



Fixed wh-expressions in classroom second language acquisition: databases of computational properties or utterance schemas?

RESEARCH

THOMAS HAMMOND

KOOK-HEE GIL

*Author affiliations can be found in the back matter of this article

WHITE ROSE
UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

ABSTRACT

This study adopts concepts from two competing approaches to second language acquisition (SLA) (usage based vs. generative) to analyse the effect of formulaic expressions (FEs) on learners' L2 syntactic development. Using spoken transcripts of the longitudinal Barcelona English Language Corpus (BELC), we identify four learned fixed wh-expressions (FEs^{wh}), which are all produced by learners in advance of respective L2 competence. We measure learners' use of these expressions and the evidence of related computational properties (e.g., wh-movement, *do*-support) and utterance schemas (e.g., [WH + AUX DO + X]) outside this use across a 7-year data collection period.

Adopting a generative analysis, we find that an earlier and more frequent use of FE^{wh} correlates with better L2 knowledge of the expressions' associated computational properties. Then, adopting a usage-based 'traceback' methodology (e.g., Lieven et al., 2003; Eskilsden, 2020), we find that learners accurately produce some L2 interrogatives that share utterance schemas of previously used FE^{wh} that appear in their production data ontogenetically. Utterance schema extraction and generalisation of model surface forms may therefore facilitate the development of the more general L2 feature specifications on the functional categories for which these surface forms exemplify. From this, we argue that such a unified account of learners' L2 development can offer a better description of the trends observed in the corpus than either usage-based or generative models can do independently.

CORRESPONDING AUTHOR:

Thomas Hammond

University of Sheffield, UK

tahammond1@sheffield.ac.uk

KEYWORDS:

formulaic expressions; L2 syntactic development; usage-based; generative

TO CITE THIS ARTICLE:

Hammond, T., & Gil, K.-H. (2023). Fixed wh-expressions in classroom second language acquisition: databases of computational properties or utterance schemas? *Journal of the European Second Language Association*, 7(1), 75–92. DOI: <https://doi.org/10.22599/jesla.100>

1. INTRODUCTION

This study examines the effect of formulaic expressions (henceforth FEs) on the development of L2 syntax in a longitudinal learner corpus. There is a large body of literature in Applied Linguistics concerned with the identification and role of FEs in SLA (e.g., Eskildsen, 2015, 2020; Horbowicz & Nordanger, 2021), as well as a growing interest in the interaction of input and usage on the acquisition of modular linguistic knowledge more generally (e.g., Lidz & Gagliardi, 2015, Truscott, 2017). Longitudinal studies investigating the relationship between FE use and later syntactic development have been more widely explored within usage-based (UBL) frameworks, regarding the extent to which learners' L2 utterances at later stages of acquisition can be traced back ontogenetically to previously used FEs which embody the same utterance schemas and or/schematic patterns (Rowland & Pine, 2000; Eskildsen, 2015; 2020). Studies of this nature feature less within the generative framework. Those few that have explored this interaction within a classroom context, however, have found evidence for L2 learners using syntactically complex FEs as building-blocks towards creative language use of a similar functionality (Myles et al., 1998). More specifically, Hammond and Gil (*in press*) recently analysed longitudinal production data and found that the use of fixed wh-expressions (henceforth FEs^{wh}) at the initial state seemed to 'bootstrap' learners into an incremental development of L2 phrase structure (i.e., from Verb Phrase (VP) to Tense Phrase (TP) to Complementiser Phrase (CP)). Learners who interacted more with these expressions showed a better L2 knowledge of functional categories T(ense) and C(omplementiser) more generally by the end of the data collection period. Studies of this kind question the consensus that has been held within the generative tradition that, despite FEs being an effective communicative tool, the creative language process develops independently of their use and/or analysis (Carroll, 2010; Bardovi-Harlig & Stringer, 2017).

The present study takes a novel approach to investigating the role of FEs on learners' syntactic development, arguing that a combination of usage-based and generative analyses as outlined above can offer a better insight into this phenomenon than either model can do independently. Analysing a subset of the Barcelona English Learner Corpus (BELC), we show that learners' initial use of memorised FEs^{wh} is facilitative on their later syntactic development both in terms of utterance-schema extraction as well as knowledge of their associated computational mechanisms more generally. We present how an application of both approaches is useful for understanding the role of input and usage on the acquisition of formal linguistic features and discuss the significant role that memorised formulaic language can play in this process.

Section 2 first outlines both generative and usage-based approaches to SLA, specifying the perceived role of FEs in each framework. Section 3 presents the data, and Section 4 analyses the identified FEs as products of abstract computational derivation (generative) and abstract utterance schemas (usage-based). Section 5 presents the results, and Section 6 gives a discussion of these. Section 7 concludes.

2. FORMULAIC EXPRESSIONS IN GENERATIVE VS USAGE-BASED APPROACHES TO SLA

2.1 GENERATIVE APPROACHES TO FES IN SLA

Under generative models, language is modular. 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), which is taken to be an innate endowment of human beings (Collins & Stabler, 2016). Generative second language acquisition (GenSLA) is largely concerned with the interplay between UG, knowledge that comes from the L1, and knowledge that comes from exposure to the target language (the L2) (Rothman & Slabakova, 2018). There are competing theories within the paradigm as to how these aspects interact. For example, there are those models that claim full transfer from the L1 at the initial stages of SLA, (known as the Strong Continuity Hypothesis) (Poeppel & Wexler, 1993) and others that assume an incremental development of phrase structure where the L2 initial state is largely lexical in nature, (known as the Weak Continuity Hypothesis) (Vainikka & Young-Scholten, 1998).

Regardless of the Strong/Weak continuity debate, how exactly L2 input and usage can trigger modular syntactic knowledge is an ongoing line of investigation. Despite an increased interest in exploring this interaction in instructional/classroom contexts (e.g., Marsden et al., 2018), there has been little focus from generative studies to investigate the role of FEs in this capacity, despite these constituting a significant proportion of L2 classroom input (Myles & Cordier, 2017). An exception is Myles and colleagues (Myles et al., 1998; Myles, 2004), who analysed spoken production data of English classroom adolescent learners of L2 French over a period of 2 years. The authors note how at the early stages of data collection, the same learners produced syntactically complex FEs such as *quel âge as tu?* [how old are you?], while at the same time producing ungrammatical sentences in similar functional environments, such as **il age frère?* [he age brother?] (how old is your brother?) that lacked wh-fronting and inversion in the L2. They then checked how learners overextended and modified these expressions over the course of the data collection period to produce similar functional structures. For example, learners were shown to add NPs such as *la fille* [the girl] to the formulaic expression (1a) which led to overextensions such as (1b) before modification led to the correct structure (1c):

- (1) a. *comment t'appelles tu?*
 how call yourself you
 'what is your name?'
 b. **comment t'appelles tu la fille?*
 how call yourself you the girl
 (lit.) 'what is your name the girl?'
 c. *comment s'appelle la fille?*
 how call-herself the girl
 'what's the girl's name?'

The authors concluded that FEs provided learners with a databank of complex structures beyond their initial state grammars, and that learners kept 'working on' these until their current generative grammar (which developed in an incremental fashion) was compatible with them.

