature this combination (Grondin and White, 1996; Prévost, 2009), which raises the question as to whether or not this sequence is fully analysed. It could be that learners are using the combination *où est* as a holistic chunk, and clipping this together with relevant NP referents such as [*la fille*], [*le homme*] and [*le dejeuner*] as seen in the examples above.

Table 8.3 also shows that the next most common manifestation of wh-interrogatives is the [wh + ESK] structure, which constitutes 31% of learners' total interrogatives (32/102). These 32 [wh + ESK] structures are observed from 6 learners only, and over 90% of these are with the wh-word *que* (what), in the sequence *qu'est ce que* (what is it that). As for *où est* above, it is ambiguous as to whether *qu'est ce que* is fully analysed or not, as this sequence has also been found to be the first to feature in emerging [wh + ESK] structures for both L1 and L2 speakers of French (Plunkett, 1999). Some examples of the [wh + ESK] structures observed at Rounds 5 and 6 are given below. Note that (28) and (29) are both examples of learners using the *qu'est ce que* sequence.

(27) Learner 9 Round 6 [wh + ESK]

où est ce qu' on peut manger le déjeneur?
where is it that one can eat the breakfast

'where can you eat breakfast?'

(28) Learner 28 Round 5 [wh + ESK]

qu' est ce qu' elle fait?
what is it that she does

'what does she do?'

(29) Learner 2 Round 6 [wh + ESK]

qu' est ce que fait la fille?
what is it that she does the girl

'what does the girl do?'

It could indeed be the case, like the FEs identified in Round 1 of data collection, and hypothesised for L2 question forms of the [wh IN SITU] kind which feature conjugations with *avoir*, that a large proportion of learners' NIs of the wh-question kind are still lexical and/or schematic in nature. The sequences *où est* and *qu'est ce-que* could be holistically taught fixed question expressions in their own right, that learners have derived from their classroom input and subsequently committed to memory, rather than instances of computational derivation.

Moving now to NIs of the yes/no question kind, 19 out of the 24 learners under analysis are shown to produce a total of 92 grammatical yes/no questions. Table 8.4 shows their manifestation.

yes/ no interrogatives			
derivation	INV	DEC	**TOTAL Y/N**
number	10	82	92
%	**11%**	**89%**	

Table 8.4: Learners' grammatical manifestations of yes/no interrogatives outside of the FEs$^{y/n}$ (NIs) at Rounds 5 and 6

Table 8.4 shows that, like in Round 1, learners' NIs of the yes/no question kind are still produced overwhelmingly in declarative [DEC] form, as exemplified by (30) and (31) below.

(30) Learner 19 Round 5 [DEC]

le fille est prés le boulangerie?
the girl is near the bakery

'is the girl near the bakery?'

(31) Learner 3 Round 5 [DEC]

elle est grand?
she is tall

'is she tall?'

Only 11% of these NIs (10/92) show inversion [INV], as exemplified by (32) and (33).

(32) Learner 7 Round 5 [INV]

est la fille assez grand?
is the girl fairly tall

'is the girl fairly tall?'

(33) Learner 28 Round 6 [INV]

avez- vous visitez le cinema?
have you visited the cinema

'have you visited the cinema?'

Of the yes/no questions in [DEC] form, compared to Round 1, a far smaller proportion of these feature personal pronouns used in conjugations with *avoir* (have) and *être* (be) (41/82 = 50%). Further, many of those learners who do produce these combinations, such as *il a* (he/she/it has) in (34), also produce [DEC] yes/no questions with other lexical verbs displaying overt morphological inflection for person/tense agreement. Learner 3 is an example, who also produces *il adore* (he loves) and *il porte* (he wears) as shown below, which indicate a relevant TP projection.

(34) Learner 3 Rounds 5 and 6 [DEC]

il a les cheveux longs
he has the hair long

'has he got long hair?'

il adore foot?
he loves football

'does he love football?'

il porte les jeans?
he wears jeans

'does he wear jeans?'

This lower proportion of yes/no questions in [DEC] form featuring *avoir* and *être* compared to Round 1, along with the higher proportion of [DEC] yes/no questions with other verbs inflected for tense and person agreement, could indicate an emerging knowledge of functional categories T and C at these later rounds of data collection. This would mean that these learners are beginning to move away from a lexically-dominant strategy of question formation as observed in Round 1.

8.1.3 Summary

At Round 1 of the data collection period, the generative analysis has found that all learners produced the FEs frequently in [wh + INV] and [DEC] form. Outside of these expressions, it is likely lexical categories that are largely relied upon to express interrogative force. This includes single items with assumed rising intonation (35) lexical items ungrammatically linked with the verb *avoir* (36) and potentially rote learned verb conjugations with subject pronouns in [wh IN SITU] and [DEC] structures (37) and (38)).

(35) Learner 28 Round 1

*les yeux?
the eyes

'what colour are your eyes?' (intended meaning)

(36) Learner 2 Round 1

*il age a frère?
he age has brother

'how old is your brother?' (intended meaning)

(37) Learner 12 Round 1 [wh IN SITU]

tu as les cheveux comment?
you have the hair how

'what style is your hair?'

(38) Learner 5 Round 1 [DEC]

il est petit ou grand?
he is small or big

'is he short or tall?'

If the small number of grammatical interrogatives including the verbs *avoir* and *être* with subject pronouns are instead taken to be products of computational derivation, this means that learners opt for the least complex derivations permitted in the target language for both wh- and yes/no questions ([wh IN SITU], [DEC]). This is *despite* these derivational options being structurally dissimilar to the learners' L1 (English). Similar to what we find with learners of the BELC, either scenario suggests an underdeveloped knowledge of L2 functional categories T and C at this initial stage of data collection. This is further indicative that the FEs are likely being produced holistically rather than via computational generation at this stage, as they are syntactically 'advanced' compared to learners' other interlanguage productions in similar environments.

Two years later, at Rounds 5 and 6 of the data collection period, the FEs^{wh} are produced far less, and are nearly always overextended, as in the example below (39).

(39) Learner 23 Round 5

*comment t'appelles tu le petite fille
how call-yourself you the little girl

'what is the little girl's name?' (intended meaning)

No learner is shown to produce the FEs$^{y/n}$ at these stages. Outside of these expressions, the analysis of NIs shows that learners now seem to prefer the [wh + INV] structure for wh-questions overall. This is structurally similar to the FEs^{wh} and the most derivationally complex option available in the target language. However, 62% of these [wh +INV] interrogatives feature the combination *où est* (where is), which raises the question of whether this sequence is fully analysed or not. Next most frequent is the [wh + ESK] wh-interrogative structure, 90% of which occur with the wh-word *que* (what) in the sequence *qu'est ce que*. Like with *où est*, it could be that this sequence is also unanalysed at this stage. It is therefore a possibility that the sequences *où est* and *qu'est ce-que* are also functioning as holistically taught fixed question expressions similar to the identified FEs, rather than being overt evidence for their associated underlying computational properties. For yes/no questions at Rounds 5 and 6, learners continue to realise these overwhelmingly in [DEC] form (89%). Far less a proportion of these contain conjugations of *avoir* and *être* used with subject personal pronouns, and many of them contain other lexical verbs with overt evidence for functional category T.

What is consistent across both wh- and yes/no interrogatives at the end of the data collection period, then, is that learners most frequently realise L2 interrogatives via structures that are exemplified by the FEs (wh-questions- [wh + INV], yes/no questions- [DEC]), despite these structures' increased derivational complexity ([wh + INV]) or dissimilarity from the L1 ([DEC]). However, a large proportion of these utterances contain linguistic elements that are ambiguous as to whether they are

products of analysed computational derivation or unanalysed holistic retrieval.

The section below now adopts a schematic usage-based traceback analysis to see whether the utterance schemas of learners' NIs correspond to those of previously produced FEs in learners' production data.

8.2 Production of the FEs and later knowledge of their schematic constructions

This Section now analyses the French production data under a usage-based schematic model. It analyses the utterance schemas of learners' wh-questions and yes/no questions outside of the FEs at the later rounds of data collection, to see how these correspond to utterance schemas of previously used FEs at the initial round of data collection. I begin with examining learners' NIs of the wh- question kind.

8.2.1 Wh-questions

Chapter 7 outlined how the FEs can function as utterance schemas for wh-interrogatives of the [wh + INV] kind (FEs^{wh}) and [wh IN SITU] kind ($FEs^{y/n}$). Out of the 102 NIs produced by 14 of the learners under analysis, 34 of these (33.33%) produced by 9 learners can be traced back to utterance schemas of previously used FEs. As adopted for the usage-based schematic analysis of the BELC learners, Tables 8.5 - 8.13 below capture these 9 learners' NIs. They show the FEs produced at Round 1 of data collection, and any FEs and NIs of the wh-question kind produced at the final rounds of data collection (Rounds 5 and 6). I have underlined instances where utterance schemas (either lexically specific or categorically specific) of wh-question NIs match those of a previously produced FE. I have used a dash '-' to indicate where learners were not shown to produce any FEs or NIs at a certain age.

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
2	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** [quel âge as] + X ***quelle est la date le ton anniversaire*** [WH + COPULA] + X	***où habites-tu***	*qu'est ce que fait la fille?* *qu'est ce que fait garçon?* *qu'est ce que la boulangerie?* *qu'est ce que la madame?* *qu'est ce que la monsieur?*	-	*quel âge a Richard?* [quel âge as] + X *où est la café?* [WH + COPULA] +X *qu'est ce que c'est?*

Table 8.5: Learner 2

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
3	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date le mon anniversaire*** [WH + COPULA] + X ***tu as un animal à la maison*** ***tu as des frères et des sœurs***	**elles* ***comment t'appelles tu*** ****comment elle t'appelles tu*** ****comment** il **t'appelles tu***	-	-	*comment est activité a Belleville?* [WH + COPULA] + X *comment est activité a Belleville?* [WH + COPULA] + X

Table 8.6: Learner 3

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
7	***tu as quel âge*** ****comment tu appelles*** ***où habites-tu*** [où habites] + X [WH + VERB] + X ***tu as un animal*** ***tu as un frère ou un sœur***	****comment t'appelle** le garçon* ****comment t'appelle** la fille*	*où habite la garçon?* *où habite la fille?* *où habite le monsieur?* *où habite la femme?* [où habite] + X [WH + VERB] + X	-	*c'est quel jour?*

Table 8.7: Learner 7

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
9	***comment t'appelles tu*** [comment t'appelles] + X [WH + REF VERB] + X [WH + REF VERB + SUBJ PRN] ***où habites-tu*** ***quel âge as-tu*** [WH + ADJ + VERB] + X [WH + ADJ + VERB + SUBJ PRN] ***quelle est la date de ton anniversaire*** [WH + COPULA] + X ***tu as un animal*** [SUBJ PRN + VERB] + X ***tu as un frère ou un sœur*** [SUBJ PRN + VERB] + X	****comment t'appelle*** *t' elle* ****comment t'appelle t'*** *il*	*où est l'homme?* *où est la femme?* *où est la garçon?* *où est la fille?* [WH + COPULA] + X *qu'est ce qu'il a faire?* *qu'est ce qu'elle a faire?* *qu'est ce qu'il a faire?* *qu'est ce qu'il porte?* *qu'est ce qu'elle a faire?* *qu'est ce qu'elle porte?* *il est comment?* *la femme elle est comment?* *il est comment?* *tu veux quel jour?* [SUBJ PRN + VERB] + X	-	*comment s'appelle le garçon?* [comment t'appelles] + X [WH + REF VERB] + X *comment s'appelle elle?* [comment t'appelles] + X [WH + REF VERB] + X [WH + REF VERB + SUBJ PRN] *quel jour as tu?* [WH + ADJ + VERB] + X [WH + ADJ + VERB + SUBJ PRN] *qu'est ce qu'il y a le monstre* *ou est ce qu'on peut mange le déjeuner?* *qu'est ce qu'on peut faire a Belleville?* *ou est ce que je rendez vous?* *qu'est ce qu'on peut faire le matin?* *ou est ce qu'on peut mange le déjeuner?* *qu'est ce qu'on peut faire l'aprés midi?* *le soir qu'est ce qu'on peut faire?* *a quelle heure tu veux un rendez vous* *comment ca s'ecrit* *comment ca s'ecrit*

Table 8.8: Learner 9

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
20	***comment t'appelles tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** [WH + COPULA] + X	****comment t'appelles tu*** *(third person reference)* ****comment t'appelles tu*** *euh (.) un fille* ****comment t'appelles tu*** *e un garçon* ****comment t'appelles tu*** *et l'homme*	*où est un garçon?* *où est la fille?* *où est le homme?* *où est le madame?* [WH + COPULA] + X *que fait un fille?* *que fait madame?*		

Table 8.9: Learner 20

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
22	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** [WH + COPULA] + X	-	*où est le dejeuner?* [WH + COPULA] + X	****quel âge as-tu*** *(third person reference)*	*où est la fille?* [WH + COPULA] + X

Table 8.10: Learner 22

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
23	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** [WH + AD] + VERB] + X ***quelle est la date de ton anniversaire***	****comment t'appelles tu*** *le (.) petit garçon* ****comment t'appelles tu*** *le petit fille* | ****comment t'appelles tu***	*qu'est ce que c'est faire le petit garçon?* *qu'est ce que tu manges?* *qu'est ce que c'est faire le petit fille?* *qu'est ce que c'est faire?* *qu'est ce que porte un Jean?* *quel âge a Paul?* *quel âge a Anne?* *quel âge a Jean?* [WH + AD] + VERB] + X	-	*a quelle heure est dénuer?*

Table 8.11: Learner 23

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
24	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** [WH + COPULA] + X ****des frères de tu des sœurs*** ***tu as les animaux***	-	*qu'est ce que elle derrière?* *qu'est ce que elle fait?* *où est la garçon?* *où est la fille?* *où est la mademoiselle?* [WH + COPULA] + X *la garçon il est comment?* *elle est comment?* *la mademoiselle elle est comment?* *la monsieur il est comment?*	***comment t'appelles tu***	-

Table 8.12: Learner 24

#	Round 1 FE	Round 5 FE	Round 5 wh Q	Round 6 FE	Round 6 wh Q
28	***où habites-tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** [WH + COPULA] + X	-	*où est le meet?* *où est le garçon?* *où est la fille?* *où est la monsieur?* [WH + COPULA] + X *qu'est ce qu'elle fait?* *qu'est ce qu'il fait?*	-	*qu'est ce que c'est le piscine a Belleville?* *qu'est ce que tu aimes le cinéma?* *où est le manger en Belleville?* *où est le manger dîner?* [WH + COPULA] + X

Table 8.13: Learner 28

The tables above show that, unsurprisingly, those wh-question NIs that share the same utterance schemas as a previously produced FE are largely of the [wh + INV] structure (33/34), as are all of the FEs^{wh}. We only find one instance of an L2 interrogative in [wh IN SITU] form which matches the utterance schema of a

previously produced FE$^{y/n}$; Learner 9 produces the [wh IN SITU] interrogative *tu veux quel jour* (what day do you want?) which shares the same [SUBJ PRN + VERB] + X utterance schema as the previously used FEs$^{y/n}$ *tu as un animal* (do you have a pet) and *tu as un frère ou un soeur* (do you have a brother or sister).

The most productive wh- question utterance schema, as was found for learners of the BELC, is [WH + COPULA] + X, extractable from the FE^{wh} *quelle est le date de ton anniversaire*, which accounts for 67% (22/33) of wh-interrogatives that follow use of this FE^{wh}. The remaining wh-interrogatives are split between [WH + ADJ + VERB] + X (5/33), [WH + VERB] + X (4/33) and [WH + REF VERB] + X structures.

However, as mentioned previously in the generative analysis of the FPC data (Section 8.1), it is likely that *où est* (where is) is functioning as a formulaic chunk separate from the FEs, as this combination features in 91% (20/22) of the [WH + COPULA] + X interrogatives identified above. It is therefore unclear whether these wh-interrogatives are instances of utterance schema extraction and generalisation as seeded from the FE^{wh} *quelle est le date de ton anniversaire*, or whether learners have memorised the lexically-specific sequence *où est* independently, and are using this as a productive fixed question expression in its own right. Section 8.2.2 now analyses learners' NIs of the yes/no question kind.

8.2.2 Yes/no questions

Chapter 7 highlighted how the FEs^{wh} can function as utterance schemas for yes/no questions of the [INV] kind, as all exemplify subject verb inversion. The FEs$^{y/n}$ can also function as utterance schemas for yes/no questions of the [DEC] kind, as they all exemplify a declarative word order. Out of 104 yes/no question NIs identified in 14 learners' transcripts, 23.08% of these (24/104) can be traced back to utterance schemas of previously used FEs. These NIs are found across 8 learners' transcripts

only, which are exemplified in Tables 8.14 - 8.21 below. The tables show the FEs produced at Round 1 of data collection, and any FEs and yes/no interrogatives produced at the final rounds of data collection (Rounds 5 and 6). I have underlined instances where utterance schemas (either lexically specific or categorically specific) of yes/no-question NIs match those of a previously produced FE. I have used a dash '-' to indicate where learners were not shown to produce any FEs or NIs at a certain age.

