UNIVERSITY OF Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: WALLINGTON, A. (2022) "Investigating the roles of feature reassembly and linguistic input in later-stage second language acquisition: a case study of aspectual development in university learners of French", University of Southampton, Department of Modern Languages and Linguistics, PhD Thesis, pp.1-248

Data: WALLINGTON, A. (2021) Dataset for: Investigating the roles of feature reassembly and linguistic input in later-stage second language acquisition: a case study of aspectual development in university learners of French. URI : https://doi.org/10.5258/SOTON/D1786

University of Southampton

Faculty of Humanities

Department of Modern Languages and Linguistics

Investigating the roles of feature reassembly and linguistic input in later-stage second language acquisition: a case study of aspectual development in university learners of French

DOI: https://doi.org/10.5258/SOTON/D1786

by

Amy Jayne Frances Wallington

ORCID ID: 0000-0003-2853-5499

Thesis for the degree of Doctor of Philosophy

February 2022

Abstract

This thesis centres on the second language (L2) development of viewpoint aspect, a grammatical property definable as the contrast between perfective viewpoint, which indicates a completed eventuality, and imperfective viewpoint, which provides an ongoing or non-complete perspective (Smith 1991, 1997). This property is well-established as challenging to acquire in an L2 for native (L1) speakers of English, due to the limited grammaticalisation of viewpoint aspect in English (Montrul & Slabakova 2002, 2003; McManus 2015). Yet studies examining L2 viewpoint aspectual development in L1 English speakers (e.g. Domínguez et al. 2011, 2017) do not always find that their results align consistently with L1 influence (frequently operationalised using the Feature Reassembly Hypothesis (Lardiere 2003, 2005, 2008, 2009), a seminal generative framework).

This project analyses the production and comprehension of L2 French viewpoint aspectual form-meaning mappings in *n* = 43 L1 English university-level learners of French, studied cross-sectionally prior to and following a "year abroad" in a French-speaking country. This L2 data is combined with a frequency-distributional analysis of L1 French from two sources: experimental production data and a >10,000-word corpus sample, the latter sourced from the freely-available CFPP2000 (http://cfpp2000.univ-paris3.fr/index.html). The aim is to assess the extent to which both Feature Reassembly predictions and frequency-distributional information from the input can respectively explain trajectories of L2 viewpoint aspectual development, and consequently evaluate whether an approach which unites generative and input-based principles can usefully contribute to our understanding of L2 acquisition as a whole. It is found that learners are indeed sensitive to frequency-distributional information – as is reflected in their production of French aspectual forms – and suggest that they are also sensitive to input changes, as shown by the unexpected lower production accuracy of the post-year abroad group. In conclusion, it appears that an approach which integrates both Feature Reassembly and input information may provide a more comprehensive picture of L2 (aspectual) development.

Keywords: L2 French, L1 English, viewpoint aspect, imperfect, input, feature reassembly, study abroad, corpus analysis

Table of Contents

Table of Tables

Table of Figures

List of Accompanying Materials

DOI: https://doi.org/10.5258/SOTON/D1786

Research Thesis: Declaration of Authorship

Print name: AMY JAYNE FRANCES WALLINGTON

Title of thesis: Investigating the roles of feature reassembly and linguistic input in later-stage second language acquisition: a case study of aspectual development in university learners of French

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature: ..... Date: 26/03/2021

Acknowledgements

I'd like to begin by thanking my supervisor, Prof. Laura Domínguez, for her constant faith and invaluable feedback at every stage of this process. Thank you for believing in me and for always encouraging me to push myself and achieve things I thought I couldn't. I feel very fortunate to be part of such a wonderful cohort at Southampton and am very grateful to everyone involved in making the past nearly-four years of my PhD so enjoyable. Special thanks to James Turner for truly invaluable (and *unbelievably* patient) statistical guidance – thank you for putting up with my endless stressed messages about R!

I would also like to say a very big thank you to Laurence Richard, Virginie Pignot-Shahov, and Laura-May Leaf for their assistance with various stages of coding and translation checking (merci mille fois!), as well as to all the participants who generously donated their time to take part in the study.

A very big thank you and endless gratitude also has to go to Union Southampton Dance and especially Ballet Society: you have been my second home for almost eight years now and I will be forever thankful for the joy and escapism that you have brought to my "out of office" hours.

Finally, thank you from the bottom of my heart to my friends and family, especially Stacey, Jan, and my parents (and not forgetting my cat, Bean). You've been my biggest cheerleaders and the brightest light in this tough year. There's so much more I could add here but I'll leave it by saying that I would quite simply not be writing these words without you.

Definitions and Abbreviations

AH = Aspect Hypothesis

AJT = Acceptability judgement task

AT (TT) = Assertion Time (= Topic Time)

BFLA = Bilingual first language acquisition

EvT = Event Time

FL = Foreign language

FR = Feature Reassembly

FRH = Feature Reassembly Hypothesis

FT/FA = Full Transfer/Full Access

IMP = Imperfect/*imparfait*

L1 = First language

L1A, FLA = First language acquisition

L2 = Second language

L2A, SLA = Second language acquisition

LEQ = Language Engagement Questionnaire

MSIH = Missing Surface Inflection Hypothesis

PC = *passé composé*

PS = *passé simple*

PoS = Poverty of the Stimulus

SA = Study abroad

UG = Universal Grammar

YA = Year Abroad

Investigating the roles of feature reassembly and linguistic input in later-stage second language acquisition: a case study of aspectual development in university learners of French

1. Introduction

The debate surrounding the precise extent and nature of the influence of a learner's first language (L1) on their acquisition of a second language (L2) has been a central question in second language acquisition (L2A, SLA) research for decades (e.g. Long & Sato 1984; Schwartz & Sprouse 1994,1996; Ringböm 2007). The L2 acquisition of aspect – the grammatical property determining the internal temporal structure of an event (Comrie 1976; Chung & Timberlake 1985; Smith 1991, 1997) - is well-suited for investigations into the impact of L1 influence in L2A, as it is demonstrated to be susceptible to variable L1 influence depending on the L1 in question (e.g. Slabakova 2001; Roberts & Liszka 2013). The L2 acquisition of viewpoint aspect – which we can think of as a distinction between perfective viewpoint, conceptualised as a "completed" perspective on an eventuality, and imperfective viewpoint, conceptualised as an "incomplete" or ongoing perspective (Smith 1991, 1997) - has been shown to be particularly challenging for L1 speakers of English, as their L1 does not fully grammaticalise the imperfective-perfective distinction (Montrul & Slabakova 2002, 2003; Howard 2005a; Labeau 2005; McManus 2015).

Research within the generative L2A paradigm has attempted to elucidate why the L2 acquisition of viewpoint aspect can be so challenging by drawing on Feature Reassembly (FR) (Lardiere 2003, 2005, 2008, 2009): an approach based on the assumption that the mappings between morphological forms and semantic meanings used in the L1 are transferred at the Initial State of L2 acquisition – following Schwartz & Sprouse's (1994, 1996) proposal of Full Transfer – and that targetlike L2 form-meaning mappings may be gradually formed during L2 development. Feature Reassembly enables the contrastive comparison of L1/L2 pairs, and thus has proven helpful in facilitating the "measurement" of the learning task in terms of the degree of remapping of semantic features onto morphological forms - form-meaning remapping - that is required for a given interpretation. It has therefore been widely used by those studying the acquisition of perfective/imperfective features (Domínguez et al. 2011, 2017; McManus 2015; Mai & Yuan 2016), as well as other areas of functional morphology (e.g. Gil & Marsden 2013; Hwang & Lardiere 2013).

However, the findings of these studies have not always been in total accordance with the central concept of Feature Reassembly: that is, that form-meaning

mappings requiring the least reassembly between the L1 and L2 will be the most straightforward to acquire (and vice versa). In their work on the L2 reassembly on Spanish (im)perfective features, Domínguez et al (2011, 2017) observed a mixed picture of results, including high accuracy in certain imperfective contexts that required relatively large degrees of reassembly. Similarly mixed findings were found in an investigation into the production of viewpoint aspectual mappings for the L1 English/L2 French pair (Wallington 2017): learners did not perform most accurately on the progressive interpretation of the imperfective, despite its one-to-one mapping between English and French, and also performed significantly more accurately with the imperfective-continuous than the imperfective-habitual, despite the two mappings requiring similar degrees of reassembly (see Section 3.2 for more details on the viewpoint aspectual mappings in question). This led to the suggestion that, in the case of the L2 acquisition of French viewpoint aspect by L1 English learners, 'feature reassembly does not act alone but instead works in tandem with additional variables', and consequently to a proposal of the potential relevance of the differing frequencies of occurrence of each imperfective interpretation, as observed in L2 production (Wallington 2017:59). This is in line with Cho & Slabakova's (2014:160) observation that 'feature reassembly may be slow to occur or may not occur at all if the relevant feature is rare or contradictory in the linguistic input', as well as with research suggesting that learners' use of viewpoint aspectual mappings may indeed reflect naturally-occurring patterns in the input to which they are exposed (e.g. McManus 2011; Domínguez et al. 2013).

This project aims to build upon the above findings with a view to assessing whether an approach to L2A that integrates predictions from Feature Reassembly – a framework rooted in generative principles (Chomsky 1965, 1980a, 1981b) - with information from the input has greater explanatory power for viewpoint aspectual acquisition in a second language: and, by extension, for second language acquisition as a whole. This project further develops the research design explored in Wallington (2017), utilising L1 English university learners of L2 French prior to and after 9-month residence abroad in a French-speaking country as a case study population to test this hypothesis. Some key differences in the research design of this project include the use of a second production task and a comprehension task (in order to gain a more representative picture on acquisition), as well as the fact that, whilst Wallington (2017) employed a longitudinal analysis, this project cross-sectionally studies two different L2

cohorts. Though a longitudinal replication of aspectual development would undeniably have been insightful, the cross-sectional design of the current project still has the means to inform on students' developing use of viewpoint aspect-expressing forms.

In addition to providing a potentially more comprehensive model of L2A, the decision to carry out an integrated exploration of naturalistic linguistic input at more advanced levels of L2 development is motivated by the fact that, as mentioned, the majority of modern languages undergraduates in UK universities participate in an extended period of residence abroad informally known as the "year abroad" (YA). Though the importance of the language learning context has been noted (Collentine 2009; Llanes 2011), research on study abroad (SA) contexts has been relatively scarce in comparison to other contexts such as immersion and instructed/foreign-language (Llanes 2011). Moreover, of the already somewhat scant existing research on L2A in study abroad contexts, fewer studies still have focused on grammatical development of a specific property, with those that have painting a contradictory picture of whether SA has a beneficial impact (e.g. DeKeyser 1991; Ryan & Lafford 1992; Guntermann 1995; Collentine 2004; Howard 2001, 2005b**,** 2006). There is thus a clear need for further research investigating the extent to which the year abroad – an arguably quintessential component of UK university-level language learning – is conducive to aspectual development, and to grammatical development as whole, which is itself fundamental to effective communication in an L2.

Moreover – and crucially for this project - the year abroad constitutes a point of interest in terms of input, as it represents a noticeable departure from university learners' previous experience with the L2, which has predominantly been in a foreign language (FL) context: an instructed language learning setting where L2 exposure occurs in fairly brief, delineated intervals, and where learners are otherwise largely surrounded by the L1. In contrast, the YA takes place by definition in an L2-speaking society, where learners are very likely to be regularly exposed to extensive naturalistic input - from a wider range of native speaker interlocutors and in situations beyond those typically encountered in the classroom - for a larger portion of their daily lives during their stay (Rehner & Mougeon 2003). In comparing learners' production and comprehension of viewpoint aspect in L2 French before their year abroad – when they have been exposed to predominantly instructed input - and after their year abroad – which constitutes an extended period of exposure to naturalistic input – this project

hopes to begin to explore the possible impact of these different input "types" on L2 French viewpoint aspectual development. It is recognised that characterising linguistic input – in terms of both quantity and content – represents significant methodological challenges, and consequently that a precise definition of "naturalistic input" or "instructed input" is beyond the scope of this project. However, an endeavour is made to partially mediate this by incorporating an analysis of not only the production data from the L1 French experimental controls, but also of an over 10,500-word sample of L1 French oral corpus data, sourced from the freely-available CFPP2000 (***http://cfpp2000.univ-paris3.fr/index.html***). The data from this large sample of conversational French is utilised to obtain more information on what kind of information on viewpoint aspectual form-meaning mappings may be available to learners in naturalistic input.

The aforementioned analysis aims principally to evaluate the impact of an extended period of residence abroad on grammatical proficiency, specifically represented here by learners' command of viewpoint aspect in L2 French. Via the novel combination of L2 French production and comprehension data with a frequency/distributional analysis of viewpoint aspectual forms in L1 French oral corpus data – which, to my knowledge, has not to date been applied to the unique learning context of university-level L2 learners – this project aims to begin addressing the question of whether the widely-reported linguistic development that follows programmes such as the year abroad (e.g. Carroll 1967; Freed 1995; Segalowitz & Freed 2004; Llanes & Muñoz 2009; Mora & Valls-Ferrer 2012) can be linked to the change in input type that learners are exposed to during this time. These findings could additionally prove generalisable to the language learning classroom: for example, if it is found that a given form-meaning mapping poses a particular challenge, and is also less naturally frequently-occurring, it may be beneficial to provide more exposure to this form in instructed settings to facilitate acquisition.

Taking a wider perspective, it could also be suggested that a more holistic approach to L2A that incorporates information from the input alongside generative Feature Reassembly-based predictions is appealing on theoretical grounds. Though the field of second language acquisition research has typically witnessed a trend of separation between generative studies employing mainstays such as Feature

Reassembly (e.g. Dominguez et al 2011, 2017; Hwang & Lardiere 2013; McManus 2015) and usage-based studies focusing on input properties (e.g. Ellis 2002; Bybee 2008; Collins et al. 2009), there is no reason why these two areas should remain so compartmentalised. Indeed, support for such a union is seen in the work of generativists such as C. Yang (2002) and Lidz & Gagliardi (2015), who have striven to conceive of models of language acquisition that, while constrained by Universal Grammar, are still able to account for the role of the input in a way that goes beyond the traditional "poverty of the stimulus" (Chomsky 1955/1975, 1978). Rankin & Unsworth (2016:564) stress that a greater focus on the role of input in generative SLA research is not only desirable but essential, warning that 'this perceived divergence [between generative and usage-based research] threatens to side-line generative SLA if the impression persists that it has nothing to say about input.'

The above also aligns with a sustained call to generative researchers (e.g. S.Carroll 1996, 2001; Felix 1986; E. Klein & Martohardjono 1999), epitomised by Gregg (1996) as the need to identify a suitable "transition theory" to accompany the "property theory" that is Universal Grammar (UG): in other words, to provide some sort of mechanism for how language acquisition actually happens. An instance of where such a mechanism could prove enlightening is within Schwartz & Sprouse's (1996) seminal Full Transfer Full Access model, which posits full availability of Universal Grammar during L2A in order for "restructuring" of the developing L2 grammar to happen: however, the manner in which this restructuring process takes place is not explicitly stated. This is particularly pertinent to the current project, given that both Full Transfer and Full Access (and, by definition, restructuring) lie at the theoretical core of Feature Reassembly. Though it must be stressed that solving the "transition theory problem" will be far from straightforward, due in main to the high degree of supposition inevitably accompanying any theorising on the structure and functioning of mental systems of language, this project aims to explore whether an approach to L2A combining UG-based and input-based information may assist in making small steps towards a better understanding of the acquisition process.

In summary, the debate surrounding the role of input within a Minimalist framework (Chomsky 1995) is an ongoing and essential one, and further exploration into this challenging and complex question has the potential to generate powerful insight into what is sometimes perceived as an area left somewhat unexplored by

generative researchers (Rankin & Unsworth 2016, Yang & Montrul 2017). This project's focus on the L2 development of viewpoint aspect within a university year abroad setting constitutes an interesting and relevant context in which to engage with this issue, notably as it involves learners who not only experience a change in the L2 input they are exposed to, but who have also already attained a relatively advanced point in their L2 development and so can be assumed to have undergone – and still be undergoing - at least some "restructuring" of the form-meaning mappings transferred from their L1. Although viewpoint aspect in French would not be categorised as a classic "poverty of the stimulus" phenomenon in itself, it could be posited that the existence within viewpoint aspect of four distinct form-meaning mappings – which have been found to occur with markedly different frequencies – makes this grammatical property an excellent candidate to explore the manner in which input properties such as frequency and distribution may influence acquisitional trajectories. This, in combination with the complex and varied reassembly task involved in acquiring L2 French viewpoint aspectual form-meaning mappings as an L1 English speaker, sets this project up as a rich terrain to explore the explanatory power of feature reassembly, input properties, and the intersection of the two.

In view of the above, the research aims of this project are as follows:

RQ1: To what extent can predictions from feature reassembly accommodate the developmental path of the viewpoint aspectual system in L2 French for L1 English university-level learners?

RQ2: What can studying the properties of naturalistic input tell us about the impact of programmes such as "the year abroad" on grammatical development, with a particular focus on the development of viewpoint aspect?

RQ3: To what extent can an approach combining information from feature reassembly with information from the input explain the process of L2 (aspectual) development at advanced levels?

This thesis will address the above questions according to the following structure. First, given the important theoretical component to the project, it will open with an explanation of theoretical assumptions in Chapter 2. Chapter 3 includes background information on viewpoint aspect in English and French, as well as a review of research investigating L2 aspectual development and the role of input on language

acquisition. The research design of the project is set out in Chapter 4, and results are presented in Chapter 5. Finally, Chapter 6 presents a discussion with regard to the three research aims and the final conclusions of the project.

2. Theoretical assumptions

In order to investigate whether a greater integration of the role of linguistic input into a generative approach to L2A would be useful, it is first necessary to be clear what is meant by "generative", and to set out the theoretical assumptions implicit in this stance. Sections 2.1 and 2.2 of this chapter will explain the basic tenets of the generative perspective - including Universal Grammar (UG), arguably its central component - and examine how generativism has evolved over the years, from the more classical Principles & Parameters (P&P) model (Chomsky 1980a, 1981a, b) to the Minimalist framework (Chomsky 1995). Importantly, this chapter will also discuss, in Section 2.3, how the assumptions instantiated by the generative perspective relate specifically to the acquisition of a second language (L2).

2.1 Universal Grammar and the generative approach to language acquisition

Though the idea that language acquisition is biologically-determined – that is, that we are physiologically adapted for language - is solidly attested (e.g. Chomsky 1959, 1965, 1981b, 1986; Aitchison 1976; Pinker 1994)**,** greater contestation surrounds whether the human capacity for language acquisition is specific to language only (domain-specific), or is borne from more general cognitive learning mechanisms that are simply applied to language (domain-general). Generative linguists advocate for a domain-specific language capacity in the form of an innate, biologically-endowed language faculty: Universal Grammar (UG). The primary argument for the existence of UG, according to generative scholars, is that the linguistic input to which children acquiring their first language are exposed does not contain enough evidence of how the target language works in order to explain the rapidity and uniformity of success of first language acquisition (L1A): this is known as the logical problem of language acquisition or the poverty of the stimulus (PoS) (Chomsky 1955/1975, 1980b; see also section 3.1 for a more in-depth discussion of the role of input in language learning). Consequently, generative linguists posit that humans are born pre-equipped with an innate language faculty (Universal Grammar), which - by setting constraints on a developing language grammar via the application of invariant universal principles and determining, within a central computational component, the kind of structural (syntactic) operations that can take place (Chomsky 1965, 1980a, 1981b; Pinker 1984, 1994) – serves to bridge the gap

between the impoverished input and successful language acquisition.

Here arises a key contrast between the generative approach and what is often positioned as its theoretical opposite: emergentist, or usage-based, approaches. The latter perspective is more closely allied with the application of domain-general cognitive skills - such as utilising statistical patterns in the input - to language learning: for Tomasello (2003:328), a prominent usage-based researcher, 'how children learn language is not a logical problem but an empirical problem'. It follows that for usage-based researchers, it is necessary to accord significantly more importance to the role of the input and linguistic experience, as under this perspective it is solely on the basis of such experience (and general cognitive skills) that language development takes place. Contrastingly, for generativists, the existence of UG is fundamentally motivated by a learnability problem related to children's proposed *lack of* linguistic experience. Given that primary linguistic data in L1A is not only limited in quantity, as mentioned, but is also underrepresentative of the complex linguistic knowledge that children nonetheless successfully acquire, the only way that this success may be accounted for is via the presence from birth of a built-in linguistic "blueprint" containing principles and properties that are universal to all languages (Baker & McCarthy 1981; Hornstein & Lightfoot 1981; White 2003). This Universal Grammar is consequently said to "constrain" a child's developing *grammar* - used here to mean the system of mental representations underpinning their knowledge of language - by establishing prior to the start of acquisition what is and is not possible in human language, and thus restricting the kind of hypotheses children may make about the target language to a much smaller subset of possibilities (Chomsky 1965, 1980a, 1981b). This is evidenced not only indirectly by the speed and uniform success of first language acquisition, but also by the fact that normally-developing children do not present with "wild" or illicit grammars containing elements impossible in any language (Goodluck 1991).

2.2 Generative developments: from Principles and Parameters to Minimalism

The generative paradigm has undergone a number of key transformations in its quest to explain language and its acquisition in a manner that is both sufficiently detailed, yet also retains simplicity and minimises unnecessary extra components. Initially, Universal Grammar was conceptualised as a blueprint for language which constrained linguistic development in two ways: firstly, via the presence of universal

language principles responsible for setting out the broad outlines of natural language grammars; and secondly, via language-specific parameters, which served to account for cross-linguistic variation (Chomsky 1965, 1980a, 1981a, b). Parameters were considered to be binary (having two settings), and to be triggered based on the input the child was exposed to: for example, the Null Subject Parameter (Hyams 1986) was proposed able to be configured, or set, to either the null subject setting (upon sufficient exposure to languages such as Spanish) or the overt subject setting (upon sufficient exposure to languages such as English). Parameters were also assumed to operate in "clusters": that is to say, that the setting of one parameter had a cascade effect on the setting of other parameters which may or may not appear related on the surface (Chomsky 1981a). This clustering effect, as well as the fact that the values of the parameter settings were predetermined by UG, was considered by generativists as the means by which the Principles & Parameters framework accounted for the logical problem of language acquisition – that is, by reducing the "volume" of the acquisition task - as well as for variation between languages (White 2003).

However, over time, perspectives within the generative field regarding the internal language architecture began to shift. The Minimalist Program (Chomsky 1995) was motivated by the theoretical appeal of a system of language which achieved its job as efficiently as possible (i.e. by implicating a minimum of components specific to language) whilst still accounting for cross-linguistic variation. Under Minimalism, the language architecture was reconfigured to comprise a central, linguistically invariant computational component (also called narrow syntax); the lexicon; and two interfaces - Phonetic Form (PF, responsible for phonological form) and Logical Form (LF, responsible for meaning) - to which linguistic information flows via narrow syntax. The introduction of a linguistically invariant computational component called for a revision of the notion of parametric variation: if syntactic operations were now considered universal, the idea of a parameter setting which captured, for example, differences in word order (e.g. SVO vs. VOS), was no longer compatible. The attention turned to the lexicon as a source of variation, with Chomsky (2001) adopting a proposal from Borer (1984) that cross-linguistic grammatical differences were associated with the properties of lexical items, particularly the heads of functional categories (such as D(eterminer) and T(ense)). The Borer-Chomsky conjecture (Baker 2008:253) forms one of the central tenets of the generative Minimalist program, de-emphasising parameters in the classical sense and placing greater focus on features - as hosted by functional category heads - as the unit

of linguistic variation (Liceras et al. 2008). Consequently, a modern generative perspective on language acquisition posits that acquiring a language essentially means acquiring its morphosyntax: in other words, selecting the appropriate features present in the language and assembling them onto (functional) morphological forms. Indeed, it has been proposed (Slabakova 2008, 2016) that once this feature mapping process has been actuated, a certain amount of relevant associated syntactic and semantic information is also automatically activated: in this way, the acquisition of morphosyntax represents a "bottleneck" to language acquisition. As can be inferred, the Borer-Chomsky conjecture has played a notable role in the development of generative theories of second language acquisition (see further discussion in section 2.3.2).

2.3 Implications for second language acquisition

2.3.1 L1 transfer and access to UG in L2A

Similarly to the generative L1A research paradigm, generative studies of second language acquisition have also been motivated by the logical problem of language acquisition, with a key question being the extent to which learners of a second language have access to the same "tools" - in other words, Universal Grammar - to fill the gaps left by impoverished input as do first language learners (Schwartz 1998; White 1989). An initial pressing query regarded the "access question": whether adult L2 learners still had access to UG during the acquisition of their second language, or whether they did not, which would mean that L1 and L2 acquisition were fundamentally different processes (Bley-Vroman 1989, 1990). In order to establish whether adult L2 learners did have access to UG during L2A, research focused on whether or not adult L2 grammars exhibited "poverty of the stimulus" (PoS) properties which - in the same way as for L1A - would attest to access to a domain-specific language faculty (UG) to instantiate grammatical information not present in the input. It has since been extensively demonstrated (e.g. Kanno 1997; Pérez-Leroux & Glass 1999; Schwartz & Sprouse 2013) that L2 grammars, albeit differing significantly from L1 grammars, do indeed contain grammatical properties that cannot be attested either to transfer from the L1 or to domain-general learning mechanisms, inferring at least some degree of access to UG in L2A. The above also highlights the other main source of linguistic information proposed

to be available to L2 learners: grammatical properties transferred from their first language, which were considered to constitute the starting point or initial state of L2 acquisition.

The interplay between L1 transfer at Initial State and UG access for subsequent grammatical development constituted a central question within the generative L2A field, with a range of theories advanced proposing varying degrees of availability of each. To contrast two examples of such theories, the Minimal Trees Hypothesis (Vainikka & Young-Scholten 1994, 1996) proposed full access to UG but partial transfer of only lexical (and not functional) categories from the L1, whereas the Failed Functional Features Hypothesis (Hawkins and Chan 1997) advocated only partial or indirect access to UG, with lexical categories considered learnable but availability of L2 functional features being restricted only to those already instantiated in the L1. Particularly influential has been Schwartz & Sprouse's (1994, 1996) Full Transfer/Full Access hypothesis, which, as the name suggests, posits that a learner's L2 Initial State consists of the entirety of their L1 grammar, and that following this there is full access to Universal Grammar to facilitate L2 development, including features/categories that were not present in the L1. The latter hypothesis has gained significant traction within generative L2A, which can be linked to the fact that it accounts for the presence of PoS properties in L2 grammars as found in the "access question" studies mentioned at the start of this section. Under the Full Transfer/Full Access Hypothesis (FT/FA), complete acquisition in L2 learners is theoretically possible - in line with the observation that L2A is not fundamentally different to L1A - although this is mediated by the L2 input to which the learner is exposed, meaning that full convergence to nativelike norms is not guaranteed (Schwartz & Sprouse 1996).

2.3.2 Feature reassembly and L2 development

As mentioned previously, the move towards Minimalism and the concomitant espousal of the feature as the principal unit of linguistic variation had a notable impact on resulting theories of second language acquisition. Several different proposals were advanced in an attempt to account for particular areas of difficulty in correctly associating the features of an L2 with their appropriate morphological forms, with some linking this to feature (un)interpretability (Hawkins & Chan 1997; Hawkins & Hattori 2006; Tsimpli & Dimitrakopoulou 2007) and others to whether the features in question

operated at “interfaces” between the grammar and external systems such as discourse or pragmatics (Sorace & Filiaci 2006; Sorace 2011).

One popularly-espoused theory which aims to shed further light on the fact that, despite FT/FA, L2 acquisition does not always converge with monolingual L1 acquisition is Lardiere’s (2003, 2005, 2008, 2009) Feature Reassembly Hypothesis (FRH), which considers the role played by L1 influence in the development of the L2 feature system (i.e. the L2 grammar). The basic premise of the FRH revolves around the Minimalist idea that language acquisition can be equated to learning the formal features of a language and, importantly, how they are bundled or assembled on the lexical items (including functional morphemes) of that language. Under FT/FA, learners of a second language begin with all of their L1 feature bundles transferred to the Initial State of their L2 grammar, and then have access to the universal repository of all features - including those not present in their L1 - to aid in the formation of the target-like feature bundles for their L2. However, it is the actual process of reconfiguring or reassembling these L2 feature bundles from the starting point of the L1 bundles that represents the onus of the learning task - particularly if the L2 groups particular features across lexical item(s) in a way that is different to the L1 (Lardiere 2009). In this way, the FRH offers a potentially more nuanced perspective on second language acquisition than other feature-based accounts such as the Failed Functional Features Hypothesis (Hawkins & Chan 1997) or the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou 2007). It is not merely a case of whether the relevant features are *available* – e.g. by being already selected in the L1 – but rather whether and to what extent the L1 and L2 bundle these features differently across various lexical items. The FRH has proven very useful in aiding generative L2A researchers not only to visualise the degree of L1 influence to be overcome for a given L1/L2 pair - based on how similarly or dissimilarly formal features are bundled in each language - but also has provided fertile terrain for researchers to make fine-grained predictions about the relative ease or difficulty of L2 acquisition of particular features, often represented by functional morphology (e.g. Hwang & Lardiere 2013; Spinner 2013; Cho & Slabakova 2014; Mai & Yuan 2016). A fruitful application of the FRH can be seen in studies of L2 aspectual development, as will be further discussed in section 3.2.3.

3. Background Information

This chapter will review the literature and provide background information on the central themes of this project. Section 3.1 explores the complex relationship between Universal Grammar and the input in studies of first and second language acquisition, and evaluates key instances where researchers have endeavoured to unite these two sources of linguistic information. Section 3.2 provides background information on aspect, both in terms of general aspectual theory (section 3.2.1) and as specific to both French (3.2.2) and French (3.2.3), before considering existing research on L2 aspectual development (3.2.4). Finally, section 3.3 presents an overview of L2 research carried out in study/residence abroad contexts, and situates the current project within this.

3.1 Input and Universal Grammar

As previously discussed, the idea that linguistic input alone is insufficient for successful language acquisition is a cornerstone of generative linguistics. A key underpinning of this stance is the aforementioned Poverty of the Stimulus (PoS) (Chomsky 1955/1975, 1978), summarised by Lasnik & Lidz (2016:1) as the fact that in language acquisition 'our experience far underdetermines our knowledge and hence… our biological endowment is responsible for much of the derived state': for generativists, this "biological endowment" is Universal Grammar (UG). Chomsky (1967) posits that the input to which a child is exposed is degenerate not only in scope - for it would be impossible for it to contain every conceivable form-meaning mapping and sentence structure - but also in quality, as 'the input itself does not contain information about the kinds of representations that should be used in building a generative grammar of the language' (Lasnik & Lidz 2016:3). Efforts to disprove the PoS argument on the grounds that the input contains sufficient "disconfirming evidence" to allow children to rule out incorrect hypotheses about the language they are acquiring - and thus that input suffices alone to arrive at the target grammar (e.g. Pullum & Scholz 2002) - have been countered by research indicating that, even if such evidence does occur in the input, its frequency is 'low enough to be considered negligible, that is, not reliably available for every human child' (Legate & Yang 2002:158). Schwartz & Sprouse (2013:138) comprehensively outline many cases of PoS in language acquisition,

concluding that the only explanation for consistent successful language acquisition is that 'the brain/mind of human children is endowed with UG, a network of domain-specific cognitive predispositions that filter the input and narrowly constrain the set of grammars that can be projected from the input', and consequently that 'the stimulus is "impoverished" only from the perspective of the expectations of a purely inductive domain-general learning hypothesis.'

Though the cardinal role played by Universal Grammar in language acquisition has been extensively demonstrated, with many positing its full availability in the acquisition of first and subsequent languages (e.g. Schwartz & Sprouse 1996; White 1989, 2003; Slabakova 2008), research has simultaneously highlighted the importance of the input. One indicator of this importance is found in bilingual first language acquisition (BFLA). Though studies (e.g. Thordardottir 2011, 2015; Cattani et al. 2014) have shown that children simultaneously acquiring two L1's appear to require less input than their monolingual peers to attain the same level in some areas, for other domains the fact that BFLA children receive less input per language appears to delay successful acquisition: an effect that may prove enduring in the case of more complex phenomena (Thomas et al. 2014; Unsworth 2014). There appears then to be a link between the level of complexity or opacity of a given linguistic property and the amount of exposure to that property necessary for its successful acquisition: a relationship amplified in the case of BFLA where the amount of input per language is less abundant (Gathercole 2002; Paradis 2010; Blom et al. 2012; Unsworth 2013, 2014). This relationship has also been attested to in L1A in Miller & Schmitt's (2012) investigation of plural morphology acquisition in two varieties of L1 Spanish. The variable and complex input provided to children acquiring Chilean Spanish meant that they took markedly longer to acquire targetlike comprehension and production of plural morphology than their Mexican counterparts, testifying powerfully to the inverse relationship between complexity and rate/ease of acquisition, even in monolingual L1A when exposure to the target grammar on the whole is abundant.

If input effects can be seen in such circumstances, we would expect to continue to see them in L2A, where input 'is not abundant, unambiguous and consistent, but messy, inconsistent and ambiguous' (de Bot 2015:263). Though the latter researcher is not associated with the generative approach, their assertion clearly aligns L2A with the same PoS conditions as in L1A, and indeed with a stimulus that (with

the exception of total immersion contexts) is even more impoverished. Consequently, it is clear that further research into the role of input in L2A – particularly regarding the acquisition of complex properties by learners such as the typical university-level languages student, whose exposure has not originated from total naturalistic L2 immersion but from a "messy, inconsistent and ambiguous" mixture of sources - is essential in order to advance our understanding of the input "conditions" required for such properties to be acquired.