In a similar study, Hammond and Gil (in press) recently analysed the spoken production data of 9 classroom longitudinal Spanish/Catalan learners of English over a period of 7 years. They found that learners across the data collection period also made extensive use of highly prototypical wh-expressions derived from their classroom input; *'what is your name?', 'where are you from?', 'how old are you?' and 'where do you live?'*. Like the anglophone learners in Myles et al., (1998), at the initial stages of data collection these expressions were produced in advance of knowledge of their associated syntactic derivations (*wh*-movement, inversion etc.). Unlike Myles' learners, however, Hammond and Gil (in press) found no evidence of learners overextending or modifying these expressions erroneously in similar functional structures. Rather, those learners that interacted more with these expressions at the initial stages of data collection (ages 10 and 12) were quicker to develop a more complex L2 grammar (e.g., VP-TP-CP). Hammond and Gil (in press) interpret syntactically complex fixed expressions as 'bootstrapping' mechanisms into higher syntactic categories, using processing models of SLA (e.g., MOGUL) to explain their results. However, the authors did not conduct a usage-based traceback analysis of the data, so it was unclear whether some of the observed syntactic development can be accounted for via utterance schema extraction and generalisation of the model FE^{wh} forms.

2.2. USAGE-BASED APPROACHES TO FES IN SLA

Rather than a dichotomy of syntax and lexicon, UBL 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 abstraction' (Tummers et al., 2005, pp. 228–229). For UBL, formulaic expressions that are high in frequency, functionality and prototypicality play a central role in SLA. It is argued that a learner's long-term knowledge of such can serve as the 'database' for their language acquisition (e.g., Ellis, 2002). The proposed usage-based learning pattern for both L1 and L2 acquisition is from formulaic expression to utterance schema (known also as semi-fixed or slot and-frame pattern) to fully productive schematic pattern (Ellis, 2012; Horbowicz & Nordanger, 2021). For example, through frequent exposure and usage of the prototypical formulaic exemplar *'where do you live?'*, learners can derive the utterance schema in (2a) before finally acquiring the fully schematic wh-question pattern in (2b):

- (2) a. [where do you VERB]
e.g., *where do you go?*, *where do you pay?*
b. [WH + AUX DO + PRN + VERB]
e.g., *what does he do?*, *when do you go?*

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 expression exemplifies is derivable from its abstract schematic construction. For instance, the utterance schema [do you + X] is equally as derivable as [where do + X] from the exemplar '*where do you live?*'. Utterance schemas can be lexically [what do + X] or categorically [WH + AUX DO + X] specific to their formulaic exemplars, where lexically specific schemas maintain some of the same lexical items and categorically specific ones the more general grammatical category sequencing. One example for L2 acquisition is Eskildsen (2015), who investigated the longitudinal development of L2 English question formation and deduced that their subjects were constructing wh-questions based on more general [WH + COPULAR + X] and [WH + AUX DO + X] utterance schemas derived from their usage. Some example utterances that exemplified these schemas are shown in (3) and (4):

- (3) [WH + COPULAR + X]
a. *what is your name?*
b. *when is your birthday?*
c. *when were you born?*
(4) [WH + AUX DO + X]
a. *where do you live?*
b. *how do you say?*
c. *where does she work?*

A significant component of L2 learning in UBL is therefore the abstraction and subsequent generalisation of FEs, which can be understood as the gradual expansion of varied utterance schema use (Roehr-Brackin, 2014). Importantly, FEs that are identified as having initiated schematic development must precede all other instantiations ontogenically in longitudinal learner data (Lieven et al., 2003). That is, learners must be shown to produce the proposed FEs in advance of any other instantiation of related utterance schemas and/or fully schematic patterns. For example, to reliably argue that '*where do you live?*' has instantiated the utterance schemas [where do + X] or [do you + X] for a particular learner, '*where do you live?*' must appear in this learner's data before all other utterances which embody these schematic frames.

2.3. RESEARCH QUESTIONS

The present study analyses a subset of the Barcelona English Language Corpus (BELC) to examine how learners' use of fixed wh-expressions (FEs^{wh}) interacts with their corresponding L2 syntactic development. To further explore the trends observed in Hammond and Gil (in press) with a novel analysis that considers both generative and usage-based frameworks, we distinguish the following research questions:

- (i) Does use of identified FEs^{wh} lead to better L2 knowledge of the expressions' underlying computational properties as conceptualised under generative frameworks?
- (ii) Can learners' L2 interrogatives be traced back to utterance schemas of FEs^{wh} in learners' production data ontogenetically?

From the results of Hammond and Gil (in press), we can predict that the current study will observe a correlation between FE^{wh} use and better L2 knowledge of the expressions' specific computational properties involved in their generation (i.e., wh-movement, T-C movement, A-movement etc), despite the general consensus amongst generative studies positing no relationship between FE use and L2 acquisition. This is because Hammond and Gil (in press) found that learners who more frequently used FEs^{wh} were the ones whose L2 grammars showed an incremental development quicker moving from a bare VP to TP to CP stage. From the results of past usage-based longitudinal studies, we can predict that learners' L2 interrogatives can be traced back to utterance schemas of previously used FEs^{wh} in their production data.

The current paper aims to bring these two analyses together to show that the most comprehensive account of learners' syntactic development seeded by FE^{wh} use is achieved by combining the results derived from both approaches.

3. METHODOLOGY

3.1. THE BARCELONA ENGLISH LEARNER CORPUS (BELC)

Our data comes from transcripts in the spoken longitudinal Barcelona English Language Corpus (BELC) (Muñoz, 2006).¹ Nine² balanced bilingual Spanish/Catalan EFL Catalanian state-school beginner learners of English participated in naturalistic L2 interview tasks across four rounds of data collection (Table 1). These rounds can be split into two groups: early years (ages 10 and 12) and later years (16 and 17), as seen in Table 1.

	AGE	HOURS OF INSTRUCTION
early years	10	200
	12	416
later years	16	726
	17	826

Table 1 The four rounds of data collection for the 9 learners under analysis and corresponding hours of classroom English instruction (accumulative).

To make an observation on the learners' progression across different rounds, nine learners were chosen for analysis out of the 55 that constitute the entire corpus, as these were the only learners that participated across at least three rounds of data collection. Spoken tasks consisted of an interview, narrative, and role-play. The interviews were semi-guided, beginning with a series of questions about the learner's family, daily life and hobbies and included a section whereby learners were required to ask questions to the interviewer. The narrative task was elicited from a series of six pictures that learners could freely look at before and during their telling of the story to the interviewer. Finally, 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 such as time setting and choice of activities.

Importantly, beginner learners 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. Controlling for these factors meant that the learners' linguistic environment was homogenous and therefore highly predictable, making them an ideal test ground for comparison.

As in Hammond and Gil (in press), we extracted the four most frequent expressions that were presented holistically to learners in spoken tasks from two local and two global EFL textbooks. These were the following wh-questions:

- (5) a. *what's/is your name?*
 b. *how old are you?*
 c. *where do you live?*
 d. *where are you from?*

3.2. LEARNER PRODUCTIONS OF THE FIXED WH-EXPRESSIONS (FESWH)

A manual analysis of the corpus revealed that all nine learners produced the extracted FE_s^{wh} and the overall distribution of them can be seen in Table 2. Note that 'NT' stands for 'no transcript' and indicates that the learner did not participate in that round of data collection. A dash '-' means that a learner participated but was not shown to produce an FE^{wh}.