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
3	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date le mon anniversaire*** [COPULA + DET + NOUN] + X ***tu as un animal à la maison*** [SUB] PRN + VERB] + X ***tu as des frères et des sœurs*** [SUB] PRN + VERB] + X	**elles* ***comment t'appelles tu*** ****comment elle t'appelles tu*** ****comment*** *il* ***t'appelles tu***	*elle va l'ecole?* *elle a au stade des foot?* *elle a les cheveux euh longs?* *il va au café?* *il a les cheveux (.) longs ou est courts?* *il adore foot?* *il adore le livre?* *il a cheveux courts ou le longs?* [SUB] PRN + VERB] + X *elle est grande?* *il est porte les (.) jeans?* *il est petit ou il est grand?* *il est une table ou le chaise?*	-	*est la (.) activité le vingt neuf?* [COPULA + DET + NOUN] + X

Table 8.14: Learner 5

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
5	***comment t'appelles tu*** ****il a quel âge*** [SUB] PRN + VERB] + X ****il habite où*** [SUB] PRN + VERB] + X ****il a un animal à la maison*** [SUB] PRN + VERB] + X		*elle a cheveux longs ou courts?* *elle a un pantalon ou un jupe?* *elle a porte des lunettes?* *elle porte un chemise?* *il a les cheveux courts ou longs?* *elle a les cheveux (.) courts ou longs?* [SUB] PRN + VERB] + X *elle est petit ou grande?* *il est petit ou grande?* *il est grande ou petit?*	-	-

Table 8.15: Learner 5

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
6	*comment t'appelles tu* *où habites-tu* *quel âge as-tu* *tu as un animal* [SUBJ PRN + VERB] + X *tu as un frère ou un sœur* [SUBJ PRN + VERB] + X	-	*il y a le monsieur?* *il est grosse?* *elle fait des courses?* [SUBJ PRN + VERB] + X	-	-

Table 8.16: Learner 6

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
9	*comment t'appelles tu* *où habites-tu* *quel âge as-tu* *quelle est la date de ton anniversaire* *tu as un animal* [SUBJ PRN + VERB] + X *tu as un frère ou un sœur* [SUBJ PRN + VERB] + X	**comment t'appelle t' elle* **comment t'appelle t' il*		-	*tu veux je répète?* [SUBJ PRN + VERB] + X *c'est possible (.) pour il?* *Richard aller avec moi, oui?*

Table 8.17: Learner 9

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
12	*comment t'appelles tu* *où habites-tu* *quel âge as-tu* *quelle est la date de ton anniversaire* *tu as un animal* [SUBJ PRN + VERB] + X	-	*il y a (.) petit?* *elle est (.) petit grand?* *elle fait le vélo?* [SUBJ PRN + VERB] + X	*quel âge as tu*	

Table 8.18: Learner 12

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
23	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** ***tu as des frères et des sœurs*** [SUBJ PRN + VERB] + X ***u as un animal à la maison*** [SUBJ PRN + VERB] + X	****comment t'appelles tu*** *le (.) petit garçon* ****comment t'appelles tu*** *le petit fille |* ****comment t'appelles tu***		-	*c'est tout?* *tu aimes richard?* *il a le cheveux bleus (.) ou vertes?* [SUBJ PRN + VERB] + X

Table 8.19: Learner 23

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
24	***comment t'appelles tu*** ***où habites-tu*** ***quel âge as-tu*** ***quelle est la date de ton anniversaire*** [COPULA + DET + NOUN] + X ****des frères de tu des sœurs*** ***tu as les animaux***	-	*est le garçon assis (.) la chaise?* *est la chaise (.) um a près des la café?* *est la fille a près de la entrance?* [COPULA + DET + NOUN] + X *la fille est asseyez?* *il est entre la voiture?* *il s'appelle la garçon?*	***comment t'appelles tu***	*ils sont la télévision?*

Table 8.20: Learner 24

#	Round 1 FE	Round 5 FE	Round 5 y/n Q	Round 6 FE	Round 6 y/n Q
26	***tu as quel âge*** [SUBJ PRN + VERB] + X ***tu as des frères ou des sœurs*** [SUBJ PRN + VERB] + X ****tu un un est animal***	****comment tu t'appelles le (.) garçon***	*il est petite ou la grande?* *il est petite?* *elle a la (.) courts a le longs?* [SUBJ PRN + VERB] + X		

Table 8.21: Learner 26

The tables above show that the majority of these instances are yes/no questions of the [DEC] structure (20/24) which match the [SUBJ PRN + VERB] + X schema of the FEs$^{y/n}$ *tu as des frères ou des soures* and *tu as un animal*. Only a handful are of the [INV] structure [4/24] which all show the [COPULA + DET + NOUN] + X utterance schema of the FE^{wh} *quelle [est la date de ton anniversaire]*.

An interesting case is Learner 5, who produces the FEs erroneously at Round 1 of data collection by replacing the second person subject pronoun *tu* (you) with *il* (he) in the examples below (40).[1]. All these erroneous productions display the interrogative utterance schema [SUBJ PRN + VERB] + X.

(40) Learner 5 Round 1 [SUBJ PRN + VERB] + X

*il a quel âge?
he has what age

'how old are you?' (intended meaning)

*il habite où?
he lives where

'where do you live?' (intended meaning)

*il a un animal à la maison?
he has a pet at the house

'do you have a pet at home' (intended meaning)

At Round 5, we see that this learner produces 6 grammatical yes/no questions with the same [SUBJ PRN + VERB] + X utterance schema as these FE productions, repeated from the tables and shown below.

(41) Learner 5 Round 5 [SUBJ PRN + VERB] + X

elle a cheveux longs ou courts?
she has hair long or short

'Does she have short or long hair?'

elle a un pantalon ou un jupe?
she has trousers or a skirt

[1]The FEs are ungrammatical in the sense that they were intended with second person reference, as was required by the interview task.

'Is she wearing trousers or a skirt?'

elle porte des lunettes?
she wears glasses

'Is she wearing glasses?'

elle porte un chemise?
she wears a shirt

'Is she wearing a shirt?'

il a les cheveux courts ou longs?
he has the hairs short or long

'Does he have short or long hair?'

elle a les cheveux courts ou longs
she has the hairs short or long

'Does she have long or short hair?'

It could therefore be that memorising the FEs^{wh} in this erroneous way has actually facilitated Learner 5's later production of yes/no questions in [DEC] form.

8.2.3 Summary

To summarise, the usage-based schematic analysis has revealed that 12 out of the 24 learners under analysis produce Novel Interrogatives at the final rounds of data collection that share the same utterance schemas of FEs that appear ontogenetically in their production data at the first round of data collection. These productions constitute 33.33% of the total NIs of the wh-question kind and 23.08% of total NIs of the yes/no question kind produced at Rounds 5 and 6. This equates to 28.15% of total NIs observed at these stages.

Matching wh-questions are found in 9 learners' transcripts only. These are over-

whelming of the [wh + INV] type, and feature the [WH + COPULA] + X utterance schema, as extractable from the FE^{wh} *quelle est le date de ton anniversaire*. However, it is unclear if these are genuine instances of utterance schema extraction and generalisation based on use of this FE^{wh}, as 91% of the these [WH + COPULA] + X wh-questions are instances of *où est* (where is). It is therefore a possibility that this combination is a lexically-specific schema in its own right, and being used independently of FE^{wh} analysis. Matching yes/no questions are found in 8 learners' transcripts. These are overwhelmingly of the [DEC] type, featuring the [SUBJ PRN + VERB] + X utterance schema of the FEs$^{y/n}$ *tu as un animal* and *tu as des fères ou des soeurs?*.

Chapter 9 now brings together the results of the BELC and FPC analyses in the Discussion.

Chapter 9

Discussion

This chapter presents the discussion. It begins with a summary of the results derived from both analyses in Section 9.1. Section 9.2 then considers the status of 'syntactically advanced' FEs at the initial state. Here, I distinguish between implicit and explicit language phenomena, and discuss the communicative benefits that come with the FEs' holistic retrieval and production. Section 9.3 then suggests how learners' interaction with FEs could support the acquisition of the expressions' underlying syntactic properties. Here, I draw on processing models of L2 acquisition and posit the FEs as scaffolding devices which can be used to 'activate' their associated L2 feature specifications. Section 9.4 considers the role of schematic learning in the observed developmental trends. This section first highlights how there is development that goes beyond utterance schema extraction and generalisation (9.4.1), and then concentrates on the development that can be accounted for by this learning strategy (9.4.2). It then offers an explanation as to how schematic learning can facilitate the acquisition of underlying syntactic properties more generally (9.4.3). Finally, Section 9.5 concludes the discussion.

9.1 Summary of results

The results of the BELC and FPC analyses reveal that both sets of learners make frequent use of memorised fixed question expressions upon appropriate contextual

cues. When initially produced by learners, these expressions show a syntactic complexity that exceeds learners' observed interlanguage competence at that particular stage of data collection. Outside of the FEs at these stages, both sets of learners largely rely on their L1 and lexical categories of the L2, and opt for the least complex derivations in similar functional environments where permitted in the target language. A common developmental trajectory is observed across both longitudinal corpora which supports a Weak Continuity view of the initial state. Over the course of the data collection periods, learners' L2 utterances outside of the FEs show a rise in syntactic complexity, with more emerging evidence for knowledge of L2 functional categories (T and C).

In the BELC, correlations are found between early and frequent FE use and a better knowledge of syntactic computations related to these functional categories, for which the fixed wh-expressions exemplify. This relationship could be causal, as learners differ significantly in L2 accuracy rates of these related computational derivations depending on their use of these expressions. In the FPC, we find that all learners produce the FEs frequently at the initial stages of learning, but tend to prefer less derivationally complex structures when producing L2 interrogatives outside of these FEs. At the end of the two-year data collection period, we see that the majority of learners' L2 interrogatives mirror the more complex structures of the FEs, which suggests an influence of these expressions on learners' knowledge of similar syntactic derivations. This influence persists despite these structures being derivationally complex [wh + INV] or dissimilar from the L1 [DEC].

The analyses of both corpora therefore suggest that learners' use of fixed question forms is influential on their acquisition of associated syntactic properties. A closer examination of this development also reveals evidence of schematic learning. 53 % of the BELC learners' and 28% of the FPC learners' L2 interrogatives can be traced back to utterance schemas of FEs present in their production data ontogenetically.

In the BELC, it is predominantly those learners who show early FE use that go on to produce L2 interrogatives with matching utterance schemas. Similarly, it is these learners that show more evidence for knowledge of the expressions' underlying syntactic properties more generally.

In the FPC, there is also evidence of lexically-specific schemas outside of the originally identified FEs. The combinations *où est* (where is) and *qu'est ce que* (what is it that) feature in a large proportion of learners' L2 interrogatives of the [wh + INV] (62%) and [wh + ESK] (90%) kind. If these combinations were to be analysed as utterance schemas in their own right, it would imply that L2 development based on this usage-based schematic strategy is more pronounced in the FPC than in the BELC. This could be because the FPC provides a more concentrated picture of learners' L2 behaviour at the initial state, covering the first two years of their language learning journey as opposed to the BELC, which follows learners for 7 years. In the FPC, such learning mechanisms are likely used to compensate for a deficit in overall L2 competence.

Together, the picture that emerges from the analyses of the longitudinal corpora is one where early use of 'syntactically complex' memorised fixed expressions seems to influence the rate in which learners pass through a Weak Continuity developmental trajectory. Schematic learning strategies likely contribute to this process. The sections below now discuss the interplay of these concepts. They attempt to demonstrate how the application of both generative and usage-based analyses can make better sense of the trends observed in the longitudinal production data than either model can do independently. More broadly, I argue that crossing research paradigms in this way is useful for understanding the role of input, usage and learning strategies in the acquisition of formal linguistic properties, and highlight the significant role that memorised fixed expressions may play in this process.

9.2 Fixed question expressions at the initial state: Production shortcuts

It is now a fairly well-established concept in studies of SLA that learners at early stages of development are able to produce strings of language that go beyond their general L2 competence (Hanania and Gradman, 1977; Peters, 1977; Myles, 2004; Wray and Fitzpatrick, 2008; Bardovi-Harlig and Stringer, 2017). This phenomena is likely indicative of an expressions' formulaicity (Bardovi-Harlig, 2009). Such a strategy offers clear benefits in both the language learning classroom and the target-language community (Jeremias, 1982; Schmitt, 2010), as it enables beginner learners to communicate minimally in predictable situations (Towell, 2012). The lengthy and complex process of acquiring morphosyntax and processing skills often compels learners to take 'shortcuts' at early stages of acquisition, which is why they often resort to FEs to engage in 'predictable routines in which they have to play a predetermined part' (Myles, 2004, p. 155).

Indeed, the analyses of the longitudinal corpora show that both sets of learners make use of this strategy to compensate for their impoverished interlanguages at the initial state, dominated outside of the FEs by L1 utterances and L2 lexical utterances only. The similarity between the fixed question forms identified for both L2 English and L2 French learners highlights the prototypicality and conventionality of these expressions. These are all closely tied to specific communicative contexts and taught to permit interaction at initial stages of classroom learning (Towell, 2012). This perhaps also reflects the nature of the communicative classroom environment in both learning contexts, where teachings methods favoured the use of role-play activities involving the repetition of learned sequences (e.g., Mitchell and Martin 1997). Therefore, we can presume that these expressions- which were presented in holistic form in the classroom- are initially retained this way by the learners under analysis (Nattinger and DeCarrico, 1992) and constitute part of their internal 'for-

mulalect', as Myles and Cordier (2017) term it. Although prototypical FEs have been identified as being mastered before knowledge of their internal parts in previous studies following immersed learners in naturalistic settings (e.g., Hanania and Gradman 1977; Schmidt 1983), it is likely that this phenomenon is more pronounced in instructed classroom contexts for beginner learners, where these expressions are the focus of explicit instruction and communicative teaching activities. Mitchell and Martin (1997), for example, when observing the FPC learners on a bi-weekly basis, stated that this teaching context favoured the rehearsal and memorisation of conversational exchanges, often involving conventional question-and-answer sequences. A question thus arises to the status of the FEs when first recalled in these contexts; that is, how do classroom learners store and produce these multi-word expressions holistically upon appropriate contextual cues, when these are clearly not products of computational generation at that particular point in time?

In order to address this question, a distinction must be made between explicit and implicit language phenomena. A variety of terms are used in the SLA literature to refer to such a distinction, albeit with sometimes differing interpretations with regards to knowledge, memory, learning and acquisition. Commonplace terminology that refers to distinctions associated with the explicit/implicit divide in these aspects includes declarative/procedural, controlled/automatic, conscious/unconscious and perceptual/modular. For clarity, I adopt the terms with the following indented meanings. I take 'implicit knowledge' as referring to the kind of linguistic knowledge that is by definition automatic, unnoticeable and incapable of being brought into consciousness and verbalized (Paradis, 2004; Anderson, 2005), in the sense of modular linguistic knowledge (Fodor, 1983), linguistic competence (Chomsky, 1965), and I (internalized) language (Chomsky, 1986). Functional categories and their associated computational properties (such as wh-movement, T-C movement etc.) constitute abstractions of implicit knowledge accessible via UG in the generative framework. These are hypothesised to be acquired incidentally as opposed to learned (Krashen,

1981; Schwartz, 1993; Paradis, 2004), sustained within procedural memory, a component of the implicit memory system (Paradis, 2009).

Conversely, I interpret 'explicit knowledge' as referring to that which a learner is consciously aware of and that can be articulated in a verbal statement (Anderson, 2005; Ellis, 2008). This kind of knowledge is perceptual and learned as opposed to acquired, characterised by a learner's deliberate attempt to memorise some material or solve a problem (Dörnyei, 2009). Explicit knowledge is subserved by declarative memory (which includes a short term/working memory component), and is responsible for the learning of metalinguistic rules and lexical material (i.e. vocabulary), in the sense of sound-meaning associations[1] (Ullman, 2004; Paradis, 2009).

It follows that the FEs for learners when initially identified in their transcripts should be analysed as fixed perceptual units. These constitute explicit knowledge that is recalled from working memory in a similar fashion to how single L2 lexical items are. This is in line with the general consensus that declarative memory likely underlies the memorization of rule-based grammar as 'chunks' (Morgan-Short et al., 2014). Like the storage of form/sound-meaning mappings between words and physical objects of the world (i.e., [kæt] for the domestic animal *cat*), the FEs can be understood as representing perceptual units that are mapped to conventional communicative situations, such as [ask name] (*comment t'appelles tu, what is your name*) and [ask age] (*quel âge as-tu, how old are you*). The establishment of the FEs as perceptual structures for learners at the initial state, rather than simply 'phonological strings of sounds' is likely aided by non-linguistic information; for example, metalinguistic knowledge of the expressions' meaning/function in certain contexts derived through classroom instruction. This concept is compatible with the idea that L2 speakers are guided (at least initially) more strongly by probabilistic or

[1]Note that knowledge of lexical items' feature specifications i.e. their morphosyntactic properties is taken to be part of implicit knowledge and hence resides in procedural memory (Paradis, 2009)

surface-level information due to problems building and manipulating abstract syntactic representations in real time (Myles, 2004; Clahsen and Felser, 2006). This is why at initial stages in learners' L2 development we find these expressions existing within a largely lexical based L2.