The observation of effects such as the above has increasingly led researchers to call for a greater focus on the role of input in language acquisition research. Miller & Schmitt (2012:277) highlight that 'one goal of language acquisition is to determine the extent to which the input provides evidence for the mappings between form and meaning that children must acquire and how that might affect the acquisition of the target grammar' and acknowledge that, while the process of form-meaning mapping itself is innate, the rate at which it occurs may depend on the frequency with which a given form and interpretation co-occur in the input. The authors look to C. Yang's (2002) Variational Model, wherein 'the grammar most closely aligned with the input will win out' (Miller & Schmitt 2012;225), for an approach to acquisition which could best accommodate their findings. A generative linguist, C. Yang (2002:24) maintains that 'while there is no doubt that innate UG knowledge must play a crucial role in constraining the child's hypothesis space and the learning process… statistical learning seems most naturally suited to modelling the gradualness of language development': a gradualness which could arguably be quite naturally extrapolated to incompleteness, a characterising feature of many L2 grammars. The Variational Model presents language acquisition as a probability-driven process, whereby grammars "compete" and are evaluated on their "fitness" based on the proportion of encountered sentences that they are respectively able to parse and analyse (C. Yang 2002, 2006). Applying this concept to the Null Subject Parameter, exposure to a language such as Spanish with a high number of sentences containing null subjects would cause a grammar which permits null subjects to be evaluated as more fitting - as it is concordant with a large proportion of sentences in the input in this respect – and a grammar which does not allow null subjects being evaluated as less suitable, as it cannot accommodate the large number of null subject sentences encountered. If the input language were

English, the opposite grammar would be favoured to analyse the input. Though the Variational Model was conceptualised during the parameter-setting era - and would consequently require reworking to accommodate a Minimalist feature-driven approach - a point of importance is the straightforwardness with which it reconciles generative and statistical concepts: for C. Yang, UG provides and constrains the hypothesis space where grammars compete, and statistical learning provides the mechanism for this competition.

Some relatively contemporary research by Mai & Yuan (2016) on the acquisition of [past], [telic] and [given] features by L1 English learners of L2 Chinese also acknowledges the appropriateness of a Yangian model of guided probabilistic learning to accommodate their findings, noting that the gradual grammatical restructuring that Yang proposes is reflected in the "uneven" reassembly of these features onto one morphological construction. A similar "unevenness" of reassembly is also seen in several studies of L2 development of viewpoint aspect (see Section 3.2.4). The inherently incremental nature of the reassembly process (Lardiere 2009) predicts gradual, "uneven" development over time, and a central reason for this is that the "restructuring" of the L1 form-meaning mappings transferred at Initial State occurs in response to difficulties in analysing the L2 input (Schwartz & Sprouse 1994, 1996). Given this, the notion that feature reassembly may be sensitive to properties of the input appears logical, and one might thus extrapolate that any change in the input encountered has the potential to impact reassembly. Crucially, the fact that some of the most advanced of Mai & Yuan's (2016) learners were able to reassemble all of the aforementioned features and showed evidence of nativelike judgements testifies to Lardiere's (2009) claim that any contrast in formal features between the L1 and L2 that can be detected by a learner is in principle acquirable. This in turn indicates that findings that are well-explained by a model of statistical learning do not preclude the usefulness of generatively-grounded frameworks such as feature reassembly.

C. Yang (2002) has not been the only generative researcher who has endeavoured to develop a model of language acquisition which makes space for both UG and input-dependent mechanisms. In a similar vein, Lidz & Gagliardi (2015) support Gregg's (1996) advocacy for both a "property" and "transition" theory of language acquisition, acknowledging (2015:349) that the UG-defined representation space 'only sets the initial conditions for learning' and that 'we must also have mechanisms for

mapping sentences onto those representations and for defining the environmental inputs that guide that mapping process.' From this proviso, Lidz & Gagliardi (2015:349) outline a model in which the language acquisition process is divided into three parts. Firstly, intake mechanisms filter the input and identify the essential information for acquisition. After this, access to UG is required to fill the disparity between the experienced and the acquired and also to allow learners to make inferences from statistical-distributional evidence in the input to the abstract grammatical representations underpinning this evidence. Finally, inference mechanisms link UG and intake and ensure that the correct abstract representation (retrieved from UG) is mapped to the correct surface form encountered in the input. This integrated model of language acquisition convincingly demonstrates the utility of incorporating the innate framework of constraints imposed by UG within a more detailed consideration of the statistical and distributional properties of the input.

Great as the potential of these models is, it is noted that they were conceptualised for first language acquisition. In a recent paper, C. Yang (2018:684) proposes that his Variational Learning Model is equally applicable to second language acquisition, postulating that 'the differences [between adult and child language acquisition] do not have to relate to differences in [the] underlying (cognitive/linguistic) mechanisms available to each.' The author looks to contrasting evidence from L2 acquisition of pro/topic-drop vs. obligatory subject languages to support his claim, juxtaposing works such as Phinney (1987) and Kanno (1997) – wherein L2 learners of Spanish and Japanese show excellent command of null subjects and null objects, respectively – with Judy's (2011) findings that even near-native L2 learners of English do not consistently use expletive subjects, the definitive sign of an obligatory subject grammar. C. Yang (2002, 2018) argues that the reason null subjects/objects in Spanish and Japanese are more readily acquired than obligatory expletive subjects in English is due to a higher incidence in the input of disambiguating evidence attesting to the former compared to the latter, and that the Variational Model can straightforwardly accommodate this state of affairs given that 'variational learning is gradual, probabilistic and quantity sensitive' (2018:685).

However, in response to this, Slabakova (2018) notes that there are elements of the L2A process that C. Yang's input-driven approach still cannot account for. Strikingly,

these shortcomings relate largely to the Minimalist concept of features, and specifically the fact that multiple features can be hosted by one functional category (e.g. TP, Tense Phrase), in the form of functional morphology (e.g. *–ed* for pastness and perfectivity, or *-s* for third person and singular number). Slabakova draws on White's (2003:187-93) seminal exposition of Lardiere's (1998 a,b)'s end-state adult L2 English learner, Patty, combining this with child data from the same L1/L2 pair (Chinese/English) (Li 2012). The data show a clear disparity between the accurate suppliance of the verbal inflection itself (*-ed*: 25.5-34.5%; *-s*, 4.5-16%), and evidence of morphosyntactic phenomena that are linked to it, such as overt subjects, nominative case, and no verb raising (accuracy at/close to ceiling). This puts input-driven approaches such as the Variational Model (C. Yang 2002) and the Tolerance Principle (C. Yang 2018) in a difficult position, as formal features associated with a morphological form (and functional category) appear to be being acquired before the form itself, and at different rates. This is challenging for models that rely heavily on the presence of (morphological) evidence in the input to reconcile, as they would predict the acquisition of the meanings hosted by a particular morphological form to be acquired at the same time and as a function of the frequency of that form in the input. Slabakova (2018) concludes that, whilst models such as the Variational Model and the Tolerance Principle are suitable in cases where there is already adequate information in the input for acquisition to occur - as is indeed the case for many areas of the acquisition process (Rothman & Slabakova 2018) - they struggle with cases where 'learner knowledge is both under-determined by the linguistic input and under-represented by learner production' (2018:780-1). Therefore, as it currently stands, input-driven models alone are argued to be unable to account for poverty-of-the-stimulus effects seen in both first and second language acquisition. This does not, however, mean that generative L2A researchers are entirely discounting a more thorough incorporation of factors such as frequency into their work. On the contrary, investigations into factors such as frequency and distribution are increasingly found in generative studies spanning a plethora of L2 acquisitional domains (e.g. Domínguez & Arche 2014; Slabakova 2015; Hopp et al. 2020), potentially in partial response to the previously-discussed calls for such an emphasis (e.g. Rankin & Unsworth 2016). Despite the importance of holding input-driven models of acquisition to rigorous standards, Slabakova (2015:25) nonetheless asserts that 'teasing apart the primacy of transfer [via the L1/UG] or frequency [via the input]' is an area that 'certainly merits further research.'

It is hoped that the above review of the literature serves to highlight the complexity of the ongoing debate surrounding the role of input in (generative) studies of language acquisition. Despite the current difficulties in fully reconciling input-driven approaches with a Minimalist approach to language acquisition, studies from both L1A and BFLA have pointed to the importance of the input in a variety of acquisition contexts. It is therefore worthwhile to continue to assess the applicability of a "UG-plus-input" approach to second language acquisition - especially given that in L2A, the "stimulus" is often all the more impoverished, rendering the input effects observed in BFLA and L1A arguably even more critical and obliging generativists to seriously evaluate whether their work should allocate a much larger role to the input than is currently the case.

3.2 Aspect

3.2.1 Theoretical background

As previously stated, this project will explore the applicability of an approach to language acquisition that combines Feature Reassembly-based predictions with information from the input, with specific regard to L2 French aspectual development. It is therefore important to be clear on the parts of language that aspect is responsible for. Under a bi-dimensional approach, "aspect" is an umbrella term depicting the interplay between two types of universal semantic information: lexical (or "situation") aspect, the inherent semantics of verbs, predicates and similar; and viewpoint (or "grammatical") aspect, the perspective – whether ongoing, complete, or repeated - from which a situation is presented (Smith 1991, 1997). However, as this project will be focusing on the development of the latter type of aspectual information, it should be noted that further references to "aspectual development", unless otherwise stated, are referring to the acquisition of the viewpoint aspectual distinction between the perfective and the imperfective. This distinction may be conceptualised in several ways, but a traditional exposition is that perfective aspect presents the entirety of a completed situation, whereas imperfective aspect opposingly presents a situation without beginning or end points, indicating that it is ongoing or incomplete (Smith 1991, 1997).

More formally, an extensive and long-spanning technical framework

espouses the idea that Aspect itself, alongside Tense, is a syntactic category (Zagona 1990; Stowell 1993, 1996, 2007), and that the heads of these categories are "ordering predicates" which take intervals of time as their arguments (W. Klein 1994; Demirdache & Uribe-Etxebarria 2000) and consequently establish relationships with them which bear on the temporal interpretation of the sentence (Enç 1987). Under such an approach, Tense and Aspect function according to the same semantic primitives (Reichenbach 1947, Demirdache & Uribe-Etxebarria 2000), and tense and aspectual distinctions are made on the basis of the intervals they take, which operate in pairs. Tense takes the Assertion Time (the time interval that the utterance is referring to, also called Topic Time (Klein 1994, 1996)) and orders it within, before, or after the Reference Time (i.e. the Speech Time), and Aspect orders the Assertion Time either within, before, after, or completely overlapping the Event Time (the interval that the whole eventuality extends over). For example, a perfective aspectual interpretation is obtained when the Assertion Time completely overlaps the Event Time (Klein 1994, Arche 2014): in other words, the eventuality coincides entirely with the time interval being referred to, in parallel with the more traditional "completed/entire event" perspective. Contrastingly, an imperfective aspectual interpretation is yielded when the Assertion Time is ordered *within* the Event Time, in a formalisation of the popular metaphorical description of an "unfinished" or "ongoing" eventuality, or of a partial (as opposed to complete) perspective on an event (e.g. Smith 1991).

Arche (2006, 2014) makes use of the above formalisation of Aspect and expands on it in her work on the three readings (progressive, habitual, and continuous) of the imperfective, thus illustrating how this framework can be usefully appropriated to set out the syntax and semantics of viewpoint aspect in fine-grained detail. As mentioned, for imperfective interpretations Asp orders the Assertion Time, AT (or Topic Time, TT) within the Event Time, EvT, whereas for a perfective interpretation, the ordering predicate within Asp is one of complete overlap. In addition to this ordering of TT with relation to EvT, Arche (2006, 2014) also proposes that the syntax contains a quantificational node, Q<occasions>, which provides information about the number of occasions in each eventuality. For the perfective interpretation, the value of this node is 1, indicating a single completed event. As all the imperfective interpretations take the same ordering predicate ("within"), they must differ in terms of the value of this quantifier. The number of occasions instantiated by Q<occasions> for the progressive interpretation is also 1, but it is distinguished from the perfective interpretation by

means of the difference in ordering predicate: the ordering of TT within EvT for the progressive yields the interpretation of the eventuality extending beyond the point being referred to, whereas for the perfective TT and EvT overlap completely, yielding the interpretation of a completed event. For the habitual, the value of Q<occasions> is >1, indicating multiple instantiations of an eventuality. Finally, the continuous interpretation results from a lack of cardinal (numerical) quantification; in the absence of such, the quantificational node is assumed to take an existential quantifier, represented by ∃. This syntactic representation of (im)perfectivity is summarised in Figure 1.

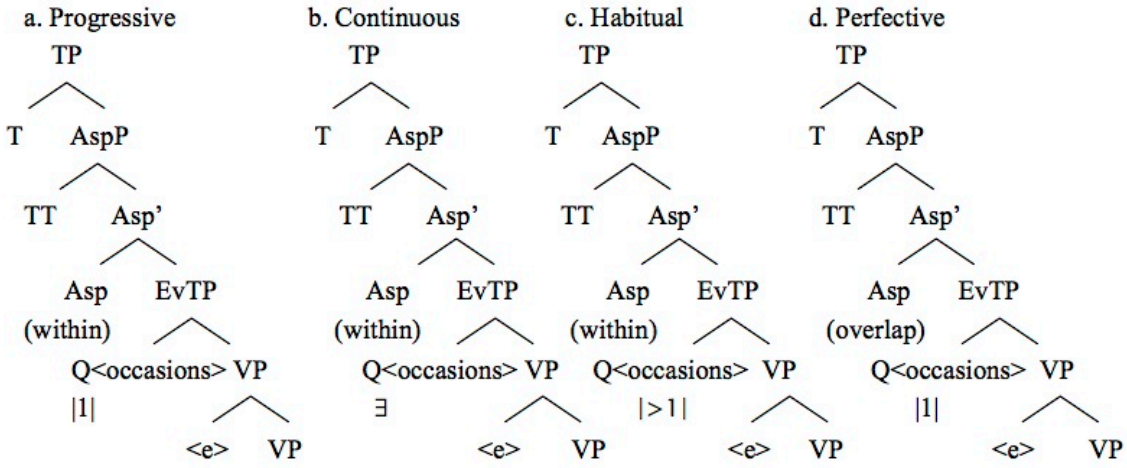


Figure 1: Syntactic structure of viewpoint aspectual interpretations as conceptualised by Arche (2006, 2014)

(Figure from Domínguez et al. 2017:4)

The application of Arche's work to L2 aspectual acquisition will be further explored in section 3.2.4, and will be discussed with specific regard to French in the following section.

3.2.2 Viewpoint aspect in French

Though French may express viewpoint aspect either morphologically or

periphrastically, it predominantly does so via tense morphology, as both viewpoint aspect and temporal reference are mapped to tense in French (Gosselin 1996; Labeau 2005): for example, in the sentence *Lucie a tricoté une écharpe* ('Lucie knitted a scarf'), the composite verbal form *a tricoté* ('knitted') encodes both past time reference and perfective aspect.

Similarly to Spanish (Salaberry 2008), aspectual distinctions in French are only obligatorily made in the past, explaining why the English present perfective *I play* and present progressive *I am playing* are both mapped to the single French present tense form *je joue.* In the past, however, French differentiates morphologically primarily between the Imperfect (*l'imparfait,* IMP) (Example 1), which expresses imperfective aspect, and the *passé composé* (PC) (Example 2), which expresses perfective aspect. French also has the *passé simple* (PS) (Example 3), a preterite-like synthetic form previously used to convey perfectivity in contrast with the composite *passé composé*, which originally expressed perfect tense only. However, the diachronic development of the PC into a host of both perfect and perfective meaning – which has happened more quickly in French than in other Romance languages (Bybee & Dahl 1989; Bybee et al 1994) – has led to the extremely diminished use of the PS in all but a very limited subset of (overwhelmingly written) discourse contexts (Wilmet 2003; Labeau 2005, 2009). Examples 1-3 summarise the viewpoint aspect-expressing forms available in French, along with possible English translations. Note that due to the extremely diminished use of the *passé simple* in modern French, it will not be further considered in this study, but is provided here for the sake of comprehensiveness.

1. *Élodie* *mangeait* *des* *gaufres*.

Élodie eat-IMP.3SG some-PL waffles-PL

'Élodie ate/used to eat/was eating waffles.'

[Imperfective]

2. *Élodie* *a* *mangé* *des* *gaufres*.

Élodie (has-AUX.3SG eat-PP)$_{PC}$ some-PL waffles-PL

'Élodie ate waffles.'

[Perfective]

3. *Élodie* <u>*mangea*</u> *des* *gaufres.*

Élodie eat-PS.3SG some-PL waffles-PL

'Élodie ate waffles.'

[Perfective]

We can observe in Example 1 that the imperfective form *mangeait* can be translated variously into either 'ate', 'used to eat', or 'was eating', illustrating that the French Imperfect (IMP) can be used to convey three different imperfective meanings, which may be labelled as the habitual, continuous, and progressive respectively (Labeau 2005, 2009). Examples 4-6 illustrate these three interpretations, contrasting them with the perfective *passé composé* (Ex. 7), used to express a completed one-time event. The habitual interpretation (Example 4) is understood as a regularly repeated past eventuality, whereas the continuous (Example 5, also called *caractérisant*, 'characterising' (Kihlstedt 1998; Labeau 2011)) is used to describe situations which are true for the entire duration of the time referred to and, importantly, extend beyond it (Labeau 2011). Lastly, the progressive interpretation (Example 6) denotes an eventuality that was ongoing, typically at the point of occurrence with another, perfective, eventuality.

4. *Pendant son enfance, Annabel* <u>*allait-IMP*</u> *au parc tous les jours.*

During her childhood Annabel go-IMP.3SG to-the park all the days

'As a child, Annabel went/would go/used to go to the park every day'.

5. *Quand elle* <u>*était-IMP*</u> *jeune, Magalie* <u>*avait-IMP*</u> *les cheveux bouclés.*

When she is-IMP.3SG young Magalie have-IMP.3SG the-PL hair-PL curly-PL

'When she was little, Magalie had curly hair.'

6. *Sophia* <u>*mangeait-IMP*</u> *un sandwich quand Charlotte est entré-PC.*

S eat-IMP.3SG a-M.SG sandwich when C [is-AUX.3SG enter-PP]$_{PC}$

'Sophia was eating a sandwich when Charlotte came in.'

7. *Coralie <u>a couru-PC</u> une fois autour du lac, puis elle <u>s'est arrêtée-PC</u>.*

C [has-AUX.3SG run-PP]$_{PC}$ one-F.SG time around of.the lake then she

[REFL-is.AUX.3SG stop-PP]$_{PC}$

'Coralie ran once around the lake, then she stopped.'

French also has available the syntactic expressions *avoir l'habitude de* + V 'to have the habit of' and *être en train de* + V 'to be in the process of', which may be used to express habituality and progressivity respectively, particularly if the speaker wishes to place particular emphasis on the habitual or progressive nature of the eventuality (McManus 2011:40).

It is worth noting at this point that previous research on French viewpoint aspect has attributed various different aspectual values to the Imperfect/IMP, beyond the habitual/continuous/progressive interpretation (based on Arche's (2006, 2014) framework) used here. Kihlstedt's (1998, 2002) work on the L2 acquisition of the French imperfective by L1 Swedish learners proposed an acquisitional cline spanning five different interpretations of IMP. First to be acquired was the *IMP caractérisant* ['characterising Imperfect'] (Kihlstedt 1998), which first appeared with stative verbs and which 'mark[ed] the characterisation of an entire time period', e.g. '*avant je <u>voulais</u> travailler avec le français'* [beforehand, I wanted-IMP to work with French] (Kihlstedt 2002:329). This was renamed in later work to *IMP de recouvrement total* ("Imperfect of total overlap", Kihlstedt 2002) to reflect the author's perception that the Event Time in such cases coincided totally with the Topic Time (referred to by Kihlstedt as the Reference Time in line with Reichenbach's (1947) original terminology). A similar idea was adopted by Howard (2005), who further divided Kihlstedt's *IMP caractérisant* into *IMP statique* ['static Imperfect'], for statives, and *IMP caractérisant*, for dynamic verbs. However, as Labeau (2011:70) points out, 'all the information the IMP gives is that the process is valid for the whole of the reference interval, but says nothing on its boundaries, and E[vent] may well have started before [Topic Time] and go on afterwards.' Indeed, as previously discussed, a complete overlap between Event Time and Topic/Assertion Time would actually correspond to a perfective interpretation under Arche's (2006, 2014) framework. It seems that what Kihlstedt and Howard chose to classify as a "characterising" imperfective fits more closely with Arche's continuous imperfective reading, represented syntactically by the ordering of the Topic(/Assertion)

Time within the Event Time, and with the number of occasions within the eventuality receiving an existential (as opposed to cardinal) quantification (Arche 2014, see Figure 1 in section 3.2.1).

Kihlstedt's (1998, 2002) classification of the habitual reading of the Imperfect is relatively similar to the classification adopted here, though she makes a further delineation between regular and irregular habitual actions which Arche (2014) does not, the latter specifying only that the number of (finished) occasions within an (unfinished) eventuality be greater than one for a habitual reading. Lastly, whereas Howard (2005a) includes the progressive as-is in their classification of IMP values, Kihlstedt approaches this somewhat differently again, distinguishing between *IMP d'inclusion brève* and *IMP aux confins*, translated as "Imperfect of short overlap" and "Imperfect 'on the limits'" respectively (2002). The examples provided by Kihlstedt (2002**:**331) for these two classifications respectively are shown in Examples 8 and 9, accompanied by their English translation, with the latter originating from Imbs (1960:62).

8. On s'est fait gentiment remettre en place parce qu'on <u>marchait pas du bon côté</u> sur le trottoir.

We [REFL-is.AUX.3SG make.PP.CAUS]$_{PC}$ kindly put-back.INF in place because-we NEG walk-IMP.3SG not on-the correct side on the-M.SG pavement

'We were kindly <u>told off</u>-PC because we <u>were not walking</u>-IMP on the right side of the pavement.'

9. Vous avez de la chance de me trouver, je <u>sortais</u>.

You.PL have.2PL some.of the.F.SG luck to me.ACC find.INF, I leave-IMP.1SG

'You were lucky to find me, I <u>was leaving</u>-IMP.'

Kihlstedt suggests (2002**:**330-1) that the Imperfect of short overlap is 'close to progressivity', and translatable via the aforementioned periphrasis *être en train de* + V, whereas the Imperfect "on the limits" exists as a means to 'solv[e] the inherent contradiction between the durative, nonlimited character of *imparfait* and punctual,

nondurative verbs', and may be better translated by the periphrasis *être sur le point de* + V "to be about to". As Labeau (2011:70) also highlights, however, there is no real distinction here in terms of a differential ordering of the Event Time within the Topic Time: the imperfective readings in Examples 8 and 9 both illustrate a single eventuality (<walk on the wrong side> or <leave>) that occurs within the reference period or Topic Time, the only contrast being that of durativity, a property of lexical aspect which itself does not impose restrictions on Arche's (2014) construction of viewpoint aspect. For this reason, we combine Kihlstedt's "short overlap" and "on the limits" interpretations of the imperfective into one single progressive reading for the purposes of this project. Table 1 below sets out the viewpoint aspectual interpretations of French as adopted for this project, setting out their characteristics under Arche's (2006, 2014) conceptualisation of (im)perfectivity and indicating where relevant how previous theoretical categorisations have been subsumed within this. The English equivalents to the French forms are also provided; these will be discussed in more detail with regard to viewpoint aspect in the following section.

Name	**Interpr-etation**	**Relationship between TT and EvT (Asp)**	**Number of occasions (Q<occasi-ons>)**	**French form**	**English form**
Perfective	Single, completed event	Complete overlap	1	*Passé composé* Angéline a acheté trois pommes	Simple Past Angéline bought three apples
Continuous (*IMP caractérisant, de récouvrement total,* Kihlstedt 1998, 2002; *IMP statique* (statives)/*caractérisant* (dynamics), Howard 2005a)	True of the duration of the time referred to, also extends beyond	TT within EvT	∃ (existential)	Imperfect Angéline voulait devenir actrice	Simple Past Angéline wanted to become an actress
Habitual (*IMP d'habitude/ d'habitude irregulière* (irregular repetitions), Kihlstedt 2002)	Eventuality unfinished, each instance completed	TT within EvT	>1 (multiple)	Imperfect Angéline jouait beaucoup de tennis	Simple Past/other constructions Angéline played/used to play a lot of tennis
Progressive *(IMP d'inclusion brève/ IMP aux confins*, Kihlstedt 2002)	Ongoing event (usually with respect to another, completed, event)	TT within EvT	1	Imperfect Angéline lisait quand le téléphone à sonné	V + -ing construction Angéline was reading when the phone rang.

Table 1: Characteristics of viewpoint aspectual forms in French, along with their English equivalents.

3.2.3 Viewpoint aspect in English

Viewpoint aspect in English may be expressed either morphologically, or syntactically via use of periphrastic constructions (Smith 1991, 1997). In terms of morphology, English utilises the Simple Past form (SP), which also marks perfect tense (Example 10), and the aspectual morpheme *–ing*, as combined in the construction *be* + V-ing (Example 11). English also makes use of the periphrasis *used to/would* + V (Example 12). Examples 10-12 indicate the range of forms available to express viewpoint aspect in English.

10. Fiona baked a cake. [Simple Past]

11. Alice was washing her hair. [V + -ing]

12. Holly used to/would eat a lot of bagels. [Periphrases]

We now consider these forms in relation to the four viewpoint aspectual meanings central to this project: perfective, imperfective-progressive, imperfective-continuous, and imperfective-habitual (see Table 1 for a summary of these meanings). As Table 1 shows, the Simple Past is used in English to express not only perfectivity, but also the continuous and habitual readings of the imperfective. The *be* + V-ing construction, in contrast, is used solely to express progressivity, and can be combined with verbal predicates in past, present and future tense (Examples 13-15), provided that they are not statives (Example 16) (Smith 1991:73).

13. I was trying to fall asleep.

14. You are wondering what to wear.

15. She will be going to London next week.

16. *We were/are/will be being happy.

Given that the progressive is the only imperfective meaning to have a dedicated morpheme (*-ing*) – as discussed, the *–ed* marker associated with the Simple Past form is not dedicated to the imperfective as it also expresses perfectivity - this infers that the main aspectual contrast made in English is between the perfective and the progressive, as opposed to languages such as French and Spanish which contrast a form solely used for the perfective with one spanning all three imperfective interpretations (Comrie 1976; Smith 1997; Kihlstedt 2002; Arche 2006, 2014; Labeau 2011). Though habituality may be expressed via the periphrases outlined in Example 12, Tagliamonte & Lawrence's (2000) corpus analysis of 1.5 million words of spoken British English revealed that the Simple Past is in fact used in nearly 70% of habitual contexts, far outstripping the more traditionally-cited markers of habituality, which occurred 19% (*used to* + V) and 6% (*would* + V) of the time (Tagliamonte & Lawrence 2000:329, 349). This attests strongly to the notion that for English speakers, the SP is not only a marker of perfectivity, but also of (imperfective) habituality (Montrul & Slabakova 2002, 2003; Slabakova & Montrul 2002; Arche 2014). In addition to the SP's use to convey habituality (Example 17), it is also used in English to express continuousness (Example

18): this final imperfective interpretation is the sole expressed by the Simple Past alone.

17. My sister visited us once a month, and she always brought fairy cakes.

18. The girl had blue eyes and wore a pink sundress.

To summarise, we can see that, in contrast to the "all-purpose" function of the French imperfective form (the Imperfect), which encompasses habituality, continuousness, and progressivity, English does not map imperfective meanings onto forms in the same way. The progressive is the only imperfective interpretation to have a dedicated form (*-ing*), while the other two meanings are predominantly (in the case of the habitual) or exclusively (in the case of the continuous) expressed via the Simple Past. Given that the latter form also expresses perfectivity, we might anticipate L1 English learners of an L2 such as French to struggle to identify and disentangle the imperfective meanings expressed on the Simple Past in their L1 and remap them onto the imperfective form in the L2. The following section will explore this and related questions further by discussing previous and current investigations into the role of the L1 on the development of viewpoint aspect in a second language.

3.2.4 L2 aspectual development

Though it is a language universal based on essentially invariant semantic primitives (Comrie 1976; Chung & Timberlake 1985; Bertinetto 1997, 2001; Smith 1991, 1997, 2006), the manner in which aspect is expressed demonstrates notable cross-linguistic variation (e.g. Comrie 1976; Smith 1991, 1997, Verkuyl et al 2005). It is therefore perhaps unsurprising that the L2 acquisition of aspect has been revealed to be susceptible to L1 influence, and that this influence varies according to the L1 of the learner (e.g. Salaberry 2000; Labeau 2005; Gabriele 2009; Domínguez et al. 2011; McManus 2015). Roberts & Liszka (2013) investigated the impact of L1 influence on L2 aspectual development in their study of advanced L1 French and German learners of English. They found that only the French learner group demonstrated implicit knowledge of L2 English aspectual distinctions, suggesting that the fact that both French and English grammaticalise aspect (unlike German (Duden 1995, Durrell 2006)) had a facilitative effect. However, if grammaticalising aspect in the L1 were the only relevant factor involved in acquiring the aspectual system of an L2, this does not fully explain why the acquisition of viewpoint aspect appears to pose such persistent problems for

English L1 learners (Coppieters 1987; Montrul & Slabakova 2002, 2003; Bartning & Schylter 2004; Howard 2005a; Ayoun 2013; McManus 2015). It therefore appears necessary to look more closely at exactly *how* the L1 and L2 grammaticalise aspect, and what differences and similarities of expression there are between the languages.

Considering a particularly relevant example to this project, Izquierdo & Collins (2008) examined the L2 development of French viewpoint aspectual distinctions in native speakers of both English and Spanish with similar levels of L2 French proficiency. Both L1 groups completed a cloze (gap-filling) test where the required forms were either the Imperfect or the *passé composé*, and a subset of participants then undertook a stimulated recall where they were asked to justify their choice of viewpoint aspectual forms on the task they had just completed. The authors' predictions that the L1 Spanish speakers would show a more accurate use of the French Imperfect, regardless of potentially confounding factors like lexical aspect, were substantiated: the Spanish-speaking learners performed significantly more appropriately regarding use of the Imperfect, appearing less easily misled in their judgements by non-prototypical combinations of lexical and viewpoint aspect, and made more explicit references to similarities in aspectual expression between French and their L1 in the stimulated recall interviews. Moreover, the significant difference in accuracy found between the L1 English and L1 Spanish groups for use of the Imperfect was not found in the case of the *passé composé*, testifying to the notion that for speakers of L1 English, perfectivity is more readily acquired whereas imperfectivity remains an area of difficulty (Harley 1978, 1992**;** Bergström 1995; Montrul & Slabakova 2002, 2003; Howard 2005a; Labeau 2005, 2011).

The enduring effect of L1 influence on viewpoint aspectual development for even highly-proficient learners can be seen with respect to French in Kihlstedt's (2002) work with the L1 Swedish/L2 French pair: for these university-level learners, whose L1 does not grammaticalise viewpoint aspect, use of the French imperfective was not fully nativelike, especially in contexts with less prototypical combinations of lexical and viewpoint aspect, such as IMP with telic verbs. It is worth mentioning at this point that a substantial amount of work on L2 aspectual development has centred on this interplay between lexical and viewpoint aspect, as set out in the predictions of the Aspect Hypothesis (AH) (Andersen & Shirai 1994, 1996; see Bardovi-Harlig & Comajoan-Colomé 2020 for a review). The AH endeavoured to trace a developmental pathway for

L2 aspectual development by proposing that learners began by acquiring the most prototypical pairings of lexical and viewpoint aspect – for example, using perfective forms with telic predicates such as achievements (e.g. *arrive at the station*) and accomplishments (e.g. *die*), and using imperfective forms with atelic predicates such as statives (e.g. *be tired, have blue eyes*) – and then moving on to more atypical pairings or 'extended uses' (Andersen & Shirai 1996;533) as their proficiency increased. The Aspect Hypothesis's categorical predictions made it a fruitful testing ground for studies of L2 aspectual development.

One such study is McManus (2013), whose work involving L1 English and L1 German university learners of L2 French set out to test if learners' acquisition of viewpoint aspectual forms was initially constrained by prototypicality effects, and whether increased proficiency permitted the extension of *passé composé* and Imperfect use to less prototypical situations (i.e. PC-atelic and IMP-telic pairings). In fact, the reverse was found to be true, with the most advanced learners showing the greatest prototypicality effects. McManus suggests (2013:318) that the reason that only the most advanced learners were influenced by prototypicality was that they were the only group who had successfully undergone the mapping of perfective aspect to the *passé composé* and imperfective aspect to the Imperfect, which also aligns with his finding that the L1 effects witnessed between the German and English groups were minimised as proficiency increased. The observation that the successful linking of morphological form with semantic meaning may not be completed until later phases of acquisition for aspectual contrasts is also concordant with Gabriele (2009). This bidirectional English/Japanese study investigating acquisition of the progressive and resultative meanings – which are mapped differentially across viewpoint aspectual forms in the two languages – indicated that the presence of these forms (the imperfective marker *te-iru* in Japanese and the *be* + V- *ing* construction and present progressive in English) in the learners' grammars preceded their ability to consistently attribute targetlike interpretations to them in a story compatibility task. This additionally highlights the importance of collecting both production and comprehension data - as is featured in the research design of this project – in order to more accurately assess aspectual development.