1 The corpus is open access and available online via <https://slabank.talkbank.org/access/English/BELC.html>.

2 We are aware that such a small sample size means that the generalisability of any results should be treated with caution. However, from an SLA perspective this is often not the goal, rather, it is sufficient to know that a phenomenon has occurred for a particular group of learners (Gass, 2013). Moreover, 9 learners is a considerably large sample when compared to similar longitudinal studies (e.g., Eskilsden, 2015; Horbowicz & Nordanger, 2021), which traditionally consist of a much smaller number of learners due to the costly and time-consuming process involved in having access to the same participants over a prolonged period of time.

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

Table 2 BELC learners' productions of the identified FEs^{wh}.

At the age learners are first shown to produce an FE^{wh}, the overwhelming majority of other L2 utterances outside of these expressions are ungrammatical and/or of a much lower syntactic complexity (6a-c, 7a-c) and they still rely heavily on the L1 (6d, 7d). Some example utterances from Learner 2 and Learner 5's transcripts are given below to demonstrate:

(6) Learner 2: Age 16: *what is your name?, how old are you?*

- a. **study*
- b. **going to excursion*
- c. **the mother (.) hm read the map*
- d. *the dog sales de (.) de las cestas* [SPANISH]
the dog comes out of of the basket
'the dog jumps out of the basket'

(7) Learner 5: Age 10: *what's your name?*

- a. **the mum it's*
- b. *I study*
- c. *girl and boy see*
- d. *mi ho pots reptir?* [CATALAN]
me you can repeat
'can you repeat it for me?'

The FEs^{wh} can therefore be confidently categorised as 'formulaic' and salient for our learners, and when first produced are of a higher syntactic complexity than the majority of other L2 utterances produced by the same learners. Section 4 now presents our analysis. It first outlines the FEs^{wh} syntactic derivation under a generative model and then presents how these would be conceptualised as abstract schematic constructions under usage-based models.

4. ANALYSIS

4.1. THE FIXED WH-EXPRESSIONS AS PRODUCTS OF COMPUTATIONAL DERIVATION

Under mainstream generative grammar, the derivation of the FEs^{wh} involves the Merging of lexical items via computational procedures driven by features on functional categories T and C. All are wh-questions, involving the computational properties A-movement, wh-movement, T-C movement, and V-raising, and 'where do you live?' also involves do- support. A syntactic tree is given in Figure 1 for 'what is your name?' to exemplify this derivation.

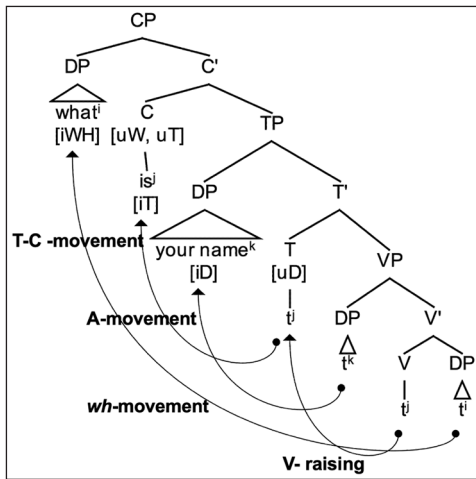


Figure 1 'what is your name' assumed syntactic structure.

These computational properties have the potential to manifest overtly via a variety of surface structures. Following Hammond and Gil (in press), Table 3 outlines the surface phenomena that we take as evidence for their manifestation.³

COMPUTATIONAL PROPERTY	wh- MOVEMENT	T-C MOVEMENT	do- SUPPORT	A- MOVEMENT
surface structure evidence	<ul style="list-style-type: none"> - wh- words occupying a clause-initial position in root interrogatives - exclamative clauses - relative clauses - interrogative complement clauses/embedded wh-clauses 	<ul style="list-style-type: none"> - inversion of the subject and (auxiliary) verb 	<ul style="list-style-type: none"> - via negation - via question formation 	<ul style="list-style-type: none"> - overt subjects used with structures that imply a TP projection, including: <ul style="list-style-type: none"> - corresponding finite verbal inflection [TNS, NUM, AGR] in declaratives (not including <i>is/are</i>) - modal/auxiliary verbs (including dummy <i>do</i>) - <i>wh</i>-movement - T-C movement - 'infinitival to'

Table 3 FEs^{wh} computational properties and reliable surface structures that evidence their manifestation.

Note that we are conservative in what we accept as surface structure evidence, to measure the manifestation of these properties as reliably as possible. A-movement, for example, is only assumed when overt subjects appear with other overt evidence for functional category T (such as an inflectional morpheme or auxiliary verb) and excluded from the count are highly frequent irregular conjugations which are often rote-learned in the EFL classroom (i.e., present simple clauses with *be* (*I am, you are*) and *have* (*you have, he has*)). We also measure learners' L2 accuracy of these properties as a relative percentage out of all production possibilities, as learners have the potential to realise a given utterance during the data collection period in the L1, via translanguaging,⁴ accurately in the L2 or inaccurately in the L2. An example with *do*-support can be used to illustrate this procedure. Say that in a learner's transcript at age 16, there were 9 contexts, as shown in (8a-i), which require *do*-support in English, and our example learner realised these as below (where the intended English output is given in squared brackets []).

- (8) a. *I don't go to school*
 b. *He not like it* [He doesn't like it]
 c. *No se* [I don't know]
 d. *No se* [I don't know]
 e. *No want eat tonight* [I don't want to eat tonight]
 f. *Want go there tonight?* [Do you want to go there tonight?]
 g. *He doesn't gustar la comida* [He doesn't like the meal]
 h. *Te gusta la musica?* [Do you like music?]
 i. *Do you speak English?*

³ Note that we do not measure learners' knowledge of 'v-raising', as surface evidence for this property in English is so limited.

⁴ We adopt the term 'translanguaging' rather than 'code-switching'. This is because, for our learners in the EFL classroom, use of the L1 in utterances such as (8-g) are likely a 'fallback' strategy used to communicate meaning, rather than constrained alternations occurring at specific points in communicative episodes (Przymus, 2023). That said, (8-g) could be classed as an instance of intra-sentential codeswitching, if looked at objectively.

Out of these 9 contexts where *do*-support should manifest, 3 of these are realised in the L1 (c, d and h), 1 via translanguaging (g) and 5 are attempted in the L2 (a, b, e, f and i). Out of these 5 L2 attempts, only 2 of these utterances are accurate (i.e., grammatical) (a and i). This learner's L2 accuracy rate of *do*-support at age 16 is therefore 22%, as they realise 2 accurate utterances in the L2 out of a possible 9 contexts.

In Section 4.2, we now analyse the FEs^{wh} as abstract schematic constructions under usage-based models and outline associated utterance schemas which are potentially extractable and generalisable across similar functional structures.

4.2. THE FIXED *wh*-EXPRESSIONS AS ABSTRACT SCHEMATIC CONSTRUCTIONS

Rather than a computational system, the level of ultimate abstractness for UBL consists of schematic knowledge of symbolic units, 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.