For example, in the BELC, when each learner is shown to produce an FE fluently for the first time, there is extremely little evidence of knowledge of the expressions' related computational properties and associated functional categories elsewhere in their interlanguages. The same can be said for learners of the FPC at Round 1 of data collection, where we find a discrepancy in syntactic complexity between the fluently produced FEs and all other attempts at question formation in the L2, which are limited to single lexical items and declarative structures involving subject pronoun-verb conjugations with *avoir* (have) and *être* only. Both corpus data therefore support the notion that L2 grammars at beginner stages largely make use of a coalition of conscious, declarative/explicit learning strategies as a substitute for poorly developed implicit knowledge (Herschensohn, 2000; Morgan-Short et al., 2014; Truscott, 2017), one strategy of which is the memorisation and reproduction of fixed expressions. It also further indicative of a Weak Continuity view of the initial state (Vainikka, 1994; Vainikka and Young-Scholten, 1998), whereby only L2 lexical categories are immediately available to the learner whilst functional categories develop later in succession.

Adopting this characterisation of the FEs, the next question is how the use of these units could support learners' acquisition of the implicit computational procedures for which their surface forms exemplify. Section 9.3 below now addresses this concept.

9.3 Use of memorised fixed expressions and the acquisition of L2 linguistic knowledge

Formulaic language for early second language learners is traditionally characterised as syntactically under-specified (eg. Myles 2004), indicating that learners can make use of these expressions in advance of a corresponding syntactic representation. Clearly, learners do not construct a syntactic representation for FEs similar to that of a native speaker, otherwise there would be nothing to learn (Fodor, 1998). However, under certain accounts of Weak Continuity, such as 'Modulated Structure Building' (Hawkins, 2001) or 'Instructional Bootstrapping' (Herschensohn, 2000), overt morphological and syntactic distributional evidence from the input are what trigger the establishment of functional categories in the L2 grammar. It could be that FEs can provide learners with this evidence, which would imply that these initially unanalysed expressions eventually become analysed (in part) and feed into/influence learners underlying L2 grammatical system. In this sense, I propose that the FEs are analogous to what some processing models of L1 and L2 acquisition would term 'perceptual intake representations' (Lidz and Gagliardi 2015 for L1 acquisition, L&G model henceforth) or 'perceptual output structure representations' (Modular Online Growth and Use of Language for L2 acquisition, MOGUL henceforth) (Truscott and Sharwood-Smith, 2004). Importantly, the 'intake' part of 'perceptual intake' is used in the sense of Carroll (2001) and Sharwood-Smith and Truscott (2014), referring to the fact that the FEs contain some linguistic information and are given some kind of a mental representation by the learners, as opposed to just being part of the raw input. As mentioned above, this is likely aided by non-linguistic information derived through explicit instruction of these expressions and their corresponding functions. This is how they have been extracted from the input and recalled upon functional contextual cues.

Under these models, during the processing of perceptual intake in declarative

memory, learners bring their implicit knowledge of the computational component of grammar guided by UG to construct a relevant syntactic representation in procedural memory, which contains information about the (L2) syntactic feature specifications. This construction will initially be based on the morphological and syntactic distributional information of the perceptual intake. The link between declarative and procedural memory is manifested differently depending on the specific model adopted. For L&G, it is by means of a processing component termed the 'inference engine', whereby acquisitional intake (supported by UG in identifying the class of representations that constrain grammars independently of linguistic experience) compares the perceptual intake against the predictions of any UG-sanctioned grammar (Lidz and Gagliardi, 2015, p. 337).

For MOGUL, the linkage between perceptual output structures and the linguistic information they contain in the computational grammar (termed the syntactic store, SS) is made possible via corresponding representations in the phonological store (PS) (Truscott, 2015), that is, via the PS/SS interface[2]. Whichever specific model is adopted, the idea is that over time, successful learners will build up a stock of stored items in declarative memory which, guided by UG, can be used in order to make stronger regularities and inferences about the underlying L2 grammar that generated it (Towell, 2014; Lidz and Gagliardi, 2015; Hicks and Dominguez, 2020). These regularities and inferences constitute implicit knowledge in procedural memory, that is, the pool of functional categories and their L2 feature specifications.

Importantly, the relationship between the explicit and implicit language processes outlined above is not one whereby explicit knowledge transforms into implicit knowledge, as is sometimes inferred (e.g. McLaughlin 1990). These two processes

[2]This model was originally based on the Parallel Architecture (Jackendoff, 1997, 2002), a modular architecture where syntax, semantics and phonology are independent generative systems with their own combinatorial principles, but are linked through interfaces. Since writing this thesis, MOGUL has now developed into a broader cognitive framework known as 'The Modular Cognition Framework'. Information can be found online at *https://www.cognitionframework.com/*.

differ in nature and underlying neural substrate (Paradis, 2004), as their representations are substantiated in separate parts of the brain. Both psycholinguistic and neurolinguistic studies point towards a dissociation of non-syntactic 'verbal' (phonological) working memory and core syntactic computations that are responsible for the processing of structural hierarchies, including filler-gap dependencies and movement operations (Waters and Caplan 2004; Santi and Grodzinsky 2007; Makuuchi et al. 2009). Neuroanatomically, verbal working memory is said to take place in the left inferior frontal gyrus (LIFG), and the computational component in the left pars operularis (LPO). This entails that one cannot become the other nor can it be converted to it (Ellis, 2005).

Rather, as argued extensively in Paradis (2004, 2009), the idea is that learners gradually shift from using controlled processes (such as metalinguistic knowledge, memorisation of prototypical FEs) to using automatic ones (implicit competence), which is not simply the 'speeding up' of controlled processes but the replacement of such by a different system entirely. For Paradis, it is the 'practicing' of a surface form that allows learners to move from the use of explicit to implicit language systems:

> 'The repeated practicing of the target form may eventually lead to the internalization of the implicit computational procedures that result in the automatic comprehension and production of that form. It is not the instruction and resulting knowledge that affect competence, but the extra practice provided by the use of the correct form'. (pp. 52-53)

'Practicing' refers to the repeated use (involving both comprehension and production) of an expression in interactive communicative situations (Paradis, 2009), in line with the notion that L2 development based on procedural memory systems occurs gradually with repeated experience and without intention (Ullman, 2005; Knowlton and Moody, 2008; Morgan-Short et al., 2014). This notion echoes the fundamental principle of the output hypothesis (e.g., Swain 1985), in which the production of L2 forms allows learners to move from semantic to syntactic processing,

thus promoting the automaticity and proceduralisation of language. It is practicing in this sense that could therefore allow learners to gradually construct an appropriate L2 syntactic representation for perceptual intake based on the linguistic information it contains. This, over time, will eventually replace explicit knowledge of the perceptual intake and allow for automatic use of the L2 via independently developed implicit means. I will now explain this process in more detail with reference to my findings specifically, adopting the MOGUL model for L2 acquisition.

MOGUL distinguishes an innate architecture split into a Perceptual Output Store (POpS) (constituting general working memory), Phonological Store (PS) and Syntactic Store (SS), where 'processing' refers to the construction of representations in each store. A representation's 'current activation' level is the extent to which it is available for current processing (Truscott, 2015), and its 'resting activation' is the level it has when not involved in processing, but this level reflects the extent of its past use. This means that with continuing use, a representation's resting activation level is raised. Importantly, for an L2 learner, processing in one language activates items in both the L1 and the L2. When an L2 utterance is being comprehended or produced, its current activation level determines which set of competing items is selected for the representation, where items' resting levels establish the starting position for competition (Truscott, 2006).

As the FEs are representations in the POpS, their linguistic information can be linked to the SS via the PS/SS interface. Each time learners interact with the FEs as perceptual output structures- via comprehension or production, based on their distributional information, both L1 and L2 SS representations will be activated. Initially, the L1 representations will dominate, and those of the L2 will be dormant, due to the processor continually dealing with repeated exposure to L1 surface forms. However, over time and with continuing use, the L2 representations will gradually increase and be in a better position to compete with those of the L1; that is, an

increased production of the FEs would lead to a higher resting activation level of their associated functional categories and L2 feature specifications. This is because a module's ability to extract information from a perceptual representation depends on the strength and durability of that representation (Sharwood-Smith and Truscott, 2014); i.e., the higher the resting activation level of the L2 syntactic representations, the better position they are in to compete with the L1 representations that initially dominate the SS. It is this competition between activation levels, that is, the degree to which the SS module can deal with an L2 POpS representation automatically, which drives the L2 acquisition process forward.

An example can be used to demonstrate the basis of competition between L1 and L2 representations in the SS. As discussed in the analysis of the BELC, English requires A-movement, a sub-operation of (Internal) Merge that is motivated by an EPP feature on T, which requires T to have a nominal syntactic subject with which it agrees in person/number as its specifier (Radford, 2009). More specifically, Alexiadou and Anagnostopoulou (1998) propose that the EPP involves categorical D feature checking in T. English checks the D feature of the EPP (or in some theoretical accounts, the D feature of Agree) by triggering movement of a subject (XP) from specifier of VP to the specifier of TP. The D feature of the EPP is [-interpretable] and attracts the [+ interpretable] D feature of the nominal subject in specifier VP (see Section 5.5.). This means that in English, subjects are always overt. Conversely, in Spanish, (the BELC learners' L1), the [-interpretable] D feature of the EPP is checked by Merging the agreement morphemes of a verb, which are nominals in their own right with a [+interpretable] D feature. This results in the raising of the verb from V to T (also known as V-raising, V-T movement), and means that Spanish subjects can be null.

All of the FEs under analysis exemplify this A-movement operation in English, as their surface forms contain arguments with overt subjects. With this in mind,

when Spanish learners of the BELC first produce the FEs as perceptual output structures in the POpS in appropraite contexts, the expressions' distributional information (i.e. the presence of an overt subject argument) will activate both L1 and L2 representations in the SS via the PS/SS interface, specifying the different strategies of D feature checking outlined above. However, at these initial stages, the L1 representation will significantly dominate, and that of the L2 will be dormant, due to the learners' long-term exposure to null subject arguments in these contexts in their L1. However, over time, with continuing use of surface forms such as the FEs which exemplify distributional evidence for A-movement in appropriate contexts, the resting activation levels of this L2 SS representation (feature checking via A-movement) will gradually increase, and thus be in a better position to compete with the initially dominant L1 representation (feature checking via V- raising), the latter of which will eventually result in parsing failures during the processing of the FEs at later ages.

This can be one explanation for why those learners who use the FEs, and related utterance schemas, at younger ages show more surface evidence for knowledge of this computational property over the course of the data collection period, that is, overt subjects used with verbs inflected for tense, person and number. For these learners, D feature checking via A-movement will be at a higher resting activation level in the SS component, thus putting it in a better position to compete with the L1 V-raising specification. Conversely, for those learners who interact less with FEs at the initial state and rely more heavily on the L1 and lexical categories of the L2, the L1 SS representation will still dominate the L2 one.

The consequence of MOGUL's processing approach is a 'no transfer/full access' variant of the Weak Continuity view of the L2 initial state. That is, there is no transfer of L1 syntactic features to the L2 lexicon and learners have full access to UG. Compared to functional categories, lexical categories and FEs are perhaps eas-

ier to perceptually encode given their high salience, contingency and prototypical functionality in the L2 classroom input (e.g., Ellis 2022). Hence, lexical categories are likely to quicker achieve higher activation levels in the L2 and thus compete with the existing L1 ones at the earliest stages of acquisition. Furthermore, in order for syntactic processor to write a particular L2 category/feature specification on SS (such as English T), it must recognize that this category is needed to handle the current L2 input. This is likely impossible without considerable syntactic and lexical distributional evidence, so therefore the acquisition of some content words and their syntactic characteristics must necessarily precede the development of L2 functional categories (Truscott and Sharwood-Smith, 2004). Applied to the data, this could explain the strong reliance on L1 utterances and lexical categories/FEs at the first two rounds of data collection, where L2 functional categories would usually be required. Interpreting the FEs as perceptual output structures, means that over time with frequent usage in appropriate contexts, the L2 syntactic information contained within these expressions eventually becomes available to the learner due to their increased activation levels. The FEs can therefore be viewed as scaffolding devices which guide learners to quicker access relevant L2 functional categories and feature specifications, through a better competition with the pre-existing, initially dominant, L1 ones. This is possibly why we find that, in the BELC, those learners who produce the FEs frequently at younger ages show more evidence for knowledge of the expressions' associated syntactic properties at the end of a 7-year data collection period than those learners who are shown to delay the production of FEs until later ages.

This overall picture that emerges from adopting MOGULs processing approach seems to be compatible with the developmental trajectories observed in both corpora. In the first rounds of data collection (i.e., the initial state), there is strong reliance on L1 utterances, and any L2 productions at these stages seem to be limited to lexical categories only, before evidence for L2 functional categories appears later.

MOGUL's processing approach hypothesises that, at these initial rounds, the resting activation levels of L2 SS representations are still extremely low, which is why learners are seen to rely on L1 utterances where functional categories are required in the L2. L2 development can therefore be seen as a gradual increase in competition between L1 and L2 SS activation levels. In an attempt to visualise the effect of FE usage on this competition, it is possible to compare across the BELC how learners realise the FEs' computational properties where required in relative terms between L1 utterances and accurate L2 utterances.

I again divide learners of the BELC into those who produce the FEs at the early ages 10 & 12 (Early FE Learners) and those who delayed production of FEs until age 16 (Later FE learners). I begin by looking at the Early FE learners. The following line graphs show their relative L1 utterances (orange line) and accurate L2 (blue line) realisations of the computational properties in all required contexts at each age of data collection. The percentages on the 'y-axis' represent all contexts where a computational property is required to manifest in the L2, and the 'x-axis' shows each computational property under investigation along with their associated phrasal category projection. Here, I have also included 'OS in VP' which refers to overt subjects used with bare VPs, that is, lexical verbs with no corresponding evidence for functional category T (i.e., no inflection for person, number or tense). These are taken to be VP projections. Figure 9.1 shows the Early FE learners' L1 utterances and accurate L2 realisations of the computational properties under analysis at age 10, the first round of data collection in the BELC.

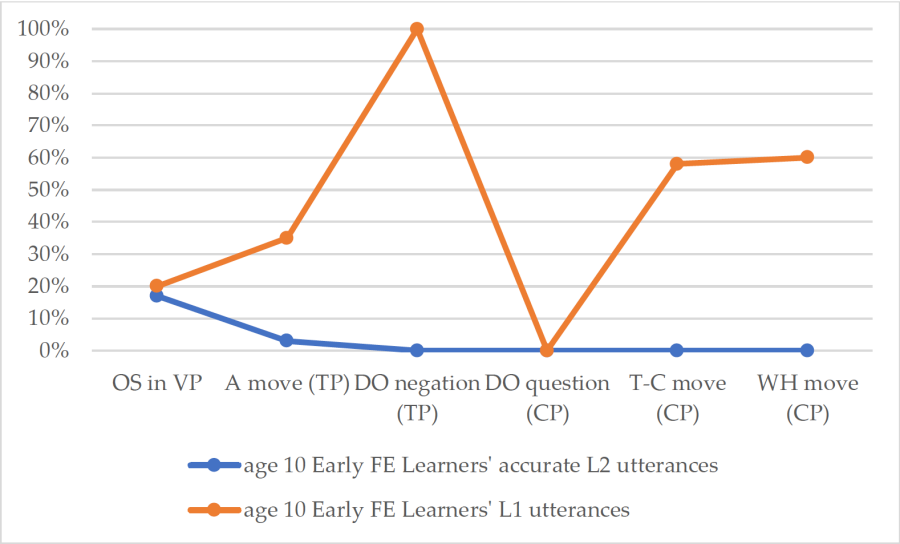

Figure 9.1: L1 realisations and accurate L2 realisations at age 10: Learners with *early* FE use

Apart from *do*-support in question formation structures (*do*-question), for which I found no required contexts at this age, Figure 9.1 shows that L1 utterances largely dominate in all computational properties' required contexts except for bare VP utterances, which are presumed to be lexical in nature. For example, the graph indicates that in all contexts of negation which required DO support in the L2 at age 10, the Early FE learners as a mean average reverted to L1 utterances in 100% of these contexts. In all contexts which required wh-movement, these learners used L1 utterances 60% of the time and L2 utterances 0% of the time. The remaining 40% of realisations in these contexts constitutes inaccurate L2 utterances and translanguaging, which were also taken into account in the analysis of learners' interlanguages (see Chapter 5). The graph in Figure 9.2 now shows learners' realisations at age 12.

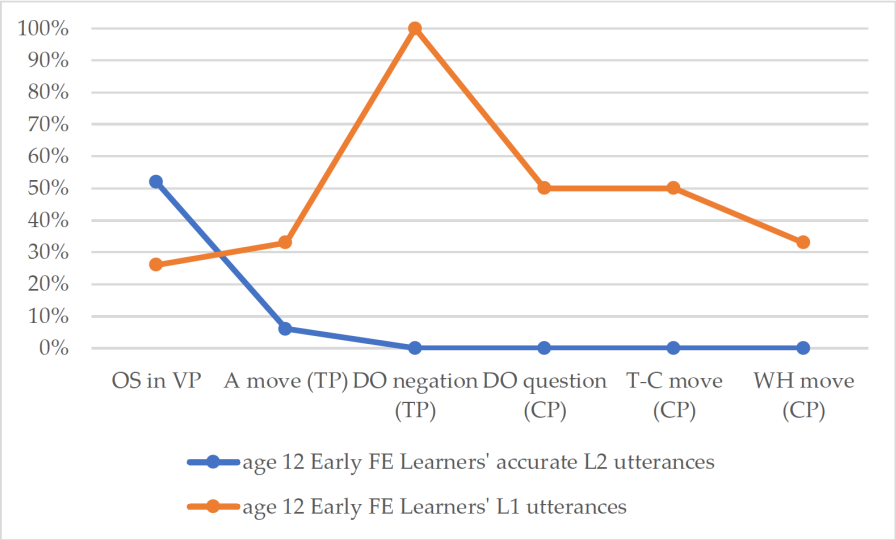

Figure 9.2: L1 realisations and accurate L2 realisations at age 12: Learners with *early* FE use

As we move to age 12, Figure 9.2 shows that overt subjects with uninflected verbs are now realised more proportionally with L2 utterances rather than L1 utterances, whilst contexts that require the computational properties under analysis are still dominated by L1 utterances. The relative gap between L1 utterances and accurate L2 ones is slightly reduced from age 10, apart from with *do*-support with negation (*do*-negation), which is still realised in the L1 in all contexts[3]. Figure 9.3 below shows learners' realisations at age 16.