McManus's findings not only provide compelling evidence contesting the AH (alongside other concordant studies such as Domínguez et al. 2013), but, like Gabriele

(2009), also highlight the importance of the process of form-meaning mapping in L2 aspectual development and in L2A more widely. It is worth mentioning here that the "prototypical pairings" of the Aspect Hypothesis were defined according to patterns observed *in L1 input* (see also Andersen's (1988, 1993) Distributional Bias Hypothesis); thus, if learners behave more prototypically as a function of proficiency, we could infer that, as L2 development progresses, learners gradually converge on the input regarding their production of (im)perfective verbal forms. This was substantiated in Domínguez's (2019) study of L2 Spanish learners, wherein the most advanced learners showed the most similar distribution of the Imperfect across lexical classes to the L1 group, and, similarly to McManus (2013), also replicated prototypical aspectual pairings most closely. Though the interaction between lexical and viewpoint aspect that is inherent to prototypicality is not a primary focus of this project, a key takeaway from these studies is the fact that learners are sensitive to the patterns with which viewpoint aspectual form-meaning mappings appear in the input. This becomes all the more relevant to the aims of this project in light of the fact that there is some evidence for differing distributional patterns of not only pairings of lexical and viewpoint aspect, but of the viewpoint aspectual mappings themselves: in Tracy-Ventura & Cuesta-Medina's (2018) corpus study of Spanish past tense verbs in L1 data, a differential frequency cline of progressive < habitual < continuous was observed for the imperfective mappings. This also parallels the frequency cline observed in the L2 French production data presented in Wallington (2017), which is striking in light of the previously-discussed aspectual similarities between Spanish and French (e.g. Izquierdo & Collins 2008). This cross-linguistic observation provides evidence – albeit small-scale – that the varying frequencies of viewpoint aspectual mappings observed in learner production may indeed reflect naturally-occurring patterns in the input (see also McManus 2011; Domínguez et al. 2013). When we consider the above, a tentative picture begins to emerge wherein L2 learners are not only sensitive to the relative frequencies and distributional patterns of viewpoint aspectual mappings in the input, but are ultimately capable of converging on them.

The shift away from the Aspect Hypothesis and its L1-independent predictions for development led researchers to focus more closely on approaches in which the L1 – and the way it mapped meaning to form – was forefronted. A

substantial body of research on L2 aspectual acquisition has aimed to measure the L1/L2 "degree of difference" and operationalise L1 influence by investigating the transition from a learner's L1 aspectual form-meaning mappings (transferred at Initial State under Full Transfer (Schwartz & Sprouse 1994,1996)) to the form-meaning mappings of the L2. Though it has been widely shown that targetlike L2 pairings develop as proficiency increases (Salaberry 1999, 2002, 2003, 2005, 2008; Montrul & Slabakova 2002, 2003), some L2 form-meaning mappings are easier to form than others (e.g. Gabriele et al 2003, 2005; Gabriele 2009; Mai & Yuan 2016). Lardiere's (2003, 2005, 2008, 2009) Feature Reassembly Hypothesis suggests that the remapping process is more difficult when there are differences in how a given form and meaning are mapped in the L1 vs. the L2 - that is, when there is some more complex process of disentangling and reassembling features from and/or onto various forms. Despite maintaining that this remapping is possible, Lardiere (2009:175) assesses it to constitute a 'formidable learning task'; it thus follows that the variability in performance observed in learners for whom the reassembly process is still ongoing may be attributable to such mapping problems (Lardiere 2000; Slabakova 2008).

An instance where Feature Reassembly (FR)-based accounts of L2A have proven especially enlightening is the acquisition of imperfective/perfective features. A key example is the work of Domínguez et al. (2017), which builds on previous work in the acquisition of the Spanish Imperfect by L1 English learners (Domínguez, Arche et al. 2011; Domínguez, Tracy-Ventura et al. 2013). This work espouses Arche's (2014) analysis of (im)perfectivity (see section 3.2.1), which proposes that, contrary to Montrul & Slabakova's (2002, 2003) model, 'imperfectivity is not described as an unanalysed single feature but as a constellation of semantic features of a specified nature' (Domínguez et al. 2017:6). As discussed, Arche's (2014) analysis divides imperfectivity into three interpretations – the habitual, continuous and progressive – which are mapped onto a range of forms in English, including the Simple Past (SP), but which can all be expressed by the Imperfect form (IMP) in both French and Spanish (see Sections 3.2.2 and 3.2.3). This analysis of (im)perfectivity enables researchers to clearly distinguish between the three interpretations of the Imperfect, as well as the perfective form (the Preterite in Spanish, (replaceable by the French *passé composé* for the purposes of this project), and to precisely depict the learning task in terms of (re)assembly of the features of Asp(ect) and the <occasions> quantifier Q (Arche 2006, 2014; see Section 3.2.1). Figure 2 models the L2 French reassembly process for an L1

English speaker.

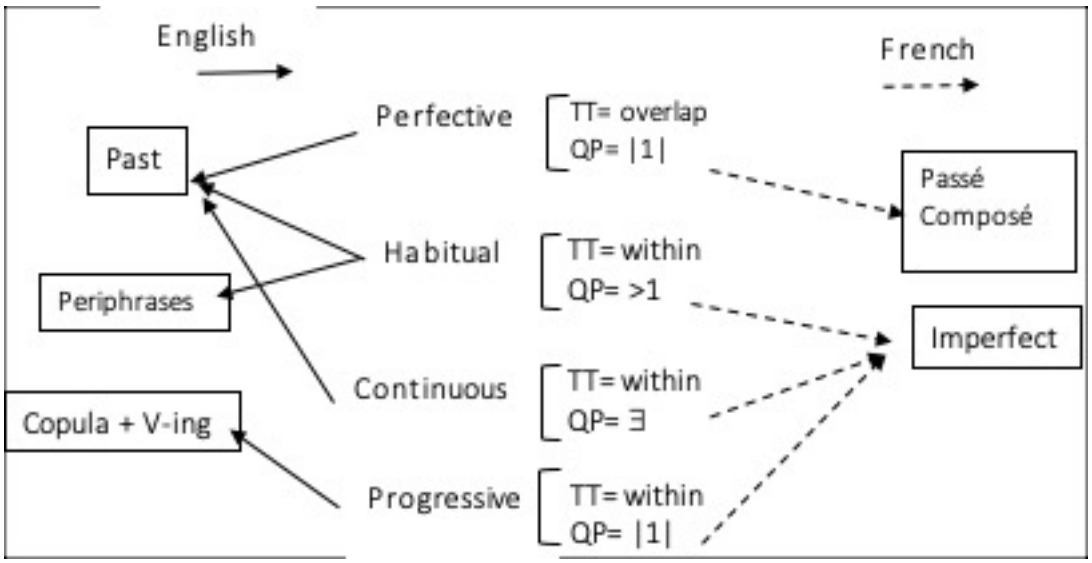


Figure 2: Mapping of features of Aspect (Asp) and the <occasions> quantifier Q for perfective and imperfective (-habitual, -continuous, and -progressive) form-meaning mappings between English and French (adapted from Domínguez et al. 2017:7)

Based on the reassembly task set out in Figure 2, we would predict the most readily-acquired imperfective interpretation to be the progressive, given the straightforward 1-to-1 mapping from *V +-ing* in the L1, which hosts only imperfective (progressive) features, to the Imperfect on the L2. More difficulty is anticipated in the cases of both the habitual and the continuous interpretations, given that in English these interpretations are either predominantly (for the habitual), or exclusively (for the continuous) expressed on the Simple Past form, to which perfectivity is also mapped. The reassembly task here is therefore more complicated: L1 English learners of a language such as French are required first to identify the continuous and habitual imperfective features expressed on the Simple Past, to disentangle them from the perfective features also mapped there, and finally to reassemble the imperfective features onto the Imperfect in the L2. It remains to be seen whether the additional presence of the periphrases *used to/would + V* in the English will facilitate reassembly for the habitual, by placing more emphasis on the distinct meaning of habituality as separate from perfectivity, or whether this will instead create an additional degree of reassembly which will render the task more difficult. Contrastingly, for the continuous, although there is technically a one-to-one mapping for this imperfective reading between the Simple Past and the Imperfect, the fact that this is the only imperfective

reading that is expressed solely by the Simple Past may render the task of disentangling continuousness from perfectivity more difficult when transferring and reassembling the relevant features.

Based on the L1 English/L2 Spanish version of Figure 2 (which as mentioned differs minimally from the French version), Domínguez et al. (2011, 2017) predicted variability - in the form of interchangeable use of imperfective and perfective forms - in imperfective contexts, and particularly in continuous and habitual contexts, where it was necessary to recruit features from a range of sources in the L1 and reassemble them onto the new L2 form. Their findings were substantiated for the continuous: though effects of proficiency were seen across all contexts in terms of increasing "correct" (Imperfect-imperfective and Preterite-perfective) acceptance, this was lowest in continuous contexts, which additionally constituted the only significant difference between the advanced group and native controls (Domínguez et al. 2011). However, not all findings were concordant with the predictions of an FR-based model, with greater accuracy shown in habitual over progressive contexts despite the greater degree of reassembly required in the former. As previously mentioned, a similarly mixed picture regarding the reassembly of imperfective form-meaning mappings was found in exploratory work utilising the same feature-based approach for the L1/L2 French pair (Wallington 2017). A notable example from this study is the result for the continuous mapping, in which learners demonstrated a level of accuracy comparable to the perfective *passé composé* at even the pre-study abroad data point. This clearly contrasts with the low continuous accuracy reported in Domínguez et al: a surprising finding given the similarities in viewpoint aspect marking between French and Spanish attested to in Izquierdo & Collins (2008). Dominguez et al. conclude most recently (2017:1) that 'the full array of interpretations associated with [the Imperfect]… is not completely acquired even at advanced levels', attributing this to 'a mapping problem of aspect-related features present in both English and Spanish onto a new form (the Imperfect).' Though their findings lie broadly within FR predictions, the authors (2017:8) advocate 'more theoretically sound research on the acquisition of the three meanings [of the Imperfect]… in a variety of contexts' to elucidate the nature of this "mapping problem" and aid in moving towards more decisive conclusions about L2 aspectual development on the whole.

When analysing the findings from Wallington (2017), it was hypothesised that the differing frequencies of each imperfective mapping in learner production may be reflective of their distribution in the French input learners had been exposed to, and that this itself may have contributed to the differential accuracy outcomes. Such a suggestion has also been attested in other areas of the L2 French literature, with Labeau (2011:70-1) proposing that 'the uses of the IMP involving concordant information from the grammatical aspect, the lexical aspect, the cotext [linguistic environment] and the context of the verbal form are more prototypical and therefore… are more frequently used and more easily acquired'. This falls in line with McManus' (2013) proposal of a close link between input, prototypicality, and proficiency. A relationship between the input and feature selection in L2A has also been inferred, in line with the notion that, following the transfer of L1 feature bundles at Initial State, relevant L2 features are added (and redundant features potentially deleted) on the basis of the presence or absence of these features in the L2 input (Lardiere 2009; Gil & Marsden 2013). With specific regard to aspectual development, Gabriele (2009:397-8) suggests that acquisition of the semantic component of aspectual forms may be especially sensitive to the input, particularly as semantic input cues may be very subtle and can crucially only be processed and incorporated into the developing grammar if the context in which they arise is sufficiently unambiguous.

In view of the differential frequencies, degrees of FR, and accuracy rates across the interpretations, it was proposed (Wallington 2017:54) that the ascending accuracy cline present in the data (habitual < progressive < continuous) may be accounted for thus: 'IMP-Habitual is not as frequent and requires reassembly; IMP-Progressive has a one-to-one form-meaning mapping but is very infrequent; and lastly IMP-Continuous, despite entailing a feature reassembly task comparable to the habitual, is considerably more frequent that both other Imperfect interpretations.' As previously mentioned, the fact that French viewpoint aspect contains multiple different form-meaning mappings that occur with a range of different frequencies makes this specific grammatical property into a rich testing ground to explore hypotheses relating both to reassembly and to input properties, and, crucially, the interaction between the two. Though the robustness of the conclusions of Wallington (2017) is limited due to dataset restrictions (particularly a lack of comprehension data and a small sample size), these findings and ideas nonetheless constitute a starting point for a more in-depth exploration of the role of input frequency on L2 aspectual development, as well as for

whether findings from a more comprehensively-designed study may be better accommodated within a generative FR framework.

3.3 Second language acquisition in the study/residence abroad context

The late-acquired nature of the French Imperfect (e.g. Bartning & Schlyter 2004; Howard 2005a; Labeau 2005; Ayoun 2013; McManus 2015) renders it a fitting candidate to analyse in a UK university-level learning context, as it is likely to be a property of the L2 that is still developing even at the relatively advanced level participants typically attain prior to their year abroad. Therefore, in addition to constituting a relatively self-contained, extended, and rich source of naturalistic input for UK languages undergraduates, whose previous exposure to the L2 has typically been dominated by instructed input – as is a central point of interest for this project - the study abroad (SA) setting is also suitable for studying later-stage L2 aspectual development. Furthermore, the applicability of research into language learning in the SA context is arguably greater than ever, given the growing popularity of these programs against global backdrop wherein university students are increasingly mobile (Allen 2010; Brooks & Waters 2010).

The learning context has been acknowledged to play an important role in L2A, given its impact on key variables such as input quantity and quality and opportunities for L2 practice (Collentine 2009; Llanes 2011). However, the majority of SLA research has, as Llanes (2011:190) notes, focused on naturalistic settings, with the lack of SA-focussed research 'particularly conspicuous' in an internationalising world where it is increasingly popular (Allen 2010). It is also worth noting that the field is to a certain degree dominated by studies of US students participating in study abroad programs (Coleman 1997; Collentine & Freed 2004; DuFon & Churchill 2006**;** Block 2007; Collentine 2009). This is significant as the American model of study abroad is arguably quite distinct, and in particular significantly more structured than other the experiences of university students from elsewhere in the world, such as Europe (Kinginger 2007). For example, American students overwhelmingly participate in university-based study programs in the country of their target language and are placed with host families, though this is beginning to diversify (Sanz & Morales-Front 2018). Contrastingly, UK students typically have the option to work – commonly as language teaching assistants in schools but also via independently-organised placements –

instead of studying (Mitchell et al. 2015; Mitchell et al. 2017) and additionally are not confined to host families for accommodation choices, with many opting to live in university halls of residence or even to rent independently (Mitchell 2015). This contrast in flexibility between US and UK study abroad programmes could be significant from an input perspective, as US students overwhelmingly continue to experience instructed or classroom-based input, whilst UK students seemingly have access to more diverse opportunities for L2 exposure and, as a result, may be exposed to more significant quantities of naturalistic input. Moreover, research on US university students entering into study abroad programs typically studies the early stages of L2 acquisition, due to the low proficiency levels of this demographic who tend to have had little to no experience in the L2 prior to college (Collentine 2009). This is in clear contrast to the average UK university modern languages undergraduate, who in general has received L2 instruction for a minimum of 4 years (often more) and consequently can be argued to have attained at least an intermediate level prior to beginning their university course. Collentine (2009:229-30) notes that the large quantity of "SLA in SA" research that focuses on lower-proficiency learners means that we know comparatively less about the impact of study abroad on more advanced learners, and that researchers should focus more on targeting this population, particularly in light of findings that a certain "threshold" of L2 knowledge may need to be obtained prior to SA in order to optimise opportunities for linguistic development (e.g. Golonka 2006; O'Brien et al 2006**;** DeKeyser 2010). This also ties in to one of the research aims of this project, which is specifically to investigate the continuing L2 development of viewpoint aspect at more advanced levels.

A noteworthy exception to the relative lack of research on UK university students in the study abroad context is the work of Mitchell and colleagues (Mitchell et al. 2017), whose research on L1 English university learners of L2 French and Spanish before, during and after their residence abroad (which constitutes the LANGSNAP corpus data as used in Wallington 2017) spans a wide range of topics ranging from acquisition of the French subjunctive (McManus & Mitchell 2015) to the impact of placement type on SA language development (Mitchell et al. 2015). Ultimately, however, Mitchell and colleagues elect to focus on L2 language development in a wider sense, frequently foregrounding socio-contextual variables such as learner identity and social networks during the year abroad (McManus et al 2014; Mitchell 2015).

To summarise, it can be surmised that research on language acquisition during study abroad conducted from a US-centric perspective can be said to reflect a more homogeneous experience than has been attested to in the European literature (e.g. Kinginger 2007; Klapper & Rees 2012; Mitchell et al. 2017), and consequently is not fully representative of the full spectrum of study abroad experiences – linguistic and social - of the wider university language learning population, particularly those at higher proficiencies. This is significant from the point of view of the input, especially given the prevalence of university study programs in the US SA model, as this means that the type of L2 input US study abroad learners receive may not differ from the at-home context as substantially as for European learners, who arguably have more regular access to a greater diversity of naturalistic input sources during their time abroad – particularly if they work as a language assistant or undertake a work placement. Additionally, among those studies which do focus on advanced European or UK university learners during SA – the latter being the focus of this study – social variables tend to come to the fore ahead of linguistic development (although see below for discussion of some exceptions).

A further motivation for investigating L2 aspectual development specifically is that a large majority of studies exploring L2A in a SA context focus on overall language proficiency (e.g. Carroll 1967; Díaz-Campos 2004; Freed 1995, 1998; Llanes & Muñoz 2009) and particularly oral proficiency (e.g. Lennon 1990; Segalowitz & Freed 2004; Juan-Garau & Pérez-Vidal 2007) (see Borràs & Llanes 2019 for a review). The overall positive influence of time spent abroad on these global measures is widely attested by such studies - and has perhaps contributed to the general perception that students return from the SA programmes with “near-nativelike” language skills (e.g. Freed 1995, Kinginger 2008) - yet, as stressed by Llanes (2011:210, emphasis AW), there is a comparative lack of research that ‘examines the effects of the SA context on L2 development *in specific areas*.’ Moreover, those that have focused on specific L2 development have not always been unanimous. Collentine (2004) and DeKeyser (1991) both focus specifically on the impact of SA on L2 grammar (as a subset of overall proficiency) in English learners of L2 Spanish, but arrive at somewhat contradictory conclusions: Collentine (2004) posits that the home environment is better suited to developing a greater number of grammatical points, whereas DeKeyser (1991)’s results

point to similar grammatical gains between the at-home vs. SA context over a six-month period. Further studies of L2 Spanish development in SA contexts have attested to a positive influence of the learning context on grammatical proficiency, including the areas of tense acquisition, verb agreement, and copula use, in both shorter single-semester stays (Ryan & Lafford 1992) and year-long residence (Guntermann 1995).

The picture arising from the SLA in SA research centering on L2 Spanish grammar is thus a fairly ambiguous one, despite the fact that a meta-analysis by J-S. Yang (2016:76) revealed that studies of English speakers learning Spanish are one of the field's major foci. Research focusing on L2 French grammar, particularly among UK university learners, appears less abundant, and again shows variation across different grammatical properties. For example, McManus & Mitchell's (2015) work on subjunctive development showed that this generally remained a source of variability even post-YA, whilst Edmonds & Gudmestad's (2018) investigation into grammatical gender found an increase in levels of targetlike gender marking over the course of the stay abroad which were maintained at a post-test. A notable body of work on L2 French grammatical development in UK university learners is that of Howard (2001, 2005b, 2006), whose investigation of Irish undergraduates acquiring L2 French found a superiority of SA over at-home contexts, specifically for past time marking and gender agreement. This work may provide an interesting point of comparison with the current project, given that the development of viewpoint aspect forms an essential component of past-time marking.

In summary, the decision to focus in this project on the impact of programmes such as the year abroad (YA) on L2 French viewpoint aspectual development aims to contribute to several areas of the "SLA in SA" field that appear relatively under-researched: namely, specific grammatical development; the L1 English/L2 French language pair; and UK university learners. The year abroad for UK university learners represents a unique learning context in that it 'allows for intensive, regular, contextualised L2-use opportunities in situ' (Rehner & Mougeon 2003 in J-S. Yang 2016:67), and therefore constitutes an important contrast with their L2 exposure prior to this point, which has been primarily concentrated in instructed/foreign-language settings. It is acknowledged there are serious challenges implicit in any concrete, precise description of how input differs in SA vs. FL contexts: however, this

project does not directly endeavour to articulate such a description. In fact, it would be erroneous to suggest that a one-size-fits-all label such as “year abroad input” could reasonably exist, given the diversity of experience and L2 exposure that is characteristic of the SA setting (Coleman 2009; Klapper & Rees 2012). This said, the experience of the UK university-level languages student nevertheless constitutes a valuable opportunity to compare linguistic performance both prior to and after exposure to an important and intensive source of naturalistic input in the target language, thus providing us with the possibility of exploring the importance of (naturalistic) input on later-stage L2 development. This is particularly relevant for areas – such as the viewpoint aspectual system – which have been documented as challenging and late-acquired for L1 English speakers, in order to explore whether a change in “input type” may have a facilitative effect on the ongoing feature reassembly process. Our findings consequently have the potential to inform us on whether the role of the input may be usefully incorporated into a perspective of L2A that is rooted in predictions from Feature Reassembly.

The following chapter sets out the research design of this project, which endeavours not only to comprehensively gather information on learners’ knowledge of viewpoint aspectual form-meaning mappings, but also to conduct a representative analysis of the kind of naturalistic French data participants may be reasonably expected to encounter during their year abroad.

4. Research Design

The research design of this project consists of two components – collection of experimental data from L1 English university-level learners of L2 French and analysis of data from a French L1 corpus - which reflect its dual theoretical approach. The following chapter details the design of each component: Section 4.1 sets out the details of the experimental design, and Section 4.2 describes the nature and rationale of the corpus analysis.

4.1 Experimental design

Before describing the participants, it should be mentioned that this project originally conceived of a longitudinal research design, wherein the same group of students would be tested at a pre- and post-year abroad data point. Unfortunately, this was unable to be achieved due to a number of logistical restrictions on sampling. As a result, this project takes a cross-sectional design, with two separate cohorts constituting the pre- and post-YA groups. It is acknowledged that the comparison of two different groups comes with some important considerations, particularly regarding individual differences, which have been indicated to play an important role in L2 development in SA contexts (e.g. Brecht et al 1995; Tokowicz et al 2004; O'Brien et al 2007). However, the non-significant between-groups difference obtained from the independent proficiency measure (see Section 4.1.1 below) suggests that proficiency differences should not have a significant bearing on results. Random intra-participant variation within the groups will also be controlled for during data analysis. It is hoped that, despite the obvious desirability of longitudinally analysing one cohort, the design employed in this project will still provide insight into L2 French aspectual development of UK university-level learners over the course of their degree programmes.

4.1.1 Participants

The L2 French participants of this project are undergraduate students studying for a modern languages degree at a UK (southern England) university. Two groups of participants were recruited: one group ("Post-YA", *n* = 23) who had just returned from the year abroad and were beginning the final year of their degree; and a second group ("Pre-YA", *n* = 20) who were nearing the end of their second year at university and will

be undertaking their year abroad in the following academic year. The "year abroad" - which in reality lasted around 9 months - took place during the third year of students' undergraduate programmes and was a mandatory component of their degree. Students spent their time in either France or (a small minority of the Post-YA group, *n* = 2) Switzerland, where they undertook either a language teaching assistantship, a work placement, or took courses at a university. All students were recruited from the same southern English university via visitation from the researcher to their French language classes; a brief description of their participation was provided, and students were left with a voluntary sign-up sheet, meaning that they self-selected for participation in the study.

In terms of proficiency, participants were generally required to obtain a minimum "A" grade (the second-highest attainable) or equivalent in French at the end of their schooling to obtain their university place, and so could be approximately characterised as high-intermediate level on arriving at university. Given that both groups were studied some time after beginning university, the cumulative instruction they had received by this point arguably situated them in line with Bartning's (1997:13) classification of *apprenants d'instruction élevée* ('advanced instructed learners', translation AW). To empirically ascertain this, participants completed a written cloze test (Tremblay 2011; Tremblay & Garrison 2010) as an independent measure of proficiency. The mean proficiency score for the pre-year abroad group was 64.7% (SD = 10.5) and for the post-year abroad group 67.2% (SD = 12.7). Though the post-YA group scored descriptively more highly than the pre-YA group, an independent samples t-test found no statistically significant between-groups proficiency difference (t = -7.16, p = .47, Cohen's d = .22).

As anticipated for a modern languages university cohort, a large proportion of participants studied a range of L2s alongside French both prior to and at university. However, all participants were studying French as part of their degree and so received the same quantity of French language instruction as per course requirements (i.e. the Pre-YA group all had the same number of French language instructed hours per week, as did the Post-YA group on their return to university). Information on other languages spoken by participants was gathered via a language background questionnaire (see section 4.1.2), which revealed that the L1 for all was English: though a small number reported additional L1s (see Table 2), all self-identified English as their dominant

language.

Control group data was also collected from *n* = 7 L1 speakers of French, all of whom were exchange students at the same university as the L2 French participants. These students were recruited via advertisement on a number of university-based social media groups, and thus were also self-selecting.

At this point, it is essential to note that it was originally intended for the L1 comprehension group to be composed of the same participants as in the L1 production group: that is, L1 French exchange students attending the same university as the L2 French students. However, at the point of analysis of the comprehension data, it was realised that the L1 group had received the instructions to the comprehension task, as well as the contextual information for each item, in English. This had the potential to yield anomalous results, given that it could not be guaranteed that the L1 French group had correctly understood the English parts of the task (though the sample were highly-educated and attending an English university at the time of data collection, no independent English proficiency measure was administered). In view of this, the comprehension data was recollected for the L1 group (after translating the AJT task fully into French and having the translation checked by another L1 French speaker). Due to time and logistical restrictions (data recollection occurred in November-December 2020 when access to university campuses in the UK was restricted), the second L1 comprehension sample was more of a sample of convenience, recruited via advertisement on social media and among the researcher's existing connections. This naturally meant that many demographic variables could not be as easily controlled for, in contrast with the L1 production group, who were much more homogenous both within the sample and compared to the L2 groups). The implications of this will be returned to in Chapter 5.

Participant information for all four groups is summarised in Table 2.

Group	**Age**	**L1**	**L2(s)**	**Proficiency Score**
Pre-YA (*n* = 20)	19-21 (Mean age = 19.9)	English Additional L1s: Portuguese (*n* = 1) Italian (*n* = 1) Armenian (*n* = 1)	French Additional L2s: *University-level:* Spanish (*n* = 12); Portuguese (*n* = 2); German (*n* = 1); Russian (*n* = 1), Arabic (*n* = 1) *Pre-university:* Spanish (*n* = 12); Italian (*n* = 1); Portuguese (*n* = 1); Latin (*n* = 1)	64.7% (SD = 10.5, 95% CIs: 59.8-69.6%)
Post-YA (*n* = 23)	21-24 (Mean age = 21.8)	English Additional L1s*:* Italian (*n* = 1) Cantonese (*n* = 1)	French Additional L2s: *University-level:* Spanish (*n* = 8); German (*n* = 4); Mandarin (*n* = 2); Portuguese (*n* = 1) *Pre-university:* Spanish (*n* = 12); German (*n* = 3)	67.2% (SD = 12.7, 95% CIs: 61.7-72.7%)
L1 production group (*n* = 7)	20-27 (Mean age = 21.7)	French	-	89.9% (SD = 8.02, 95% CIs: 82.4-97.3%)
L1 comprehension group (*n* = 14)	24-59 (Mean age = 34.1)	French	-	Not collected

Table 2: Participant demographic information.

4.1.2 Task design and data collection

The following section details each of the experimental tasks in the order that participants completed them. Note that both L2 French groups completed all tasks, whereas the L1 French controls completed only the oral tasks, the proficiency test, and the comprehension task.

Production Task 1: Cat Story

Oral production data was gathered across two tasks with differing degrees of structuredness in an endeavour to gain a more reliable perspective on learners' ability to accurately produce French viewpoint aspectual form-meaning mappings.

The first production task completed was a structured impersonal narrative retell task entitled Cat Story: a French adaptation of a task originally conceptualised for the SPLLOC corpus (splloc.soton.ac.uk), data from which has been used in L2 Spanish studies of viewpoint aspectual development (e.g. Domínguez et al 2011, 2017). The French version of the task was also used with learners in the LANGSNAP project (langsnap.soton.ac.uk), from which the data analysed in Wallington (2017) were sourced. In the Cat Story task, participants were presented with a series of pictures (and a few written prompts in French), and are asked to tell the story of the pictures in French to the researcher (see Appendix A.1 for sample pictures from the task). The picture series began with the prompt *Tous les matins étaient pareils* ("Every morning was the same"), followed by a series of daily activities of a little girl and her cat: this targeted the imperfective-habitual mapping. Following this was the prompt *Mais il est arrivé un jour* ("But one day…"), which preceded the main storyline: a series of events wherein the little girl and her cat cannot find each other for some time, but are eventually reunited. This section was designed to elicit a range of past tense forms encompassing perfective and imperfective aspectual meanings.

Participants were given approximately five minutes prior to the task to familiarize themselves with the pictures, and indicated to the researcher when they were ready to begin. They were requested to tell the story in the past and to incorporate the written prompts into their narrative: this was in an attempt to reduce the number of participants who may otherwise have told the story in the present, despite both written prompts containing past tense forms. Participants were also permitted to ask the researcher for any vocabulary items needed, either prior to or during the task, given that the focus was not on vocabulary size. The retell lasted on average approximately 3-5 minutes per participant, and was audio recorded.

Production Task 2: Conversation

In addition to the relatively structured Cat Story task, participants also completed a less structured task in the form of an informal interview "conversation" led by the researcher. The interview lasted on average 15 minutes and covered a range of familiar topics including university life, personal interests, interviewees' motivations for studying languages (or for studying in the UK for the control group), and their experiences of either preparing for the year abroad (pre-YA group), the year abroad itself (post-YA group), or studying in the UK (control group). Though participants could respond as they chose, the questions asked endeavoured primarily to encourage elicitation of past tense forms spanning perfective and imperfective meanings. A principal motivation for adding the less structured, more participant-led interview to the battery of oral tasks was to encourage participants to speak freely and at some length over a range of more personal topics, which was not achievable through Cat Story alone. This range of degrees of structure in the oral production tasks aimed to provide a more comprehensive picture of participants' production of viewpoint aspectual mappings across a range of communicative contexts and consequently reduce task-related bias (Labeau 2005, 2011). The interview was also audio recorded.

Proficiency test

As previously mentioned, all participants completed a written proficiency test in order to assess their overall French proficiency. The test was a cloze or gap-filling written task developed by Tremblay (2011; Tremblay & Garrison 2010); the deleted words were balanced across content (lexical) and functional (grammatical) words and thus tested proficiency across a range of grammatical properties as well as vocabulary size. The text from which the test was derived was a 314-word article on global warming written for French newspaper *Le Monde* and aimed at a general readership; on this basis, Tremblay (2011) judged it to be suitable for university-age and -level students. Participants completed the proficiency test on paper and in controlled conditions (i.e. in the presence of the researcher, without access to resources.) No time limit was imposed, but in general participants took between 15-40 minutes to complete the test. The test was scored out of 45, and the raw scores converted to percentages to give an overall indicator of proficiency. The control group also completed the proficiency test to

act as a baseline. The mean results of the proficiency test per group are reported in Table 2.

Language Engagement Questionnaire (LEQ)

Participants also completed the Language Engagement Questionnaire (LEQ), as devised by LANGSNAP researchers (langsnap.soton.ac.uk). The LEQ is designed to assess how much time (measured on a 6-point scale from "Every day" to "Never") participants spend using each of their languages across different activities, ranging from academic contexts (e.g. seminars, language classes) to social activities (e.g. service encounters, watching television, texting). A full list of the LEQ contexts rated by participants can be found in Appendix A.2. The LEQ aims to gain an overall qualitative picture of the time students spend exposed to English and French before and during the year abroad (NB: the LEQ for the post-YA group emphasised that participants should answer based on their language engagement *during* the YA), and whether/how this changed. The purpose of gathering this data was to help explore, both qualitatively and quantitatively, the amount of exposure to the L2 learners receive both in their at-home instructed context (Pre-YA responses) and in the naturalistic YA context (Post-YA responses), and in what contexts (social/work/leisure etc.), although an in-depth exploration of the latter unfortunately falls outside of the scope of this project. The data from the LEQ will be considered alongside participant production and comprehension data to aid in assessing whether: a) participants were indeed exposed to more L2 input during the YA; and b) if so, whether this increased exposure could be linked to any developments to the L2 aspectual system. In view of the centrality of the L1 in a Feature Reassembly-focussed approach to L2A, learners' exposure to English in both settings will also be considered and compared.

Following completion of the above tasks in the presence of the researcher, participants were later sent a link to complete the final two tasks of the study, which were presented together via a survey software package (iSurvey, https://www.isurvey.soton.ac.uk/). Participants completed the tasks online at a time of their convenience. The below section details each of these online activities (Language background questionnaire and Acceptability judgement task) in the order participants

completed them.