- (9) 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]

Usage-based models posit an acquisition of fully schematic constructions and/or utterance schemas through the analysis and subsequent generalisation of prototypical, formulaic expressions that exemplify these constructions. Due to their saliency, prototypicality and formulaicity for all learners under analysis, the FEs^{wh} are good candidates for acquisitional seeds in this proposed developmental sequence. They are also all produced in isolation and in advance of any other grammatical L2 utterance of a similar complexity (see Section 3). Adopting this learning strategy, for example, learners could gradually move from the FE^{wh} [*what is your name?*] to a derived utterance schema (a fixed part and open slot) [*what is + PossDET + NOUN*], to the fully schematic construction [WH + COPULA + PossDET + NOUN], as schematised in Figure 2.

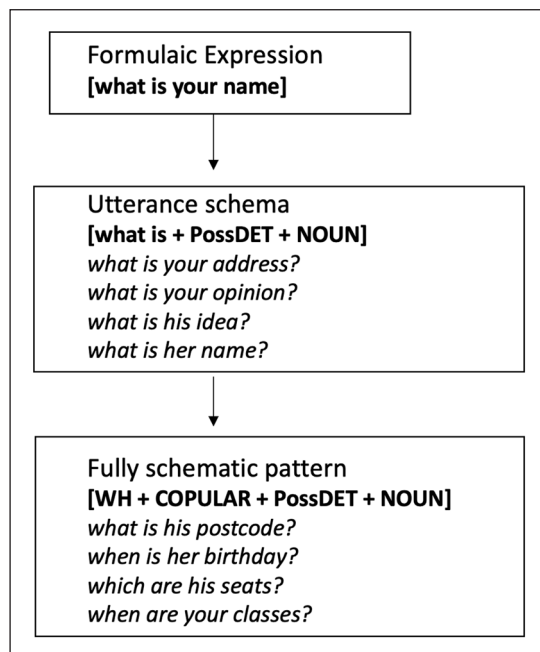


Figure 2 A usage based developmental trajectory of the schematic construction [WH + COPULA + PossSUBJ + NOUN] derived from the formulaic exemplar *what's your name*.

Equally, as past studies on English L2 interrogative development have suggested (see Section 2.2), learners can use FEs^{wh} to derive more general 'wh-question' utterance schemas. Utterance schemas based on fixed wh-questions traditionally comprise the [WH + VERB] element, based on evidence that a learner'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 (Rowland & Pine, 2000; Eskildsen, 2015). Based on the FEs^{wh}, this would give for the following utterance schemas, which have the potential to be lexically (10) and/or categorically (11) specific.

- (10)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]
- (11)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]

As any utterance schema is potentially extractable from formulaic exemplars, learners could also extract the FE^{wh} [VERB + SUBJ] utterance schemas and omit the wh-element to derive yes/no questions. These lexically and categorically specific yes/no question utterance schemas are shown in (12) and (13) respectively.

- (12)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]
- (13)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]

To examine whether learners' L2 questions shared an utterance schema/fully schematic pattern of a previously used FE^{wh} in their production data, we adopted a traceback methodology and created individual learner tables documenting their FE^{wh} productions and L2 questions across the four rounds of data collection (ages 10, 12, 16 and 17). Underneath each FE^{wh} and L2 question, we specified their lexically (i) and categorically (ii) specific utterance schemas, as well as their fully schematic patterns (iii). We then underlined instances where these of a L2 question matched those of a previously produced FE^{wh}. Learner 13's wh-questions can be seen in Table 4 as an example. Note that where FE^{wh} are not shown for a certain age, this means that the learner did not produce an FE^{wh} at this age. '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 wh-questions in the L2 at this stage.

Table 4 Traceback methodology: Learner 13's FE^{wh} productions and L2 wh-questions.

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

Table 4, for example, shows that one L2 wh-question in Learner 13's transcripts share 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, sharing the same fully schematic pattern [WH + AUX DO + PRN + VERB].

5. RESULTS

Section 5 presents the results of both the generative and usage-based analyses of the data, before bringing these together in Section 6. We begin with the generative analysis.

5.1. FE^{wh} USE AND LATER KNOWLEDGE OF ASSOCIATED COMPUTATIONAL PROPERTIES

Although all learners are shown to produce the FE^{wh} across the data collection period, they differ in their frequency of FE^{wh} productions and age they first produce an FE^{wh}. We test the effect of these two variables on learners' L2 accuracy of associated computational properties at the later stages of data collection (ages 16 and 17). 'Age of first FE^{wh} production' refers to the age in which a learner first produces an FE^{wh} in the corpus (e.g., 10, 12, 16 or 17) and 'frequency of FE^{wh} production' refers to the number of FE^{wh} learners produce at the early ages (ages 10 & 12), not including repetitions. We measure learners' L2 accuracy of the computational properties at the later ages as a mean average between their relative accuracy score at age 16 and that of age 17. Table 5 demonstrates this with Learner 47.⁵

LEARNER 47					
AGE	wh -MOVE	T-C MOVE	do - SUPPORT	A- MOVE	MEAN
16	(2/4) 50%	(5/7) 71%	(6/11) 55%	(36/52) 69%	61%
17	(3/3) 100%	(6/6) 100%	(7/7) 100%	(29/29) 100%	100%
mean	75%	86%%	78%	85%	81%

Table 5 Learner 47
 L2 accuracy rates of computational properties at the later ages.

These are discussed Section 5.1.1. and Section 5.1.2. respectively below.

5.1.1. Age of first FE^{wh} production

Figure 3 displays a scatterplot showing the learners' age of first FE^{wh} production (y-axis) and their mean L2 computational accuracy rates in all required contexts (calculated as a combined average between wh-movement, T-C movement, A-movement and do- support) at the end of the data collection period (x-axis, ages 16 and 17). The scatterplot shows a negative slope regression line, which indicates an amount of linearity between a younger age of first FE^{wh} production and a higher L2 computational accuracy rate at the later ages. Those learners who produce an FE^{wh} for the first time at age 16 are clustered towards accuracy rates between 20–40%, whereas those who produce them at ages 10 and 12 are largely between 80–100%.

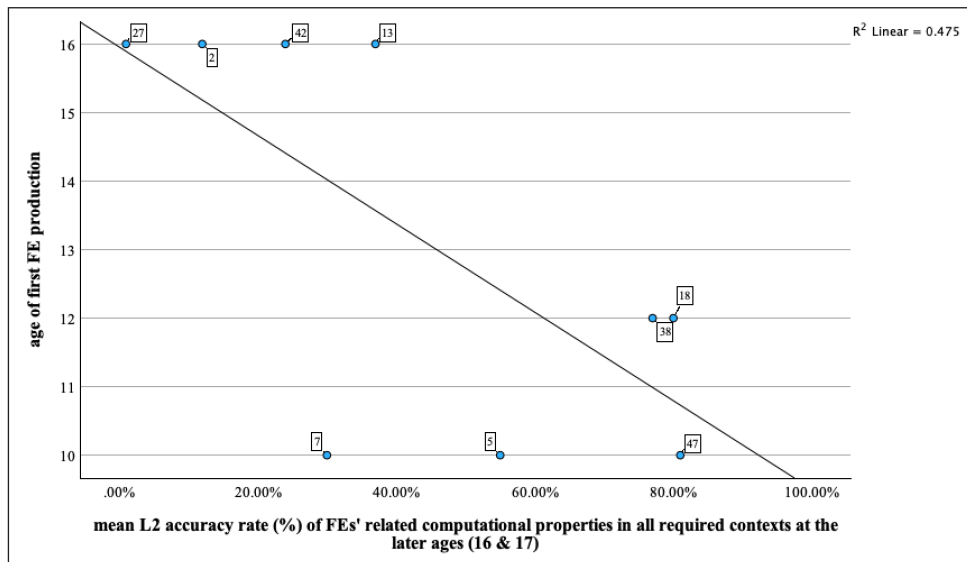


Figure 3 Scatterplot showing learners' age of first FE^{wh} production and mean L2 accuracy of computational properties at later ages (16 & 17).