[3]This is likely a result of the high proportion of *no lo sé* (I don't know) utterances produced by learners at this age, which could be attributed to a code-switching discourse strategy. However, this issue is not pursued further here.

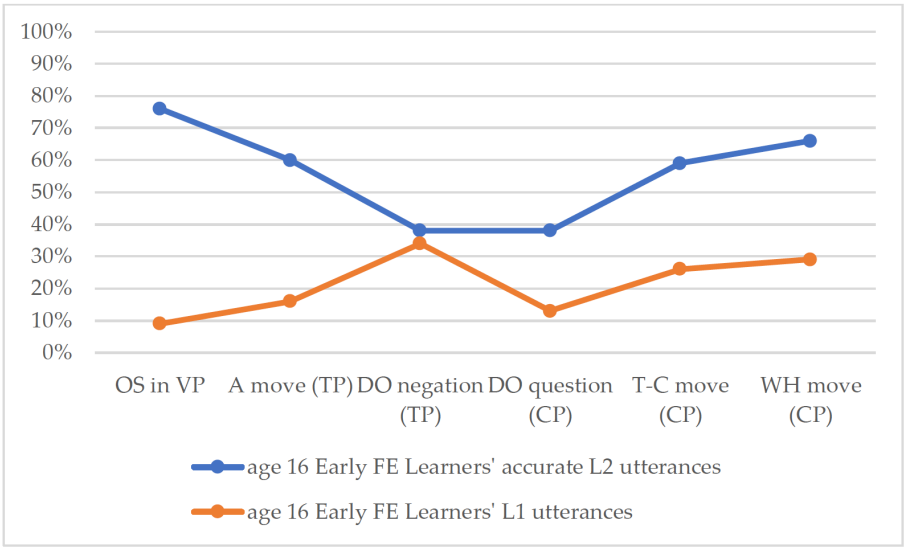

Figure 9.3: L1 realisations and accurate L2 realisations at age 16: Learners with *early* FE use

At age 16, we see that accurate L2 realisations have now overtaken L1 utterances in all required contexts in relative terms. The proportional gap is largest with OS in VP and A-movement, and is smaller with *do*-support, T-C movement and wh-movement, reflecting how L2 knowledge of functional categories seems to develop later than lexical ones. *Do*-support with negation is realised equally in L1 and L2 utterances, likely due to learners' frequent use of *no lo sé* (I don't know) as a code-switching discourse marker. Finally, Figure 9.4 below shows realisations at age 17, the final age of data collection.

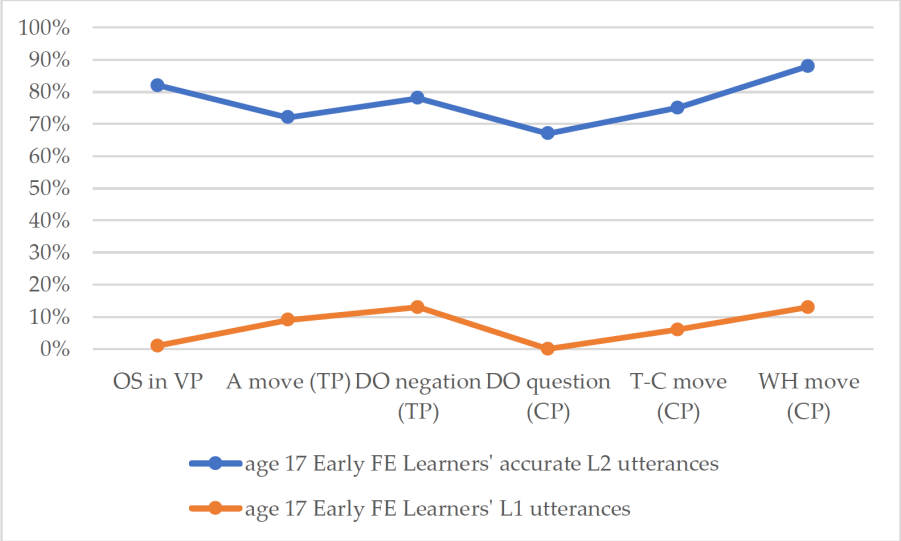

Figure 9.4: L1 realisations and accurate L2 realisations at age 17: Learners with *early* FE use

At age 17, Figure 9.4 shows that accurate L2 utterances now dominate L1 utterances in all contexts that require the manifestation of the computational properties under analysis.

Moving now to examine the realisations of the Later FE learners (i.e. those who delayed production of the FEs until age 16), Figure 9.5 below shows their comparative L1 and L2 realisations at age 10.

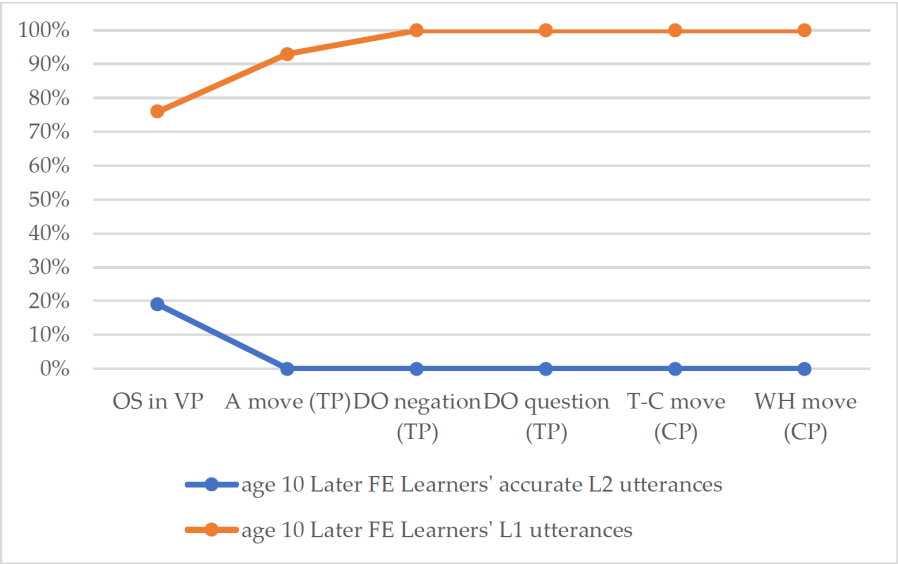

Figure 9.5: L1 realisations and accurate L2 realisations at age 10: Learners with *later* FE use

Figure 9.5 shows a similar pattern to that observed for the Early FE learners at this age, albeit with a more prominent use of L1 utterances which dominate in all required contexts at this age. A similar pattern is found for these learners at age 12, as Figure 9.6 below shows.

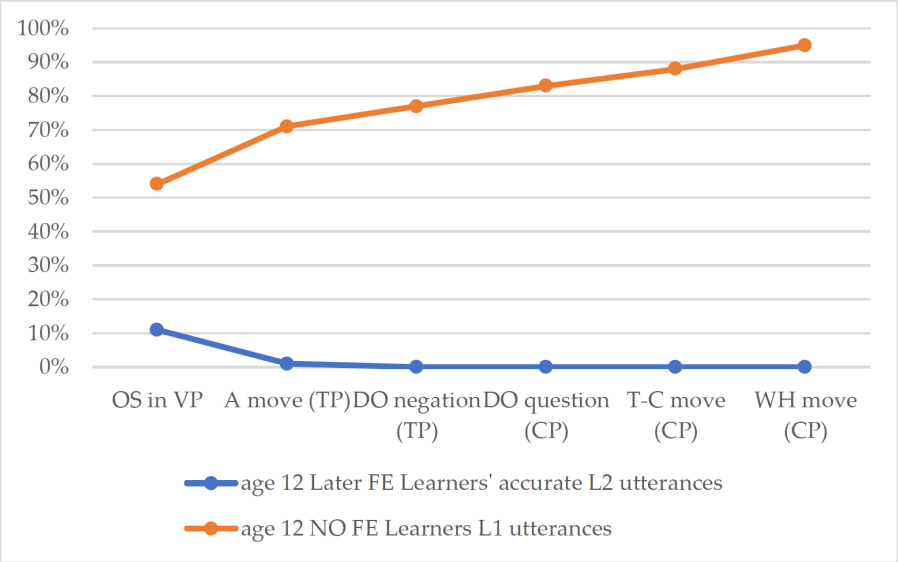

Figure 9.6: L1 realisations and accurate L2 realisations at age 12: Learners with *later* FE use

Figure 9.6 shows a slight reduction in the gap between relative L1 and accurate L2 utterances, but the L1 utterances still remain dominant. It should be noted that the relative gap is more reduced with properties assumed to stem from functional category T (A-movement, DO support negation) than from functional category C (T-C and *wh*-movement). This observation reflects the incremental development of L2 phrase structure observed longitudinally in the corpus. The Later FE learners' realisations at age 16 can be seen below in Figure 9.7.

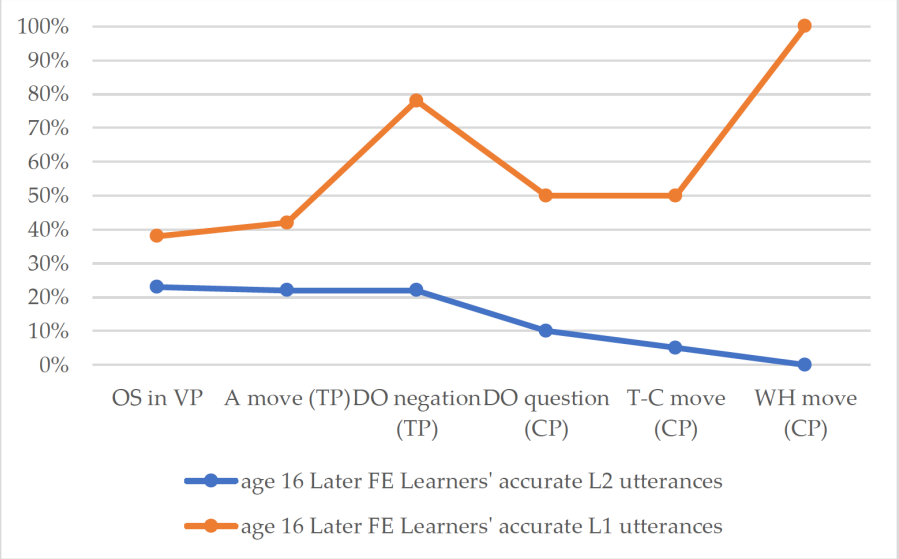

Figure 9.7: L1 realisations and accurate L2 realisations at age 16: Learners with *later* FE use

Unlike the Early FE learners at age 16, whose L2 utterances now dominate L1 ones in relative terms, Figure 9.7 shows that for the Later FE learners at this age, L1 realisations continue to dominate L2 ones in relative terms in all contexts which require the computational properties under analysis. OS in VPs and A-movement show the closest gaps, whilst *do* negation and those properties associated with C are realised more predominantly with L1 utterances. Figure 9.8 below finally shows the Later FE learners' L1 utterances and L2 realisations at age 17, the final round of data collection.

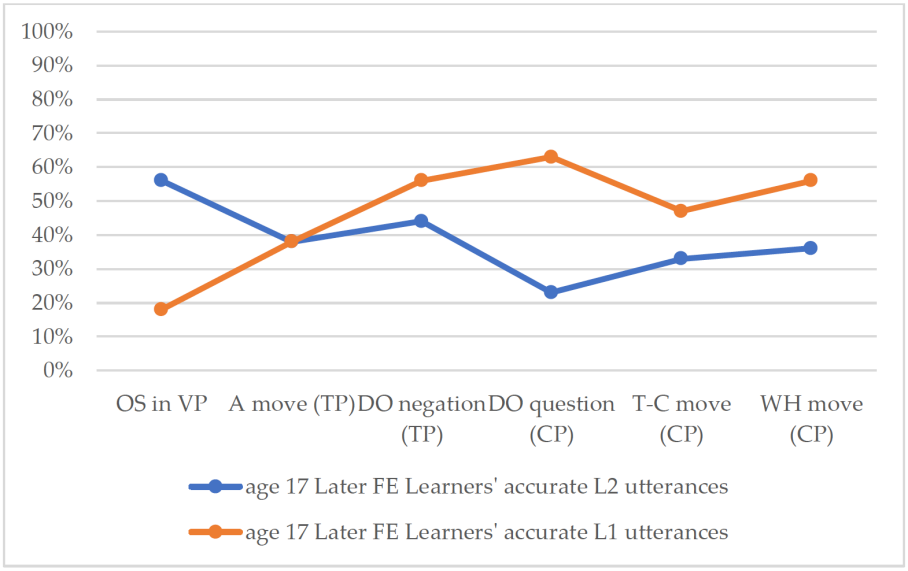

Figure 9.8: L1 realisations and accurate L2 realisations at age 17: Learners with *later* FE use

Differing from the Early FE learners, Figure 9.8 shows that L1 utterances continue to dominate in most required contexts for the Later FE learners at age 17. The only relatively higher L2 accuracy rates are observed with OS in VPs, and A-movement shows an equal split. All other computational properties under analysis are still realised more via L1 utterances in relative terms, and the gap between these and learners' accurate L2 realisations is larger with those properties associated with functional category C (*do*-support question, T-C movement, *wh*-movement) than those associated with functional category T (*do*- support negation), which again shows how L2 knowledge of higher functional categories seems to develop in succession.

The line graphs in Figures 9.1-9.8 above illustrate the interaction between the incremental development of L2 phrase structure (i.e. from lexical to functional category acquisition) and learners' competing L1 and L2 syntactic activation levels. For both sets of learners (Early and Later FE), utterances with OS in VPs are the first to be realised proportionally more in the L2 than in the L1, before the replacement of those utterances involving A-movement follows. Unlike the Later FE learners, the Early FE Learners also manage to replace their L1 utterances with accurate L2

ones in all contexts associated with functional categories T and C by age 16. Their overall transition from the L1 to accurate L2 utterances proceeds much quicker than that of the Later FE learners. I use this as evidence to suggest that learners' L2 feature specifications are at higher activation levels as a result of earlier and more frequent usage of syntactically complex FEs.

Section 9.4 discusses how utterance schema extraction and generalisation could potentially facilitate this process.

9.4 Schematic learning and the acquisition of L2 linguistic knowledge

This section now considers how the usage-based schematic learning trajectory is applicable to some of the syntactic development observed in both corpora. It begins by emphasising that there is evidence of development that goes beyond schematic learning (9.4.1). It then concentrates on the development schematic learning can account for (9.4.2), and offers an explanation as to how this strategy could interact with the acquisition of underlying syntactic knowledge more generally (9.4.3).

9.4.1 Observed development beyond schematic learning

A fundamental notion in usage-based models of SLA is that language learning is initially exemplar-based, and development proceeds from the formulaic to the schematic (Horbowicz and Nordanger, 2021). In these models, the L2 learner acquires constructions whilst using language (i.e. engaging in conversation), and L2 syntax emerges 'from the memories of the utterances in their history of language use and the abstraction of regularities within them' (Ellis, 2005, p. 306). In other words, implicit knowledge is not conceptualised as a modular computational system, rather, it is posited that learners are tuned to the frequencies of form-function map-

pings, allowing them to acquire schematic and categorisational information based on default patterns that emerge from prototypical exemplars (Tomasello, 2003; Ellis, 2012). Learners encounter such exemplars in the input and identify regularities amongst them, which leads to the extraction and generalisation of utterance schemas to derive similar functional structures (Roehr-Brackin, 2014; Eskildsen, 2015, 2020). Hence, linguistic structure emerges through reinforcement of commonalities inherent in multiple experiences (Langacker, 2000).

Therefore, a key component of longitudinal usage-based studies is the search for developmental sequences for particular constructions that are seeded by particular memorized formulaic phrases (Ellis, 2012). However, as discussed in Section 9.3, there is L2 development observed in both corpora beyond utterance schema extraction and generalisation, which can also be linked to learners' use of identified FEs. In the BELC, those learners who frequently use FEs at earlier ages seem to arrive more quickly at L2 accuracy of the functional categories and corresponding computational properties that the expressions exemplify. Knowledge of these properties manifests through a variety of surface structures that are superficially unrelated, rather than as schematic extensions of the FEs in similar functional environments.