Language background questionnaire

The online component began with a questionnaire which aimed to collect background information about learners' language learning, as well as a rough quantification of the amount of L2 exposure they had received at various point of their language learning career. This may help paint a broad picture of learners' exposure to the L2 input prior to their year abroad. After some basic demographic information, the questionnaire elicited information about which L2(s) participants had studied at several key points of UK schooling (GCSE level, A-level, and university-level), and the approximate number of hours per week participants spent using/studying these languages in and out of the classroom. In order to gather contextual information, participants from the post-YA group were asked about the location and duration of their YA, the placement type they undertook (teaching assistant, Erasmus university exchange, or workplace internship), and the type of accommodation they stayed in. A summary of these responses (excepting YA location) is presented in Table 3.

	Pre-YA (*n* = 20)	**Post-YA (*n* = 23)**
French exposure per week at GCSE level/lower secondary school (Mean, [SD])	1-6 hours (3.32, [1.26])	1-10 hours (3.71, [2.32])
French exposure per week at A-level/upper secondary school (Mean, [SD])	3-10 hours (6.09, [2.21])	3-20 hours (6.84, [4.05])
French exposure per week at university (Mean, [SD])	3-12 hours (6.94, [2.89])	3-17.5 hours (6.86, [4.20])
Duration of YA (Mean, [SD])	N/A	7-10 months (9.1, [0.70])
YA placement type	N/A	Teaching assistant *n* = 3 University exchange *n* = 20 Work internship *n* = 0
Time spent at placement per week (Mean, [SD])	N/A	3.5-32 hours (10.45, [5.62])
YA accommodation type	N/A	University accommodation *n* = 10 Shared student accommodation *n* = 7 Shared non-student accommodation *n* = 2 Solo accommodation *n* = 3 Host family *n* = 1

Table 3: Summary of language background information for Pre- and Post-YA learner groups.

It can be seen that, on average, both learner groups received similar amounts of exposure to French at each stage of their foreign-language education, and that the university setting did not provide significantly more weekly exposure than the A-level/upper secondary stage. Considering the YA characteristics of the Post-YA group, we observe a very uniform length of residence, but more variety in terms of the number of hours engaged in a placement per week. This is perhaps linked to the variety of placements undertaken: although the majority of students participated in university exchange programmes, a small number undertook teaching assistantships. Although it may appear from this that most of the YA students were simply having an “instructed

experience abroad", more akin to the American model of study abroad (see Section 3.3), we also observe that on average, YA students only participated in their placements for 10.45 hours per week. We can infer from this that the majority of students' language exposure therefore occurred outside of their placement hours, i.e. in a more naturalistic setting, as anticipated.

Acceptability judgement task

Collecting both comprehension and production data in L2A research is essential to gaining insight into whether there is a disparity between learners' understanding of how a given form is used and their ability to produce it. This is supported with specific respect to the current project by studies of L2 aspectual development wherein learners have performed very differently between production and comprehension tasks (e.g. Domínguez et al. 2011, 2017). The comprehension task used in this project is an acceptability judgement task (AJT). Acceptability judgement tasks have been used extensively in studies where there is a semantic component to be acquired as well as a morphosyntactic form (e.g. Domínguez et al. 2011, 2017; Mai & Yuan 2016; Guo 2020) as they enable researchers to precisely assess whether a particular semantic interpretation of a given form has been acquired, based on the learner's evaluation of the acceptability of a sentence containing that form in a range of different contexts. Moreover, gathering comprehension data alongside production data is widely regarded as desirable in L2A research, as it enables a more comprehensive picture of learners' L2 knowledge to be gathered. This is linked to the additional cognitive demands associated with "on-line" tasks like producing spontaneous speech, which are heightened in an L2 even for very advanced speakers (e.g. Hahne 2001, Hopp 2006). This may mean that learner's production alone is underrepresentative of their actual L2 knowledge. Contrastingly, untimed comprehension tasks like the AJT used in this project allow learners the opportunity to access their "off-line" knowledge, and the data from this kind of task can be triangulated with production data to give a more complete perspective on the state of learners' L2 representations.

In this project, each item on the AJT followed the same format. Firstly, the learner read a "context" of a few sentences long, which provided background information in English (e.g. "My sister was invited to a concert but she got there late. When she finally arrived, the pianist had already started playing."). Following this was a sentence in French (e.g. *Le pianiste a joué du piano quand ma soeur est arrivée* "The pianist played

the piano when my sister arrived.") which participants were required to rate on a scale of 1 (definitely not appropriate for the context) to 5 (definitely appropriate for the context). In this instance, the presence of the verb *jouer* in the perfective *passé composé* form (*a joué*, "played") is at odds with the progressive nature of the context (the Imperfect (*jouait*, "was playing") would have been more appropriate): therefore, the targetlike response would have been to give the sentence a low rating (< 3) for acceptability in the context. Prior to beginning the AJT, learners were given descriptors of each number on the rating scale, as well as two example context/sentence pairs – which contrasted appropriate and inappropriate present and future tense use, in order to minimise priming effects – to show them how the rating system worked in practice. L1 French data was also collected, using a version of the AJT that was identical except for the fact that the instructions and contexts, as well as the sentences to rate, were presented in French.

The contexts and sentences used in the AJT were adapted from previous work by Dominguez et al. (2017) on viewpoint aspectual development in L2 Spanish, and therefore were already well-suited to the present project; however, a few minor alterations were made to ensure all items were suitable for current university learners of L2 French. These alterations included "localisation" of typically Spanish names (such as Juan and Ana) for more typically French or English names. In addition, an item was removed that referred to the death of a celebrity that was both recent and highly-publicised at the time of Domínguez et al.'s 2017 study, but which may have been less familiar to participants of this project a few years later on. This item was one for which the target form was the perfective; as the perfective was already the most frequent aspectual condition in the AJT relative to the other aspectual conditions, the item was not replaced. Finally, Domínguez et al. also collected judgement data for the progressive condition with achievement verbs, but did not report on this due to 'the peculiar semantic properties of these events in this context' (Domínguez et al. 2017:454). In line with this, the AJT in the present project did not include items with this combination of viewpoint and lexical aspect. The remaining items were conserved, and the relevant parts were translated from Spanish to French by the researcher and subsequently checked by a native speaker of French to ensure they elicited the desired interpretation.

After adaptation, the AJT contained 22 contexts, each of which were presented

twice: once in combination with a French sentence containing a targetlike viewpoint aspectual form (*passé composé* for perfective, Imperfect for imperfective), and once with an identical sentence but with the opposite (non-target) verb form. This meant that in total the AJT contained 44 items, which were presented in a randomised order. However, it should be noted that due to a software error, 1 context (2 items) in the progressive-eventive condition was presented with the same verb form twice; consequently, these items were removed from analysis, meaning that the total number of analysed items per participant was 42.

The contexts were balanced as far as was possible with regard to the possible combinations of viewpoint and lexical aspect (see Appendix A.3 for a full list of AJT items – note that in the Appendix the contexts for the items are presented in both English and French, but participants saw the contexts either in English (for the L2 groups) or French (for the L1 group), not both languages).

Table 4 (below) displays a summary of the AJT design, including the number of items for each viewpoint aspectual condition (habitual, continuous, progressive, and perfective), subdivided according to lexical aspect (stative vs. eventive). Table 5 presents examples of each condition. It should be noted that the AJT contained no fillers or distractor items. This decision was made on consideration that, as the AJT already contained a fairly large number of items, adding further items may have resulted in a loss of concentration from participants due to the length of the task, resulting in responses that were not truly representative of their knowledge. However, it is possible that by not including distractor items, participants – perhaps especially in the L2 groups – may have realised that the focus of the task was the imperfective/perfective contrast. This is explored further in Chapter 5.

		Number of items		
Viewpoint aspectual condition *(Target_Aspect)*		*Lexical_Aspect* = Stative (Sta)	*Lexical_Aspect* = Eventive (Ev)	Total
Imperfective	Habitual (Hab)	3	3	6
	Continuous (Cont)	4	0	4
	Progressive (Prog)	0	2*	2
Perfective (Perf)		3	5	8
Total conditions		10	12	21
Total items		20	22	42

Table 4: Summary of AJT design.

*NB: *Condition where 1 context (2 items) was removed from analysis due to software error.*

Viewpoint aspect condition	**Lexical aspect condition**	**Example**
Habitual	Stative	***Context:*** Martine has moved to different flat in a much quieter part of town. Before, she was too close to a train station and couldn't sleep well at all. ***Sentence to rate:*** *Target*: Martine entendait les trains le matin. (Martine heard-IMPERF trains in the morning). *Non-target:* Martine a entendu les trains le matin. (Martine heard-PERF trains in the morning).
	Eventive	***Context*:** I was always a bit lazy when I was in secondary school, and it was always difficult for me to wake up early on school days. ***Sentence to rate:*** *Target:* J'arrivais en classe en retard. (I arrived-IMPERF late to class). *Non-target:* Je suis arrivé(e) en classe en retard. (I arrived-PERF late to class).

Continuous	Stative	***Context*:** My husband and I have moved to the south of France looking for some sun. Although we liked Scotland, we were a bit tired of the cold weather. ***Sentence to rate:*** *Target*: En Écosse, il faisait très froid. (In Scotland, it was-IMPERF very cold). *Non-target:* En Écosse, il a fait très froid. (In Scotland, it was-PERF very cold).
	Eventive	N/A
Progressive	Stative	N/A
	Eventive	***Context***: We went to the teachers' room to look for Mademoiselle Dupont, the new French language assistant, but she wasn't there. Instead, Ms Robinson the English teacher was there, working on our final exam. ***Sentence to rate:*** *Target*: La professeure d'anglais préparait l'examen final. (The English teacher prepared-IMPERF the final exam). *Non-target:* La professeure d'anglais a prepare l'examen final. (The English teacher prepared-PERF the final exam(.
Perfective	Stative	***Context***: Rachel's grandma is normally very healthy. However, last winter she caught a cold that became very complicated and she ended up in hospital for a month. ***Sentence to rate:*** *Target*: Sa grand-mère a été très malade. (Her grandmother was-PERF very ill). *Non-target:* Sa grand-mère était très malade. (Her grandmother was-IMPERF very ill).
	Eventive	***Context:*** It was so warm and nice that Jean decided to go out for a walk during his break and have lunch outdoors. ***Sentence to rate:*** *Target*: Jean a mangé au parc. (Jean ate-PERF in the park). *Non-Target:* Jean mangeait au parc. (Jean ate-IMPERF in the park).

Table 5: Examples of viewpoint and lexical aspectual conditions in the AJT.

Procedure

Data was collected in two sessions: firstly, participants met with the researcher in person to complete both production tasks (Cat Story and interview), the proficiency test, and the Language Engagement Questionnaire, in that order. This typically lasted around an hour in total. Following this, participants were sent a link to complete the language background questionnaire and comprehension task online; this could be completed in participants' own time, so duration varied, but on average took around 20-25 minutes. Following completion of the online component, participants were reimbursed for their time. Data from the Pre-YA group was collected between 1-2 months prior to the end of students' second year of university (after which they departed for their year abroad, at the start of the following academic year), whilst data for the Post-YA group was collected between 1-2 months after the start of students' final year of their undergraduate studies. Due to the fact that not all stays abroad were of equal length (see Table 3), and also to small differences in precisely when each student started and finished their year abroad, the amount of time elapsed between students' return from the YA and their data being collected was also not totally uniform across the Post-YA sample. Taking the mean YA duration (9 months) as well as the most common month of departure (September), it can be estimated that the majority of Post-YA students had returned from their year abroad approximately 4-5 months prior to their data being collected. It is acknowledged that this is a fairly delayed testing, but was largely inevitable given that students were recruited in person from their university campus, and it was therefore necessary to wait for the start of the new academic year to begin testing.

As previously mentioned, the data from the additional L1 French comprehension group was collected at a later point (approximately 1.75 years after the L1 production data was collected). Due to the unforeseen nature of collecting data for this group, recruitment was carried out via convenience sampling using both social media and pre-existing connections of the researcher.

Coding

Data from the production tasks was transcribed and coded manually by the researcher, following the protocols of the SPLLOC project for analysing data from similar tasks that also aimed to elicit information on learners' L2 viewpoint aspectual systems

(see Domínguez et al. 2013). Coding involved first identifying each instance where one of the viewpoint aspectual form-meaning mappings – imperfective-habitual, imperfective-continuous, imperfective-progressive, and perfective – should have been used (based on the surrounding context), and secondly assessing which form was actually used in each instance. Forms used were categorised into PC (passé composé), IMP (Imperfect), PRES (Present), and OTHER (all other forms). The same coding process was applied to the corpus data (presented in Section 4.2 below).

A random sample of the coded data was checked by a second rater and native speaker of French to ensure inter-coder reliability. The reliability index was close to ceiling (≈ 90.0%). In general, identifying the expected mapping in a given context was relatively straightforward, as, even in the less-structured Conversation task and corpus data, the surrounding discourse facilitated identification. Additionally, despite the aural similarity of many French grammatical morphemes – such as the endings of *manger* (eat-INFINITIVE), *mangé* (ate-PAST-PARTICIPLE) and *mangeait* (ate-IMPERFECT) - identifying the forms used in the recorded data also did not pose significant problems. In the small number of ambiguous cases that were identified, each was discussed by the raters, and excluded from further analysis if a consensus could not be reached.

4.2 Corpus analysis

4.2.1 About the corpus data

The corpus data was sourced from a freely-available online corpus of adult L1 French, the CFPP2000 (***http://cfpp2000.univ-paris3.fr/index.html***). The corpus is composed of oral interview data gathered in the 2000s in Paris and its suburbs, and collected via open questions about Parisian life. Though data from Paris alone may not be fully representative of the input encountered elsewhere in France or in Switzerland, the CFPP2000 sample can nonetheless be said to broadly represent the kind of naturalistic data learners encountered during their year abroad.

The corpus contains 41 interviews: of these, a subset of 10 was randomly selected for use in this study. The first ≈ 1000 words of each interview was analysed, creating a final corpus sample of n = 10,641 words.

4.2.2 Corpus analysis in second language acquisition research

Legate & Yang (2002) note that a successful defense of the poverty of the stimulus (PoS) argument – and consequently, one could argue, of generative linguistics - calls on generativists to incorporate both corpus linguistics and mathematical learning theory into their approach. The authors go on to utilise data from the CHILDES corpus in their rebuttal of Pullum & Scholz (2002), illustrating the negligible frequency of disconfirming evidence in the input encountered by English children during L1A. Though this project will use corpus data for a different purpose, it aims to do so in a similarly principled way, by analysing the relative frequencies of the four viewpoint aspectual form-meaning mappings of interest in the corpus data sample outlined previously. This frequency-distributional information will be compared against learners' viewpo0int aspectual performance in the production and comprehension tasks, with a view to assessing whether there is any relationship between the frequency at which a particular form-meaning mapping occurs in the input and its ease of acquisition.

Legate & Yang (2002) undertake their convincing counter-argument to Pullum & Scholz (2002) by employing a comparative approach to input frequencies. The authors highlight the pertinent point that, even if it is impossible to precisely state that exposure to *x* examples will result in the acquisition of *y* feature, it is nonetheless possible to model for this indirectly. For example, in order to assess whether the input contained sufficient disconfirming evidence to enable L1 English-acquiring children to rule out the incorrect "first auxiliary hypothesis" in their acquisition of yes/no interrogatives - thus rendering redundant the PoS - Legate & Yang (2002) compared this type of disconfirming evidence with the disconfirming evidence found for the subject drop phenomenon, given that realisation of overt subjects and targetlike SAI structure appear at around the same time in L1 English development. As the point of acquisition for both grammatical properties could be considered as fixed/equivalent, the frequency of subject drop disconfirming evidence could thus be utilised as 'an independent yardstick to quantitatively relate the amount of relevant linguistic experience to the outcome of language acquisition' (Legate & Yang 2002:156).

The corpus analysis undertaken in this project is arguably based on an inverse version of Legate & Yang's (2002) indirect modelling of the relationship between frequency and acquisition, albeit based on the same logic. We do not presuppose that the four viewpoint aspectual mappings are acquired at the same point - indeed, a

central motivation of this project is linked on the fact that they are not - but instead aim to explore if the differing frequencies with which they occur in the French input appears to be in any way correlated with their ease of acquisition in for L2 learners of French. An important caveat should be made here that, as has been highlighted elsewhere in the L2 aspectual acquisition literature (e.g. Gabriele 2009), the presence of a form in the input – even if it is frequent – does not guarantee that acquisition will take place: in order for form-meaning mapping to occur, there must be sufficient extralinguistic context available in the input for a learner to link a form to its meaning. We acknowledge that this may not always be the case, especially given the added difficulty that many distinct morphological forms in French, which encode different meanings, (e.g. *allait* "went-IMP", *allé* "went-PP", *aller* "went-INF") have near-identical pronunciations. These factors should certainly be borne in mind; however, as Gabriele (2009:397) suggests, 'a careful consideration of the input will allow for better understanding of the circumstances in which learners are able to successfully converge on the target… and the circumstances in which convergence is not possible.' Even if our investigations into the distribution of French viewpoint aspectual form-meaning mappings in the input can only be exploratory in nature, they may permit us to evaluate whether the frequency of a given mapping in the input is a "circumstance" that is facilitative of acquisition.

In short, the combination of the L1 French corpus analysis described here alongside a feature-based approach to (im)perfectivity hopes to explore the potential cumulative explanatory power of both Feature Reassembly (an indicator of qualitative differences in learnability between form-meaning mappings) and quantitively-focussed statistical-distributional information from the input on L2 viewpoint aspect development, the data for which is sourced from the experimental design outlined in section 4.1. The results of these analyses are presented in the following chapter.

5. Presentation of results

As a reminder, the research aims of this thesis are three-fold: firstly, to analyse the extent to which Feature Reassembly can predict the L2 French viewpoint aspectual development of L1 English speakers (RQ1); secondly, to explore how studying naturalistic input can inform us on the impact of programmes such as the year abroad on grammatical development (RQ2); and lastly, to assess the extent to which an approach which combines information from these two sources can explain the process of L2 aspectual development at advanced levels (RQ3). In this section, we present the results of analyses that combine to address these three questions.

Firstly, the performance of the L2 groups and the L1 controls in oral production will be analysed (Section 5.1), beginning with an overview of the data (Section 5.1.1), a closer look at the variables underpinning production accuracy (5.1.2) and subsequently a more fine-grained analysis of the proportions of different verbal forms used by learners in each task and aspectual condition (Section 5.1.3). These analyses will pave the way towards answering RQ1 (from a production perspective): the production results will be considered alongside predictions from Feature Reassembly, evaluating how learners converge on or diverge from these predictions. Information on learners' exposure to French and English before and during their year abroad (gathered via the Language Engagement Questionnaire (LEQ)) is also presented, and the potential relationship this has to their production accuracy is discussed.

Section 5.2 features a frequency-distributional analysis of the L2 and L1 production data presented in Section 5.1, in comparison with a 10,641-word sample of L1 French conversational corpus data. This analysis involves first presenting the frequency and distribution of viewpoint aspect-expressing verbs in the corpus sample (Section 5.2.1), then comparing this with the relative frequencies of each aspectual mapping in the L2 and L1 experimental production data (Section 5.2.2). These analyses, as well as the LEQ data presented in the previous section, provide the basis for addressing RQ2.

Lastly, Section 5.3 presents the comprehension component of the experimental data, which aims to access learners' underlying representations of French viewpoint aspect and consequently permit more concrete discussion on the acquisitional element

of RQ1. The ensemble of the analyses presented in this chapter will be used collectively to address RQ3, which will be discussed in detail in Chapter 6.

5.1 Production Data

5.1.1 Overview of production data

Before presenting any inferential analysis of the production data, it is useful to provide some descriptive context. Table 6 summarises the total number of viewpoint aspectual tokens in the production data, subdivided by aspectual condition (habitual/continuous/progressive/perfective), task (Cat Story/Conversation/overall production) and group (Pre-YA/Post-YA/L1). Each cell presents tokens per aspectual condition as both a raw count and as a proportion of the total number of viewpoint aspectual tokens produced by the group.

		Tokens per viewpoint aspectual condition				
Task	***Group***	*Habitual*	*Continuous*	*Progressive*	*Perfective*	*Total*
Cat Story	Pre-YA (*n* = 20)	96 19.1%	116 23.1%	8 1.6%	282 56.2%	502 100.0%
	Post-YA (*n* = 23)	91 21.0%	108 24.9%	10 2.3%	224 51.7%	433 100.0%
	L1 (*n* = 7)	57 28.2%	50 24.8%	13 6.4%	82 40.6%	202 100.0%
Conversation	Pre-YA (*n* = 20)	16 1.9%	407 47.6%	6 0.7%	426 49.8%	855 100.0%
	Post-YA (*n* = 23)	68 4.2%	988 60.9%	12 0.7%	554 34.2%	1622 100.0%
	L1 (*n* = 7)	8 1.8%	205 45.2%	14 3.1%	226 50.0%	453 100.0%
Overall Production	Pre-YA (*n* = 20)	112 8.3%	523 38.5%	14 1.0%	708 52.2%	1357 100.0%
	Post-YA	159	1096	22	778	2055

	(*n* = 23)	7.7%	53.3%	1.1%	37.9%	100.0%
	L1 (*n* = 7)	65 9.9%	255 38.9%	27 4.1%	308 47.0%	655 100.0%

Table 6: Summary of viewpoint aspectual tokens produced, divided by task and group.

Though a more detailed discussion of the relative proportions of viewpoint aspectual tokens per condition, group, and task will be provided in Section 5.2, a brief look at Table 6 yields the following observations. Looking at overall production, it can be seen that production is variable across the four viewpoint aspectual contexts. Most strikingly, there is a clear frequency discontinuity between, on one hand, the progressive (1.0-4.1% of total tokens across the three groups) and habitual (7.7-9.9% of tokens), and on the other, the continuous (38.5%-53.3% of tokens) and perfective (37.95-52.2% of tokens). Regarding the total number of tokens produced by each group, it can be seen that the Post-YA group produce more tokens (2055 total) than the Pre-YA group (1357 total), and that these appear to be mainly concentrated in the continuous condition. Proportionally, all three groups produce relatively similar numbers of habitual and progressive tokens, and the Pre-YA group also produce similar proportions of continuous and perfective tokens to the L1 group. Here again, the Post-YA group differ slightly, producing proportionally more tokens of the continuous and proportionally fewer tokens of the perfective than the other two groups.

Briefly considering production by task, it can be seen that participants produced markedly more tokens in the Conversation task than the Cat Story task: this is unsurprising, given the differing durations and degrees of open-endedness of each task. Proportionally, participants across all groups produced more habitual tokens in the more-structured Cat Story task than in the less-structured Conversation task, in tandem

with fewer Continuous tokens. Proportions of progressive and perfective tokens remained fairly unaffected across each task. Finally, it can be observed that the large "surge" of continuous tokens observed for the Post-YA group appears to originate from their production in the Conversation task only, with proportions of continuous tokens in the Cat Story task remaining similar across all three groups.

5.1.2 Accuracy in L2 French viewpoint aspectual production

Overview

In order to examine how the L2 French viewpoint aspectual system develops in L1 English learners, it is useful to begin by considering how accurately the two learner groups in this project can produce viewpoint aspectual form-meaning mappings, and how this compares to the production of the L1 French control group. It is also interesting to compare production accuracy between the two L2 groups as we might anticipate that, having spent an extended time in a French-speaking country, the Post-YA group are more accurate at producing French aspectual mappings. To recap, the form-meaning mappings in question are the habitual, continuous, and progressive imperfective meanings, for which the target form is the Imperfect, and the perfective meaning, for which the target form is the *passé composé.* An accurate use is considered to be when the appropriate form is produced for a given meaning.

As mentioned in the previous chapter, participants completed two production tasks: the shorter, more structured Cat Story task, and the longer, less-structured Conversation task. Figure 3 shows the mean accuracy per group for overall production (Cat Story and Conversation combined), while Figure 4 shows mean accuracy per task.

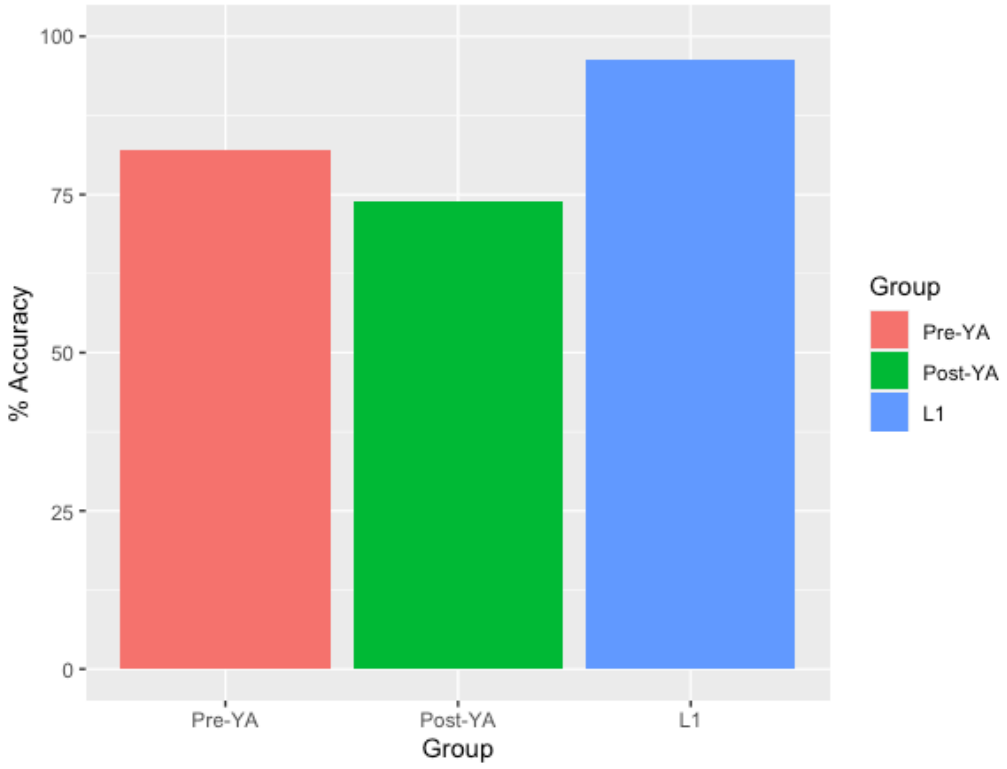


Figure 3: Mean accuracy (%) in overall oral production per group.

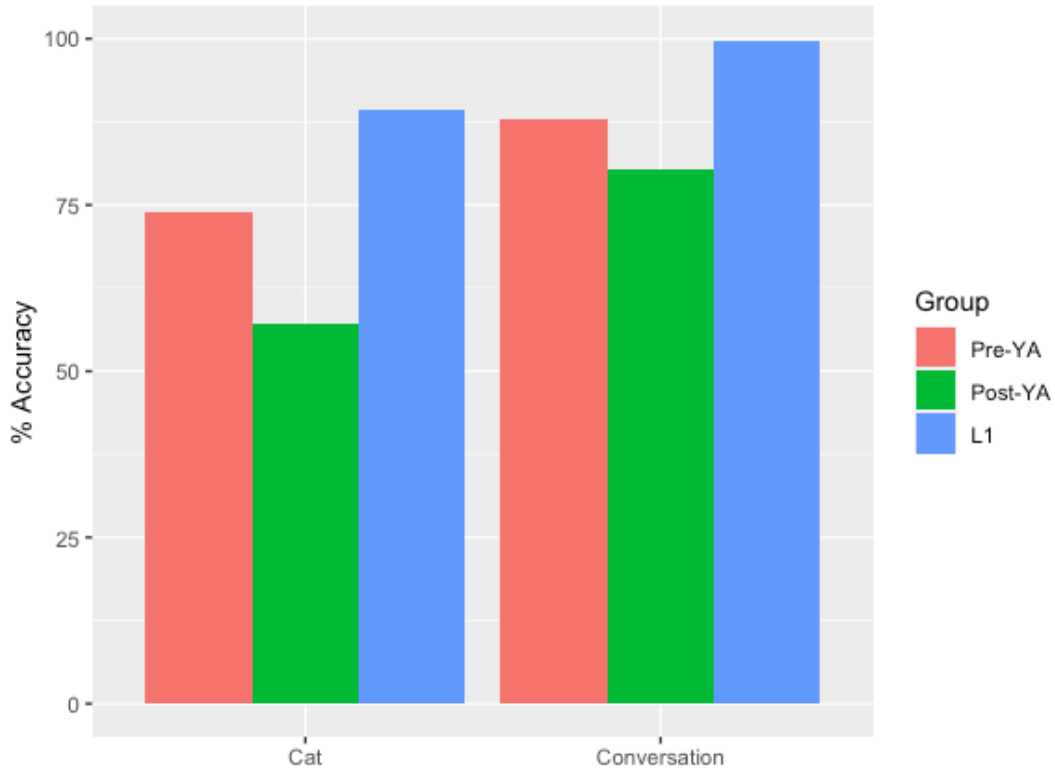


Figure 4: Mean production accuracy (%) per group and per task.

Focusing first on the overall accuracy trend as shown in Figure 3, it can be seen that the Pre-YA group had a mean overall production accuracy of 82.0% (SD = 0.38), the Post-YA group a mean accuracy of 74.0% (SD = 0.44), and the L1 group a mean accuracy of 96.3% (SD = 0.19). As anticipated, the L1 group perform at close to ceiling. The mean accuracy of both L2 groups is also quite high – reflecting their overall fairly advanced proficiency – but the Post-YA group are less accurate overall than the Pre-YA group. This runs counter to the expectation – and general public perception – that an extended time spent in the country of the L2 will be beneficial to grammatical development. Although the two L2 groups are different cohorts, the independent proficiency measure demonstrated that, in terms of overall proficiency, the groups did not differ significantly, and that the Post-YA group even obtained a slightly higher score (see Table 2 in section 4.1.1). It is therefore possible to conclude with reasonable confidence that the differences seen between the Pre- and Post-YA groups here are not simply due to the Post-YA group being generally less proficient.

The same accuracy cline across groups holds when the two production tasks are considered separately: the L1 group score the highest (Cat 89.4% [SD = 0.31], Conversation 99.8% [SD = 0.05]), followed by the Pre-YA group (Cat 73.8% [SD = 0.44], Conversation 87.8% [SD = 0.33]) and lastly the Post-YA group (Cat 57.2% [SD = 0.50], Conversation 80.3% [SD = 0.40]). In addition, all three groups are more accurate in the Conversation task than in the Cat Story task. Though the effect of task will be examined more closely at a later point, it is suggested that this accuracy difference may be linked to the less-structured nature of the Conversation task, which would have permitted learners in particular to avoid or circumvent structures or forms that they felt less confident with.

It is also useful at this point to give some consideration to the individual results across both of the production tasks, to provide more detailed insight into any cross-task and cross-group differences. Figures 5 and 6 illustrate the range of scores per group in the Cat Story and Conversation tasks, respectively. (A full table of every participant's scores across both production and comprehension tasks can be found in Appendix B.)

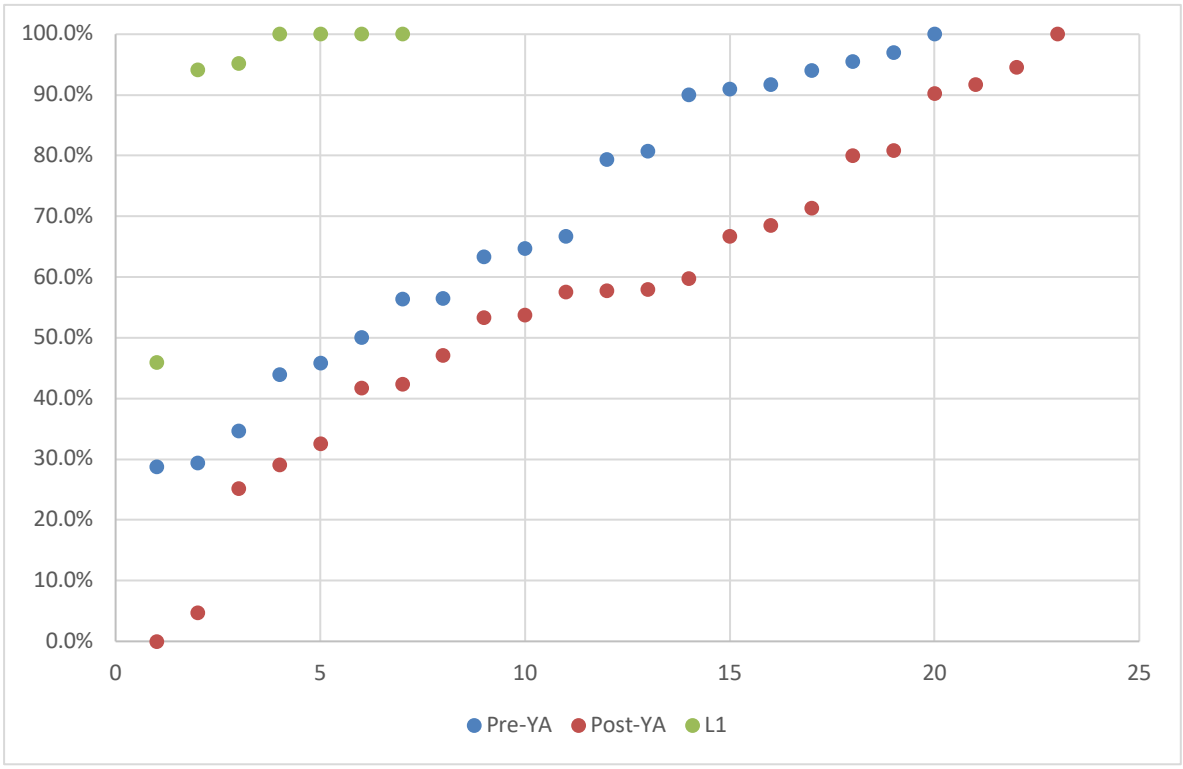


Figure 5: Graphical representation of individual Cat Story scores, by group.