To investigate this linearity further, we ran correlations between these variables, shown in Table 6. Correlations were run between age of first FE^{wh} production and each computational property individually, as well as with these individual accuracy rates combined as a mean average (as in the scatterplot above). Following recent developments in the application of statistics in SLA which question assumptions of significance traditionally derived by *p* values (Paquot & Plonsky, 2017; Larson-Hall & Mizumoto, 2020), we have included confidence intervals (CIs) in tandem

⁵ A table displaying all learners' first age/frequency of FE^{wh} production along with raw scores and relative L2 accuracy rates at ages 16 and 17 can also be found in Appendix 1.

with bootstrapping to give a more accurate picture of the r effect sizes. We have also adjusted the alpha level to .15 (from the traditional .05) to compensate for small SLA data samples (Stevens, 1996; Pallant, 2010), and measure effect sizes for SLA following Plonsky & Oswald (2014) as $r = .2$ as a small effect, $r = .4$ as a medium effect and $r = .6$ as a large effect.

L2 ACCURACY (AGES 16 & 17)	AGE OF FIRST FE ^{wh} PRODUCTION	SIG. (2-TAILED)	BOOTSTRAP (BCA) 95% CONFIDENCE INTERVAL
Mean	-.689	*.040	**-.944/-.265
wh- movement	-.683	*.062	**-.954/-.294
T-C movement	-.593	*.093	**-.903/-.028
do- support	-.600	*.088	**-.946/-.099
A- movement	-.779	*.013	**-.985/-.320

The negative effect sizes indicate that a learner’s earlier production of the FE^{wh} shows strong, significant correlations with their later L2 accuracy of all related computational properties and these combined as a mean average. Taken together, these figures show that those learners who produce the FE^{wh} for the first time at younger ages show a higher L2 accuracy rate of their associated computational properties at the end of the data collection period.

5.1.2. Frequency of FE^{wh} production

Figure 4 shows a scatterplot of the learners’ frequency of FE^{wh} production at the early ages (y-axis) and their mean L2 computational accuracy rates in all required contexts (calculated as a combined average between wh-movement, T-C movement, A-movement and do- support) at the end of the data collection period (x-axis, ages 16 and 17). The scatterplot shows a positive slope regression line, indicating linearity between a higher number of FE^{wh} produced at the early ages and a higher L2 accuracy of their associated computational properties at the later ages.

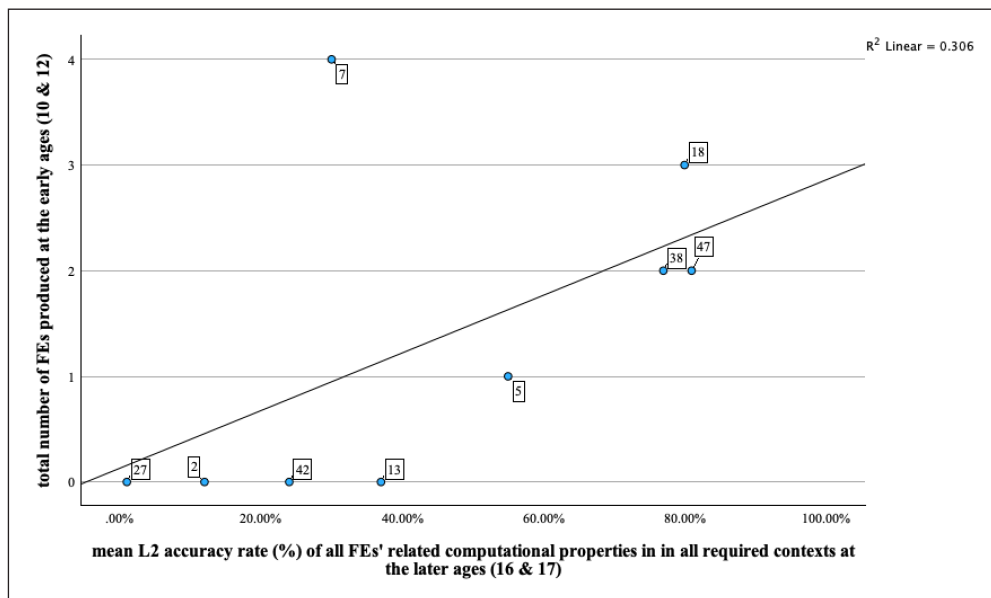


Table 6 Correlation coefficient between age of first FE^{wh} production and L2 accuracy of computational properties at the later ages.

*Significant at the adjusted $p < .15$ for small sample sizes in SLA.

**CI effect for a relationship among variables.

Figure 4 Scatterplot showing learners’ frequency of FE^{wh} production at the early ages (10 & 12) and mean L2 accuracy of computational properties at later ages (16 & 17).

Correlations were run to investigate this relationship further, which compare frequency of FE^{wh} production at the early ages with L2 accuracy at the later ages of each computational property individually and then these as a mean average. These are shown in Table 7. A learner’s higher number of FE^{wh} productions at the early ages shows strong significant correlations with their later L2 accuracy of wh- movement, T-C movement, and the four computational properties as a mean average. Individually, A- movement and do- support show medium correlations, and fail to reach significance ($p = .156$, $p = .263$).

L2 ACCURACY (AGES 16 & 17)	NUMBER OF FE ^{wh} PRODUCTIONS (AGES 10 & 12)	SIG. (2TAILED)	BOOTSTRAP (BCA) 95% CONFIDENCE INTERVAL
Mean	.578	*.123	** .001/.967
wh- movement	.615	*.104	** .143/.968
T-C movement	.520	*.150	-.168/.966
do- support	.418	.263	-.199/.969
A- movement	.515	.156	-.095/.945

Taken together, learners' higher L2 accuracy of the FE^{swh} associated computational properties at the later ages (16 and 17) correlates with a younger age of first FE^{wh} production and a higher number of FE^{wh} productions at the early ages (10 and 12). Note that this relationship between learners' FE^{wh} use and L2 accuracy of associated computational properties seems to be developmental; that is, 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. For example, 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 5 shows a scatterplot with a relatively flat regression line, and Table 8 shows that overall frequency of FE^{wh} production across the four rounds of data collection shows no correlation with later L2 accuracy of any associated computational property individually or these as a mean average.