For example, Learner 18 shows early and frequent use of the identified FEs. Over the course of development, this learner demonstrates a high accuracy rate of L2 utterances assumed to be manifested by underlying computational properties related to the FEs, albeit with different surface structures. Their knowledge of wh-movement in the L2, for example, surfaces via relative/interrogative complement clauses, as well as wh-questions. These structures are not deducible from the FEs' schematic constructions as exemplified by their surface forms, all of which display the wh-word in clause-initial position as a question operator followed by the copula or auxiliary. In addition to Learner 18, Learners 7 and 47 also produce relative/interrogative complement structures in the L2 at later ages in the BELC, all of whom use the

FEs at early ages. Table 9.1 below shows how these learners differ from the others in their production of these structures in the L2. FEs produced at the early ages are in shown in bold and '-' indicates the absence of FEs/relative/interrogative complement clauses in learners' transcripts at that point of data collection.

learner	age 10: FE	age 12: FE	age 16: relative/interrogative complement clause	age 17: relative/interrogative complement clause
2	-	-	-	-
5	***what's your name***	-	-	-
7	***what's your name***	***what's your name how old are you where do you live***		*I need some food and somebody who play the music*
13	-	-	-	-
18	-	***what's your name how old are you *what do you live***	*when I have homework I do the homework and [\] and go [\\] or watch TV* *when the mother is [/] (.) is (.) telling <what *is the>[/] (.) what *is the [/] (.) the *street they are [//] (.) have <to to> [/] to go the dog (.) came to [//] <into the> [/] into the basket>* *when they (.) arrive in the mountain they have a surprise that (.) the dog *(.) eat <all the> [/] all the (.) food*	*and when they: [\] they go to eat the breakfast the dog they see that the dog *eat [\\] eats the: [\]*
27	*no transcript*	-	-	-
38	*no transcript*	***what's your name where you live***	-	-
42	*no transcript*	-	-	-
47	***how old are you what's your name***	*no transcripts*	-	*when the sister and brother arrive to the [\] to the forest they looked that her dog mm *_was [\] was appeared*

Table 9.1: Learners' early FE use and corresponding later L2 productions of relative and interrogative complement clauses

Similarly, knowledge of A-movement, measured by learners' use of overt subjects in clauses demonstrating overt knowledge of functional category T (i.e. via auxiliary verbs, inflectional morphology, infinitival 'to'), also goes beyond the schematic analysis of the FEs' surface forms. Notably, none of the FEs offer overt indications of categorical and/or schematic properties of auxiliary verbs or inflectional morphology on V, as they contain either second person conjugations of the irregular copula verb 'is' or a dummy auxiliary 'do'. Yet, learners who make use of these expressions at the early ages show significantly greater knowledge of these surface structure phenomena at later ages, as Table 9.2 below reiterates. Here, it should be noted that the provided example utterances for each learner are representative of

their understanding of similar structures. Elements of learners' utterances providing overt evidence for functional category T used with overt subjects are underlined.

learner	age 10: FE	age 12: FE	age 16: overt subjects with overt evidence for FC T	age 17: overt subjects with overt evidence for FC T
2	-	-	*because _ is my [*] birthday Friday | *the mother (.) hm read the map	*hm _is a big room and hm _ have two beds | *the dog eat all food
5	**what's your name**	-	if you stay in the house (.) I can make the party? | the childrens see that the dog eats	*I wanna to have a party in the house | can you leave me the house this weekend?
7	**what's your name**	**what's your name how old are you where do you live**	the boy is going to a tree or something | I study or I'll have my homework	I'm going to make a party at home | can I do it?
13	-	-	*he make (.) the food for … | *he eaten the food and they surprise	*the mother say goodbye | *the boy and the girl say that the dog eat a lot of the cake and sandwiches
18	-	**what's your name how old are you *what do you live**	there is a girl and a boy and they are (.) preparing the breakfast | I will go to the town	the brother and sister are saying goodbye | they can go to the house to (.) to prepare someone to eat
27	no transcript	-	*the dog (.) into the bag | *_(.) go to the house	*the mother say goodbye and the child look | *the child surprised the dog is the bag
38	no transcript	**what's your name where you live**	we are going to do the party in my bedroom | can you do something for the party (.) to eat?	I wanna do a economical career because my brother is a teacher | I've been at this school half of my life
42	no transcript	-	*altra is the childrens and mother and dog | *_listen to music	*the childrens in the street | *mother say goodbye and a dog is on the basket
47	**how old are you what's your name**	no transcripts	Sunday afternoon, I played football match | she's showing to his child the map	I suppose I will go to the university | on Saturday night we arrived to Barcelona

Table 9.2: Learners' early FE use and representative examples of later L2 productions in contexts which require A-movement

Therefore, in the BELC, over a 7-year period, we find correlations between early FE use and later knowledge of grammatical properties that cannot be solely deduced from a schematic/categorical analysis of the FEs' surface forms. The connection between the FEs and grammatical utterances, as illustrated in Tables 9.1 and 9.2, lies in the underlying syntactic properties assumed in their generation, rather than in schematic patterns derived solely from their surface structures. This observation, coupled with the dominance of lexical categories in all learners' interlanguages at early stages outside of the FEs, implies an incremental acquisition of L2 properties extending beyond the schematic analysis of prototypical surface forms. It instead indicates a gradual and successive acquisition of L2 functional categories more generally, potentially influenced by early and frequent use of the FEs.

Similarly in the FPC, over the three rounds of data collection, there is a noticeable shift in how learners construct L2 wh-interrogatives. Initially, learners tend to

use lexical/declarative structures, but there is a progression toward more complex derivations involving the computational properties exemplified by the FEs. This shift occurs even though these properties are optional for question formation structures in the L2, and extends beyond schematic learning. One case is Learner 9, who produces four FE^{wh}s in Round 1 of data collection of the [wh + INV] kind which all exhibit wh-movement and T-C movement (1). Outside of these expressions, this learner only produces one other wh-interrogative in the L2, which is lexical/verbless (2).

(1) Learner 9 FEs: Round 1, age 11-12

comment t'appelles tu

où habites- tu

quelle est la date de ton anniversaire

quel âge as tu

(2) Learner 9 L2 interrogatives outside of FEs: Round 1, age 11-12

*quelle coleur pantalon noir blanc?
what colour trousers black white

'what colour are his trousers, black or white?' (intended meaning)

At Round 5, Learner 9 produces a variety of interrogative forms in the L2. Some of these show wh-movement and T-C movement, mirroring the structures observed in the previously produced FEs. However, these interrogatives feature the combinations *où est* and *qu'est ce que* (3), which could be lexically-specific utterance schemas in their own right.

(3) Learner 9: Example L2 interrogatives with *où est* and *qu'est ce que* (Round 5, ages 12-13)

où est la garçon?
where is the boy

'where is the boy?'

qu' est ce qu' on peut faire a Belleville
what is it that one can do at Belleville

'what can you do at Belelville?'

Outside of these instances, the learner opts for [wh IN SITU] structures for wh-interrogatives (4), which is the least derivationally complex option as it does not involve wh-movement or T-C movement.

(4) Learner 9 L2 interrogatives outside of FEs, *où est* and *qu'est ce que*

(Round 5, ages 12-13) [wh IN SITU]

il est comment?
he is how

'what is he like?'

la femme elle est comment?
the girl she is how

'what is the girl like?'

tu veux quel jour?
you want what day

'what day do you want?'

However, by Round 6, apart from the FEs and lexically-specific schemas *où est* and *qu'est ce que*, all of Learner 9's L2 wh-questions now show wh-movement, with some also featuring T-C movement.

Round 6 **wh Q**	**Structure**
comment s'appelle le garçon? *comment s'appelle elle?* *quel jour as tu?*	[wh + INV]
où est ce qu'on peut mange le déjeuner? *où est ce que je rendez vous?* *où est ce qu'on peut mange le déjeuner?*	[wh + ESK]
a quelle heure tu veux un rendez vous *comment ca s'ecrit* *comment ca s'ecrit*	[wh + NO INV]

Table 9.3: Learner 9 L2 interrogatives outside of FEs, *où est* and *qu'est ce que* (Round 6, age 13-14)

Therefore, as with learners of the BELC, a generative analysis helps make sense of the developmental trajectory of the FPC learners' wh-questions in the L2. Outside of fixed question forms, learners gradually move from preferring less derivationally complex structures to more complex ones involving the maturation of functional categories exemplified by the FEs. A description of the development of L2 interrogatives based on utterance schema extraction and generalisation alone does not capture the fact that learners' wh-interrogatives outside of the FEs are initially lexical and/or in declarative word order form. Positing the acquisition of complex underlying syntactic derivations driven by emerging functional categories can also explain why their interrogatives gradually develop to exhibit wh-movement in the majority of cases, most of which do not share the same utterance schemas of previously produced FEs (eg. the [wh + ESK] and [wh + NO INV] kind).

9.4.2 Observed development that schematic learning can account for

However, there is clearly evidence of schematic learning in both corpora, which can account for a considerable proportion of the observed longitudinal development

based on FE use. Some learners are shown to produce L2 interrogatives that share utterance schemas of identified FEs that appear in their production data ontogenetically. Learner 18 is one example from the BELC, where it can be argued, based on L2 utterances observed in their longitudinal production data (5), that they move from the FE *what is your name* (5-a) to the lexically-specific utterance schema [what is] + X (5-b) to the categorically-specific utterance schema [WH + COPULA] + X (5-c).

(5) Learner 18

a. Age 12: *what's your name* [Formulaic Expression]

b. Age 16: *what is your job* [what is] + X

c. Age 17: *what are you studying* [WH + COPULA] + X

An example in the FPC is Learner 7, who seems to extract the lexically-specific utterance schema [où habite] + X (where lives + X) from the FE *où habites-tu* and extends this to similar functional structures with third person reference.

(6) Learner 7

a. Round 1, Ages 11-12

[où habites]-tu (where do you live)

b. Round 5, Ages 12-13

[où habite] la garçon (where does the boy live)

[où habite] la fille (where does the girl live)

[où habite] le monsieur (where does the man live)

[où habite] la femme (where does the girl live)

Lexically-specific utterance schemas are particularly prominent in the FPC. Outside of the FEs, the example utterances below demonstrate how the majority of learners consistently utilise the schemas [où est] + X (where is + X) and [qu'est ce-que] + X (what is it that + X). Note that some of these instances involve overextensions, as

seen in (8-d) (8-e) (8-f), providing additional evidence that these are lexically 'fixed' forms.

(7) Learner 20

a. *[où est] un garçon* (where is the boy)

b. *[où est] la fille* (where is the girl)

c. *[où est] le homme* (where is the man)

d. *[où est] le madame* (where is the woman)

(8) Learner 2

a. *[qu'est ce que] c'est* (what is it that it is)

b. *[qu'est ce que] fait la fille* (what is it that the girl does)

c. *[qu'est ce que] fait garçon* (what is it that the boy does)

d. **[qu'est ce que] la boulangerie* (*what is it that the bakery)

e. **[qu'est ce que] la madame* (*what is it that the woman)

f. **[qu'est ce que] la monsieur* (*what is it that the man)

The comparison of both corpora also supports the idea that learners gradually shift from the formulaic to the schematic (i.e. from fixed lexically-specific expressions/schemas to more abstract schemas). This transition is a fundamental notion at the heart of usage-based SLA. For example, at the latter stages of the BELC, most of the utterance schemas related to the FEs are more abstract (i.e. categorically-specific: [WH + COPULA] + X), whereas in the FPC, we see more reliance on FEs and lexically-specific schemas (i.e. *où est* and *qu'est ce que*). This would be predicted by the usage-based learning trajectory, as the FPC provides a more concentrated view of learners' L2 behaviour during the first two years of classroom instruction only (i.e., offering more data within a smaller window of time). Conversely, in the latter rounds of the BELC, we gain a more general picture of learners' development after a 6/7 year period of instruction, where more abstract schemas are expected beyond the initial state.

The analysis of both corpora also revealed that those learners who show the most evidence for schematic learning are also those with the highest accuracy rates of related underlying L2 syntactic properties more generally. The example of Learner 9 from the FPC as mentioned above is a case in point, who initially relies on FEs at Round 1 of data collection, moves to lexically-specific utterance schemas by Round 5 ([*où est* + X], [*qu'est ce que* + X]) and finally adopts wh-movement for all wh-interrogatives by Round 6, manifesting via a variety of optional structures permitted in the target language ([wh + INV], [wh + ESK] and [wh + NO INV]). Similarly, in the BELC, those learners who show earlier and more frequent use of the FEs and related utterance schemas are also those learners who demonstrate the most evidence of related L2 syntactic knowledge more generally at the later ages. This suggests that schematic learning strategies can be facilitative on the acquisition of underlying L2 syntactic properties more generally.

9.4.3 The influence of schematic learning on the acquisition of underlying syntactic properties

Therefore, to provide a comprehensive description of the developmental trends observed in both corpora, schematic learning strategies must be incorporated. In doing so, I posit that schematic learning strategies could interact with a Weak Continuity view of the initial state and L2 development thereafter in light of the processing view of SLA as posited by MOGUL. I now document what this process might look like.

If learners can extract utterance schemas from prototypical formulaic exemplars and extend them to similar functional structures, it allows them to produce/interact with more complex surface forms that exemplify the same L2 syntactic and morphological distributional information. A consequence of this under MOGUL, is that,

with each production or interaction involving similar complex surface forms, the resting activation levels of L2 linguistic information (i.e. the functional categories and their feature specifications) in the SS component are further raised. In other words, schematic learning strategies could help to keep L2 SS activation levels high enough to compete with the existing L1 levels. For example, learners in the BELC, such as those in (9) and (10), are shown to extract the utterance schemas [WH + COPULA] + X from the FE *what is your name* and then generalise this to produce similar functional structures in the L2 at later ages.

(9) Learner 38

a. Age 12: *what's your name* [WH + COPULA] + X

b. Age 16: *why are you doing this kind?* [WH + COPULA] + X

c. Age 17: *why are you doing this work?* [WH + COPULA] + X

(10) Learner 18

a. Age 12: *what's your name* [WH + COPULA] + X

b. Age 16: *what is your job* [WH + COPULA] + X

c. Age 16 *what are you studying* [WH + COPULA] + X

Every time these learners produce similar wh-questions in the L2 via utterance schema extraction and generalisation, they are in a better position to (unconsciously) license conclusions about the underlying syntactic structures that generated them. This is based on their higher activation levels in the SS component derived from the FEs' morphological/syntactic distribution information. This is why learners who show more evidence of schematic learning (such as Learner 38 and 18 above) also show more surface evidence for knowledge of these underlying properties at the later ages more generally, as exemplified with utterances such as those below.

(11) Learner 38: [T-C movement, wh-movement]

a. Age 16: *can I make a party with my friends*

b. Age 16: *can you do something for eat*

c. Age 16: *what you wanna say*

(12) Learner 18: [wh-movement]

a. Age 16: *when the sister is telling what is the street they are on*

b. Age 16: *when they go to eat the breakfast they see that the dog eats the...*

Utterance schema extraction and generalisation from model FEs under this interpretation can therefore be analysed as an implicit learning strategy subserved by part of the domain-general cognitive procedural system (Ullman, 2004). Although implicit in nature, this process is necessarily separate from modular linguistic knowledge (i.e. the computational component) and is instead driven by learners' tuning to the form-frequencies of form-function mappings (Ellis, 2005, 2012). This sensitivity to statistical-distributional information of perceptual intake (such as that found in the FEs) can then feed forward for inferences about L2 grammatical structure (Lidz and Gagliardi, 2015), raising the resting activation levels of these structures in the SS to better compete with those of the L1. I suggest that it is in this sense that schematic learning can facilitate the independent acquisition of underlying L2 syntactic knowledge more generally.

Importantly, reliance on FEs and other lexically-specific schemas is most prominent at initial stages of SLA, when implicit knowledge of L2 functional categories is underdeveloped. Indeed, we observe this to be the case in the FPC, in comparison to the BELC. In absence of this knowledge, the learner must rely on declarative knowledge and general cognitive processes to 'unpack' the information derived from the memorisation of the FEs. This generalisable information deduced from a schematic analysis of perceptual intake (e.g. the FEs) enhances learners' L2 output accuracy by allowing them to construct well-formed L2 utterances based on the deconstruction of these model forms. The repeated practicing of these utterances and the

subsequent raising of their underlying L2 syntactic activation levels is what could then drive the move from L2 lexical to functional category acquisition.

9.4.4 On the role of explicit instruction

A final note on the potential role of explicit instruction within the processes described above is warranted here to round up the discussion. As both sets of learners under analysis are formal classroom EFL learners, it is possible that explicit instruction of the prototypical FEs, and potentially, their related functional utterance schemas, has facilitated syntactic development based on learners' usage and analysis of these elements. As mentioned previously, it is commonplace and now well established in these learning contexts that prototypical expressions (such as the FEs identified in this study) are explicitly taught in holistic form, and indeed retained this way by learners (Nattinger and DeCarrico, 1992; Myles et al., 1998, 1999; Towell, 2012, 2014), potentially through drilling exercises. This is because memorisation of these expressions allows learners to participate in conventional task-based/role-play activities that are typical of communicative classroom settings, which further allow learners to practice with these expressions in appropriate contexts. Indeed, this was confirmed to be the case during the bi-weekly classroom observations carried out by the FPC creators (Mitchell and Martin, 1997).