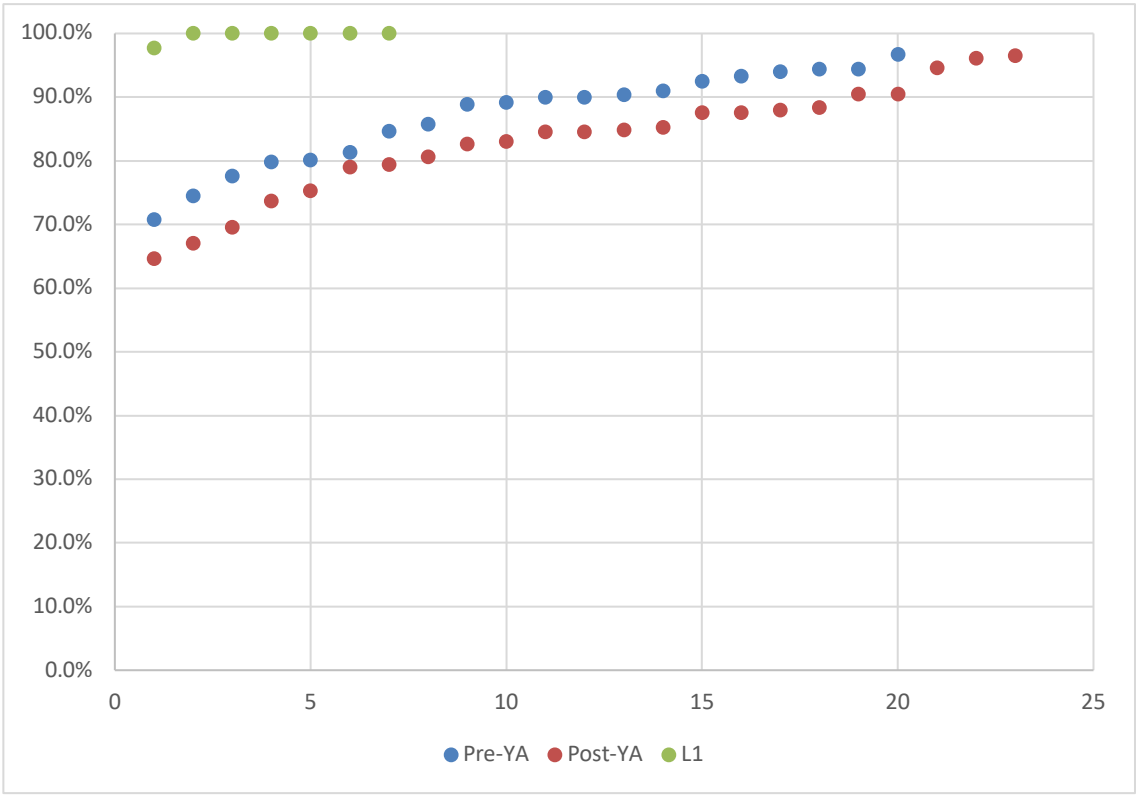


Figure 6: Graphical representation of individual Conversation scores, by group.

When Figures 5 and 6 are compared, the cross-task differences mentioned above are immediately apparent. Whilst the vast majority of L2 participants scored 70% accuracy or higher in the Conversation task (Figure 6), there is a much wider range of scores in Cat Story (Figure 5), with over a quarter of Pre-YA participants and over a third of Post-YA participants scoring 50% or lower. In addition, two very low-scoring Post-YA participants scored under 10% in Cat Story. Whilst outliers are being considered, it is also notable that there is one low scorer from the L1 group in Cat Story: this participant switched to recounting the narrative in the Present tense halfway through the task. Other than this, it can be seen that the L1 group performed essentially at ceiling in both production tasks. Considering the overall trend for the L2 groups, the steeper gradient of Pre-YA scores in both production tasks indicates that – despite both groups containing a range of accuracy scores – there were a larger number of higher accuracy scores in the Pre-YA group, indicating that this group performed more accurately than the Post-YA group overall in production. This difference is more marked in Cat Story, whereas in the Conversation task the scores of the L2 groups pattern markedly more closely together.

In order to evaluate cross-task consistency for the L2 groups, the top scoring participants for each task were identified and compared. To be classified as a "top scorer" in the case of this analysis, participants were required to have an accuracy score of 90% or higher. Table 7 presents the top scorers from highest to lowest across both L2 groups in the two production tasks. The IDs of participants who appear in the top scoring list for both Cat Story and Conversation are presented in bold, underlined font.

TOP SCORERS (90% + ACCURACY)			
Cat Story		*Conversation*	
Participant ID	Score	Participant ID	Score
202	100.0%	218	96.7%
404	100.0%	**410**	96.5%
210	97.0%	422	96.1%
207	95.5%	417	94.6%
408	94.5%	**202**	94.4%
206	94.0%	**207**	94.4%
214	91.7%	203	94.0%
410	91.7%	226	93.3%
212	91.0%	220	92.5%
427	90.2%	**214**	91.0%
201	90.0%	424	90.5%
		427	90.5%
		213	90.4%
		206	90.0%
		215	90.0%

Table 7: Top scoring L2 participants in Cat Story vs. Conversation.

On studying the highest-scoring participants presented in Table 7, it can be seen that four participants from the Pre-YA group (202, 206, 207, 214) and two participants from the Post-YA group (410, 427), were categorised as "top scorers" in both Cat Story and Conversation. Looking broadly at the total number of participants falling into the "top scorer" category (13 total Pre-YA, 7 total Post-YA), this shows that roughly one-third of the total number of top scorers in each L2 group were categorised as such in both production tasks. This could be taken to suggest that cross-task consistency for production is not particularly high. A possible explanation for this may be the previously-attested more challenging nature of the Cat Story task, or the fact that the latter may target a different area of participants' production abilities: for example, participants who are comfortable with holding an extended informal conversation may not necessarily be as skilled at completing a more structured narrative task requiring specific vocabulary and grammar choices, and vice versa. Finally, it can be noted that more Pre-YA participants than Post-YA participants fell into the top scorer category for both production tasks, lending further support to the patterns indicated in Figures 5 and 6 regarding the overall more accurate production from the Pre-YA group.

Given that one of the central areas of interest in this project is the extent to which Feature Reassembly can explain the trajectory of L1 English/L2 French viewpoint aspectual development, it is also crucial to analyse how learners' production accuracy differs by viewpoint aspectual condition, i.e. which viewpoint aspectual form-meaning mapping is required in a given instance. Figure 7 displays the mean percentage accuracy for each of the groups, subdivided by aspectual condition.

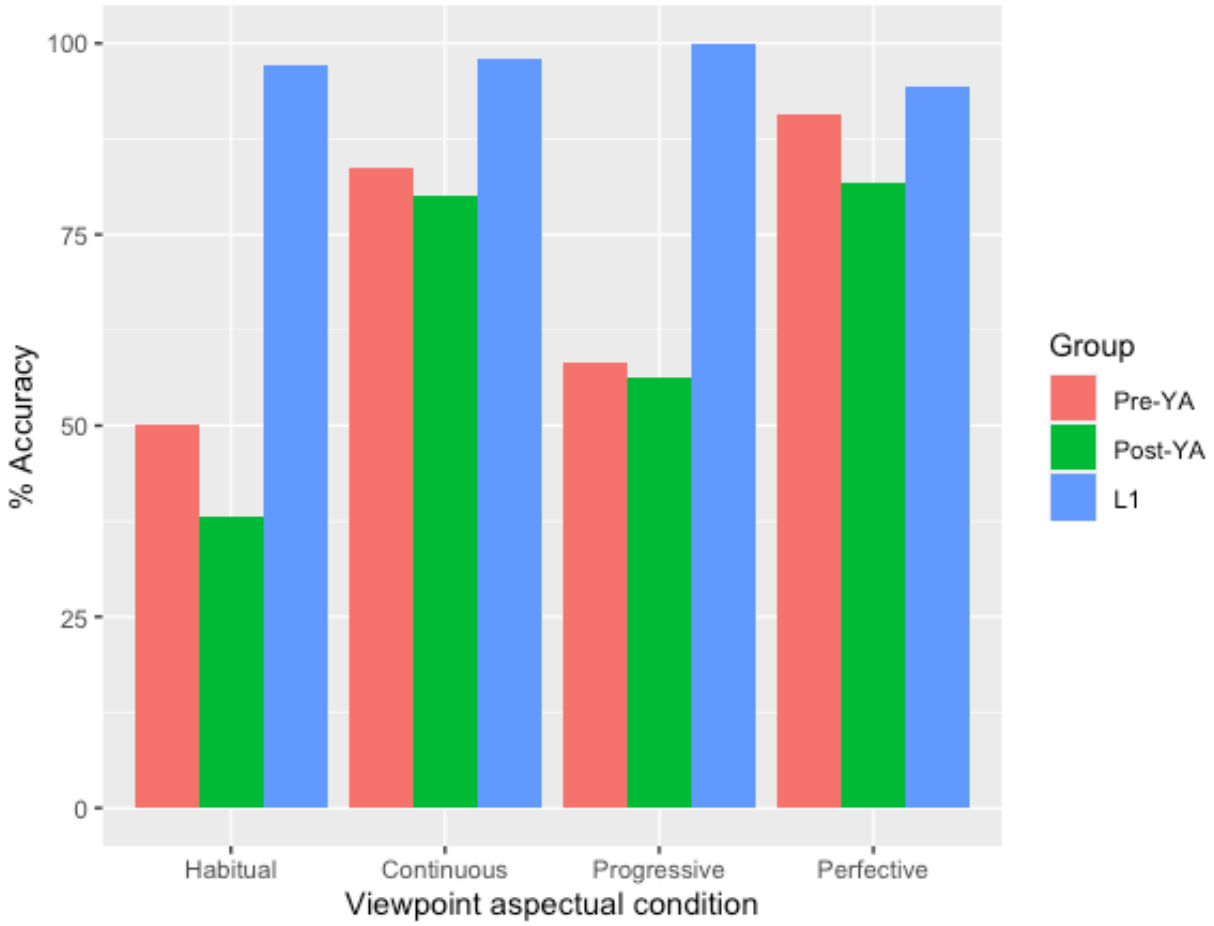


Figure 7: Mean production accuracy (%) by group and aspectual condition.

Overall, Figure 7 shows that, whilst the L1 group remain close to ceiling across all conditions, there is more variation in the performance of the L2 groups. There is a clear discontinuity in accuracy between the perfective and continuous conditions, in which learners display accuracy rates of around 80-90%, and the habitual and progressive conditions, in which learner accuracy ranges from approximately 40-60%. A

full summary of the mean percentage accuracies shown in Figure 5 can be found in Table 8:

Form	**Imperfect**				**Passé composé**
Meaning	*Habitual*	*Continuous*	*Progressive*	*Overall*	*Perfective*
Pre-YA	50.0% [0.50]	83.7% [0.37]	58.3% [0.50]	74.3% [0.44]	90.7% [0.29]
Post-YA	38.2% [0.49]	79.9% [0.40]	56.4% [0.50]	69.9% [0.46]	81.7% [0.39]
Control	97.0% [0.17]	98.1% [0.14]	100.0% [0.0]	98.0% [0.14]	94.5% [0.23]

Table 8: Summary of mean production accuracies (%, [SD]) per group and aspectual condition.

Before analysing these results from a feature reassembly perspective, it is helpful to briefly re-outline what Feature Reassembly would predict as to the ease of acquisition for each of the form-meaning mappings, based on the L1 English/L2 French pairing. Both the perfective and the imperfective-progressive are predicted to be easily acquirable from a reassembly perspective, as both involve a one-to-one mapping between English and French (from V +*ing* to the Imperfect in the case of the progressive, and from the Simple Past to the *passé composé* in the case of the perfective). The habitual would be expected to cause more difficulty, as while habituality can be expressed periphrastically in English (e.g. via *would/used to* + V), it is often expressed using the Simple Past, and so its reassembly task involves a degree of "disentangling", given that the Simple Past also hosts perfectivity. Lastly, the imperfective-continuous is predicted to be the most challenging aspectual mapping to reassemble into French for the learners in this project, as continuousness is the only imperfective meaning to be expressed solely via the Simple Past in English (that is, no additional periphrases are available), and so involves the greatest degree of disentangling from the perfective meaning before reassembly onto the Imperfect can be achieved.

In view of these predictions, the results presented in Table 8 prove somewhat surprising from a feature reassembly perspective, as learners' accuracy does not appear to fully align with the degree of reassembly required for each form-meaning mapping. Though learners' high accuracy in the perfective condition and lower accuracy in the habitual is in keeping with feature reassembly predictions, one would not expect the imperfective-progressive, with its one-to-one form-meaning mapping between English and French, to pose such problems for learners if feature reassembly were the only factor at play. Neither would one expect such high accuracy levels for the continuous, given that this is arguably associated with the most complex reassembly task between English and French. However, when the learners' accuracy for imperfective meanings overall is compared with that of their accuracy for the perfective, the latter is more accurate: this reflects the much more complex reassembly task for L2 French imperfectivity as a whole for these L1 English learners. Taken with the more or less as-expected findings for the habitual and perfective conditions, this preliminary overview suggests that Feature Reassembly is able to accommodate L2 viewpoint aspectual production to at least some extent. In order to probe more deeply into the role of aspectual condition on production accuracy, however, it is necessary to examine this alongside other variables using mixed-effects regression.

Accuracy modelling

Mixed-effects modelling was chosen as the production data included a range of multi-level independent variables (such as aspectual condition, group, and task type) to be modelled as fixed effects, and because it was also desirable to control for random variation in the sample (e.g. by participant) using random effects. As accuracy was coded as a binary variable in the data (1 = accurate, 0 = inaccurate), logistic regression was required, and the type of model used was a generalised linear mixed model. The logistic mixed-effects regression was conducted with the *lme4* package (Bates et al. 2015, version 1.1-26) in R (R Core Team 2018, version 4.0.3).

When carrying out logistic mixed-effects regression on the production data, the statistical analysis was motivated by several core aims. Firstly, it was important to understand the extent to which learners' production accuracy was affected by aspectual condition, given that one of the research questions of this project (RQ1) was based on

Feature Reassembly, which makes specific predictions about the respective difficulty of acquisition of each L2 French viewpoint aspectual form-meaning mapping for L1 English speakers. For this reason, one of the independent variables entered into the model was *Target_Aspect*, a multi-level variable composed of four levels, one for each aspectual condition (habitual, continuous, progressive, perfective). This variable was coded using Helmert coding, as this permitted the model to make comparisons between the levels of *Target_Aspect* in a way that reflected their predicted level of difficulty according to Feature Reassembly. Therefore, the comparisons were made in the following order:

- Perfective vs. [Progressive + Habitual + Continuous]
- Progressive vs. [Habitual + Continuous]
- Habitual vs. Continuous

Given the focus in RQ2 on exploring the properties of naturalistic input, partly through programmes such as the year abroad, it was also important to explore how production accuracy varied by group: this involved making comparisons not only between the Pre- and Post-YA L2 groups, but also between these groups and the L1 French control group, in order to assess whether learners' production differed significantly from that of native French speakers, and whether this was different before vs. after an extended stay abroad. For this reason, the *Group* variable was also entered into the regression model, subdivided according to each of the participant groups involved in this project (Pre-YA, Post-YA, L1 (control)). The *Group* variable was dummy coded, with the L1 controls set as the reference level.

Finally, the effect of task type on production accuracy was also of interest, as, although not directly specified in the research questions, understanding if and how participants' production accuracy varied by task was considered key to gaining a comprehensive picture of the state of their L2 French viewpoint aspectual system. For example, did participants perform more or less accurately in a more structured task (Cat Story), or a less structured task (Conversation)? The descriptive information presented in Figure 4 already suggests that the latter was the case for all three groups, but entering *Task* into the regression model (with two levels, *Cat* and *Conversation*) allows a more powerful exploration into the strength of the relationship between task type and production accuracy. In addition, the model contained a random effect of *Partic_ID,* to control for random variation in how each individual participant responded.

Utilising the above variables, the first phase of model-fitting involved fitting a full model, featuring interactions between all fixed effects, to explore the influence of all of the aforementioned variables on production accuracy. All possible random slopes were also added for all of the fixed effects, to create a maximal random structure as recommended by Barr et al. 2013. As this initial model did not converge, the model was first simplified (as recommended by Linck & Cunnings 2015) by gradually removing random slopes (none could be accommodated within a convergent model), and then by removing the three-way interaction (Group*Task*Target_Aspect). From here, the best-fitting model was ascertained by making manual comparisons with AIC to ensure goodness of fit, and also by using likelihood ratio tests (obtained via the *drop1* function in *lme4*).

The formula for the best-fitting accuracy model, as well as the statistical output for the model, is presented below, with output presented in Table 9:

Formula: Accuracy ~ Group + Task + Group:Task

+ Target_Aspect + Task + Target_Aspect:Task

+ Group + Target_Aspect + Group:Target_Aspect + (1|Partic_ID)

Fixed effects:	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	6.6736	4.3900	1.520	0.128468
GroupPre-YA	-5.7461	4.3952	-1.307	0.191093
GroupPost-YA	-6.4646	4.3939	-1.471	0.141225
TaskConversation	3.8846	1.0490	3.703	0.000213 ***
Target_Aspect1	-4.4453	4.3882	-1.013	0.311048
Target_Aspect2	7.9764	11.6871	0.682	0.494926
Target_Aspect3	0.2244	0.4563	0.492	0.622852
GroupPre-YA:TaskConversation	-3.6816	1.0504	-3.505	0.000457 ***
GroupPost-YA:Task Conversation	-3.0647	1.0446	-2.934	0.003347 **
TaskConversation:Target_Aspect1	0.4549	0.1892	2.404	0.016205 *
TaskConversation:Target_Aspect2	0.1557	0.3843	0.405	0.685386
TaskConversation:Target_Aspect3	-0.5967	0.1223	-4.880	1.06e-06 ***
GroupPre-YA:Target_Aspect1	5.6024	4.3900	1.276	0.201893
GroupPost-YA:Target_Aspect1	5.1557	4.3897	1.174	0.240196
GroupPre-YA:Target_Aspect2	-8.3467	11.6880	-0.714	0.475150
GroupPost-YA:Target_Aspect2	-8.0733	11.6888	-0.691	0.489763
GroupPre-YA:Target_Aspect3	-0.8864	0.4689	-1.890	0.058711 .
GroupPost-YA:Target_Aspect3	-0.7475	0.4668	-1.601	0.109292

Table 9 : Output for all-groups production accuracy model (mixed-effects logistic regression).

Likelihood ratio tests showed that all three interactions in the formula (now henceforth referred to as *Group*Task*, *Task*Target_Aspect*, and *Group*Target_Aspect*, for simplicity) significantly contributed to the goodness-of-fit of the model (*Group*Task*: χ^2 (df = 2) = 31.11, p<.0001; *Task*Target_Aspect*: χ^2 (df = 3) = 31.84, p<.0001; *Group*Target_Aspect*: χ^2(df = 6) = 44.22, p<.0001).

Looking more closely at the output of the model in Table 9, there is a significant main effect of *Task*, showing that, compared to the reference level (Cat Story), accuracy is significantly higher in Conversation (z = 3.703, p = .0002). There is also a significant interaction effect of *Task* for both L2 groups, showing that both the Pre-YA and Post-YA groups behave significantly more accurately in the Conversation task than in the Cat Story task (*GroupPreYA:TaskConversation*: z = -3.505, p = .0005; *GroupPostYA:TaskConversation*: z = -2.934, p = .003). This confirms the significance of the trend presented in Figure 4. A further exploration of the effect of task, and how this may relate to learners' viewpoint aspectual development, will be returned to later.

Within the Conversation task, some significant interaction effects related to aspectual condition are also observed. Firstly, there was a significant interaction of *Target_Aspect1* (the comparison of accuracy in the perfective vs. all imperfective conditions) on accuracy in the Conversation task (z = 0.1892, p = .016). This suggests that accuracy was significantly higher when producing perfective over imperfective form-meaning mappings – even in the Conversation task, where learners were already shown to behave more accurately overall. This confirms the significance of the imperfective vs. perfective accuracy differences presented in Figure 7 and Table 8, and also aligns with the well-attested finding (e.g. Bergtröm 1995; Kihlstedt 1998; Labeau 2005, 2011) that perfective forms tend to be less challenging than their imperfective counterparts.

In addition, there was a significant interaction between the third level of *Target_Aspect* comparisons (habitual vs. continuous) and performance in the Conversation task (z = -4.880, p < .0001). This supports the descriptive data shown in Table 8, wherein it can be seen that accuracy is markedly higher in continuous over habitual conditions for the L2 groups.

Contrastingly, the interaction between *Task(Conversation)* and the second level of *Target_Aspect* comparisons (accuracy in the progressive vs. combined habitual

and continuous accuracy) is not significant (z = 0.405, p = 0.69). This could be attributed to the fact that there is a substantial difference between average production accuracy in the habitual and continuous conditions respectively (i.e. in the region of 30-40% for the L2 groups, as shown in Figure 7), and so the average of these two conditions is likely to be close to the mean accuracy in the progressive condition, given that learners' production accuracy in the latter falls between that in the habitual and continuous.

It is worth recalling at this point that the *Target_Aspect* variable was Helmert coded according to the *anticipated* difficulty of each aspectual condition, based on the predictions of the Feature Reassembly Hypothesis: that is, in order of increasing difficulty, Perfective > Progressive > Habitual > Continuous. The reason behind the lack of significant main effects of *Target_Aspect* on production accuracy, therefore, may be linked to the fact that the degree of feature reassembly does not fully correlate with learners' production accuracy. This discussion will be explored in more depth in Chapter 6.

Bringing the results of this analysis back to the research questions of this project, it can be seen that, with regard to RQ1, several of the descriptive patterns shown in Figure 7 and Table 8 are substantiated: in particular, the significance of accuracy in perfective vs. imperfective conditions, and that between accuracy in habitual and continuous conditions. Whilst the former does support what Feature Reassembly would predict (given the greater reassembly task overall for imperfective mappings), the fact that production accuracy is significantly higher in continuous over habitual conditions – despite the continuous entailing a more complex reassembly task – suggests that Feature Reassembly cannot entirely accommodate the trajectory of viewpoint aspectual acquisition in this case. Of course, to be able to explore this more effectively, it is necessary to fit a model containing only the L2 production data, which will be the focus of the following subsection. This will also allow a closer inspection into RQ2, as the two L2 groups can be directly compared.

Finally, it is worth noting that, despite the descriptive differences between both learner groups and the L1 French group shown in Table 8, these do not translate to significant main effects in the model, where the L1 French group constitutes the reference category. This may be due to the very varied performance of the L2 groups across the four aspectual conditions (for example, Table 8 shows that there is not a

large difference in the performance of learners vs. native speakers in the perfective condition), in combination with the way that the *Target_Aspect* variable is coded.

Modelling L2 production accuracy: a role for language exposure?

When modelling production accuracy for the L2 data alone, it was necessary to include two additional variables, which aimed to capture some detail on the characteristics of the input to which learners were exposed both before and during their year abroad. This relates directly to an element of RQ2, which aims to investigate how studying the properties of naturalistic input can inform our understanding of the impact of programmes such as the year abroad on L2 (viewpoint aspectual) development. An in-depth analysis of representative naturalistic input from L1 French data will be carried out in Section 5.2, but an associated point of interest prior to this is to explore the extent to which learners' exposure to both their L1 (English) and their L2 (French) differs in an instructed vs. naturalistic setting, and how this may predict their L2 production accuracy.

Learners' exposure to French and English was self-assessed using the Language Engagement Questionnaire (LEQ). As mentioned, the LEQ data was collected with a view to providing crucial background to the discussion of the impact of the year abroad and similar programmes. A key motivation behind the collection of the LEQ data was to ascertain whether learners did in fact receive input that was in some way different during their year abroad. In order to make this claim, this necessitated an exploration of whether and how levels of exposure to French and English differed between the Pre- and Post-YA groups (the latter of whom were instructed to respond to the questionnaire based on their experiences *during* their year abroad.)

The two exposure variables were labelled *LEQFren* and *LEQEng* respectively, and were categorical variables whose value could be either Low, Medium or High. Learners were categorised according to the number of their LEQ responses that fell into the three most high-frequency categories ("Every day", "Several times a week", and "A few times a week"): High exposure learners were considered those with 15+ responses (out of a possible 26) in these categories; Medium exposure to be 10-15 responses; and

Low exposure under 10 responses.[1] Figures 8 and 9 show the distribution of Pre- and Post-YA learners across the three exposure categories for French and English respectively.

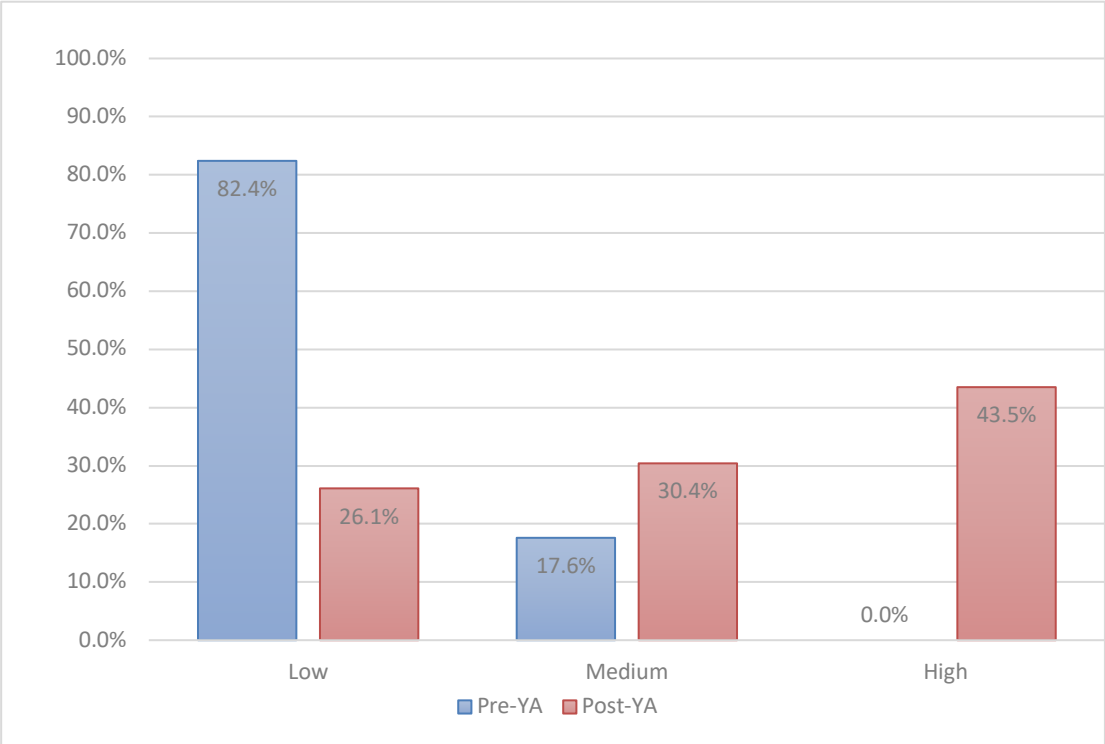


Figure 8: Distribution of LEQFren categories (Low, Medium, High) across L2 groups (Pre-YA, Post-YA).

[1] A full list of the exposure items scored by participants on the LEQ can be found in Appendix A.3.

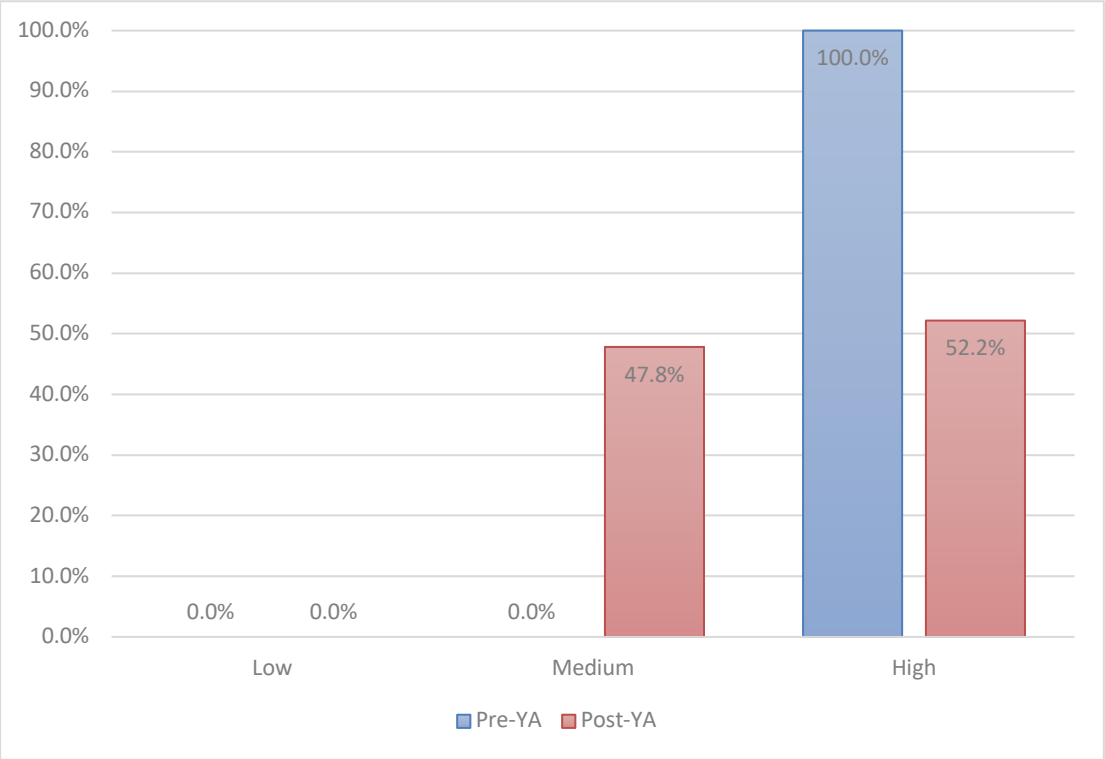


Figure 9: Distribution of LEQEng categories (Low, Medium, High) across L2 groups (Pre-YA, Post-YA).

In total, *n* = 17 Pre-YA participants completed the LEQ (*n* = 3 did not), alongside the full sample of *n* = 23 Post-YA participants. Though the assignment of the criteria for the exposure categories was somewhat arbitrarily determined, it nonetheless seems reasonably fit for purpose when the distribution of learners across the categories is considered (bearing in mind, again, that the Post-YA group were instructed to complete the questionnaire based on their exposure *during* the year abroad). It can be seen in in Figure 8 that, whilst the Post-YA learners are concentrated in the High and (to a slightly lesser extent) Medium exposure categories for French, the Pre-YA group are predominantly categorised as Low exposure, with just a few learners reaching Medium exposure. This aligns with the idea of the Pre-YA learners receiving French input mainly/exclusively in the classroom, whilst the Post-YA learners were exposed to French more frequently and from a wider range of sources. However, it should be acknowledged that around a quarter of Post-YA learners also fell into the Low French exposure category; this highlights the important reality of the heterogeneity of year

abroad experiences (e.g. Kinginger 2007; Klapper & Rees 2012) and that increased exposure to the target language is not guaranteed.

Turning now to Figure 9, it can be seen that no learners fell into the Low exposure category for English – this is unsurprising given that even while abroad, learners would have continued to remain in contact with friends and family in the UK, and may also have used English as a lingua franca or with other L1 English-speaking coursemates (e.g. Coleman 2015; McManus 2019). All of the Pre-YA respondents fell into the High exposure category, again providing support for the validity of the exposure criteria, while the Post-YA group were divided near-equally between the Medium and High exposure categories. Overall, the exposure patterns seen in Figures 8 and 9 permits the exploratory conclusion that learners generally experienced more exposure to French and (somewhat) less exposure to English in the year abroad vs. instructed setting. The latter is worth bearing in mind when exploring the feature reassembly component of learners' aspectual development: even though learners experienced richer and more frequent exposure to French during their year abroad, this did not necessarily coincide with a reduction in their exposure to English, and so learners may have continued to be influenced by their L1 in this setting. Nonetheless, the exposure patterns do suggest that, in general, "year abroad input" is not identical to "pre-year abroad input" – this will be key to future analyses and comparisons of the L2 groups.

Having introduced the new variables *LEQFren* and *LEQEng*, the formula for the best-fitting L2 production model is presented below, along with the model output in Table 10. The same coding schemes for the *Group, Task,* and *Target*_Aspect variables were used as in the all-groups production model. Of the two new variables discussed above, *LEQFren* was Helmert coded, allowing the following comparisons to be made:

1. Low exposure vs. [Medium + High exposure]

2. Medium vs. High exposure

As shown in Figure 9, *LEQEng* had only two levels in reality, as no participants reported Low exposure. This variable was dummy coded, with Medium exposure as the reference category, but did not in any case feature in the best-fitting model for the data.