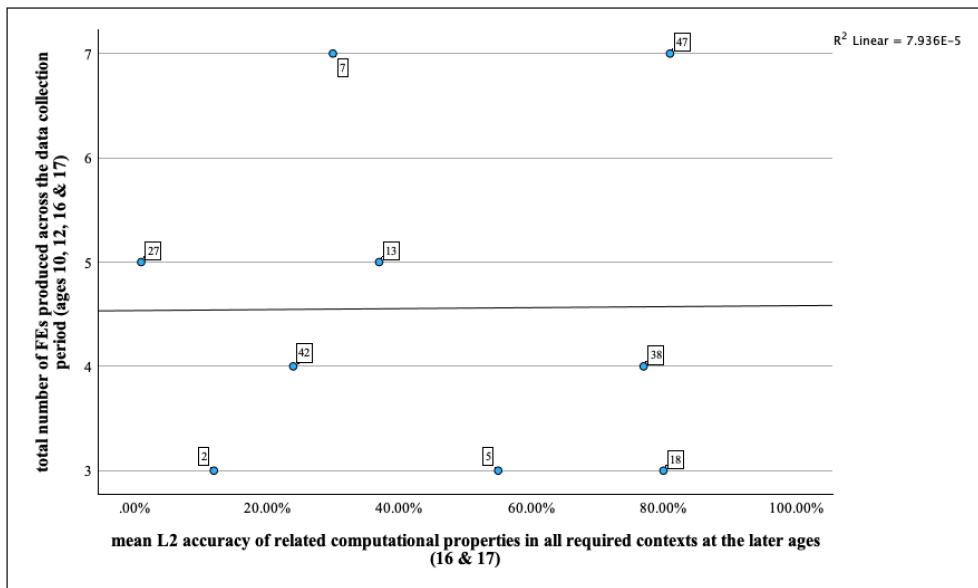


Table 7 Correlation coefficient between number of FE^{swh} produced at the early ages (10 & 12) and L2 accuracy of computational rules at the later ages.

*Significant at the adjusted p < .15 for small sample sizes in SLA.

**CI effect for a relationship among variables.

Figure 5 Scatterplot showing learners' frequency of FE^{wh} production across all ages and mean L2 accuracy of computational rules at later ages (16 & 17).

L2 ACCURACY (AGES 16 & 17)	TOTAL NUMBER OF FE ^{swh} PRODUCED (ACROSS ALL AGES)	SIG. (2-TAILED)	BOOTSTRAP (BCa) 95% CONFIDENCE INTERVAL
Mean	.009	.982	-.693/.663
wh- movement	.062	.883	-.682/.840
T-C movement	-.018	.963	-.817/.680
do- support	.018	.962	-.689/-.665
A- movement	.009	.982	-.711/.655

Table 8 Correlation coefficient between total number of FE^{swh} produced across all ages and L2 accuracy of computational rules at the later ages (16&17).

Therefore, a better L2 accuracy of associated computational properties seems to correlate specifically to a more frequent production of the FE^{wh} at early stages of data collection, rather than a frequent production of the expressions across the entire data collection period. This is suggestive of a more developmental relationship between early use of these expressions and a better knowledge of related computational derivations.

5.2. LEARNERS' USE OF THE FE^{wh} AND LATER KNOWLEDGE OF THEIR SCHEMATIC CONSTRUCTIONS

Moving now to test if the usage-based developmental sequence is applicable to the present dataset, we identified all learners' L2 root interrogatives across the data collection period to see if they embodied the same schematic patterns/utterance schemas of previously produced FE^{wh}, starting with learners' wh-questions.

5.2.1. Wh-questions

As discussed previously, the FE^{wh} have the potential to represent lexically and categorically specific wh-question utterance schemas and fully schematic patterns. Following the procedure outlined in Section 4.2, our usage-based analysis reveals that a total of 20 wh-questions are produced by all 9 learners across the data collection period. Out of these 20 wh-questions, 17 appear after an FE^{wh} ontogenically in learners' production data. Of these 17, 9 embody the same categorically specific utterance schemas of a previously produced FE^{wh}, 3 of which also embody the same lexically specific utterance schemas and 4 of which show the same fully schematic patterns. This accounts for 53% of learners' total wh-questions that follow FE^{wh} use in the longitudinal data. An example is Learner 38, who 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 respectively, which all 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 an ungrammatical wh-question at age 16 '*what you wanna say?'. Their productions across the data collection period are presented in Table 9.

Table 9 Traceback methodology: Learner 38's FE^{wh} productions and L2 wh-questions.

#	AGE 10	AGE 12: wh-Q	AGE 12: FE ^{wh}	AGE 16: wh-Q	AGE 16: FE ^{wh}	AGE 17: wh-Q	AGE 17: FE ^{wh}
38	NT		what's your name i. [what's] + X ii. [WH+COPULA] + X iii. [WH+COPULA]+ A]+PossDET+ NOUN]	<i>why are you doing this kind?</i> i. [why + are] + X ii. [WH + COPULA] + X iii. [WH+COPULA +PRN+ VERB - ing+DET+NOU N] <i>*what you wanna say?</i>	what's your name i. [what's]+X ii. [WH+COPULA] + X iii. [WH+COPULA]+ PossDET+NOUN]	<i>why are you doing this work?</i> 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]
		* where you live i. [where + you] + X ii. [WH + PRN] + X iii. [WH + PRN + VERB] + X		i. [what + you] + X ii. [WH + PRN] + X iii. [WH + PRN + AUX + VERB]			

5.2.2. Yes/No questions

As well as the wh-question utterance schemas presented above, the FE^{wh} have the potential to represent lexically and categorically specific 'yes/no-question utterance schemas'. A total of 23 yes/no questions are produced by all 9 learners across the data collection period. Of these 23 yes/no questions, 21 follow an FE^{wh} in learners' data ontogenically, out of which 11 embody the same categorically specific utterance schemas as a previously produced FE^{wh} (53%). All 11 of these yes/no questions also share the same lexically specific utterance schemas as the FE^{wh}. An example is Learner 18, who makes use of the [are you] + X utterance schema in 'are you studying?' at age 17 after producing the FE^{wh} 'how old are you?' at age 12. They also produce the erroneous FE^{wh} '*what do you live?' at age 12 and continue 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?'. Their production data is shown below in Table 10.

#	AGE 10: Y/N Q	AGE 10: FE ^{wh}	AGE 12: Y/N Q	AGE 12: FF ^{wh}	AGE 16: Y/N Q	AGE 17: Y/N Q
18	-	-	<i>do you have any pets?</i> i. [do you] + X ii. [AUX DO + PRN] + X <i>do you like pets?</i> i. [do you] + X ii. [AUX DO + PRN] + X	<i>what's your name</i> i. [is your] + X ii. [COPULA+ PossDET] + X <i>how old are you</i> i. [are you] + X ii. [COPULA + PRN]+X <i>*what do you live</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X	<i>do you born in Spain?</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X <i>do you like your job?</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X	<i>are you studying?</i> i. [<u>are you</u>] + X ii. [COPULA + PRN] + X <i>do you live in Barcelona?</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X <i>do you live on your own?</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X <i>do you have any brothers or sisters?</i> i. [<u>do you</u>] + X ii. [<u>AUX DO + PRN</u>] + X

To summarise, 53% of L2 (20/38) interrogatives (both *wh*- and yes/no) produced by all learners under analysis can be traced back to utterance schemas of previously produced FE^{wh}. All these utterance schemas are categorically specific to preceding FE^{wh} (20/20), 70% are lexically specific (14/20) and 44% of learners' *wh*-interrogatives (4/9) share the same fully schematic patterns. 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 L2 interrogatives can be linked back to utterance schemas of previously used FE^{wh}.

Table 10 Traceback methodology: Learner 18's FE^{wh} productions and L2 yes/no questions.