The explicit teaching of related grammar could manifest in two slightly different ways in these contexts. Traditional grammar teaching methods (often associated with older, grammar-translation approaches) comprise the explicit instruction of individual 'rule-based' grammatical forms. Indeed, this method, if carried out in the two learning contexts under analysis, would likely facilitate learners' acquisition of certain grammatical aspects, and could have potentially contributed to learners success with various structures (including those related to the FEs) at later ages. More modern methods are based on communicative language approaches. Since the rise

of 'lexicogrammar' (e.g., Sinclair, 1991; Halliday, 1994), what have become known as 'lexical approaches' to grammar instruction explicitly teach functional grammar in holistic, lexical form (e.g., Lewis 1993). If carried out in these learning contexts, it is clear how a lexical approach to grammar teaching specifically would facilitate the developmental trends identified across both corpora, and this method is indeed advocated by the study's findings. This would involve explicit instruction of the FEs as well as their internal functional elements, such as word ordering/patterning in [WH + AUXILIARY/COPULA] combinations, essentially facilitating learners' extraction and generalisation of utterance schemas across multiple meaningful units. This in turn would better encourage learners to approach L2 data in a systematic, analytical way, and perhaps lead to more instances of them being able to produce FEs and deconstruct related schematic material in meaningful interactions. Therefore, explicit grammatical instruction via a 'lexical-approach' should be viewed as beneficial to learners' ability to produce and generalise grammatical sequences from formulaic material, which in turn leads to a quicker acquisition of related underlying syntactic properties through more repeated interaction and output of these forms. The explicit instruction is ancillary to increased production in promoting underlying syntactic development in the sense of Paradis (2004), as quoted previously, in that 'it is not the instruction and resulting knowledge that affect competence, but the extra practice provided by the use of the correct form' (pp. 52-53). To be sure, the FPC and BELC learners compared in this study were all exposed to the same level of explicit instruction in the same classroom contexts. Thus, the interlanguage developmental differences observed and described above relate to learners *usage* of FEs and related utterance schemas in appropriate interactions/contexts- irrespective of how they have been taught.

Section 9.5 below now concludes the discussion of the developmental trends observed in both corpora.

9.5 Summary of discussion

This discussion has used MOGUL to present a unified account of classroom learners' L2 development as observed across two longitudinal corpora capturing learners' L2 productions at the initial state and a period of time thereafter. By adopting a generative framework of SLA whilst also drawing on usage-based notions of schematic learning, common developmental trends have been uncovered from the datasets that would otherwise have been missed if relying on one theoretical framework alone. I have argued for a Weak Continuity view of the initial state, whereby learners' early production data presents little evidence for L2 functional category knowledge. Instead, this shows evidence that learners rely on L2 lexical categories and fixed expressions only. The use of prototypical fixed expressions such as those identified in both corpora is perhaps more salient in instructed contexts than immersive ones, particularly when these expressions are explicitly taught to aid communicative 'role-play' activities in the classroom. The analyses of both corpora shows evidence that learners gradually move from the formulaic to the schematic, shifting from these fixed expressions and lexically-specific utterance schemas to more abstract categorically-specific schemas. Outside of schematic learning, we discover evidence for emerging knowledge of L2 functional categories more generally across the data period, particularly with those learners who interact more with fixed expressions and associated utterance schemas. Using MOGUL's processing approach to SLA, I have suggested that schematic learning can facilitate underlying L2 syntactic development by keeping the resting activation levels of L2 syntactic structures high enough to compete with existing L1 ones in the Syntactic Store (SS). The L2 syntactic and morphological distributional information provided through use of the FEs and related schematic constructions could better equip learners to make inferences about the L2 feature specifications on the functional categories involved in their generation. This could therefore be a contributing factor which drives the move from lexical to functional category acquisition in the L2.

Chapter 10 now presents the conclusion.

Chapter 10

Conclusion

This study has analysed two longitudinal learner corpora to investigate the relationship between formulaic language and the development of L2 syntax. It has adopted generative and usage-based frameworks of grammar and notions of acquisition to discover developmental trends in the datasets and provide a multiparadigmatic account of such. This Chapter now concludes the study. Section 10.1 gives a summary of major findings. It focuses first on the BELC (10.1.1), and then the FPC (10.1.2), before bringing the findings of both corpora together (10.1.3). Section 10.2 discusses the contribution of this study to the field of SLA, and Section 10.3 outlines its limitations. Finally, Section 10.4 suggests some implications for further research based on these contributions and limitations.

10.1 Summary of main findings

The analysis of both longitudinal corpora present evidence that learners' use of FL is influential on their L2 syntactic development. At initial rounds of data collection, learners are found to produce strikingly similar conventional wh-expressions. Outside of these, interlanguage analyses indicate that learners' L2 knowledge is largely lexical in nature. This is mostly in the form of single word utterances and non-finite verb phrases, with limited evidence of surface phenomena related to L2 functional categories. The FEs are therefore analysed as fixed lexical units that learners recall

upon appropriate contextual cues. Evidence for knowledge of L2 functional categories begins to emerge incrementally towards the later rounds of data collection. This developmental trajectory fits a Weak Continuity view of the L2 initial state.

In both corpora, those learners who show the highest levels of L2 syntactic knowledge by the end of the data collection period are those who make the most use of formulaic material. Within this development, there is evidence for schematic learning. Similar utterance schemas are identified across both corpora, some of which can be traced back to the FEs in learners' production data ontogenetically. MOGUL's processing approach to SLA is drawn upon to make sense of the interaction between schematic learning and the incremental development of L2 phrase structure. It is posited that utterance schema extraction and generalisation, as an implicit domain-general cognitive learning strategy, keeps the activation levels of the L2 syntactic structures high enough to compete with the existing L1 structures. A more frequent use of complex formulaic material via schematic learning strategies could therefore facilitate a quicker transition from lexical to functional category knowledge in the L2.

Sections 10.1.1 and 10.1.2 below now summarise the results of the BELC and FPC analyses specifically, and Section 10.1.3 brings these together.

10.1.1 The Barcelona English Language Corpus

In the BELC, all 9 learners under analysis were shown to produce conventional wh-expressions derived from representative textbooks in their learning context. These expressions are below in (1).

(1) Conventional wh-expressions in the BELC

a. *what is your name*

b. *how old are you*

c. *where are you from*

d. *where do you live*

For all learners, there is a discrepancy between their first fluent production of these syntactically complex expressions and their interlanguage competence. For example, Learner 5 was shown to produce the fluent conventional wh-expression below in (2-a) at the same time as typical utterances such as those in (2-b) - (2-e).

(2) Learner 5; Age 10

a. *what's your name*

b. *no sé que dir-te* [CATALAN]

'I don't know what to say to you'

c. *it's dog eat*

d. *the mum it's*

e. *I study*

This confirmed the fixed nature of these conventional wh-expressions for all learners of the BELC. The generative analysis tracked learners' interlanguages over the four data collection rounds (age 10, 12, 16 & 17) outside of the FEs, and observed an incremental development of L2 knowledge. At the early ages, learners largely rely on the L1 and lexical categories of the L2, including NPs and non-finite VPs. Evidence for knowledge of L2 functional categories T and C begins to emerge at the later ages. Importantly, correlations are found between a learner's earlier and more frequent production of the FEs and a higher accuracy of their L2 computational properties at the later ages (16 & 17). That is, those learners who produce the FEs more frequently at a younger age demonstrated more evidence for L2 knowledge of wh-movement, T-C movement, A-movement and *do*- support at the later ages.

A usage-based traceback analysis of the BELC also found evidence for schematic learning, potentially instantiated by the FEs. Specifically, 53% (20/38) of L2 interrogatives could be traced back to utterance schemas of previously used FEs. The

most productive for wh-questions was the abstract utterance schema [WH + COPULA] + X, as derived from *what is your name* and *where are you from*. For yes/no questions, it was the lexically specific schema [do you] + X, as generalised from *where do you live*.

10.1.2 The French Progression Corpus

In the FPC, all 24 learners were shown to produce conventional question expressions derived from their classroom input in the first round of data collection. These can be seen below in (3) and (4).

(3) Conventional wh- expressions

a. *comment t'appelles tu?* (what is your name?) [wh + INV]

b. *quel âge as-tu?* (how old are you?) [wh + INV]

c. *où habites- tu?* (where do you live?) [wh + INV]

d. *quel est le date de ton anniversaire?* (when is your birthday?) [wh + INV]

(4) Conventional yes/no expressions

a. *tu as un animal?* (do you have a pet?) [DEC]

b. *tu as des frères ou des soures?* (do you have any brothers or sisters?) [DEC]

As can be seen above, the generative analysis found that all conventional wh-expressions display wh-movement and T-C movement [wh + INV] and all conventional yes/no-expressions are in declarative form [DEC]. At Round 1 of data collection, there is a discrepancy between learners' fluent productions of these conventional question forms and other L2 interrogatives. It is likely that learners are relying on lexical categories only to express interrogative force in other similar environments. This includes single NPs (5), lexical items ungrammatically linked with

avoir ('to have') (6) and potentially rote-learned verb conjugations with subject pronouns in [wh IN SITU] (7) and [DEC] (8) structures.

(5) Learner 28 Round 1

*les yeux?
the eyes

'what colour are your eyes?' (intended meaning)

(6) Learner 2 Round 1

*il age a frère?
the age has brother

'how old is your brother?' (intended meaning)

(7) Learner 12 Round 1 [wh IN SITU]

tu as les cheveux comment?
you have the hair how

'what style is your hair?'

(8) Learner 5 Round 1 [DEC]

il est petit ou grand?
he is small or big

'is he short or tall?'

This suggests that the conventional question forms are formulaic in nature, and that outside of these, L2 functional categories responsible for generating wh-movement and T-C movement are underdeveloped. Two years later, at Rounds 5 and 6, only 14 out of the 24 learners are shown to produce grammatical question forms outside of these FEs. The majority of wh-questions now display wh-movement ([wh + INV], [wh + ESK], [wh + NO INV]) and T-C movement ([wh + INV]). This is similar to

the FEs produced previously, and is indicative of emerging knowledge of L2 functional categories T and C.

A usage-based analysis of the FPC revealed that 28.15% of interrogatives in learners' production data could be traced back to one of the FEs produced previously. It is also likely that two other combinations are being used as lexically-specific utterance schemas independent of FE analysis. 62% of [wh + INV] interrogatives feature the combination *où est* ('where is'), and 90% of the [wh + ESK] are with the sequence *qu'est ce que* ('what is it that').

10.1.3 Bringing the corpora together

In both corpora, all learners display a similar developmental trajectory in terms of L2 syntactic development, which fits a Weak Continuity view of the initial state. Individual differences between learners' interlanguage development seems to relate to their use of memorised FEs derived from their classroom input. There is evidence that usage of these expressions has instantiated schematic learning of similar functional structures, as well as knowledge of their underlying syntactic properties more generally. In the FPC, there is more evidence of lexically-specific material outside of the FEs (i.e. *où est* and *qu'est ce que*). This is predicted by usage-based models, as the FPC data is a more concentrated picture of learners' L2 productions at the initial stages of learning, where reliance on formulaic material is presumed to be at its highest.

Syntactically complex FEs as 'bootstrapping' mechanisms into L2 knowledge is analysed under the MOGUL framework. Through more frequent use of formulaic material and schematic generalisation across similar functional structures at the initial state, learners become better equipped to make inferences about the L2 syntactic feature specifications involved in its generation. This is a result of their higher activation levels which compete with the existing L1 structures, which quicker allows

them to move from lexical to functional category acquisition.

10.2 Contributions of the current research

This research offers several contributions to the field of SLA. Most broadly, it highlights the importance of prototypical formulaic language in classroom SLA, both as functional communicative tools at the initial state and gateways to grammatical development thereafter (Fillmore, 1979; Myles et al., 1998, 1999; Ellis, 2012). Related to this, it demonstrates how detailed interlanguage analysis of longitudinal production data can bring evidence to bear on the nature of the L2 initial state, in terms of the Strong vs. Weak continuity debate within generative SLA. In particular, it highlights the importance of identifying potentially formulaic/lexically-specific material as a consequence for this debate, and provides a systematic methodology for doing so.

Secondly, the study brings evidence to support the usage-based notion that some L2 development proceeds from the formulaic to the more schematic. In this regard, the analysis is unique by comparing two complimentary longitudinal corpora that differ in the target language and time span that learners are followed. This has allowed for a wider-ranging discovery of the kind of formulaic/schematic material that is typical of certain developmental stages in classroom SLA- irrespective of the target language being acquired. This is an interesting contribution to usage-based traceback studies who typically focus on an in-depth analysis of one or two learners acquiring the same target language over a small period of time.

Thirdly, and perhaps most significantly, the research has offered a methodological demonstration of how the gap between generative and usage-based approaches to SLA can be bridged to investigate the interplay between input, usage and L2 syntactic knowledge. Through drawing on both frameworks, the study has been able to provide a more comprehensive description of the developmental trends observed in

the longitudinal datasets than an analysis with either framework could have achieved independently. Usage-based notions of formulaicity and schematisation have allowed for the identification of formulaic material at initial data collection rounds, and related schematic material in the later rounds. Grammatical frameworks based on generative linguistic theory have allowed for the identification of a common developmental trajectory for both sets of learners based on shared underlying syntactic properties outside of schematic learning. They have also allowed to discover an effect of formulaic language on L2 syntactic acquisition that goes beyond surface-level schematic development. In this regard, the study also offers a theoretical suggestion as to how this effect might happen. It demonstrates that processing approaches to SLA, in particular MOGUL, are ideal frameworks to try and make sense of the interaction between L2 usage and modular syntactic development.

Section 10.3 now moves to addresses some limitations of the methodology and analysis and any implications this has for my proposed interpretation of the observed developmental trends.

10.3 Limitations

As this study's focus has been on the relationship between use of formulaic language and L2 syntactic development, it has concentrated primarily on describing the interplay of these concepts. However, there are other factors that could have contributed to the developmental trends observed in the corpora that the study cannot account for but must address here. Previous research on individual differences in SLA is vast, and several factors have been identified as potential contributors to differential levels of L2 development. These include language aptitude, motivation and memory capacities (see for example Dörnyei 2009). Whilst variables such as age, linguistic environment, quality of input/instruction and L2 task type were more or less homogeneous for learners in both corpora, it could be that their differing uses of FEs/utterance schemas reflects other cognitive and/or affective factors that are

responsible for the observed differences in development.

For example, the FEs at the initial state have been analysed as perceptual lexicalised units, constituting explicit knowledge recalled from declarative/working memory. Via procedural memory, learners can extract and generalise related utterance schemas across similar functional structures. Through increased practicing, this can eventually result in implicit knowledge of these surface forms' underlying computational properties, which allows for the automatic use of these forms in comprehension and production. There is now a large body of evidence showing that working/phonological memory capacities are indeed correlated with L2 grammar acquisition both for L2 children (Ellis, 1996; Verhagen et al., 2015; Verhagen and Leseman, 2016) and L2 adults (Tagarelli et al., 2011). More specifically, declarative memory abilities are considered a significant predictor of development at early stages of acquisition, and procedural memory abilities a significant predictor of development subsequently after this (Morgan-Short et al., 2014). These notions underpin the assumption that the temporary storage of strings of language in working memory provides a database of structures from which learners can generalise and abstract grammatical patterns (Speidel, 1993; Ellis and Sinclair, 1996; Ellis, 2012). It is clear how these concepts align with the developmental trends observed in the longitudinal corpora. Greater working memory capacities are a possible explanation as to why some learners were able to produce the FEs/ lexically-specific utterance schemas more frequently at early stages in the BELC and FPC. Likewise, greater procedural memory abilities are a possible explanation as to why certain learners of both corpora were subsequently able to extract material from these expressions and extend these to similar functional structures. The question, though, is to whether learners' FE use and their later grammatical development are both *independent* consequences of learners' differing memory capacities/abilities. That is, greater abilities in declarative/procedural memory would predict better progress in the L2 over both a 2 year period (FPC) and seven year period (BELC), irrespective of use of the FEs.

Similarly, motivation is of central importance in SLA as it provides the initial interest into L2 learning and the driving force to sustain this process thereafter (Dörnyei, 2009). It could likewise be the case that the observed differences in L2 development are down to learners' differing levels of motivation (whether intrinsic or extrinsic) throughout the language learning process.

In fact, there are an abundance of person variables, both cognitive (intelligence, language learning aptitude, memory capacities and speed) and affective (motivation, anxiety, emotion) that have the potential to differentiate learners regarding their success in instructed SLA. To consider even half of these based on longitudinal corpus data alone would be superfluous and speculative at best, which is far beyond the scope of the current study. Working memory capacity is typically measured via serial word/non-word repetition/recall tasks, which were not carried out with any of the corpus learners under analysis. Access to affective measurements is equally unavailable; it is impossible to know, for example, how motivated or anxious each learner was throughout the recording periods.

The principle aim of this study is to discover patterns and trends in L2 development based on analyses of the longitudinal corpora, and consequently explain them with reference to established grammatical frameworks and concepts of acquisition. The developmental trajectory found in both the BELC and FPC proceeds from lexical to functional category knowledge, and it is clear that learners who show more productive use of FEs and associated utterance schemas are those that show the most grammatical development within this trajectory, irrespective of the factors underlying learners' differing uses of these expressions or language capabilities in general. Therefore, this study's account of how use of complex FEs can potentially bootstrap learners into an acquisition of L2 syntactic knowledge, remains the same, regardless of whether this effect has cognitive or affective antecedents.