In addition to a random intercept for *Partic_ID*, an attempt was made to fit random slopes for each of the fixed effects in the formula (following Barr et al. 2013): first in combination, and then progressively simplifying the random slopes structure by removing individual interactions. However, none of the models featuring random slopes converged, so the random structure was simplified to include just the random intercept for *Partic_ID.*

Formula: Accuracy ~ Group+Task+Group:Task+
Target_Aspect+Task+Target_Aspect:Task+
Group+Target_Aspect+Group:Target_Aspect+
Group:Task:Target_Aspect+
Target_Aspect+LEQFren+Target_Aspect:LEQFren+
(1|Partic_ID)

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	0.66492	0.30931	2.150	0.03158 *
GroupPost-YA	-0.42105	0.39064	-1.078	0.28110
TaskConversation	0.21423	0.28493	0.752	0.45213
Target_Aspect1	1.28627	0.25635	5.018	5.23e-07 ***
Target_Aspect2	-0.83278	0.50162	-1.660	0.09688 .
Target_Aspect3	-0.33521	0.14638	-2.290	0.02202 *
LEQFren1	-0.22620	0.19757	-1.145	0.25225
LEQFren2	-0.03740	0.12487	-0.299	0.76457
GroupPost-YA:TaskConversation	0.64300	0.35392	1.817	0.06925 .
TaskConversation:Target_Aspect1	0.07704	0.33789	0.228	0.81963
TaskConversation:Target_Aspect2	-0.16309	0.67704	-0.241	0.80965
TaskConversation:Target_Aspect3	-0.59588	0.24523	-2.430	0.01510 *
GroupPost-YA:Target_Aspect1	-0.66246	0.31657	-2.093	0.03638 *
GroupPost-YA:Target_Aspect2	0.74050	0.63095	1.174	0.24054
GroupPost-YA:Target_Aspect3	-0.24678	0.18737	-1.317	0.18781
Target_Aspect1:LEQFren1	0.27878	0.14595	1.910	0.05613 .
Target_Aspect2:LEQFren1	-0.36803	0.30891	-1.191	0.23350
Target_Aspect3:LEQFren1	-0.04790	0.07413	-0.646	0.51819
Target_Aspect1:LEQFren2	-0.16818	0.08318	-2.022	0.04320 *
Target_Aspect2:LEQFren2	-0.20480	0.17571	-1.166	0.24380
Target_Aspect3:LEQFren2	0.13075	0.04472	2.924	0.00346 **
GroupPost-YA:TaskConversation:Target_Aspect1	0.64242	0.41878	1.534	0.12502
GroupPost-YA:TaskConversation:Target_Aspect2	0.57630	0.85065	0.677	0.49810
GroupPost-YA:TaskConversation:Target_Aspect3	-0.03572	0.28557	-0.125	0.90045

Table 10: Output for L2 production model (mixed-effects logistic regression)

When modelling using only the L2 production data, the two-way interactions *Group*Task, Target_Aspect*Task,* and *Target_Aspect*LEQFren* were all found within the best-fitting model for the data, as seen in the formula above. Model fit was established via manual AIC comparisons, beginning with a maximal fixed-effects structure and progressively simplifying this by removing individual interactions. The three-way interaction *Group*Task*Target_Aspect* was found via likelihood ratio test to border on significance (χ^2 (df = 3) = 7.484, p = .058) with regard to contributing to goodness of fit in the model; however, manual AIC comparisons indicated that removing the three-way interaction worsened model fit, and so the *Group*Task*Target_Aspect* interaction was conserved, despite not yielding significant effects in Table 10. Following the suggestion of Cunnings 2012, this finding was confirmed using the *anova* chi-square function to compare the fit of the model with vs. without the three-way interaction: the model containing the three-way interaction yielded a borderline significant result (χ^2 (df = 3) = 7.484, p = .058), as was the case when the contribution of the three-way interaction was tested via the likelihood ratio test.

Considering the output of the model in Table 10, there is a main effect for *Target_Aspect1* (perfective vs. all imperfective) (z = 5.018, p < .0001), adding further weight to the previous evidence seen in the all-groups model and showing that learners specifically were significantly less accurate in producing imperfective over perfective forms. There is also a significant effect for *Target_Aspect3* (comparing accuracy in habitual vs. continuous conditions, z = -2.290. p = .02), further substantiating the trend shown both in Figure 7 and partly aligning with the all-groups model. *Target_Aspect2* (progressive vs. [habitual + continuous]) is not a significant effect (z = -1.160, p = .09); as mentioned, this is likely to be linked to the fact that, in the L2 data, accuracy in the progressive condition fell between accuracy for the habitual and continuous conditions.

It is worth noting here that, in the all-groups model, the significant effects relating to aspectual condition were seen as interaction effects in combination with *TaskConversation*, whereas for the L2 model, *Target_Aspect1* and *Target_Aspect3* are significant main effects. In the L2 model, there was no significant main effect of Task when the Conversation task was compared to the reference category of Cat Story (z = -0.432, p = 0.67), suggesting that learners did not behave significantly more accurately

overall in one task over the other; however, there was a significant interaction effect between *TaskConversation* and *Target_Aspect3* (z = -2.430, p = 0.02). Taken together, the L2 production model suggests that learners' production accuracy is significantly directly impacted by aspectual condition: this is most clear in terms of perfective vs. imperfective accuracy, but learners are additionally significantly less accurate in habitual over continuous conditions, particularly during the Conversation task. This has the same implications for RQ1 as discussed with relation to the all-groups model: namely, that feature reassembly can only partly accommodate the accuracy cline shown in the data.

The second research question of this project requires comparison between the two L2 groups, as it relates to the impact of programmes such as the year abroad on viewpoint aspectual development. In this L2 model, it can be seen that there is no significant difference in overall production accuracy between the groups, as seen by the lack of main effect of *GroupPostYA* when compared to the reference category, *GroupPreYA* (z = -1.078, p = 0.28). However, there is a significant interaction effect of *Target_Aspect1* on *GroupPostYA* (z = -2.093, p = .04). This suggests that the Post-YA group's production accuracy is significantly lower than that of the Pre-YA group when comparing perfective vs. imperfective performance. Taken in combination with the data shown in Figure7, there is therefore some evidence to suggest that the Post-YA group are less accurate at producing French viewpoint aspectual mappings than the Pre-YA group, despite performing descriptively more accurately on the global written cloze test. This raises the question of the role played by the year abroad on learners' L2 viewpoint aspectual development. Given that RQ2 pertains specifically to the interplay between programmes such as the year abroad and properties of naturalistic input, a logical next step is to begin exploring the influence of the input to which the Pre- and Post-YA learners were exposed.

As discussed at the start of this subsection, a descriptive exploration of the LEQ data revealed that - despite some (anticipated) heterogeneity within both groups - learners were shown to generally experience greater exposure to French and somewhat lesser exposure to English during their year abroad, when compared with the pre-year abroad setting. However, to what extent did learners' exposure to English and French predict their production accuracy of viewpoint aspectual forms? Returning to the L2 production model, it should first be stated that *LEQEng* did not feature in the best-

fitting model, suggesting that learners' level of exposure to English did not substantially influence their performance. Contrastingly, Table 10 shows two significant interaction effects pertaining to *LEQFren*. Firstly, there was a significant interaction effect of the second comparison of *LEQFren* (Medium vs. High exposure) on the first comparison of *Target_Aspect* (perfective vs. combined imperfective accuracy), z = -2.022, p = .04. There was also a significant interaction effect of the same *LEQFren* comparison (Medium vs. High) on *Target_Aspect3*, the comparison between habitual and continuous accuracy (z = 2.924, p - .003). To facilitate the interpretation of these interaction effects, the effect plot for the *Target_Aspect*LEQFren* interaction is shown below in Figure 10.

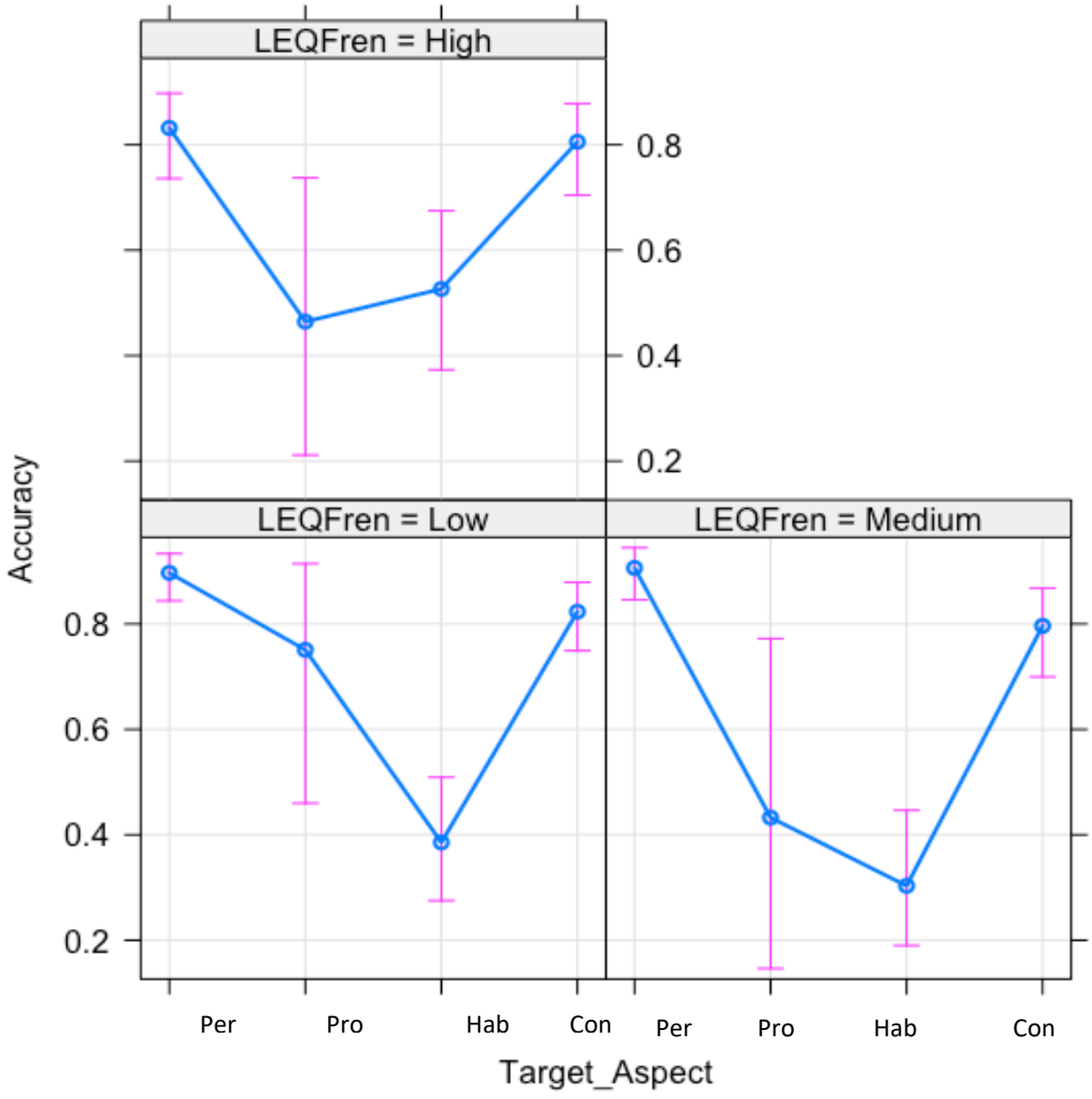


*Figure 10: Effect plot of Target_Aspect*LEQFren interaction in L2 production model*

Beginning with the first significant interaction effect, it can be seen that accuracy is fairly uniform and high in the perfective condition for both the Medium and High exposure groups, but there is more variety in the imperfective conditions. Although there appears to be little difference in production accuracy in the continuous condition between the Medium and High exposure groups, this difference is larger in the progressive condition, and larger still in the habitual condition. Taken together, the data shown in Figure 10, in combination with the significant interaction effect, provide some evidence that the group with the highest exposure to French (composed of around 44% of the Post-YA participants and none of the Pre-YA participants), produce imperfective viewpoint aspectual forms significantly more accurately than the group with medium exposure to French (composed of around 30% of Post-YA participants and around 18% of Pre-YA participants). The habitual condition appears most influenced by level of exposure, followed by the progressive condition, whilst the continuous and perfective conditions appear relatively unaffected. The sensitivity of the habitual to level of exposure (and the relative stability of the continuous) is supported by the second significant interaction effect, which holds as mentioned between accuracy in habitual vs. continuous conditions when the Medium and High exposure groups are compared.

This is interesting to consider in light of Figure 7, where it can be seen that the habitual and (to a lesser degree) progressive conditions – which appear to be sensitive to level of exposure – are the conditions where both L2 groups are least accurate. Further discussion and surrounding this can be found in Chapter 6. It should also be stressed at this point that high exposure does not correlate entirely with the year abroad, as under half of year abroad participants fell into this category on the LEQ. This reinforces the previously-highlighted point that year abroad experiences can vary enormously, and not all participants will have been exposed to French (or English) to the same extent during their time abroad. With this in mind, even if higher exposure to French does predict higher production accuracy for imperfective mappings in this data, this cannot be used to suggest that participating in a year abroad programme will itself positively influence production accuracy. This helps to explain the patterns shown in Figure 7, as well as the above-mentioned significant interaction effect indicating the

Post-YA group's significantly lower accuracy in comparison with the Pre-YA group when perfective vs. imperfective accuracy is compared.

Although RQ2 relates specifically to the impact of the year abroad from an input perspective, it seems logical to explore the impact of differing input across the board within the L2 data, to establish a basis for comparison. In addition, the role of exposure to English should also be taken into account: although *LEQEng* was not found to be a significant predictor of L2 production accuracy, it maintains relevance from a theoretical perspective, given the roles of both the L1 and L2 within Feature Reassembly. Therefore, this section will be closed with an exploration into how L2 production accuracy varies by level of French and English exposure. Figure 11 below displays how learner accuracy varies according to the levels of *LEQFren*.

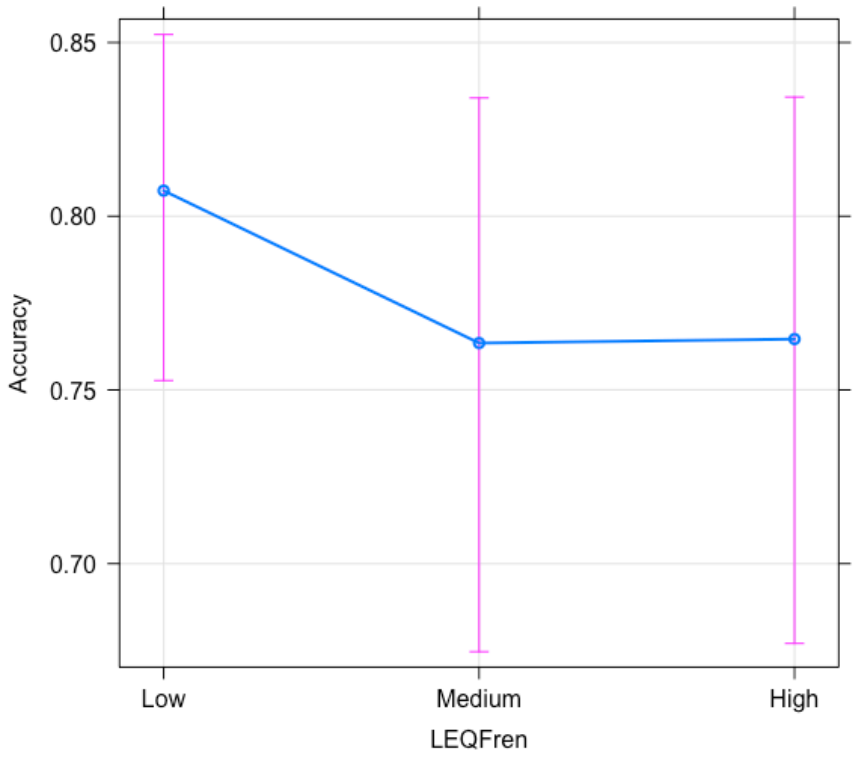


Figure 11: Effect plot from logistic regression analysis displaying how L2 production accuracy varied according to LEQFren category.

As can be seen, learners categorised as having medium or high French exposure had nearly identical accuracy scores on average (around 76.0%), whereas the

Low exposure group had a higher average accuracy (around 81.0%). It should also be noted that error bars for the Medium and High exposure groups are very wide, suggesting a lot of variance in accuracy within these groups, whereas the narrower error bars for the Low exposure group suggests less variance.

What explanation can be given for the fact that the Low exposure group are most accurate on average? Returning to the distribution shown in Figure 8, recall that the Low exposure group was composed predominantly of Pre-YA learners, who were found to be (descriptively) more accurate than the Post-YA group on the whole. Given that the Pre-YA group were receiving their French input mainly from the language classroom, it may be tentatively suggested that the kind of "targeted" instructed input received by these learners might be more facilitative of reinforcing aspectual form-meaning mappings than the more diverse, naturalistic input received by the Post-YA group (i.e. mainly Medium and High exposure groups) during their year abroad. This will be explored in further detail later.

Next, Figure 12 displays how learners' production accuracy varies according to level of English exposure (expressed as *LEQEng*).

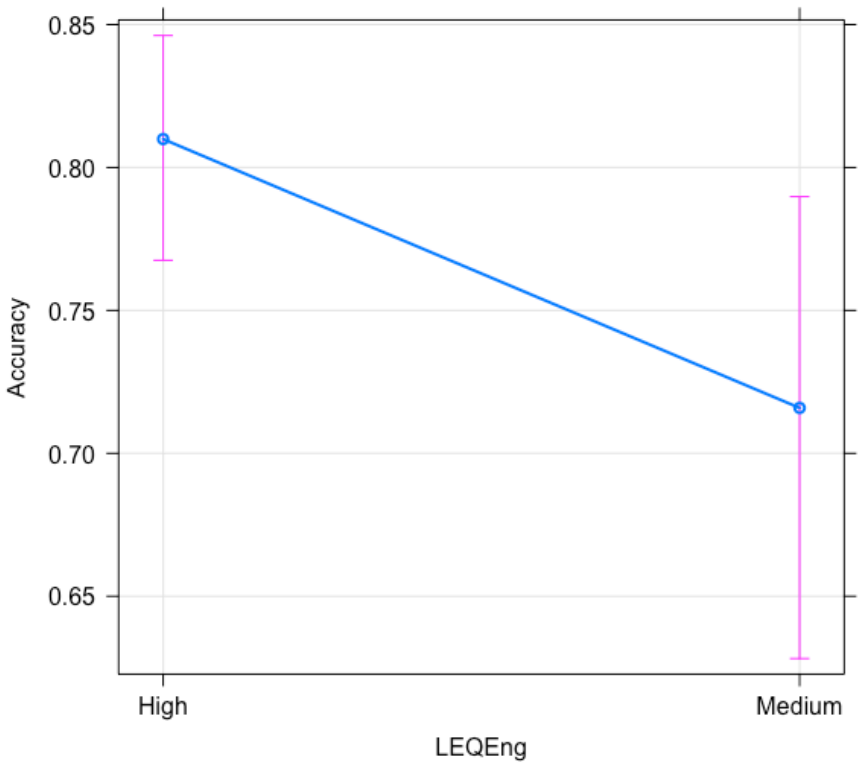


Figure 12: Effect plot from logistic regression analysis displaying how L2 production accuracy varied according to LEQEng category.

Similarly to Figure 11, the findings in Figure 12 arguably run counter to expectations, as the group with the highest English exposure are more accurate than the group with less English exposure (although it should not be assumed that less English exposure automatically translated to more French exposure in every case). Regardless, in view of the fact that the entirety of the Pre-YA group was in the high English exposure category, and that the Medium exposure category was composed solely of members of the Post-YA group, this pattern appears to align with that seen for *LEQFren*.

To summarise, the analysis of the L2 production data revealed that, whilst aspectual condition did predict learners' production accuracy of viewpoint aspectual forms, this once again did not fully align with Feature Reassembly predictions. Whilst some predictions, such as the significantly more accurate performance in perfective over imperfective conditions, were aligned, degree of reassembly could not account for learners' significantly higher accuracy in continuous over habitual conditions, or why the progressive was not found to be significantly more accurate than the combined accuracy in habitual and continuous conditions. Though the lack of significant main effect of *Group* seemed initially to suggest a lack of significant overall difference between the pre- and post-year abroad cohorts, the significant interaction between *GroupPostYA* and the first comparison (perfective vs. imperfective) of *Target_Aspect* provides some evidence to the contrary. Exploring this further, the exposure variables *LEQFren* and *LEQEng* were analysed: firstly, to gain background information on learners' exposure to French and English prior to and during the year abroad; and secondly, to explore whether this was a predictor of production accuracy. The significant interactions from the L2 production model pointed towards a beneficial effect of increased French exposure on production accuracy, particularly with regard to the habitual and progressive imperfective mappings. However, this should be considered alongside the important caveat that under half of Post-YA learners were categorised as receiving high French exposure during their year abroad, with the rest split across the other two exposure categories. Finally, the relationship between accuracy and both French and English exposure was directly explored. This yielded the unexpected finding that the low French exposure category and high English exposure category corresponded with the highest production accuracy. This unusual discovery may be explained by the fact that the primary members of the Low French and High English exposure categories were in the Pre-YA group, who performed descriptively more

accurately overall. The findings from the L2 production analysis were therefore somewhat ambiguous with regard to the impact of programmes such as the year abroad on L2 viewpoint aspectual development; this may be linked to the high degree of variety in year abroad experiences, which makes analysing their impact a complex task.

In view of the unexpected nature of some of the above production results, an avenue of investigation that may provide informative is an examination of the full range of verbal forms utilised by participants in viewpoint aspectual conditions. This has the potential to provide further insight on the feature reassembly process, as well as on whether any particular strategies external to feature reassembly are being employed. This exploration of verbal form will be presented in the following subsection.

5.1.2 Examining verbal forms used in the expression of viewpoint aspect

In addition to assessing learners' accuracy in production of French aspectual form-meaning mappings, the verbal forms that were used by learners (and the control group) across the four aspectual conditions were also recorded. The following section will present the proportions of forms used across the four aspectual conditions by both L2 groups and the control group, considering not only the accurate suppliance of the target form per condition, but also which alternative forms were used and their frequency. Given that task type was shown to significantly impact production accuracy, the results for Cat Story and Conversation are presented separately, taking stock of any differences in the forms used between the tasks.

The aim of analysing the forms used by participants is motivated by several aims. Of course, it is possible to see how often the target form is produced in each aspectual condition, which can be equated to accuracy. However, looking at the forms used in 'non-target' usage can also be informative in several ways. Firstly, this can be used as a measure of how successfully learners have overcome their L1 form-meaning mappings: for example, both continuity and habituality are mapped to the Simple Past (host of perfectivity) in English, and so learners who use the *passé composé* in habitual or continuous contexts are likely to be affected by sustained L1 influence. This information can contribute towards RQ1's aim of exploring how well FR can

accommodate L2 aspectual development trajectories: based on this theory, persistent non-target usage of the *passé composé* in habitual and (especially) continuous contexts would not be surprising, whilst it would be unexpected in progressive conditions (as progressivity has a dedicated form in English and is not mapped to the Simple Past).

In addition to use of the two forms of direct interest to this project (the Imperfect and the *passé composé*), use of the Present, as well as of 'other' forms,[2] was recorded. Recording rates of suppliance of the target form is naturally of interest in order to provide information on learners' developing L2 French viewpoint aspectual systems: for example, upcoming analyses in this section will compare not only the performance between the two L2 groups, but also contrast this with the L1 group. This will provide information not only on whether one L2 group is able to produce viewpoint aspectual mappings in a significantly more targetlike way than the other, but also whether either group has attained a level of targetlike suppliance that approximates that of native French speakers. The decision to further explore usage of the Present specifically was motivated by an initial observation that this form was unexpectedly prevalent in learners' production, despite often arising in contexts where a past time reference was required. The analyses pertaining to the Present are thus somewhat exploratory in nature, but nonetheless could be considered an extension of RQ1: as using the Present cannot feasibly be linked to a reassembly problem, its use must be linked to something external to Feature Reassembly. Therefore, exploring the use of non-reassembly-linked forms in learner production aids in defining the limits of what Feature Reassembly can explain regarding L2 viewpoint aspectual development.

The forms used in Cat Story are presented in Figure 13. Form usage was subcategorised into IMP (Imperfect), PC (*passé composé*), PRES (Present) and OTHER (all other forms). As a reminder, the target form for all three imperfective conditions (habitual, continuous, and progressive) was the Imperfect, and the target form for the perfective condition was the *passé composé*. A table of the number of tokens of each form used by each group in each condition and production task can be found at the start of this chapter.

---

[2] The other forms used were relatively minimal (35 total in Cat Story and 45 total in Conversation), and mainly spanned a handful of tokens each of forms such as the infinitive, the conditional, the future, and the pluperfect.

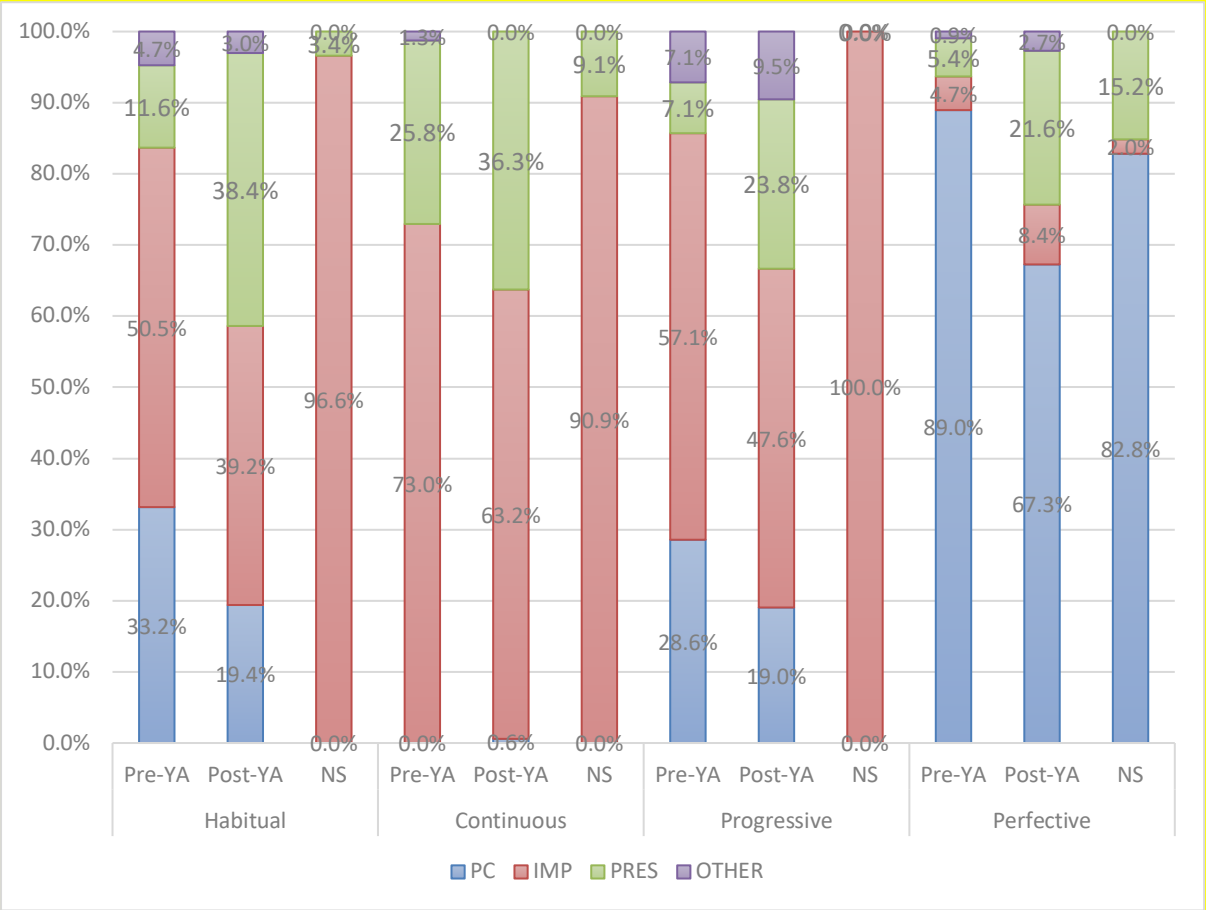


Figure 13: Proportions of forms used by aspectual condition by Pre- and Post-YA L2 groups and the control group (NS) in the Cat Story task.

Figure 13 shows that the target form was supplied to varying extents by each group across each of the aspectual conditions, as has previously been discussed.

Due to some form categories containing very few tokens, it was not feasible to use a regression model to evaluate which variables significantly impacted the suppliance of a given form. Instead, a series of one-way ANOVAs (with Bonferroni correction) was used to assess if there were significant differences in the proportions of each form (PC/IMP/PRES/OTHER) supplied by each group (Pre-YA/Post-YA/L1) in each aspectual condition (habitual/continuous/progressive/perfective). Firstly, in order to provide baseline information on the performance of each group, a series of ANOVAs were run to establish the effect of the independent variable *Group* on suppliance of the target form in each aspectual condition during Cat Story. The output of these tests is presented in Table 11.

Aspectual condition	**Group**	**Mean**	**SD**	**Std. Err.**	**95% CI for mean**	**F**	**p-value**	**Effect size (ω^2)**
Imperfective (combined)	Pre-YA (n = 20)	60.0	30.3	6.77	[45.8, 74.2]	6.337	.004 **	.176
	Post-YA (n =23)	49.4	33.7	7.03	[34.8, 64.0]			
	L1 (n = 7)	95.8	7.8	2.95	[88.6, 103.0]			
Continuous	Pre-YA (n = 20)	74.1	24.8	5.55	[62.5, 85.7]	3.452	.040 *	.089
	Post-YA (n =23)	64.4	32.5	6.79	[50.3, 78.5]			
	L1 (n = 7)	95.5	11.8	4.46	[84.6, 106.4]			
Habitual	Pre-YA (n = 20)	51.1	40.4	9.03	[32.2, 70.0]	6.182	.004 **	.175
	Post-YA (n =23)	38.3	41.0	8.73	[20.2, 56.5]			
	L1 (n = 7)	96.4	9.4	3.57	[87.7, 105.2]			
Progressive	Pre-YA (n = 12)	54.2	49.8	14.38	[22.5, 85.8]	3.794	.035 *	.153
	Post-YA (n =13)	42.3	44.9	12.46	[15.2, 69.5]			
	L1 (n = 6)	100.0	0.00	0.00	[100.0, 100.0]			
Perfective	Pre-YA (n = 20)	87.1	19.4	4.35	[78.0, 96.2]	2.831	.069	.068
	Post-YA (n =23)	68.1	31.8	6.63	[54.3, 81.8]			
	L1 (n = 7)	85.7	32.9	12.43	[55.3, 116.1]			

Table 11: Output of ANOVA analyses exploring the effect of the independent Group variable on suppliance of target forms per aspectual condition

As might be anticipated, a significant main effect of Group (F = 6.337, p = .004, ω^2 = .176) was found for suppliance of the target Imperfect form when the results for the three imperfective mappings were combined. Post-hoc analyses revealed that this effect was significant between the L1 French control group and both the Pre-YA (p = .029, 95% CI [-68.7, -2.9]) and Post-YA (p =.003, 95% CI [-78.7, -14.0]) L2 groups, but not between the two learner groups themselves. In contrast, no significant main effect of Group was found for suppliance of the target *passé composé* in perfective contexts (p = 0.69). This indicates that the production of the L2 groups was comparable to that of

the control group in the perfective condition, but that this was not the case for their production of imperfective mappings.

Focusing on the L2 groups themselves, Figure 13 displays a lower suppliance of the target form in every condition for this group when compared with Pre-YA target suppliance. This aligns with the previously-stated overall lower accuracy for the Post-YA group. These differences lie approximately within the 10% range for the three imperfective mappings (9.5% progressive, 9.8% continuous, 11.0% habitual), and 21.7% for the perfective. Though the latter difference appears pronounced, it is important to remember the very high accuracy of the Pre-YA group in the perfective condition, relative to the imperfective conditions. As shown in Table 11, the ANOVA analyses established a significant main effect of Group on suppliance of the Imperfect in each of the individual imperfective conditions: habitual (F= 6.182, p = .004, ω^2 = .175); continuous (F =3.452, p = .04, ω^2 = .089); and progressive (F = 3.794, p =.035, ω^2 = .153). This main effect was found to be significant between the control and Post-YA groups for every imperfective mapping (p(habitual) = .003, 95% CI [17.0, 99.2]; p(continuous) = .036, 95% CI [1.5, 60.7]; p(progressive) = .033, 95% CI [3.8, 111.6]), but only for the habitual condition (p = .02, 95% CI [-86.9, -3.8]) between the control and Pre-YA groups. The difference in targetlike Imperfect suppliance between the two L2 groups did not reach statistical significance (p = .772) – however, the very fact that the Post-YA group performed less accurately as seen in Figure 13, despite displaying descriptively higher global proficiency, runs counter to expectations, particularly in view of the positive perception of time spent abroad on L2 development (e.g. Freed 1995; Llanes 2012; Llanes & Muñoz 2009).