6. DISCUSSION

In Section 5.1 we adopted a generative model to address research question (i), finding that higher L2 accuracy rates of the FE^{wh} associated computational properties at the end of the data collection period correlates with a younger age of first FE^{wh} production and a higher number of FE^{wh} productions at the early ages. This supports the trends observed in Hammond and Gil (*in press*), whereby those learners who interacted more with the FE^{wh} were quicker to move from VP-TP-CP based grammars. In Section 5.2, we adopted a usage-based schematic model to address research question (ii) and discovered that 53% of learners' L2 interrogatives can be traced back to utterance schemas of previously used FE^{wh} ontogenetically in their spoken transcripts. This supports those longitudinal usage-based studies who have been able to trace back productive use of complex L2 utterances to model formulaic exemplars in learners' production data.

The discussion now compares how each model can account for the observed L2 development over the longitudinal data collection period, and argues that the most comprehensive description is achieved by combining the results of both analyses.

6.1. FIXED WH- EXPRESSIONS: DATABASES OF COMPUTATIONAL PROPERTIES OR SCHEMATIC PATTERNS?

Both generative and usage-based analyses of the longitudinal data can distinguish a relationship between learners' use of identified FE^{wh} and associated L2 syntactic development, which highlights the central role that formulaic language can play in L2 development. It can be said that conceptualising the FE^{wh} as databases for the acquisition of more general associated computational properties can account for a larger range of corresponding L2 development, rather than limiting the expressions to databases for the acquisition of L2 interrogative utterance schemas only. This is somewhat unsurprising, given that these properties have the potential to manifest via a larger range of related surface structures. For example, a gradual acquisition of the underlying syntactic mechanisms necessary to construct interrogatives in the L2 can account for 100% of learners' L2 interrogatives across the corpus, including the 47% that constitute different utterance schemas than those of the FE^{wh}. An acquisition of the FE^{wh} computational properties can also, of course, account for grammatical L2 utterances outside of learners' interrogatives. For example, an acquisition of the feature specifications necessary to constrain *wh*-movement in the L2, as influenced by early and frequent FE^{wh} use, is also

exemplified by learners' comparative use of relative clauses and interrogative complement clauses. Table 11 shows that the only learners who produce these structures in the L2 are those that show early FE^{wh} usage.

LEARNER	AGE 10: FE	AGE 12: FE	AGE 16: RELATIVE/INTERROGATIVE COMPLEMENT CLAUSE	AGE 17: RELATIVE/INTERROGATIVE COMPLEMENT CLAUSE
2	-	-	-	-
5	<i>what's your name</i>	-	-	-
7	<i>what's your name</i>	<i>what's your name how old are you where do you live</i>		<i>I need some food and somebody who play the music</i>
13	-	-	-	-
18	-	<i>what's your name how old are you *what do you live</i>	<i>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</i>	<i>and when they: [] they go to eat the breakfast the dog they see that the dog *eat [\] eats the: []</i>
27	<i>no transcript</i>	-	-	-
38	<i>no transcript</i>	<i>what's your name where you live</i>	-	-
42	<i>no transcript</i>	-	-	-
47	<i>how old are you what's your name</i>	<i>no transcripts</i>	-	<i>when the sister and brother arrive to the [] to the forest they looked that her dog mm *_was [\] was appeared</i>

However, utterance schema extraction and generalisation based on previous FE^{wh} use is clearly a productive learning strategy, as this can account for over half of learners' total interrogatives produced in the L2 across the corpus. Therefore, the most unified account of the observed syntactic development must incorporate this strategy within the development of associated underlying syntactic mechanisms more generally. Section 6.2 now discusses some theoretical concepts which are compatible with this combination of results derived from both approaches.

6.2. THE INTERACTION OF USAGE-BASED AND GENERATIVE APPROACHES TO SLA

We posit that the usage-based notion of utterance schema extraction and generalisation can facilitate the acquisition of the underlying computational properties for which their surface forms exemplify. The FE^{wh} for all learners are first produced in advance of associated L2 competence, so must be taken as memorised products of holistic retrieval via working/phonological memory. This is also an indication that the FE^{wh} constitute learners' *intake* rather than *input* (Carroll, 2001), as they are the expressions that learners rely on upon functional contextual cues. At these initial stages, the FE^{wh} as recalls from working memory are analogous to what some models of L1/L2 acquisition term 'perceptual intake' (Lidz & Galiardi, 2015) or 'perceptual output structures' (Truscott & Sharwood-Smith, 2004). Importantly, when processing these perceptual strings, learners construct an associated linguistic representation which contains information about the L2 syntactic feature specifications. Thus, an increased interaction with the FE^{wh} may quicker engender a restructuring of learners' L1 grammar based on this new L2 linguistic information, as they are better exposed to this in model form.

It follows that if learners can extract utterance schemas from prototypical formulaic exemplars (via general cognitive means) and extend these to similar functional structures, it allows them to interact with more surface forms which exemplify the same L2 linguistic information, leading to a better identification of the abstract representations realised in these surface forms of L2. In our data, utterance schema extraction and generalisation has likely facilitated the production of a large proportion of L2 interrogatives (53%), which exemplify the L2 functional categories

Table 11 Learners early FE^{wh} use and later L2 productions of relative and interrogative complement clauses.

and related computational properties necessary to construct interrogatives in L2, irrespective of their specific schematic surface forms ([WH + COPULAR, WH+AUX DO] etc.). With every interaction, or more specifically, with every use of these complex L2 surface forms, the learner is better equipped to make inferences about the underlying grammar that generated them, based on their pre-existing (UG-based) knowledge of the computational component. This concept of increased interaction and usage of L2 surface forms is compatible with Paradis's (2004, pp. 52–53) notion of 'practicing'. They state:

"The repeated practicing of a 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."

We believe that utterance schema extraction and generalisation is best analysed as a productive learning strategy that can lead to the further, repeated 'practicing' of complex L2 surface forms, allowing for a quicker restructuring of corresponding L2 linguistic representations in the computational component. Utterance schema extraction alone based on previous FE^{wh} use cannot account for all L2 question forms in our corpus, nor can it account for the development of related syntactic phenomena outside of interrogatives. This strategy is best interpreted as facilitating the gradual acquisition of related underlying computational properties and corresponding feature specifications more generally. It is in this scaffolding, facilitative sense that we propose the usage and deconstruction of classroom input-derived formulaic expressions can interact with the development of modular syntactic knowledge.

Finally, a note on the limitations of the data is in order. It is necessary to reiterate that the BELC is only a snapshot of these learners' L2 capabilities at particular points in time. As for all corpus studies, a learner's production of a specific form at these recorded intervals may not necessarily reflect their L2 competence and, as such, the absence of a form does not entail a learner's lack of knowledge (Grondin & White, 1993; Lozano, 2021). Similarly, learners will likely have been exposed to other holistically taught prototypical expressions in their EFL classroom that did simply not surface in their transcripts. However, the aim of this paper is to account for the developmental trends that *are* observable in the available production data, using both generative and usage-based frameworks. The identified FE^{wh} are clearly salient, as all learners are shown to produce these same expressions upon the same contextual cues, and initially in advance of associated L2 competence. No other formulaic material was identifiable in the transcripts alongside these expressions. This salience in learners' production data, along with their inherent prototypicality and functionality, place the FE^{wh} as prime candidates for acquisitional seeds under all usage-based accounts, and we observe a clear relationship between early and frequent use of these expressions and learners' corresponding L2 development.