In addition, it is necessary to reiterate that both corpora are only a snapshot of these learners' L2 capabilities at particular points in time. In corpus studies, the production of a specific form may not necessarily reflect a learner's competence; the absence of a form does not entail a lack of knowledge on the learner's part (Grondin and White, 1996; Lozano, 2021). As such, any links drawn between corpus data and the minds of learners should be tentative (Siyanova-Chanturia, 2015). Similarly, learners may have been exposed to other holistically taught prototypical expressions in their EFL classroom that they did not repeat/produce in any of the data collection rounds. This study initially extracted FEs from a limited number of representative textbooks and over the first half of spoken tasks only. It is extremely likely that, along with these expressions, others were presented to learners in holistic form in different sections of these textbooks and in different, more intermediate level teaching materials as learners moved through proficiency levels. The study is also based on the analysis of oral production data only, which means that it cannot account for any formulaic material that may be present in learners' written output. This entails that grammatical L2 interrogatives at the latter stages of data collection, or indeed any complex grammatical L2 utterance at these stages, could potentially be patterns derived from the schematic analysis of other constructions not observable in the available production data. Therefore, they might not necessarily reflect evidence for learners' knowledge of underlying syntactic properties.

However, the central point remains unchanged. The aim of the present study is to describe and account for developmental trends that *are* observable in the available production data. Furthermore, the homogeneity and predictability of all learners' linguistic environment (regardless of the L2), help to override this issue. In both corpora, the identified FEs are clearly salient, as all learners are shown to produce the same conventional question expressions upon the same contextual which initially precede associated L2 competence. This salience, along with their inherent

prototypicality and functionality, place these expressions as prime candidates for 'acquisitional seeds' (Ellis, 2012; Ellis et al., 2015; Myles and Cordier, 2017). In the BELC, little other formulaic material was identified in learners' transcripts at these stages, and in the FPC, only two other combinations were identified as potential chunks in the latter rounds (*où est* and *qu'est ce-que*). In both corpora, there is a clear correlation between learners' use of these expressions and their later associated grammatical development.

It could also the case that each transcript is somewhat indicative of a learners' L2 behaviour/ability in general at that specific point in time. For example, those learners who manage to recall FEs at the early ages of the BELC are likely those who are interacting more frequently with similar complex expressions inside the EFL classroom, just like those learners of the FPC who move on from the FEs and interact with other potential chunks such as *où est* and *qu'est ce que*. Extending this concept to all data collection rounds would mean it would be these learners who are also capable of producing other complex L2 surface forms that were not captured in the later data collection periods. That is, the results of the analysis and subsequent implications for the role of FEs in the development of L2 knowledge would remain the same, only on a larger scale.

A related point is the issue of learners' proficiency levels in both corpora across the data collection periods. It is fairly well established in SLA that proficiency level and syntactic complexity, specifically in terms of degree of subordination, are closely linked (e.g., Ortega 2003). Learners have been shown to move from co-ordination through subordination to phrasal elaboration as proficiency rises (Wolfe-Quintero et al., 1998). This relationship has largely been investigated with written data across a variety of L2s with learners from different L1s (e.g., Kuiken and Vedder 2019). Although both the FPC and BELC learners can be classified as 'beginners' at the initial data collection rounds, unfortunately, the corpus metadata did not specify

learners' proficiency levels as the data collection progressed, and no such measures were carried out in the present study. However, the comparison between the trends I have identified and discussed here and various measurements of learners' proficiency (lexical and/or syntactic) presents an interesting avenue to be investigated further, particularly with regard to oral/speaking abilities. It is clear from both analyses that some learners show a rise in proficiency level across the data collection periods. For example, they are shown to move from producing simple and co-ordinated structures to more subordinate ones, and shown to use more diverse lexis. It would be interesting to quantify this development at each round of data collection in terms of proficiency measures, and investigate how this interplays with learners' usage of FEs, utterance schemas and knowledge of underlying syntactic properties. Furthermore, the comparison of different proficiency measures in this regard (i.e., lexical vs. grammatical vs. global) would surely lead to interesting insights regarding the role of formulaic material in the L2 progression of learners more generally.

10.4 Implications for future research

In the previous sections, the contributions of the present study to the field of SLA as well as its limitations were discussed, which lead to implications for future research of a similar kind. I first state some implications based on the study's contributions, and then some based on its limitations.

Firstly, the study advocates a multiparadigmatic approach to SLA, encouraging researchers to incorporate methodologies and concepts from different linguistic frameworks in order to uncover developmental patterns. In this regard, I believe that the study of formulaic language and its relationship with emerging L2 creativity provides an ideal test ground for the integration of usage-based and generative models of SLA, which can help to better understand the interplay between input, usage and L2 linguistic knowledge. The study has offered a theoretical point of

depature in this regard. If it is that the conceptualisation of FEs as 'perceptual units' (in whichever specific model) is on the right track, the exact point in which they become available for analysis and the measurement of their potential influence on core properties of grammar remain interesting but open areas of enquiry.

The results of the study obtained as a consequence of a multiparadigmatic approach, as well as highlighting the strengths of both generative and usage-based models, also bring to light areas in which the traditional methodologies/analyses of each framework could be improved. It shows how generative studies can benefit from longitudinal corpus-based methodologies when looking to investigate what learners actually *do* with the input they receive. It also suggests that the concept of formulaic language, traditionally viewed as a peripheral phenomenon alongside the independent development of L2 syntax, is an important aspect of determining the nature of the L2 initial state and learners' developmental trajectory thereafter. For usage-based analyses, the study demonstrates how traceback methodology could also focus on the nature of learners' interlanguage outside of particular formulaic and schematic constructions. This could bring to light common developmental trends that occur alongside of schematic development, for which the framework could provide an independent explanation for or perhaps link to more abstract properties of schematic constructions that are yet to be developed.

Based on its limitations, the study suggests that a future line of research could complement longitudinal classroom production data with in-depth analysis of the same learners' corresponding input/classroom activities. As well as elicited spoken tasks, learners' classroom activities/behaviour could be followed and documented throughout the same data collection period. This would give for a more precise understanding of the interplay between schematic learning and L2 development, specifically, how learners' manipulate classroom input as a gateway to creative L2 use. It would also give more indication as to how other variables (motivation, lan-

guage learning aptitude, anxiety etc.) contribute to individual differences in L2 syntactic development, for which production data alone cannot capture. Similarly, in order to capture the development of L2 knowledge more comprehensively, longitudinal production data could be complimented with longitudinal comprehension and experimental data from the same learners over the same period of time. This would be a fantastic resource for researchers across all SLA frameworks, but to my knowledge, such a dataset is yet to exist.

Bibliography

Aijmer, K. (1996). *Conversational routines in English*. Longman, London.

Aijmer, K. (2009). So er i just sort of i dunno i think it's just because...: A corpus study of 'i don't know' and 'dunno' in learner spoken English. In Jucker, A. H., Schreier, D., and Hundt, M., editors, *Corpora: Pragmatics and Discourse*, volume 68, pages 151–168. Rodopi, Amsterdam.

Alexiadou, A. and Anagnostopoulou, E. (1998). Parametrizing AGR: Word order, V-movement and EPP- checking. *Natural Languages and Linguistic Theory*, 16(3):491–539.

Anderson, N. (2005). L2 learning strategies. In Hinkel, E., editor, *Handbook of Research in Second Language Acquisition*, volume 1, pages 757–771. Routledge.

Backus, A. (2003). Units in codeswitching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41(1):83–132.

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language and Learning*, 59(4):755–795.

Bardovi-Harlig, K. (2019). Formulaic language in second language pragmatics research. In Siyanova-Chanturia, A. and Pellicer-Sánchez, A., editors, *Understanding Formulaic Language. A Second Language Acquisition Perspective*, volume 1, pages 97–114. Routledge.

Bardovi-Harlig, K., Bastos, M., Burghardt, B., Chappetto, E., Nickels, E., and Rose, M. (2010). The use of conventional expressions and utterance length in L2 pragmatics. *Pragmatics and Language Learning*, 12:163–186.

Bardovi-Harlig, K. and Stringer, D. (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research*, 33(1):61–90.

Bhatt, R. and Hancin-Bhatt, B. (2002). Structural minimality, CP and the initial state in second language acquisition. *Second Language Research*, 18(4):348–392.

Biber, Douglas, Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman, London.

Biber, D., Conrad, S., and Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3):371–405.

Bohn, O. (1986). Formulas, frame structures, and stereotypes in early syntactic development: Some new evidence from L2 acquisition. *Linguistics*, 1:185–202.

Bridges, K. and Van Lancker-Sidtis, D. (2013). Formulaic language in Alzheimer's disease. *Aphasiology*, 27(7):799–810.

Burton, G. (2019). *The canon of pedagogical grammar for ELT: a mixed methods study of its evolution, development and comparison with evidence on learner output*. PhD thesis, University of Limerick.

Bybee, J. and Eddington, D. (2006). A usage-based approach to Spanish verbs of 'becoming'. *Language*, 82(2):323–355.

Carroll, S. (2001). *Input and Evidence: The raw material of second language acquisition*. John Benjamins.

Carroll, S. (2010). Explaining how learners extract 'formulae' from L2 input. In Véronique, G., editor, *The Processing of Input in Second Language Acquisition*, volume 1, pages 229–250. John Benjamins.

Carroll, S. and George, A. (2018). What absolute beginners learn from input: From laboratory to classroom research. *Instructed Second Language Acquisition*, 2(2):112–136.

Cazden, C., Cancino, H., Rosansky, E., and Shumann, J. (1975). Second language acquisition sequences in children, adolescents, and adults. Final report, National Institute of Education.

Celaya, M. L. (2019). The emergence and development of syntactic patterns in EFL writing in a school context: A longitudinal study. *Languages*, 4(2):41.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1975). Questions of form and interpretation. In Austerlitz, R., editor, *The Scope of American Linguistics*, pages 159–196. De Gruyter Mouton.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3(1):1–15.

Chomsky, N. (1981). *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Chomsky, N. (1996). *Minimalism*. MIT Press, Cambridge, MA.

Chomsky, N. (2002). *On Nature and Language*. Cambridge University Press.

Clahsen, H. and Felser, C. (2006). Grammatical processing in language learners. *Psycholinguistics*, 27(1):3–42.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale, NJ.

Collins, C. and Stabler, E. (2016). A formalization of minimalist syntax. *Syntax*, 19(1):43–78.

Conklin, K. and Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than non-formulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1):72–89.

Conklin, K. and Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32(1):45–61.

Coulmas, F. (1981). *Conversational Routine. Explorations in prepatterned speech and standardized communication situations*. The Hague, Mouton.

Criado, R. and Sanchez, A. (2009). English Language Teaching in Spain: Do textbooks comply with the official methodological regulations? A sample analysis. *International Journal of English Studies*, 9(1):1.

Cutting, J. and Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, 25(1):57–71.

Damien, L. (2018). Pedagogical shortcomings of Global EFL textbooks and Spanish speaking learners. *Lenguas Modernas (Santiago)*, (51):9–28.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.

Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition*. Oxford University Press.

Doughty, C. and Long, M. (2003). The scope of enquirey and goals of SLA. In Doughty, C. and Long, M., editors, *The Handbook of Second Language Acquisition*, pages 3–16. Blackwell, Malden.

Durrleman, S., Marinis, T., and J, F. (2016). Syntactic complexity in the comprehension of wh-questions and relative clauses in typical language development and autism. *Applied Psycholinguistics*, 37(6):1501–1527.

Edmonds, A. (2014). Conventional expressions: Investigating pragmatics and processing. *Studies in Second Language Acquisition*, 36(1):69–99.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition*, 18(1):91–126.

Ellis, N. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit langauge acquisition. *Studies in Second Language Acquisition*, 24(2):297–339.

Ellis, N. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2):297–339.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2):305–352.

Ellis, N. (2008). Implicit and explicit knowledge about language. In Cenoz, J. and Hornberger, N., editors, *Encyclopedia of language and education: Vol 6. Knowledge about language*, pages 119–131. Springer.

Ellis, N. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32(17):17–44.

Ellis, N. and Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3):370–385.

Ellis, N. and Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, 7(1):188–221.

Ellis, N., M, O., Romer, U., Gries, S., and Wulff, S. (2009). Measuring the formulaicity of language. In *AAAL 2009, the Annual Conference of the American Association of Applied Linguistics*, Denver, CO.

Ellis, N., Simpson-Vlack, R., Romer, U., M, O., and Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. In Granger, S., Gilquin, G., and Meunier, F., editors, *The Cambridge Handbook of Learner Corpus Research*, pages 357–378. Cambridge University Press, Cambridge.

Ellis, N. and Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax. *Quarterly Journal of Experimental Psychology*, 49(1):234–250.

Ellis, R. (2002c). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition*, 24(2):223–236.

Eskildsen, S. (2008). *Constructing a second language inventory-the accumulation of linguistic resources in L2 English.* PhD thesis, University of Southern Denmark.

Eskildsen, S. (2009). Constructing another language: Usage-based linguistics in second language acquisition. *Applied Linguistics*, 30(3):335–357.

Eskildsen, S. (2015). What counts as a developmental sequence? Exemplar-based L2 learning of English questions. *Language Learning*, 65(1):33–62.

Eskildsen, S. (2020). Creativity and routinisation in L2 English: Two usage-based case-studies. In Lowie, W., Michel, M., Rousse-Malpat, A., Keijzer, M., and Steinkrauss, R., editors, *Usage-based dynamics in second language development*, pages 107–129. Multilingual Matters.

Eskildsen, S. and Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In Nenonen, M. and Niemi, S., editors, *Collocations and Idioms, 1: Papers from the First Nordic Conference on Syntactic Freezes*, pages 86–99. Joensuu University Press.

Felix-Brasdefer, J. and Hasler-Barker, M. (2015). Complimenting in Spanish in a short-term study abroad context. *System*, 48(1):75–85.

Field, A., Miles, J., and Field, Z. (2012). *Discovering statistics using R*. Sage.

Fillmore, C. (1979). On fluency. In Fillmore, C., Kempler, D., and Wang, W. S.-Y., editors, *Individual Differences in Language Ability and Language Behaviour*, pages 85–101. Elsevier, 1 edition.

Fillmore, C., Kay, P., and O'connor, M. (1988). Regularity and idiomaticity in grammatical constructions. The case of let alone. *Language*, 64(3):501–538.

Fletcher, P. (1985). *A Child's Learning of English*. Blackwell, Oxford, UK.

Fodor, J. (1983). *The modularity of mind*. MIT Press.

Fodor, J. (1998). Parsing to learn. *Journal of Pscholinguistic Research*, 27(1):339–374.

Gass, S. (2013). *Second Language Acquisition: An Introductory Course*. Routledge, New York and London.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In Cowie, A., editor, *Phraseology: Theory, Practice and Applications*, pages 145–160. Clarendon Press.

Granger, S. (2021). Have learner corpus research and second language acquisition finally met? In Le Bruyn, B. and Paquot, M., editors, *Learner corpus research meets second language acquistion*, pages 243–257. Cambridge University Press, Cambridge.

Grondin, N. and White, L. (1996). Functional categories in child L2 acquisition of French. *Language Acquisition*, 5(1):1–34.

Grüter, T. (2006). Another take on the L2 initial state: Evidence from comprehension in L2 German. *Language Acquisition*, 13(4):287–317.

Hamann, C. (2006). Speculations about early syntax: The production of wh-questions by normally developing French children and french children with SLI. *Catalan Journal of Linguistics*, 5(1):143–189.

Hanania, E. and Gradman, H. (1977). Acquisition of English structures: A case study of an adult native speaker of Arabic in an English-speaking environment. *Language Learning*, 27(1):75–91.

Hawkins, R. (2001). *Second Language Syntax: A Generative Introduction*. Wiley Blackwell.

Hawkins, R. (2008). The nativist perspective on second language acquisition. *Lingua*, 118(4):465–477.

Haznedar, B. and Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In Hughes, E., Hughes, M., and Greenhill, A., editors, *Proceedings of the 21st Annual Boston University Conference on Language Development*, volume 21, pages 257–268.

Herschensohn, J. (2000). *The Second Time Around: Minimalism and L2 Acquisition*. John Benjamins, Amsterdam-Philadelphia.

Hicks, G. and Dominguez, L. (2020). A model for L1 grammatical attrition. *Second Language Research*, 36(2):143–165.

Hopper, P. (1987). Emergent grammar. *Annual Meeting of the Berkelehy Lingusitics Society*, 13:139–157.

Horbowicz, P. and Nordanger, M. (2021). Epistemic constructions in L2 norwegian: A usage-based longitudinal study of formulaic and productive patterns. *Language and Cognition*, 13(3):438–466.

Hulk, A. and Zuckerman, S. (2000). The interaction between input and economy: Acquiring optionality in French wh-questions. In Howell, S., Fish, S., and Keith-Lucas, T., editors, *Proceedings of the 24th Annual Boston University Conference on Language Development*, pages 438–449, Somerville MA. Cascadilla Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. MIT Press, Cambridge, MA.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, USA.

Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically-developing and SLI learners of L1-french. *Lingua*, 121(3):339–51.

Jeremias, J. (1982). The 'social' component in the English classroom: Semantic and pragmatic considerations. *Anglo-American Studies*, 2(1):63–73.

Juffs, A. and Fang, S. (2022). A generative approach to the instructed second language acquisition of Spanish 'se'. *Language Learning*, 72(1):83–124.

Keller, M. (2020). Compositionality: Evidence from code-switching. In Piirainen, E., Filatkina, N., Stumpf, S., and Pfeiffer, C., editors, *Formulaic Language and New Data*, volume 3, pages 197–222. De Gruyter.