In view of this, further analyses were carried out to explore which forms learners were using when they were *not* providing the target form. The first series of additional ANOVAs aimed to assess whether there was a significant effect of Group on suppliance of the Present. As mentioned, this direction of investigation had an exploratory basis, after it was observed that learners produced a fairly large number of tokens of the Present compared to the entirety of forms in the 'Other' category. For example, during Cat Story, the Pre-YA group produced 81 Present tokens vs. 15 'other' tokens, whilst the Post-YA group produced 228 Present tokens vs. 18 'other' tokens (full information on all tokens produced can be found at the start of the chapter). Given that use of the Present cannot be linked to L1 influence (given that all of the viewpoint

aspectual form-meaning mappings in this project instantiate past time reference), significant use of this form may indicate an alternative strategy or deficit that is affecting learner performance in a manner not directly linked to the reassembly process. Moreover, if Present usage varies significantly between the L2 groups, this could be said to be linked to the French input that each group has recently been exposed to, thus providing potential insight into RQ2. These ideas will be explored in depth in Chapter 6.

The output for the analyses exploring the effect of Group on suppliance of the Present in the Cat Story task is presented in Table 12. Note that the numbers in each group are smaller in the progressive condition, as some participants did not produce any progressive tokens. This may also partly explain the non-significance and small effect size of this test in the series.

Aspectual condition	**Group**	**Mean**	**SD**	**Std. Err.**	**95% CI for mean**	**F**	**p-value**	**Effect size (ω^2)**
Imperfective (combined)	Pre-YA (n = 20)	15.9	16.9	3.78	[8.0, 23.8]	4.835	.012 *	.133
	Post-YA (n =23)	34.7	34.7	7.23	[19.7, 50.0]			
	L1 (n = 7)	4.2	7.8	2.95	[-3.0, 11.4]			
Continuous	Pre-YA (n = 20)	24.7	25.2	5.64	[12.9, 36.5]	3.350	.044 *	.086
	Post-YA (n =23)	35.2	32.9	6.85	[21.1, 49.5]			
	L1 (n = 7)	4.5	11.8	4.47	[-6.5, 15.4]			
Habitual	Pre-YA (n = 20)	12.7	27.6	6.16	[-0.25, 25.6]	3.948	.026 *	.107
	Post-YA (n =23)	37.0	42.5	9.07	[18.2, 55.9]			
	L1 (n = 7)	3.6	9.4	3.57	[-5.2, 12.3]			
Progressive	Pre-YA (n = 12)	8.3	28.9	8.33	[-10.0, 26.7]	1.127	.338	.008
	Post-YA (n =13)	23.1	43.9	12.16	[-3.4, 50.0]			
	L1 (n = 6)	0.0	0.0	0.00	[0.0, 0.0]			
Perfective	Pre-YA (n = 20)	5.5	13.0	2.91	[-0.5, 11.6]	1.688	.196	.027
	Post-YA (n =23)	18.5	27.2	5.67	[6.8, 30.3]			
	L1 (n = 7)	12.6	30.9	11.67	[-15.9, 41.1]			

Table 12: Output of ANOVA analyses exploring the effect of the independent Group variable on suppliance of the Present form in Cat Story

As seen in Table 12, a significant main effect of Group was established for the suppliance of the Present in both habitual (F = 3.948, p =.026, ω^2 = .107) and continuous (F = 3.350, p=.044, ω^2 = .086) conditions, as well as when suppliance across all three imperfective mappings was combined (F = 4.835, p= .012, ω^2 = .133). Post-hoc analyses confirmed that the significant differences were between the L1 group and the Post-YA group for the continuous (p= .042, 95% CI [0.9, 60.8]) and overall imperfective (p = .029, 95% CI [2.4, 58.6]) (no significant post-hoc results were obtained for the habitual).

A series of repeated-measures ANOVAs (one per participant group) were used to examine within-groups differences and specifically explore the effect of Target_Aspect as an independent variable on Present suppliance. The output of these analyses is presented in Table 13.

Group	**Aspectual condition**	**Mean**	**SD**	**df (hypothesis)**	**df (error)**	**F**	**p-value**	**Effect size (η^2)**
Pre-YA (n = 20)	Continuous	25.9	23.5	3	9	2.588	.07	.190
	Habitual	10.1	25.4					
	Progressive	8.3	28.9					
	Perfective	9.2	16.0					
Post-YA (n = 23)	Continuous	37.3	36.2	3	10	3.418	.027 *	.222
	Habitual	46.5	41.9					
	Progressive	23.1	43.9					
	Perfective	18.3	27.7					
L1 (n = 7)	Continuous	5.2	12.8	3	15	1.172	.353	.190
	Habitual	0.0	0.0					
	Progressive	0.0	0.0					
	Perfective	14.7	33.2					

Table 13: Output of repeated-measures ANOVA analyses exploring the effect of the independent Target_Aspect variable on suppliance of the Present in Cat Story

The analyses shown in Table 13 indicate that, while aspectual condition did not significantly influence percentage suppliance of the Present for either the L1 or Pre-YA groups, it did for the Post-YA group (F = 3.418, p= .027, partial η^2 = .222). Estimated marginal means pairwise comparisons revealed that the significant difference was specifically found between Present suppliance in perfective vs. habitual conditions (p = .037, 95% CI [1.4, 54.9]).

Taken together, the analyses presented in Tables 12 and 13 indicate that the Post-YA group produced the non-target Present tense significantly more than either other group, and significantly more so than the L1 group in imperfective conditions overall, and particularly the continuous condition. They were also significantly more likely to produce the Present in the habitual condition compared to the perfective condition. This non-targetlike usage was not found to a significant extent in the Pre-YA group. These findings provide some potential support for the above-mentioned idea that learners may be employing (consciously or subconsciously) a particular strategy when producing viewpoint aspectual mappings, that does not initially appear to relate directly to Feature Reassembly. Given that it is in the Post-YA group that this particularly marked use of the Present is seen, it may be suggested that this unexpected behaviour could be in some way linked to learners' linguistic experience during the year abroad.

Before developing these ideas further, data from the second production task, Conversation, should be analysed. The main descriptive trends of this dataset have already been summarised at the beginning of the chapter, but to recap, this data relates to a production task that is both longer – resulting in markedly more overall tokens (2930 vs. 1137) of viewpoint aspectual forms – and less structured than Cat Story. As previously discussed, in the absence of explicit scaffolding, the frequency of the habitual mapping is much lower in the Conversation task (1.8-4.2% across all groups vs. 19.1-28.2% in Cat Story). Another key observation is the large 'surge' in tokens for the continuous mapping, specific to the Post-YA group (constituting 60.9% of Post-YA tokens vs. 24.9% in Cat Story). The relative proportions of perfective and progressive mappings are relatively unchanged between the two tasks across all groups.

Moving away from the proportions of viewpoint aspectual mappings, the proportions of forms used per aspectual condition in the Conversation task are presented in Figure 14. As in Cat Story, it can be seen that there is variety in the proportions of suppliance of the target form by the learner groups in each condition: for

example, in the habitual condition, learners supply the Imperfect in around 37-47% of instances, while both the perfective and continuous see the target form supplied in around 82-92% of cases. Usage of the *passé composé* is quite prevalent in the habitual condition, accounting for approximately 45% of intended imperfective-habitual tokens in both L2 groups. Usage of the Present is consistently in the region of 15% across all imperfective conditions for the Post-YA group, whereas its use is more varied in the Pre-YA group. As anticipated, the L1 French group perform at ceiling throughout.

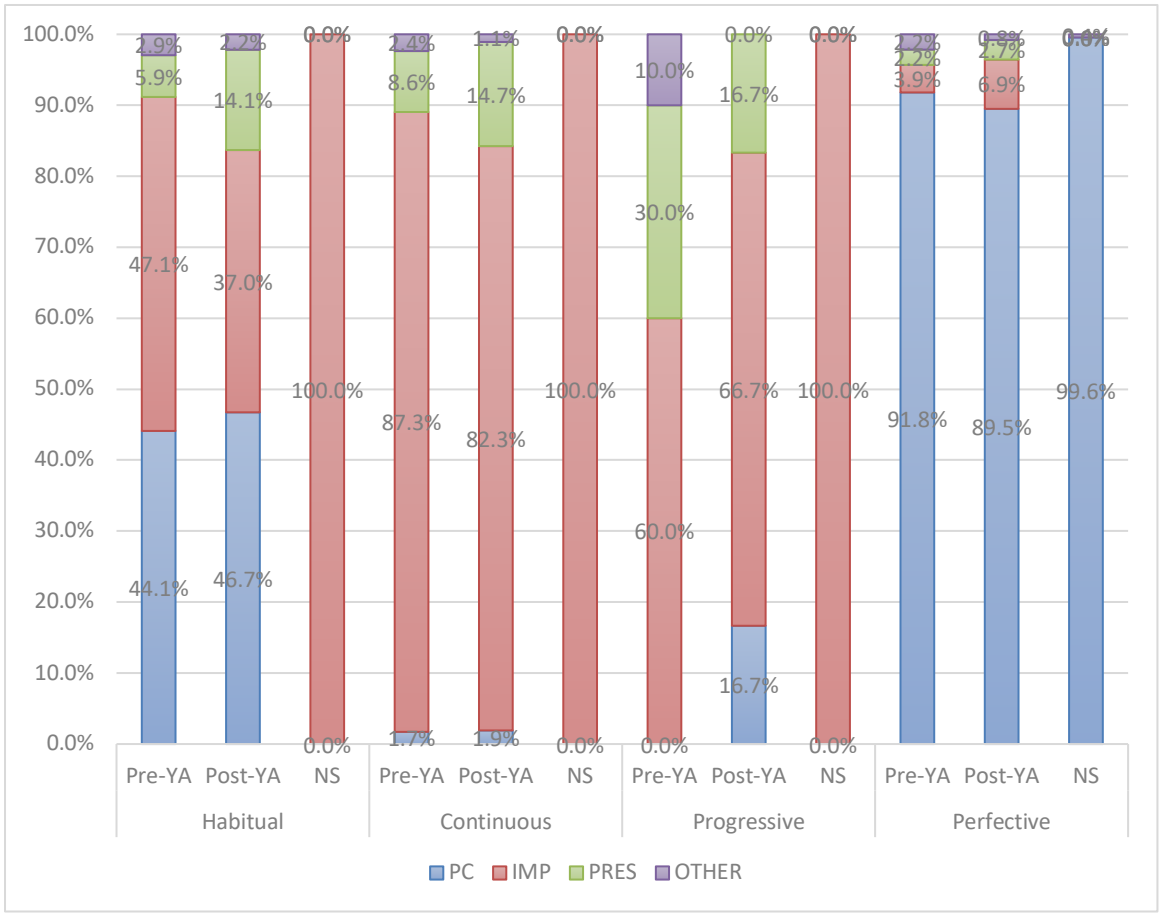


Figure 14: Proportion of forms used by aspectual condition by Pre- and Post-YA L2 groups and the L1 group (NS) in the Conversation task.

In order to establish a baseline for comparison, a series of one-way ANOVAs aimed to assess the effect of the independent variable Group on suppliance of the target form in each condition (Imperfect for the continuous, habitual and progressive conditions; *passé composé* for the perfective condition). The output of these analyses is

presented in Table 14. Again, the reduced numbers in each group for the progressive condition is due to the fact that not every participant produced progressive tokens in the Conversation task.

Aspectual condition	**Group**	**Mean**	**SD**	**Std. Err.**	**95% CI for mean**	**F**	**p-value**	**Effect size (ω^2)**
Imperfective (combined)	Pre-YA (n = 20)	82.5	12.5	2.80	[76.6, 88.3]	9.206	< .001 **	.247
	Post-YA (n =23)	75.6	15.4	3.21	[69.0, 82.3]			
	L1 (n = 7)	100.0	0.0	0.00	[100.0, 100.0]			
Continuous	Pre-YA (n = 20)	85.5	13.1	2.92	[79.4, 91.6]	5.567	.007 **	.154
	Post-YA (n =23)	81.1	14.9	3.10	[74.6, 87.5]			
	L1 (n = 7)	100.0	0.0	0.00	100.0, 100.0]			
Habitual	Pre-YA (n = 12)	41.9	43.4	12.5	[14.4, 69.5]	5.763	.007 **	.200
	Post-YA (n =22)	36.6	31.7	6.75	[22.6, 50.7]			
	L1 (n = 4)	100.0	0.0	0.00	[100.0, 100.0]			
Progressive	Pre-YA (n = 7)	71.4	48.8	18.44	[26.3, 116.6]	1.461	.257	.040
	Post-YA (n =10)	61.0	45.8	14.49	[28.2, 93.8]			
	L1 (n = 5)	100.0	0.0	0.00	[100.0, 100.0]			
Perfective	Pre-YA (n = 20)	91.4	7.7	1.72	[87.8, 95.0]	4.592	.015 *	.126
	Post-YA (n =23)	90.8	6.7	1.40	[87.9, 93.7]			
	L1 (n = 7)	99.4	1.7	0.64	[97.8, 100.9]			

Table 14: Output from ANOVA analyses exploring the effect of the independent variable Group on suppliance of the target form in the Conversation task.

As in Cat Story, for the Conversation task a main effect of Group is found on the suppliance of the target Imperfect form when the three imperfective conditions are combined (F = 9.206, p <.001, ω^2 = .247), as well as in habitual (F = 5.673, p= .007, ω^2 = .200) and continuous (F = 5.567, p= .007, ω^2 = .154) conditions. Post-hoc analyses confirm that this effect holds between the L1 group and both L2 groups (overall

imperfective: *p*(Pre-YA) = .012, 95% CI [-31.9, -3.13], *p*(Post-YA) <.001, 95% CI [-38.5, -10.2]; habitual: *p*(Pre-YA)= .019, 95% CI [-108.2, -7.9], *p*(Post-YA) = .005, 95% CI [-110.6, 16.2]; continuous: *p*(Pre-YA) = .047, 95% CI [-28.8, -0.2], *p*(Post-YA) = .005, 95% CI [-33.0, -4.9]). However, unlike Cat Story, no main effect of Group is seen on suppliance of the Imperfect (or indeed any other form) in the Conversation task for the progressive condition (*p* = .257). Additionally, the progressive condition of the Conversation task is the only instance where the Post-YA group are seen to supply the target form more than their Pre-YA counterparts, as shown in Figure 14. However, as will be discussed in the following section, the number of occurrences of the progressive condition is extremely low overall, meaning that this result (as well the unexpected 16.7% increase in *passé composé* usage that is unattributable to L1 influence) should arguably not be over-analysed.

In a manner that also differs from Cat Story, a significant effect of Group is also found for the perfective (*F* = 4.592, *p* = .015, ω^2 = .126), and is found to hold between both L2 groups and the L1 group (*p*(Pre-YA) = .031, 95% CI [-15.3, -0.58]; *p*(Post-YA) = .015, 95% CI [-15.8, -1.4]), as was the case for the other aspectual conditions (save the progressive). Given that the Conversation task provided a larger data sample in terms of number of tokens than the Cat Story, and is also more representative of natural conversational French, it is likely to present a more realistic depiction of the production of all three groups. Therefore, looking at the Conversation data and the results of the analyses in Table 14, it can be inferred that, although learners may show high targetlike suppliance in certain viewpoint aspectual conditions in this task (such as the perfective and continuous), they are generally not yet fully converging on nativelike production of viewpoint aspectual forms.

In order to further examine the supplementary exploratory angles initially formulated while analysing the Cat Story data, it was also necessary to assess whether Group significantly affected suppliance of the Present in the Conversation data. The results of these one-way ANOVA analyses are presented in Table 15. An additional direction of interest related to the unusual Present usage observed in the Cat Story data is the extent to which aspectual condition predicts this behaviour; consequently, a series of repeated-measures ANOVAs for each participant group were carried out to assess the effect of Target_Aspect on Present suppliance. The results of these analyses are presented in Table 16 (note that the analyses for the L1 group could not be carried

out due to the fact that no Present tokens were produced in any condition by this group).

Aspectual condition	**Group**	**Mean**	**SD**	**Std. Err.**	**95% CI for mean**	**F**	**p-value**	**Effect size (ω^2)**
Imperfective (combined)	Pre-YA (n = 20)	10.2	10.6	2.37	[5.3, 15.2]	5.063	.010 **	.140
	Post-YA (n =23)	15.6	13.6	2.83	[9.7, 21.4]			
	L1 (n = 7)	0.0	0.0	0.00	[0.0, 0.0]			
Continuous	Pre-YA (n = 20)	10.2	10.2	2.28	[5.4, 15.0]	5.235	.009 **	.145
	Post-YA (n =23)	15.9	14.0	2.91	[9.8, 21.9]			
	L1 (n = 7)	0.0	0.0	0.0	[0.0, 0.0]			
Habitual	Pre-YA (n = 12)	2.9	6.8	2.00	[-1.5, 7.2]	2.146	.132	.057
	Post-YA (n = 22)	12.3	18.8	4.00	[3.9, 20.6]			
	L1 (n = 4)	0.0	0.0	0.00	[0.0, 0.0]			
Progressive	Pre-YA (n = 7)	14.3	37.8	14.29	[-20.7, 49.2]	0.514	.606	-.046
	Post-YA (n =10)	17.0	33.4	10.55	[-6.9, 40.9]			
	L1 (n = 5)	0.0	0.0	0.00	[0.0, 0.0]			
Perfective	Pre-YA (n = 20)	2.2	4.1	0.91	[0.3, 4.1]	1.783	.179	.030
	Post-YA (n =23)	3.2	4.3	0.90	[1.3, 5.1]			
	L1 (n = 7)	0.0	0.0	0.00	[0.0, 0.0]			

Table 15: Output of ANOVA analyses exploring the effect of independent variable Group on suppliance of the Present in the Conversation task

Group	**Aspectual condition**	**Mean**	**SD**	**df (hypothesis)**	**df (error)**	**F**	**p-value**	**Effect size (η^2)**
Pre-YA (*n* = 20)	Continuous	5.7	9.0	3	3	0.729	.550	.127
	Habitual	7.0	7.5					
	Progressive	16.7	40.8					
	Perfective	1.9	3.2					
Post-YA (*n* = 23)	Continuous	15.9	22.5	3	7	0.902	.453	.091
	Habitual	16.1	16.7					
	Progressive	17.0	33.4					
	Perfective	3.2	2.8					
L1 (*n* = 7)	Continuous	0.0	0.0	-	-	-	-	.-
	Habitual	0.0	0.0					
	Progressive	0.0	0.0					
	Perfective	0.0	0.0					

Table 16: Output of repeated-measures ANOVA analyses exploring the effect of independent variable Target_Aspect on suppliance of the Present in the Conversation task

As shown in Table 15, a main effect of Group is also found in the Conversation data on learners' suppliance of the Present, this time in the continuous condition (F = 5.235, p= .009, ω^2 = .145) and in the combined suppliance for imperfective conditions (F = 5.063, p= .010, ω^2 = .140). Again, this effect was found to hold only between the control group and the Post-YA group (p(continuous) = .008, 95% CI [3.5, 28.3] p(imperfective) = .009, 95% CI [3.3, 27.9]). These findings pattern with those from Cat Story, further supporting the idea that the Post-YA group are using the Present in (certain) imperfective contexts in a manner that is significantly different to that of the Pre-YA group. However, the distribution of Present usage in the Conversation task is less-clear cut. Whilst the Cat Story showed a significant main effect of aspectual condition – specifically relating to the habitual – on Post-YA non-target Present suppliance, no such main effect is found for the Conversation data for either group (see Table 16). It can be concluded from this that, while non-targetlike Present usage is apparent in the Conversation task for the Post-YA learners and (among other things)

significantly differentiates them from the control group, it does not occur in any specific aspectual condition significantly more than others.

Finally, it is a point of interest from a Feature Reassembly perspective whether there is an effect of Group on suppliance of the *passé composé* in imperfective conditions – particularly the habitual and continuous, given that habituality and continuity are at least partly mapped to the perfective Simple Past in English. A series of one-way ANOVAs were carried out to examine this, with the outputs of these analyses found in Table 17. To explore this question from an alternative perspective, a series of repeated-measures ANOVAs were used to analyse the effect of aspectual condition on suppliance of the target Imperfect form by each group in imperfective contexts; the results of these analyses are presented in Table 18. Note that the full statistical test in Table 18 could not be reported for the L1 group, due to the low prevalence of both habitual and progressive tokens within this already small group; nonetheless, it can be seen that the L1 group performed at ceiling in all conditions.

Aspectual condition	**Group**	**Mean**	**SD**	**Std. Err.**	**95% CI for mean**	**F**	**p-value**	**Effect size (ω^2)**
Imperfective (combined)	Pre-YA (n = 20)	5.0	5.1	1.13	[2.7, 7.4]	6.691	.003 **	.185
	Post-YA (n =23)	7.6	5.3	1.11	[5.3, 9.9]			
	L1 (n = 7)	0.0	0.0	0.00	[0.0, 0.0]			
Continuous	Pre-YA (n = 20)	2.1	3.6	0.82	[0.4, 3.9]	1.395	.258	.016
	Post-YA (n =23)	1.9	2.8	0.57	[0.7, 3.2]			
	L1 (n = 7)	0.0	0.0	0.00	[0.0, 0.0]			
Habitual	Pre-YA (n = 12)	53.6	45.7	13.18	[24.5, 82.6]	4.119	.025 *	.141
	Post-YA (n =22)	49.7	28.6	6.10	[37.0, 62.4]			
	L1 (n = 4)	0.0	0.0	0.00	[0.0, 0.0]			
Progressive	Pre-YA (n = 7)	0.0	0.0	0.00	[0.0, 0.0]	1.612	.226	.053
	Post-YA (n =10)	22.0	41.6	13.15	[-7.7, 51.7]			
	L1 (n = 5)	0.0	0.0	0.00	[0.0, 0.0]			

Table 17: Output from ANOVA analyses exploring the effect of the independent Group variable on suppliance of the passé composé in imperfective contexts in the Conversation task

Group	**Aspectual condition**	**Mean**	**SD**	**df (hypothesis)**	**df (error)**	**F**	**p-value**	**Effect size (η^2)**
Pre-YA (*n* = 20)	Continuous	89.4	11.5	2	4	3.493	.071	.411
	Habitual	39.4	40.0					
	Progressive	66.7	51.6					
Post-YA (*n* = 23)	Continuous	80.3	17.8	2	8	4.714	.023 *	.344
	Habitual	45.5	34.3					
	Progressive	61.0	45.8					
L1 (*n* = 4)	Continuous	100.0	0.0	-	-	-	-	-
	Habitual	100.0	0.0	-	-			
	Progressive	100.0	0.0	-	-			
	Perfective	100.0	0.0	-	-			

Table 18: Output from repeated-measures ANOVA analyses exploring the effect of aspectual condition on suppliance of the target Imperfect form in the Conversation task

Whereas the results from Cat Story did not show a significant main effect of Group on use of the *passé composé* (which may be indicative of enduring L1 influence)in either the habitual or continuous condition, the analyses of the Conversation data in Table 17 do reveal a main effect of Group for the habitual condition (F = 4.119, p = .025, ω^2 = .141), which is significant between the L1 group and both L2 groups (p(Pre-YA) = .029, 95% CI [4.4, 102.7], p(Post-YA) = .032, 95% CI [3.4, 96.0]). There is also a significant main effect of Group found for *passé composé* suppliance in the overall imperfective (F = 6.691, p = .003, ω^2 = .185), which was found to hold only between the Post-YA group and the L1 group (p = .002, 95% CI [2.4, 12.8]. Taken together, these findings suggest that both L2 groups produce the non-target *passé composé* in imperfective conditions significantly more than the L1 group: this is particularly the case for the habitual condition, which suggests an effect of L1 influence as predicted by Feature Reassembly, but the Post-YA group also appear to be affected more broadly across the imperfective conditions.

Moreover, the results of the repeated-measures ANOVA analyses in Table 18 demonstrate that the Post-YA group were significantly affected by aspectual condition

with regard to supplying the Imperfect in imperfective conditions (F = 4.714, p = .023, η^2 = .344**),** whilst the Pre-YA group were not (p = .071). Post-hoc tests showed that the significant effect was found specifically between the Post-YA group's suppliance of the Imperfect in habitual vs. continuous conditions (p=.0074, 95% CI [12.9, 56.9]). Considering the group means, it can be inferred that, despite the continuous entailing a more challenging reassembly task, the Post-YA group supply the target Imperfect significantly less in habitual over continuous conditions.

In summary, the analyses of forms used by the L2 groups reveals that, while there is no statistically significant difference in how much each learner group supplies the target Imperfect form in the imperfective conditions (the post-hoc testing carried out following the analyses in Table 14 found significant differences only between each L2 group vs. the L1 group), the Post-YA group is nonetheless significantly different in its use of the Present in the Conversation data, as shown in Table 15 and the subsequent discussion. Adding this observation to the similar findings from the Cat Story data, there is growing support for the fact that the Post-YA group use the Present in a manner that significantly differentiates them from both the Pre-YA and L1 groups. This non-target Present usage is particularly apparent in imperfective conditions – specifically the habitual and (to a slightly lesser extent) continuous - and in the Cat Story task. In addition, the analyses presented in Table 17 (and subsequent post-hoc testing) demonstrate a persistent L1 influence for both L2 groups for the habitual, as evidenced by their significant use of the *passé composé* in this condition – although, interestingly, no such influence is seen for the continuous.

What do the ensemble of these findings mean in relation to both the primary research aims of this project, as well as the additional exploratory avenues of investigation set out in this section? Returning first to RQ1, there is once again a mixed picture with regard to Feature Reassembly – though the prevalence of the *passé composé* used by learners in the habitual condition aligns neatly with what the approach would predict, it does not explain why similar results are not seen for the continuous – where, if anything, one might expect to see an even greater prevalence of non-target *passé composé* due to the fact that continuity is mapped solely to the Simple Past in English. In addition, Feature Reassembly cannot account for the striking use of the Present – in contexts where past time reference is required – by the Post-YA group. With reassembly ruled out as a source of this unexpected behaviour, this lends support

to the idea that learners may be utilising the Present in place of the correct aspectual mapping for a different purpose. Given that this behaviour is only observed to a significant extent in the Post-YA group – who have been shown to be otherwise comparable to the Pre-YA group in terms of overall French proficiency – this poses the question of whether learners' time spent abroad has impacted their aspectual production in some way. This is a key idea, related to RQ2 and the role played by programmes such as the year abroad, which will be returned to in detail in Chapter 6.

To close this section, a summary of the distribution of forms produced in the production data as a whole is presented in Figure 15.

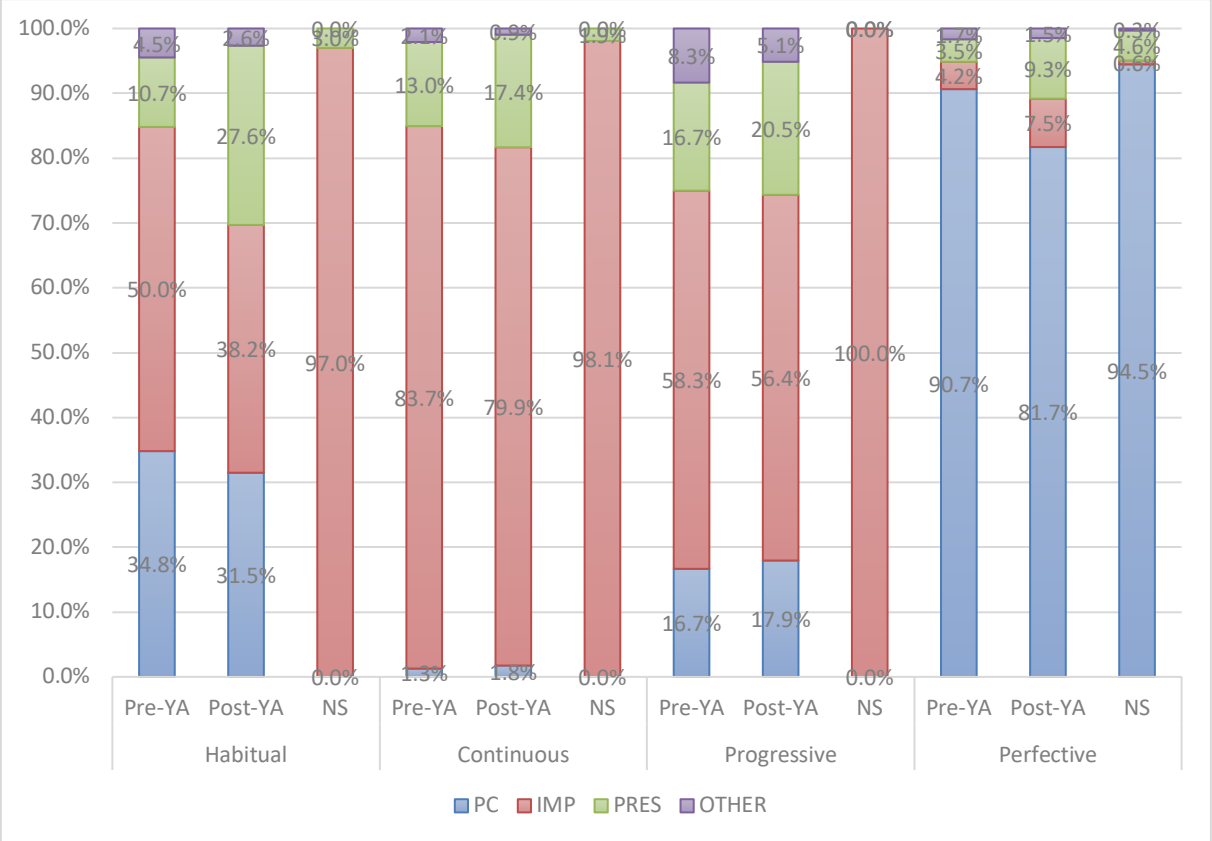


Figure 15: Proportion of forms used by aspectual condition by Pre- and Post-YA learners and the L1 control group (NS) in overall production (Cat Story + Conversation).

Considering the trends shown in Figure 15, it can be seen that for the L2 groups, suppliance of the target form follows the same cline of Habitual < Progressive < Continuous < Perfective as that observed in Wallington (2017), where it was suggested that learners' differential development across French viewpoint aspectual form-

meaning mappings may be influenced by both feature reassembly and by the frequency of each mapping in the input learners were exposed to. This idea has been developed for this project into RQ3, which aims specifically to explore the extent to which an approach combining information relating to feature reassembly with information from the input can account for L2 (aspectual) development, particularly at more advanced levels. Of course, looking at learner production alone is not sufficient to be fully informative on the input that learners have received, and so the following section will analyse the distribution of viewpoint aspectual mappings in two sources of L1 French input: the production data presented in this section, and a sample of corpus data.

5.2 Frequency-distributional analysis of viewpoint aspectual form-meaning mappings in L2 and L1 French

This section will present frequency data for each of the form-meaning mappings in the French viewpoint aspectual system. This information has been obtained not only from the L2 and L1 production data presented in the previous section, but also from a *n* = 10,641-word sample of an L1 French oral corpus, the CFPP2000, in order to provide a more comprehensive perspective on the distribution of viewpoint aspectual mappings in L1 French. This section will first present the corpus data, and then will move on to comparing this with the production data, considering any similarities and differences in the relative distributions of aspectual form-meaning mappings in each dataset.

5.2.1 Presentation of L1 French corpus data

In analysing the corpus data, the aim was two-fold: firstly, to establish the frequency of occurrence of viewpoint aspect-expressing forms (i.e. the Imperfect and the *passé composé*) as a whole; and secondly, to analyse the distribution of the four form-meaning mappings (perfective and imperfective-habitual, -continuous, and -progressive) across the ensemble of the viewpoint aspect tokens. The intention in doing this was to allow an investigation of whether the differences in production accuracy/targetlike suppliance seen in the L2 data were in any way reflected in similar frequency-distributional patterns in the L1 input (as represented by the corpus data).

With regard to this first point of enquiry, the overall composition of verbal forms within the corpus sample should be examined. Of the 10,641-word sample, a

total of 1145 verbs were identified, meaning that verbs (of all types) constituted 10.8% of the sample. Of these 1145 verb tokens, 400 (or 34.9% of the verbs) expressed viewpoint aspect. Taken together, this means that 3.8% of the sample consisted of viewpoint aspect-expressing tokens. This alone cannot necessarily inform on whether there is substantial evidence for viewpoint aspect in L1 French input; indeed, establishing what counts as "substantial evidence" is in itself a challenging methodological and theoretical issue. That said, it is perhaps enlightening to compare the frequency of viewpoint aspectual tokens in the sample with the frequency of another verb form, such as the present tense. The latter was found to comprise 478 of the 1145 verb tokens in the sample (or 41.8% of verbs), constituting 4.49% of the corpus data overall. These exploratory findings suggest that viewpoint aspect is less frequently represented in the input, especially when compared to forms such as the Present, which occurs around 1.2 times more often. This observation is even more noteworthy in view of the fact that French viewpoint aspect (as defined in this project) encompasses four separate form-meaning mappings. If viewpoint aspect as a whole is infrequent when compared to forms such as the Present, any specific aspectual form-meaning mapping will necessarily be more infrequent still. Figure 16 displays the distribution of the four viewpoint aspectual form-meanings as a proportion of the total number of viewpoint aspect tokens in the corpus sample.