7. CONCLUSION

This study has adopted a novel approach by combining usage-based and generative analyses of longitudinal learner production data to discover an effect of formulaic expressions on L2 syntactic development. We have argued that positing the FE^{wh} as databases for the underlying computational properties/L2 feature specifications for which they exemplify can account for a large range of corresponding syntactic development, and that utterance schema extraction and generalisation is a productive learning strategy that is likely facilitative on this process. More generally, we believe that the study of FEs and their relationship with syntactic development is 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 modular linguistic knowledge.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix 1.** All learners' L2 accuracy rates of computational properties at the later ages.
DOI: <https://doi.org/10.22599/jesla.100.s1>

For detailed comments and advice that improved the quality of the article, we would like to thank two anonymous JESLA reviewers as well as the Journal Editor, Sarah Liszka. Work on this article has been supported by the Arts and Humanities Research Council (AHRC) Studentship Award. This support is gratefully acknowledged.

AUTHOR AFFILIATIONS

Thomas Hammond  orcid.org/0009-0003-2722-6413

University of Sheffield, UK

Kook-Hee Gil  orcid.org/0000-0002-7786-4451

University of Sheffield, UK

REFERENCES

- Bardovi-Harlig, K., & Stringer, D.** (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research*, 33(1), 61–90. DOI: <https://doi.org/10.1177/0267658316641725>
- Carroll, S. E.** (2001). *Input and Evidence: The raw material of Second Language Acquisition*. Language Acquisition & Language Disorders. John Benjamins. DOI: <https://doi.org/10.1075/lald.25>
- Carroll, S. E.** (2010). Explaining how learners extract ‘formulae’ from L2 input. *Language, Interaction and Acquisition*, 1(2), 229–250. DOI: <https://doi.org/10.1075/lia.1.2.05car>
- Collins, C., & Stabler, E.** (2016). A formalization of minimalist syntax. *Syntax*, 19(1), 43–78.
- Ellis, N. C.** (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 297–339. DOI: <https://doi.org/10.1111/synt.12117>
- Ellis, N. C.** (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. DOI: <https://doi.org/10.1017/S0267190512000025>
- Eskildsen, S.** (2015). What counts as a developmental sequence? Exemplar based L2 learning of English questions. *Language Learning*, 65(1), 33–62. DOI: <https://doi.org/10.1111/lang.12090>
- Eskildsen, S. W.** (2020). Creativity and routinisation in L2 English: Two usage-based case-studies. In W. Lowie, M. Michel, A. Rousse–Malpat, M. Keijzer, & R. Steinkrauss (Eds.), *Usage-based dynamics in second language development* (pp. 107–129). Multilingual Matters. DOI: <https://doi.org/10.21832/9781788925259-008>
- Gass, S.** (2013). *Second language acquisition: An introductory course*. New York and London: Routledge.
- Grondin, N., & White, L.** (1993). Functional categories in child L2 acquisition of French. *McGill Working Papers in Linguistics*, 9, 121–145.
- Hammond, T., & Gil, K. H.** (in press). Fixed expressions as ‘bootstrapping’ mechanisms into L2 phrase structure: Examining trends in a longitudinal learner corpus.
- Horbowicz, P., & 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. DOI: <https://doi.org/10.1017/langcog.2021.9>
- Larson-Hall, J., & Mizumoto, A.** (2020). Using statistical analysis software (R, SPSS). In J. McKinley & H. Rose, (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 385–397). Routledge. DOI: <https://doi.org/10.4324/9780367824471-33>
- Lidz, J., & Gagliardi, A.** (2015). How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics* 1(1), 333–352. DOI: <https://doi.org/10.1146/annurev-linguist-030514-125236>
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M.** (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30(2), 333–370. DOI: <https://doi.org/10.1017/S0305000903005592>
- Lozano, C.** (2021). Generative approaches. In N. Tracy–Ventura & M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Corpora* (pp. 213–227). New York and London: Routledge. DOI: <https://doi.org/10.4324/9781351137904-19>
- Marsden, H., Whong, M., & 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. DOI: <https://doi.org/10.1017/S0272263117000018>
- Muñoz, C.** (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters. DOI: <https://doi.org/10.21832/9781853598937>
- Myles, F.** (2004). From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society*, 102(2), 139–168. DOI: <https://doi.org/10.1111/j.0079-1636.2004.00133.x>

- Myles, F., & 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. DOI: <https://doi.org/10.1017/S027226311600036X>
- Myles, F., Hooper, J., & Mitchell, R.** (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323–363. DOI: <https://doi.org/10.1111/0023-8333.00045>
- Pallant, J.** (2010). *SPSS Survival Manual: A step by step guide to data analysis using SPSS*. Open University Press.
- Paquot, M., & Plonsky, L.** (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1), 61–94. DOI: <https://doi.org/10.1075/ijlcr.3.1.03paq>
- Paradis, M.** (2004). *Neurolinguistic theory of bilingualism*. John Benjamins. DOI: <https://doi.org/10.1075/sibil.18>
- Plonsky, L., & Oswald, F.** (2014). How big is ‘big’? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. DOI: <https://doi.org/10.1111/lang.12079>
- Poeppl, D., & Wexler, K.** (1993). The full competence hypothesis of clausal structure in early German. *Language*, 69, 1–33. DOI: <https://doi.org/10.2307/416414>
- Przymus, S. D.** (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*. DOI: <https://doi.org/10.1080/13670050.2023.2220880>
- Rizzi, L.** (2009). Some elements of syntactic computations. In D. Bickerton, E. Szathmari & J. Lupp (Eds.), *Biological foundations and origin of syntax*, (pp. 63–89). MIT Press. DOI: <https://doi.org/10.7551/mitpress/8468.003.0009>
- 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. DOI: <https://doi.org/10.1111/lang.12081>
- Rothman, J., & 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. DOI: <https://doi.org/10.1017/S0272263117000134>
- Rowland, C., & Pine, J. M.** (2000). Subject-auxiliary inversion errors and wh-question acquisition: ‘what children do know?’* *Journal of Child Language*, 27(1), 157–181. DOI: <https://doi.org/10.1017/S0305000999004055>
- Stevens, J.** (1996). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates.
- Truscott, J.** (2017). Modularity, working memory, and second language acquisition: A research program. *Second Language Research*, 33(3), 313–323. DOI: <https://doi.org/10.1177/0267658317696127>
- Truscott, J., & Sharwood-Smith, M.** (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, 7(1), 1–20. DOI: <https://doi.org/10.1017/S1366728904001178>
- Tummers, J., Heylen, K., & Geeraerts, D.** (2005). Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory*, 1(2), 225–261. DOI: <https://doi.org/10.1515/clt.2005.1.2.225>
- Vainikka, A., & Young-Scholten, M.** (1998). The initial state in the L2 acquisition of phrase structure. In S. Flynn, G. Martohardjono & W. O’Neil (Eds.), *The Generative Study of Second Language Acquisition* (pp. 17–34). Erlbaum.

TO CITE THIS ARTICLE:

Hammond, T., & Gil, K.-H. (2023). Fixed wh-expressions in classroom second language acquisition: databases of computational properties or utterance schemas? *Journal of the European Second Language Association*, 7(1), 75–92. DOI: <https://doi.org/10.22599/jesla.100>

Submitted: 18 April 2023
Accepted: 01 December 2023
Published: 27 December 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of the European Second Language Association, is a peer-reviewed open access journal published by White Rose University Press.