Knowlton, B. and Moody, T. (2008). Procedural learning in humans. In Bryne, J., editor, *Learning and memory: A comprehensive reference*, pages 321–340. Oxford.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Pergamon Press Inc.

Krashen, S. and Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning*, 28(2):283–300.

Kuiken, F. and Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and l1 writing in dutch, italian and spanish. *International Journal of Applied Linguistics*, 29(2):192–210.

Kuiper, K. (1996). Smooth talkers: The linguistic performance of auctioneers and sportscasters. In Leeds-Hurwitz, W. and Sigman, S., editors, *Everyday communication: Case studies of behavior in context*, pages 14–109. Erlbaum, Mahwah, NJ.

Kuiper, K., Van Egmond, M., Kempen, G., and Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2(3):313–357.

Langacker, R. (1987). Nouns and verbs. *Language*, 63(1):53–94.

Langacker, R. (2000). *Grammar and Conceptualization*, volume 14. Walter de Gruyter.

Langacker, R. (2013). *Essentials of Cognitive Grammar*. Oxford University Press.

Lardiere, D. (1998a). Case and tense in the 'fossilized' steady state. *Second Language Research*, 14(1):1–26.

Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(4):359–375.

Lardiere, D. (2008). Feature assembly in second language acquisition. In Liceras, J., Zobl, H., and Goodluck, H., editors, *The Role of Formal Features in Second Language Acquisition*, pages 106–140. Lawrence Erlbaum Associates, New York.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. Routledge, New York, 1 edition.

Larson-Hall, J. and Herrington, R. (2009). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, 31(3):368–390.

Larson-Hall, J. and Mizumoto, A. (2020). Using statistical analysis software (R, SPSS). In McKinley, J. and Rose, H., editors, *The Routledge Handbook of Research Methods in Applied Linguistics*, chapter 32, pages 385–397. Routledge, London, 1 edition.

Larson-Hall, J. and Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(1):127–159.

Le Bruyn, B. and Paquot, M. (2021). Learner corpus research and second language acquisition: An attempt at bridging the gap. In Le Bruyn, B. and Paquot, M., editors, *Learner Corpus Research Meets Second Language Acquistion*, pages 1–9. Cambridge University Press.

Leech, G. (1983). *Principles of pragmatics*. Longman, New York.

Lesonen, S., Steinkrauss, R., Suni, M., and Verspoor, M. (2020). Lexically specific vs productive constructions in L2 finnish. *Language and Cognition*, 12(3):526–563.

Lewis, M. (1993). *The Lexical Approach*. Hove: Language teaching publications.

Lidz, J. and Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, 1(1):333–353.

Lieven, E., Behrens, H., Speares, J., and Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30(2):333–370.

Lieven, E., Salomo, D., and Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3):481–507.

Lindholm, C. and Wray, A. (2011). Proverbs and formulaic sequences in the language of elderly people with dementia. *Dementia*, 10(4):603–623.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press, Edinburgh.

Littlejohn, A. (2011). The analysis of language teaching materials: Inside the trojan horse. In Tomlinson, B., editor, *Materials Development in Language Teaching*, pages 179–211. Cambridge University Press, Cambridge, 2 edition.

Long, M. (1996). The role of linguistic environment in second language acquisition. In W. C, R. and T. K, B., editors, *Handbook of second language acquisition*, pages 413–468. Academic Press, San Diego, CA.

Lorenz, G. (1999). *Adjective Intensification: learners versus native speakers: a corpus study of argumentative writing*. Rodopi.

Lozano, C. (2021). Generative approaches. In Tracy-Ventura, N. and Paquot, M., editors, *The Routledge Handbook of Second Language Acquisition and Corpora*, pages 213–227. Routledge, New York and London.

Lyons, J. (1969). *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. (2009). Segregating the core computational faculty of human language from working memory. In *Proceedings of the National Academy of Sciences*, volume 106, pages 8362–8367.

Marsden, H., Whong, M., and Gil, K. (2018). What's in the textbook and what's in the mind: Polarity item 'any' in learner English. *Studies in Second Language Acquisition*, 40(1):91–118.

McDonough, J., Shaw, C., and Masuhara, H. (2013). *Materials and Methods in ELT. A Teacher's Guide*. Wiley Blackwell, Oxford, UK.

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11(2):113–128.

Menkabu, A. and Harwood, N. (2014). Teachers' conceptualization and use of the textbook on a Medical English course. In Harwood, N., editor, *English Language Teaching Textbooks. Content, Consumption, Production*, pages 145–177. Palgrave.

Mitchell, R. and Martin, C. (1997). Rote learning, creativity and understanding in classroom foreign language teaching. *Language Teaching Research*, 1(1):1–27.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Scheutz, K., Carpenter, H., and Wong, P. C. (2014). Declarative and procedural memory as individual differences

in second language acquisition. *Bilingualism: Language and Cognition*, 17(1):56–72.

Muñoz, C. (2006). *Age and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.

Myles, F. (2004). From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society*, 102(2):139–168.

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4):373–391.

Myles, F. (2015). Second language acquisition theory and learner corpus research. In Granger, S., Gilquin, S., and Meunier, F., editors, *The Cambridge Handbook of Learner Corpus Research*, volume 3, pages 309–332. Cambridge University Press.

Myles, F. and Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition*, 3(28):3–28.

Myles, F., Hooper, J., and Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3):323–363.

Myles, F., Mitchell, R., and Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition*, 21(1):49–80.

Nattinger, J. and DeCarrico, J. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.

Newmayer, F. (2008). Universals in syntax. *Linguistic Review*, 25(1):1–48.

O'Grady, W. (2011). Syntax: The analysis of sentence structure. In O'Grady, W., Archibald, J., and Katamba, F., editors, *Contemporary Linguistics: An introduction*. Pearson Education Limited, 2 edition.

Ortega, L. (2003). Syntactic complexity measures and their relationship to l2 proficiency: A research synthesis of college-level l2 writing. *Applied Linguistics*, 24(4):492–518.

Pallant, J. (2010). *SPSS Survival Manual: A step by step guide to data analysis using SPSS*. Open University Press, 4 edition.

Paquot, M. and Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32(32):130–149.

Paquot, M. and Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1):61–94.

Paradis, M. (2004). *Neurolinguistic Theory of Bilingualism*. John Benjamins, Amsterdam-Philadelphia.

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. John Benjamins.

Peters, A. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language*, 53(1):560–573.

Plonsky, L. and Oswald, F. (2014). How big is 'big'? Interpreting effect sizes in l2 research. *Language Learning*, 64(4):878–912.

Plunkett, B. (1999). Targeting complex structure in French questions. In Greenhill, A., Littlefield, H., and Tano, C., editors, *Proceedings of the 23rd Annual Boston University Conference on Language Development*, pages 764–775, Somerville MA. Cascadilla Press.

Poeppel, D. and Wexler, K. (1993). The full competence hypothesis of clausal structure in early German. *Language*, 69(1):1–33.

Porte, G. K. (2002). *Appraising research in second language learning: A practical approach to critical analysis of quantitative research*. John Benjamins, Philadelphia, PA.

Prévost, P. (2009). *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. John Benjamins.

Prévost, P., Strik, N., and Tuller, L. (2014). Wh-questions in child L2 french: Derivational complexity and its interactions with L1 properties, length of exposure, age of exposure, and the input. *Second Language Research*, 30(2):225–250.

Prévost, P., Tuller, L., Scheidnes, M., Ferré, S., and Haiden, M. (2010). Computational complexity effects in the acquisition of wh-questions in child L2 french. In Dominguez, L. and Guijarres-Fuentes, P., editors, *New directions in langauge acquisition: Romance langauges in the generative perspective*, pages 415–443. Cambridge Scholar Publisher.

Przymus, S. (2023). Code-switching is metaphor, translanguaging is metonymy: A transdisciplinary view of bilingualism and its role in education. *International Journal of Bilingual Education and Bilingualism*, 27(4):1–17.

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

Radford, A. (2009). *Analysing English Sentences: A Minimalist Approach*. Cambridge University Press.

Radford, A., Atkinson, M., Britain, D., Clahsen, H., and Spencer, A. (1999). *Linguistics: An introduction*. Cambridge University Press.

Rankin, T. (2009). Verb second in advanced L2 english: A learner corpus study. In *Proceedings of the 10th generative approaches to second language acquisition conference*, Somerville MA. Cascadilla Proceedings Project.

Rastelli, S. and Gil, K. (2018). No fear of George Kinglsley Zipf: Language classroom, statistics and universal grammar. *Instructed Second Language Acquisition*, 2(2):242–264.

Ricci, V. (2005). Fitting distributions with R. Retrieved from https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf.

Richards, J. (2005). The role of textbooks in a language program.

Rizzi, L. (2009). Some elements of syntactic computations. In Bickerton, D., Szathmary, E., and Lupp, J., editors, *Biological Foundations and Origin of Syntax*, volume 3, pages 63–89. MIT Press.

Roehr-Brackin, K. (2014). Explicit knowledge and processes from a usage-based perspective: The developmental trajectory of an instructed L2 learner. *Language Learning*, 64(4):771–808.

Ross, J. (1967). *Constraints on Variables in Syntax*. PhD thesis, Massachusetts Institute of Technology.

Rothman, J., Bayram, F., Cunnings, I., and Gonzalez Alonso, J. (2019). Formal linguistics approaches to adult second language acquisition and processing. In Schwieter, J. and Benati, A., editors, *The Cambridge Handbook of Language Learning*. Cambridge University Press.

Rothman, J. and Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, 40(2):417–442.

Rowland, C. and Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquistion: 'what children do know?'*. *Journal of Child Language*, (27):157–181.

Rowland, C. and Pine, J. M. (2003). The development of inversion in wh-questions: A reply to Van Valin. *Journal of Child Language*, 30(1):197–212.

Samian, S. and Tavakoli, M. (2012). The relationship between Iranian EFL learners rote learning strategy use and their level of proficiency. *Journal of Language Teaching and Research*, 3(4):625–631.

Santi, A. and Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *Neuroimage*, 37(1):8–17.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In Wolfson, N. and Judd, E., editors, *Sociolinguistics and Language Acquisition*, volume 137, pages 137–174. Rowley, MA.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Springer.

Schmitt, N. and Carter, R. (2004). Formulaic sequences in action. In Schmitt, N., editor, *Formulaic sequences: Acquisition, Processing and Use*, pages 1–22. John Benjamins, Philadelphia, PA.

Schmitt, N., Grandage, S., and Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In Schmitt, N., editor, *Formulaic sequences: Acquisition, Processing and Use*, pages 127–151. John Benjamins.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behaviour. *Studies in Second Language Acquisition*, 15(2):147–163.

Schwartz, B. and Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1):40–72.

Sharwood-Smith, M. and Truscott, J. (2014). *The multilingual mind: A modular processing perspective*. Cambridge University Press.

Simpson-Vlack, R. and Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4):487–512.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Lingiustics and Linguistic Theory*, 11(2):285–301.

Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2):1–22.

Slabakova, R. (2016). *Second Language Acquisition*. Oxford University Press.

Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language*, 13(37):69–91.

Sportiche, D., Koopman, J., and Stabler, E. (2014). *An Introduction to Syntactic Analysis and Theory*. Wiley Blackwell, West Sussex, UK.

Sprenger, S., Levelt, W., and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2):161–184.

Staudte, R. and Sheather, S. (1990). *Robust Estimation and Testing*. Wiley Blackwell, New York.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Mahwah, NJ, 3rd edition.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass, S. and Madden, C., editors, *Input in second language acquisition*, pages 235–253. Newbury House, Rowley, Mass.

Tagarelli, K., Borges Mota, M., and Rebuschat, P. (2011). The role of working memory in the implicit and explicit learning of languages. In Carlson, I., Holscher, C., and Shipley, T., editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 2061–2066.

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

Tomasello, M. (2009). Universal Grammar is dead. *Behavioral and Brain Sciences*, 32(5):470–471.

Tomasello, M. and Abbot-Smith, K. (2002). A tale of two theories: Response to fisher. *Cognition*, 83(2):207–214.

Tomlinson, B. (1998). Glossary of basic terms for materials development in language teaching. In Tomlinson, B., editor, *Materials Development in Language Teaching*, pages 9–18. Cambridge University Press, Cambridge.

Towell, R. (2012). Complexity, accuracy and fluency from the perspective of psycholinguistic second language acquisition research. In Housen, A., Kuiken, F., and Vedder, I., editors, *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, volume 32, pages 47–71. John Benjamins.

Towell, R. (2014). Formulaic and proceduralised language in the initial and advanced stages of learning French. In Lindquist, C. and Bardel, C., editors, *Acquisition of French as a Second Language*, pages 95–115. John Benjamins.

Tracy-Ventura, N. and Myles, F. (2015). The importance of task variability in the design of learner corpora for SLA research. *International Journal of Learner Corpus Research*, 1(1):58–95.

Truscott, J. (2006). Optionality in second language acquisition: A generative, processing-oriented account. *International Review of Applied Linguistics in Language Teaching*, 44(4):311–330.

Truscott, J. (2015). Consciousness in SLA: A modular perspective. *Second Language Research*, 31(3):413–434.

Truscott, J. (2017). Modularity, working memory, and second language acquisition. A research program. *Second Language Research*, 33(3):313–323.

Truscott, J. and Sharwood-Smith, M. (2004). Acquisition by processing: A modular

perspective on language development. *Bilingualism: Language and Cognition*, 7(1):1–20.

Tummers, J., Heylen, K., and Geeraerts, D. (2005). Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Lingiustics and Linguistic Theory*, 1(2):225–261.

Turner, J. (2014). *Using statistics in small-scale language education research: Focus on non-parametric data*. Routledge, New York.

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1):231–270.

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In Sanz, C., editor, *Adult Second Language Acquisition*, pages 141–178. Georgetown University Press.

Underwood, G., Schmitt, N., and Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt, N., editor, *Formulaic sequences: Acquisition, Processing and Use*, pages 153–172. John Benjamins, Amsterdam.

Vainikka, A. (1994). Case in the development of English syntax. *Language Acquisition*, 3(1):257–325.

Vainikka, A. and Young-Scholten, M. (1998). The inital state in the L2 acquisition of phrase structure. In Flynn, S., Martohardjono, G., and O'Neil, W., editors, *The Generative Study of Second Language Acquisition*, pages 17–34. Erlbaum.

Van Lancker-Sidtis, D., Choi, J., Alken, A., and Sidtis, J. (2015). Formulaic language in Parkinson's disease and Alzheimer's disease: Complementary effects of subcortical and cortical dysfunction. *Journal of Speech, Language and Hearing research*, 58(5):1493–1507.

Van Lancker-Sidtis, D. and Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and Communication*, 24(3):207–240.

VanPattern, B. and Rothman, J. (2015). What does current generative theory have to say about the explicit-implicit debate? In Rebuschat, P., editor, *Implicit and Explicit Learning of Languages. Studies in Bilingualism*, pages 91–116. John Benjamins.

Verhagen, J. and Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquistion of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, 141(1):65–82.

Verhagen, J., Messer, M., and Leseman, P. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, 65(2):417–448.

Verspoor, M., Lowie, W., and Wieling, M. (2021). L2 developmental measures from a dynamic perspective. In Le Bruyn, B. and Paquot, M., editors, *Learner Corpus Research Meets Second Language Acquistion*, pages 172–190. Cambridge University Press.

Waters, G. and Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *Quarterly Journal of Experimental Psychology*, 57(1):129–163.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2):180–205.

White, L. (1989). *Universal Grammar and Second Language Acquisition*. John Benjamins, Amsterdam-Philadelphia.

White, L. (1990). Second language acquisition and Universal Grammar. *Studies in Second Language Acquisition*, 12(2):121–133.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge University Press.

White, L. (2012). *Second Language Acquisition and Universal Grammar*. Cambridge University Press.

White, L. and Juffs, A. (1998). Constraints on wh-movement in two different contexts of non-native language acquisition: Competence and processing. In Flynn, S., Martohardjono, G., and O'Neil, W., editors, *The Generative Study of Second Language Acquisition*, pages 111–130. Erlbaum, Hillsdale, NJ.

Wilcox, R. (2003). *Applying contemporary statistical techniques*. Academic Press, San Diego, CA.

Wolfe-Quintero, K., Inagaki, S., and Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* University of Hawai'i Press, Honolulu, HI.

Wong-Fillmore, L. (1976). *The second time around: Cognitive and social strategies in second language acquisition*. PhD thesis, Stanford University, Palo Alto, CA.

Wray, A. (2002). *Formulaic Langauge and the Lexicon*. Cambridge University Press, Cambridge.

Wray, A. (2008). *Formulaic Langauge: Pushing the boundaries*. Oxford University Press.

Wray, A. and Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a guage of native-like competence. In Meunier, F. and Granger, S., editors, *Phraseology in Foreign Language Learning and Teaching*, pages 123–147. John BenjaminsBenjamins.

Yang, C. (1999). Unordered merge and its linearization. *Syntax*, 2(1):38–64.

Yang, C. and Montrul, S. (2017). Learning datives: The tolerance principle in monolingual and bilingual acquisition. *Second Language Research*, 33(1):119–144.

Zyzik, E. (2009). The role of input revisited: Nativist versus usage-based models. *L2 Journal: An electronic refereed journal for foreign and second language educators*, 1(1):42–61.