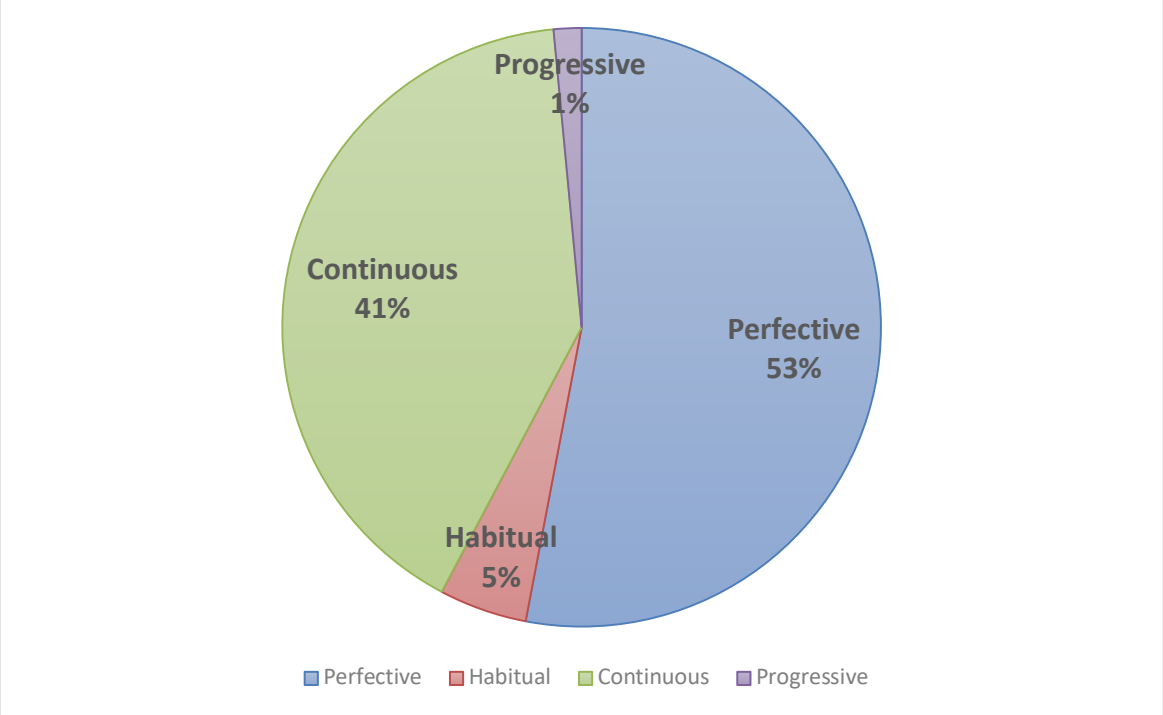


Figure 16: Frequency distribution of viewpoint aspectual mappings in L1 French corpus data, out of a total number of n = 400 viewpoint aspect-expressing tokens.

It can be seen that, of the four mappings, the perfective is decidedly the most frequent, constituting over half of all the tokens. The imperfective-continuous is also frequent, at over 40% of tokens; collectively, then, the perfective and continuous alone account for 94% of the viewpoint aspect tokens in the sample. This establishes a clear frequency divide between these two mappings and both the imperfective-habitual and -progressive, which make up 5% and 1% of the total viewpoint aspect tokens respectively. Viewed from the perspective of the entire sample, habitual tokens constitute 0.18% and progressive tokens just 0.06%, whilst the *passé composé* constitutes 1.99%, making it three times as frequently-occurring as the habitual and over thirty-three times as frequently-occurring as the progressive overall. These simple calculations go some way in illustrating the scale of the differing frequencies with which these four form-meaning mappings naturally occur in the L1 input. The potential relationship of these frequency differences to L2 viewpoint aspectual production will be examined in the following section.

5.2.2 A frequency-distributional perspective on L2 vs. L1 French viewpoint aspectual production

To summarise the findings of the previous section, a large disparity was observed between the frequencies in the L1 corpus data of the perfective and continuous mappings on one hand, and the much less frequent habitual and progressive mappings on the other. Are these differential distributions also visible in the production data presented earlier in this chapter?

Tables 19 and 20 show the respective frequencies of each viewpoint aspectual condition in L2 production in Cat Story (divided into the Pre- and Post-YA groups), along with the same information from the L1 Cat Story and L1 corpus data. Note that this project distinguishes between *contexts* (when a given aspectual mapping "should" have been produced, based on the surrounding context; Table 19) and *tokens* (when the appropriate form was actually supplied in that context; Table 20). As each experimental group contained a different number of participants, the number of contexts is given first in raw form and then as an average per participant (the average for the corpus data is given per file). Finally, "%VA" expresses each mapping as a percentage of the total number of viewpoint aspectual contexts/tokens produced by that group. Finally, it

should be noted that the corpus data column presented in each of Figures 19-22 is always the same data.

	CAT STORY: VIEWPOINT ASPECT CONTEXTS (in descending order of frequency)			
	L2 French production: Pre-YA	**L2 French production: Post-YA**	**L1 French production**	**L1 French corpus data**
	Imperf. (Overall) **Contexts:** 363 **Avg:** 18.2 **% VA**: 53.4%	Imperf. (Overall) **Contexts:** 424 **Avg:** 18.4 **% VA:** 56.0%	Imperf. (Overall) **Contexts:** 127 **Avg:** 18.2 **% VA:** 56.2%	Perfective **Contexts:** 212 **Avg:** 21.2 **%VA:** 53.0%
	Perfective **Contexts:** 317 **Avg:** 15.9 **% VA***:* 46.6%	Perfective **Contexts:** 333 **Avg:** 14.5 **% VA:** 44.0%	Perfective **Contexts:** 99 14.2 **% VA:** 43.8%	Imperf. (Overall) **Contexts:** 188 **Avg:** 18.8 **% VA:** 47.0%
	Habitual **Contexts:** 190 **Avg:** 9.5 **% VA***:* 27.9%	Habitual **Contexts:** 232 **Avg:** 10.1 **% VA:** 30.6%	Habitual **Contexts:** 59 **Avg:** 8.4 **% VA:** 26.1%	Continuous **Contexts:** 163 **Avg:** 16.3 **% VA:** 40.8%
	Continuous **Contexts:** 159 **Avg:** 8.0 **% VA***:* 23.4%	Continuous **Contexts:** 171 **Avg:** 7.4 **% VA:** 22.6%	Continuous **Contexts:** 55 **Avg:** 7.9 **% VA:** 24.3%	Habitual **Contexts:** 19 **Avg:** 1.9 **% VA:** 4.8%
	Progressive **Contexts:** 14 **Avg:** 0.7 **% VA**: 2.1%	Progressive **Contexts:** 21 **Avg:** 0.9 **% VA:** 2.8%	Progressive **Contexts:** 13 **Avg:** 1.9 **% VA:** 5.8%	Progressive **Contexts:** 6 **Avg:** 0.6 **% VA:** 1.5%
Total VA contexts	680	757	226	400
Mean contexts /person	34.0 (*n* = 20)	32.9 (*n* = 23)	32.3 (*n* = 7)	40.0 (per file, *n* = 10)
Total words /group	7552	7379	2213	10641
Mean file length (words)	337.6	320.8	316.1	1061.4

Table 19: Frequencies of viewpoint aspect contexts in the Cat Story task compared with L1 French corpus data, presented in descending order of frequency.

	CAT STORY: VIEWPOINT ASPECT TOKENS (in descending order of frequency)			
	L2 French production: Pre-YA	**L2 French production: Post-YA**	**L1 French production**	**L1 French corpus data**
	Perfective **Tokens:** 282 **Avg:** 14.1 **% VA**: 56.2%	Perfective **Tokens:** 224 **Avg:** 9.7 **% VA:** 51.7%	Imperf. (Overall) **Tokens:** 120 **Avg:** 17.1 **% VA:** 59.4%	Perfective **Tokens:** 212 **Avg:** 21.2 **%VA:** 53.0%
	Imperf. (Overall) **Tokens:** 220 **Avg:** 11.0 **% VA***:* 43.8%	Imperf. (Overall) **Tokens:** 209 **Avg:** 9.1 **% VA:** 48.3%	Perfective **Tokens:** 82 **Avg:** 11.7 **% VA:** 40.6%	Imperf. (Overall) **Tokens:** 188 **Avg:** 18.8 **% VA:** 47.0%
	Continuous **Tokens:** 116 **Avg:** 5.8 **% VA***:* 23.1%	Continuous **Tokens:** 108 **Avg:** 4.7 **% VA:** 24.9%	Habitual **Tokens:** 57 **Avg:** 8.1 **% VA:** 28.2%	Continuous **Tokens:** 163 **Avg:** 16.3 **% VA:** 40.8%
	Habitual **Tokens:** 96 **Avg:** 4.8 **% VA***:* 19.1%	Habitual **Tokens:** 91 **Avg:** 4.0 **% VA:** 21.0%	Continuous **Tokens:** 50 **Avg:** 7.1 **% VA:** 24.8%	Habitual **Tokens:** 19 **Avg:** 1.9 **% VA:** 4.8%
	Progressive **Tokens:** 8 **Avg:** 0.4 **% VA**: 1.6%	Progressive **Tokens:** 10 **Avg:** 0.4 **% VA:** 2.3%	Progressive **Tokens:** 13 **Avg:** 1.9 **% VA:** 6.4%	Progressive **Tokens:** 6 **Avg:** 0.6 **% VA:** 1.5%
Total VA tokens	502	433	202	400
Mean tokens /person	25.1 (*n* = 20)	18.8 (*n* = 23)	32.3 (*n* = 7)	40.0 (per file, *n* = 10)
Total words /group	7552	7379	2213	10641
Mean file length (words)	337.6	320.8	316.1	1061.4

Table 20: Frequencies of viewpoint aspect tokens in the Cat Story task compared with L1 French corpus data, presented in descending order of frequency.

Beginning with the production data, it can be seen that in terms of contexts (Table 19), both L2 groups and the L1 control group show the same frequency cline: the perfective is the most frequent individual mapping, comprising around 45% of contexts for each group in approximate accordance with its frequency in the corpus data. The production data diverges from the frequency cline of the corpus data with regard to imperfective mappings, however. Whilst the progressive is the least frequent mapping in both, the habitual is the most frequent imperfective mapping in the production data, constituting 26.1-30.6% of contexts, whilst in the corpus data the continuous is the most frequent imperfective mapping and occurs over eight times as frequently as the habitual. This contrast between the production and corpus data highlights the extent to which a more structured production task such as Cat Story can manipulate the proportions of viewpoint aspectual mappings that are elicited.

It is also important to consider how closely each of the production groups meet the contextual requirements of Cat Story, i.e. how closely the frequencies of tokens of the target forms match with the contexts requiring those forms. Looking at Table 20, it is observable that, fairly unsurprisingly, the learner groups produce more perfective tokens (i.e. uses of the *passé composé* in perfective contexts) than imperfective tokens (i.e. uses of the Imperfect in habitual, continuous, and progressive contexts). This aligns with the higher accuracy seen in perfective over imperfective contexts. It is to be noted within the imperfective mappings that, although the habitual is the most frequent imperfective context in Cat Story, it is not the most frequent type of imperfective token: learners produce more continuous tokens than habitual ones. Crucially, the higher incidence of continuous tokens in learner production mimics the corpus data, where the continuous is by a large margin the most frequent imperfective mapping. Moreover, the high frequency of continuous tokens in L2 production aligns with the high accuracy for this mapping relative to the other imperfective mappings that was observed at the start of the chapter. To further explore this potential link between accuracy and frequency, the data from the Conversation task (Tables 21 and 22), whose structure more closely resembles that of the corpus data, will be considered next.

	CONVERSATION: VIEWPOINT ASPECT CONTEXTS (in descending order of frequency)			
	L2 French production: Pre-YA	**L2 French production: Post-YA**	**L1 French production**	**L1 French corpus data**
	Imperf. (Overall) **Contexts:** 510 **Avg:** 25.5 **% VA**: 52.4%	Imperf. (Overall) **Contexts:** 1402 **Avg:** 61.0 **% VA:** 69.4%	= Imperf. (Overall) **Contexts:** 227 **Avg:** 32.4 **% VA:** 50.0%	Perfective **Contexts:** 212 **Avg:** 21.2 **%VA:** 53.0%
	Continuous **Contexts:** 466 **Avg:** 23.3 **% VA***:* 47.8%	Continuous **Contexts:** 1200 **Avg:** 52.2 **% VA:** 59.4%	= Perfective **Contexts:** 227 **Avg:** 32.4 **% VA:** 50.0%	Imperf. (Overall) **Contexts:** 188 **Avg:** 18.8 **% VA:** 47.0%
	Perfective **Contexts:** 464 **Avg:** 23.2 **% VA***:* 47.6%	Perfective **Contexts:** 619 **Avg:** 26.9 **% VA:** 30.6%	Continuous **Contexts:** 205 **Avg:** 29.3 **% VA:** 45.2%	Continuous **Contexts:** 163 **Avg:** 16.3 **% VA:** 40.8%
	Habitual **Contexts:** 34 **Avg:** 1.7 **% VA***:* 3.5%	Habitual **Contexts:** 184 **Avg:** 8.0 **% VA:** 9.1%	Progressive **Contexts:** 14 **Avg:** 2.0 **% VA:** 3.1%	Habitual **Contexts:** 19 **Avg:** 1.9 **% VA:** 4.8%
	Progressive **Contexts:** 10 **Avg:** 0.5 **% VA**: 1.0%	Progressive **Contexts:** 18 **Avg:** 0.8 **% VA:** 0.9%	Habitual **Contexts:** 8 **Avg:** 1.1 **% VA: 1**.8%	Progressive **Contexts:** 6 **Avg:** 0.6 **% VA:** 1.5%
Total VA contexts	974	2021	454	400
Mean contexts/person	48.7 (*n* = 20)	87.9 (*n* = 23)	64.9 (*n* = 7)	40.0 (per file, *n* = 10)
Total words/group	38592	48732	19827	10641
Mean file length (words)	1872.6	2114.9	2832.4	1064.1

Table 21: Frequencies of viewpoint aspect contexts in the Conversation task compared with L1 French corpus data, presented in descending order of frequency.

	CONVERSATION: VIEWPOINT ASPECT TOKENS (in descending order of frequency)			
	L2 French production: Pre-YA	**L2 French production: Post-YA**	**L1 French production**	**L1 French corpus data**
	Imperf. (Overall) **Tokens:** 429 **Avg:** 21.5 **% VA**: 50.2%	Imperf. (Overall) **Tokens:** 1068 **Avg:** 46.4 **% VA:** 65.8%	Imperf. (Overall) **Tokens:** 227 **Avg:** 32.4 **% VA:** 50.0%	Perfective **Tokens:** 212 **Avg:** 21.2 **%VA:** 53.0%
	Perfective **Tokens:** 426 **Avg:** 21.3 **% VA**: 49.8%	Continuous **Tokens:** 988 **Avg:** 43.0 **% VA:** 60.9%	Perfective **Tokens:** 226 **Avg:** 32.3 **% VA:** 50.0%	Imperf. (Overall) **Tokens:** 188 **Avg:** 18.8 **% VA:** 47.0%
	Continuous **Tokens:** 407 **Avg:** 20.4 **% VA**: 47.6%	Perfective **Tokens:** 554 **Avg:** 24.1 **% VA:** 34.2%	Continuous **Tokens:** 205 **Avg:** 29.3 **% VA:** 45.2%	Continuous **Tokens:** 163 **Avg:** 16.3 **% VA:** 40.8%
	Habitual **Tokens:** 16 **Avg:** 0.8 **% VA**: 1.9%	Habitual **Tokens:** 68 **Avg:** 3.0 **% VA:** 4.2%	Progressive **Tokens:** 14 **Avg:** 2.0 **% VA:** 3.1%	Habitual **Tokens:** 19 **Avg:** 1.9 **% VA:** 4.8%
	Progressive **Tokens:** 6 **Avg:** 0.3 **% VA**: 0.7%	Progressive **Tokens:** 12 **Avg:** 0.5 **% VA:** 0.7%	Habitual **Tokens:** 8 **Avg**: 1.1 **% VA:** 1.8%	Progressive **Tokens:** 6 **Avg:** 0.6 **% VA:** 1.5%
Total VA contexts	855	1622	453	400
Mean tokens /person	42.8 (*n* = 20)	70.5 (*n* = 23)	64.7 (*n* = 7)	40.0 (per file, *n* = 10)
Total words/group	38592	48732	19827	10641
Mean file length (words)	1872.6	211.9	2832.4	1064.1

Table 22: Frequencies of viewpoint aspect tokens in the Conversation task compared with L1 French corpus data, presented in descending order of frequency.

As predicted, some different distributional patterns across all production groups can be seen in the Conversation task. Compared to in Cat Story, the L1 production group converge more in this task to the frequency cline of the corpus data – that is, close-to-equal proportions of perfective and (overall) imperfective contexts, a high frequency of continuous contexts, and markedly lower and relatively equal numbers of progressive and habitual contexts. This indicates that the Conversation task does indeed fairly closely resemble the format of the corpus data, i.e. natural conversational French. It is also apparent from the Conversation data that when the habitual mapping is not specifically targeted (as it was in Cat Story), it is not naturally very frequent, and patterns more closely with the progressive.

As for the L2 groups, Table 21 shows that, while both groups display the same frequency cline of contexts as each other, there are some interesting differences regarding the relative proportions with which each mapping occurs. Whereas the Pre-

YA group produce perfective and continuous contexts in fairly equal numbers (around 47% of total viewpoint aspect contexts each, similarly to the L1 production proportions), the Post-YA group produce a very high incidence of continuous contexts: that is over twice as many per participant as the Pre-YA group, constituting 60% of their total viewpoint aspect contexts. This contrast is upheld when the tokens (i.e. actual production) of the groups are considered (Table 22). This spike of continuous contexts/tokens represents quite a divergence from the distributional patterns of the native data by the Post-YA group.

When the trends in frequency and distribution explored in the above section are viewed in combination with the trends in production accuracy across viewpoint aspectual conditions, some potential links begin to appear, including some evidence that appears to align with the accuracy discontinuity separating the habitual and progressive from the continuous and perfective for L2 learners of French. Specifically, the high frequency of the continuous when compared with the other two imperfective mappings appears to be reflected not only in the relative proportions of imperfective tokens produced by learners (sometimes to a disproportionate extent), but also in their production accuracy. There is also support for the idea that task type can manipulate the relative proportions of aspectual mappings produced – particularly in a more structured task such as Cat Story – whereas in less structured tasks such as Conversation, a distributional pattern which is more closely aligned to "natural" French emerges. This highlights the importance of triangulating frequency-distributional data across a range of tasks with different degrees of structuredness, as has previously been advocated for aspectual research (Domínguez 2019). However, to be able to speak more decisively on whether factors such as frequency have actually impacted learners' underlying aspectual representations, it is necessary to look at comprehension data. This will be presented in the following section.

5.3 Comprehension Data

5.3.1 About the task

To recap, the comprehension component of the experimental design was in the form of an acceptability judgement task (AJT), which tested participants' understanding of the four viewpoint aspectual form-meaning mappings. This was intended to capture

a more exact picture of participants' underlying aspectual competence, which cannot be fully commented on from production data alone due to the added cognitive demands of "on-line tasks" e.g. producing spontaneous speech in an L2. Contrastingly, for the AJT, participants were free to take as much time as they needed to complete the task and access their "off-line" knowledge. The AJT required participants to read a context providing background information (in English for the L2 groups and in French for the control group), and then to rate from 1-5 a French sentence that followed it, depending on how appropriate they felt the sentence was for the context (with 5 being the most appropriate). The AJT contained 21 contexts, each of which was presented twice, each time with a sentence to rate that differed only in the form of the verb (Imperfect vs. *passé composé*). The AJT was therefore composed of 42 items in total, presented in a randomised order and spanning perfective and imperfective-habitual, -continuous, and -progressive meanings.

5.3.2 Rating the appropriateness of viewpoint aspectual forms in context

Overview

The analysis of the AJT data will first focus on how participants rated the two viewpoint aspectual forms of interest (the Imperfect and the *passé composé*) across the four aspectual conditions of interest (habitual, continuous, progressive, and perfective). This permits more decisive commentary on learners' underlying representations of the French aspectual system, based on the extent to which they accept the target form in a given condition (indicated by higher ratings), along with the extent to which they reject the target form (indicated by lower ratings). A key focus is on whether or not learners can appropriately distinguish when to use the Imperfect and when to use the *passé composé* with regard to each aspectual condition – that is, to what extent they have successfully mapped a given meaning to the appropriate form.

Table 23 summarises the mean ratings on the AJT task, divided by group and aspectual condition and subcategorised according to the form being rated. The highlighted columns represent the target forms for the aspectual condition, where targetlike ratings would be expected to be > 3, and the unshaded columns are the non-target forms for the condition, where targetlike ratings would be expected to be < 3.

	HABITUAL		CONTINUOUS		PROGRESSIVE		PERFECTIVE	
Form->	IMP	PC	IMP	PC	IMP	PC	IMP	PC
GROUP	**Mean rating (on a scale of 1-5)**							
Pre-YA	4.18 [1.17]	2.42 [1.45]	3.68 [1.44]	2.95 [1.53]	4.22 [1.18]	2.20 [1.42]	2.79 [1.53]	4.04 [1.23]
Post-YA	4.00 [1.31]	2.44 [1.46]	3.73 [1.44]	2.85 [1.56]	4.10 [1.37]	2.17 [1.36]	2.78 [1.59]	4.01 [1.33]
L1	3.22 [1.74]	2.19 [1.45]	3.10 [1.68]	1.92 [1.29]	4.10 [1.35]	1.97 [1.36]	2.53 [1.69]	3.86 [1.55]

Table 23: Mean ratings (and [SDs]) on the AJT task by group, aspectual condition and form rated. The rating scale is 1-5, with 5 being the most appropriate.

Looking at the data in Table 23, it can be seen that the differences in average rating per condition for the two L2 groups are quite minimal, ranging from a difference of 0.03-0.18 for target forms and within an even narrower range of 0.01-0.1 for non-target forms. This tentatively suggests that there is not a major between-groups difference in learners' comprehension of aspectual form-meaning mappings, in somewhat of a contrast from the production data, where there some larger (albeit statistically non-significant) differences between the L2 groups.

Additionally, it can be seen that the L1 control group near-consistently rate lower on average than the L2 groups, for both target and non-target forms. This is not necessarily a cause for concern: as noted by Dillon & Wagers (2019:2), quantitative differences in ratings provided on a Likert scale (as in this AJT) 'do not have any inherent meaning: they are filtered through an additional participant's interpretation of the response scale.' With this in mind, the main question of importance when analysing the AJT data is whether each group differentiated between the target and non-target form for each condition in their ratings. A graphical depiction of each group's ratings of the target vs. non-target form is presented in Figure 17.

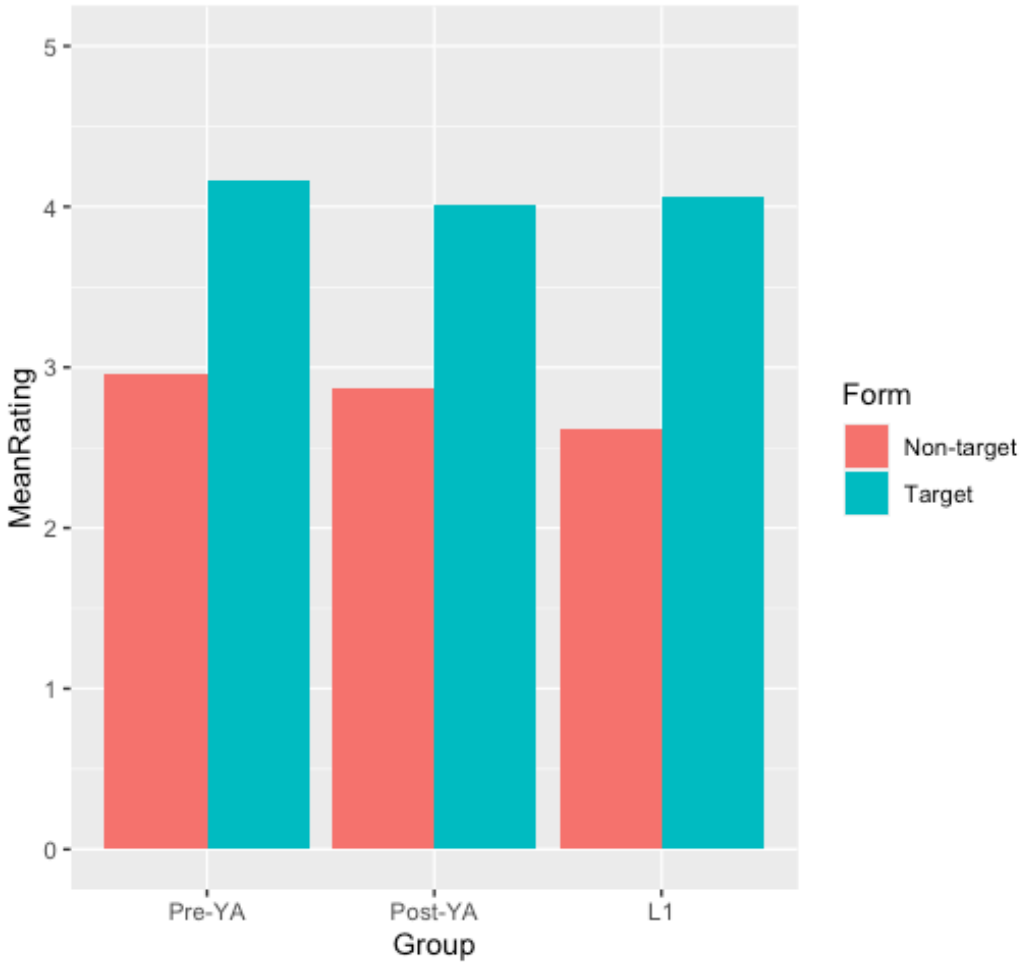


Figure 17: Graph of mean AJT ratings of target and non-target forms per group.

Considering the overall picture depicted in Figure 17, all three groups do indeed appear to differentiate between the target and non-target form in terms of mean ratings across all aspectual conditions. The L1 group display the largest difference in mean rating between non-target and target forms, whereas both L2 groups show a similar and smaller difference in rating. The relationship between group, form, and rating will be explored further via ordinal regression analyses later in this section.

Nonetheless, the lower ratings from the L1 group remain a point of interest: from a methodological standpoint, it is arguably worth exploring the impact of certain variables, both related to the participants and to the task itself, and how these may have predicted rating behaviour in a manner that is external to the actual aspectual contrast in question. Some discussion of these variables will therefore also be found in

this section, with a view to highlighting some of the methodological considerations that should be undertaken during AJT design and data collection.

Rating Modelling

The principal questions that the rating modelling aimed to answer were whether each group rated the target form significantly differently from the non-target form, and whether this varied according to the aspectual condition of the AJT item. This feeds directly back to RQ1, as it aims to broaden understanding of learners' developing viewpoint aspectual systems and how this interfaces with feature reassembly. An additional key question of interest is whether or not different groups behaved differently in terms of rating: in particular, if there were differences between the L2 groups, this may provide insight into the role played by the year abroad on viewpoint aspectual development, falling under RQ2. The inclusion of additional 'background' variables (that will be introduced shortly) were more exploratory in nature, and aimed to elucidate the potential roles played by these additional factors on participants' rating behaviour, which in turn may be informative of patterns in the results that are *not* attributable to level of aspectual development.

As discussed in the previous section, participants responded to each item on the AJT by giving each sentence presented a rating from 1-5, with '1' corresponding to the lowest level of appropriateness for the context, and '5' corresponding to the highest level of appropriateness. *Rating* is therefore treated as an ordinal variable, and the type of modelling undertaken was ordinal regression via cumulative link mixed modelling, carried out using the *ordinal* package (Christensen 2019), version 2019.12-10 in R.

In addition to the key variables of *Group* (Pre-YA/Post-YA/L1), *Target_Aspect* (habitual/continuous/progressive/perfective), and *Form* (target/non-target) - necessary in order to explore each group's rating of target and non-target forms across the four aspectual conditions – some additional variables were also included in the rating analysis. Given the large number of studies that have previously linked lexical and viewpoint aspect (see Section 3.2.4), the lexical aspect of the verb in the rated sentence (variable name: *Lexical_Aspect,* values: Eventive, Stative) was recorded, in order to explore whether this influenced participants' comprehension of the viewpoint aspectual mappings of interest.

The other two variables explored were related to the structure of the AJT itself. Whilst the AJT in this project presented every context twice – once followed by a sentence containing the Imperfect, and once with a sentence containing the *passé composé*, the SPLLOC project (Mitchell et al 2008; splloc.soton.ac.uk) from which the AJT items for this project were originally sourced presented each context only once, showing the two contrasting sentences simultaneously. It is possible that the decision to separate the two sentences for each context may have influenced how participants responded: as a result, the variable *Seen_First* was used during modelling. This is a binary variable which indicates whether a given sentence was seen first or second in its "context pair". (Recall that the order of questions in the AJT was randomised, so this will have been different for every participant). The final variable, labelled *Semantic_connectedness*, is another binary variable that expresses whether or not the background context of a given AJT item was considered to be semantically congruent with the sentence being rated. The congruence coding was added post-hoc, and was decided on with the assistance of an L1 French speaker during a walkthrough of the AJT. The below items contrast a context with a congruence rating of 0 (= not considered semantically well-connected to the sentence to rate) and a context with a congruence rating of 1 (= considered semantically well-connected to the sentence to rate). For both contexts, the target aspectual condition was the imperfective-habitual and so the target form was the Imperfect.

- ***Semantic_connectedness* = 0:**

  Context: Jean says that he has fond memories of his childhood, especially when he went on picnics with his grandparents.

  Sentence to rate: Jean mangeait/a mangé au parc. [Translation: Jean ate-IMP/-PC in the park].

  (Basis for lack of contextual congruence: The sentence does not specify whether Jean was with his grandparents in the park or if they were having a picnic.)

- ***Semantic_connectedness* = 1:**

  Context: When Anne-Marie was a child she had a very close friend, Amélie, and she liked to spend a lot of time at her house after school.

  Sentence to rate: Anne-Marie allait souvent/est souvent allé chez Amélie à la sortie du college. [Translation: Anne-Marie often went-IMP/-PC to Amélie's house at the end of the school day].

The post-hoc nature of the *Semantic_connectedness* variable is due to the fact that it wasn't anticipated that there would be semantically-motivated differences in how participants responded, especially given that the AJT items used were originally sourced from previous work on viewpoint aspectual contrasts for which the AJT ratings had patterned as expected (e.g. Domínguez et al. 2011, 2017). The lower-than-expected ratings by the L1 group prompted a further investigation into this, and some post-hoc discussions with L1 participants revealed a possible semantic component to ratings. It was in an endeavour to control for this unexpected finding that the semantic connectedness/congruence variable was conceptualised.

As in the production data, *Target_*Aspect was Helmert coded in the comprehension data – with comparisons made in order of increasing difficulty as predicted by Feature Reassembly – and Group was dummy coded, with the L1 group constituting the reference category. In addition to the previous fixed effects, all models contained random effects for *Partic_ID* (to control for random variation in how each individual participant responded), and for *AJT_Context* (to control for variation across AJT items), as recommended by Baayen et al (2008).

The model-building strategy employed was similar to that for the production data: modelling began by endeavouring to fit a relatively maximal model: interactions were fitted between key variables of interest (*Group, Form, Target_Aspect*), as well as 'background variables' which may also play a role in task performance (*Semantic_connectedness, Seen_First, Lexical_Aspect*). Model structure was subsequently gradually simplified to obtain a converging model. Variables that did not improve model fit were identified and removed using the *drop1* function in *lme4*, and final model selection was made by manual AIC comparisons. With regard to random

slopes, the same principle was applied, beginning with an attempt to fit random slopes for every possible interaction and variable in the model. Ultimately, it was only possible to fit a random slope for *Semantic_connectedness* within the best-fitting converging model.

The formula for the best-fitting model is provided below, with output presented in Table 24.

Formula: Rating ~ Target_Aspect + Form + Target_Aspect:Form +
Group + Target_Aspect + Group:Target_Aspect +
Lexical_Aspect + Group + Lexical_Aspect:Group +
Semantic_connectedness +
(1 + Semantic_connectedness | Partic_ID) +
(1 + Semantic_connectedness | AJT_context)

Coefficients	**Estimate**	**Std. Error**	**z-value**	**p-value**
Target_Aspect1	0.70970	0.19931	3.561	0.000370 ***
Target_Aspect2	0.41000	0.31278	1.311	0.189919
Target_Aspect3	0.39649	0.20709	1.915	0.055543
FormTarget	1.73895	0.08855	19.637	< 2e-16 ***
GroupPre-YA	0.98906	0.24660	4.011	6.05e-05 ***
GroupPost-YA	0.74170	0.23212	3.195	0.001397 **
Lexical_AspectStative	1.13560	0.29755	3.817	0.000135 ***
Semantic_connectedness0	-0.44228	0.19738	-2.241	0.025038 *
Target_Aspect1:FormTarget	-0.18016	0.12290	-1.466	0.142687
Target_Aspect2:FormTarget	0.79319	0.16103	4.926	8.40e-07 ***
Target_Aspect3:FormTarget	0.32048	0.11316	2.832	0.004624 **
Target_Aspect1:GroupPre-YA	-0.56627	0.16605	-3.410	0.000649 ***
Target_Aspect2:GroupPre-YA	-0.96518	0.24741	-3.901	9.58e-05 ***
Target_Aspect3:GroupPre-YA	-0.48879	0.16322	-2.995	0.002748 **
Target_Aspect1:GroupPost-YA	-0.34104	0.15986	-2.133	0.032893 *
Target_Aspect2:GroupPost-YA	-0.66425	0.23719	-2.800	0.005103 **
Target_Aspect3:GroupPost-YA	-0.34037	0.15614	-2.180	0.029268 *
GroupPre-YA:LexicalAspectStative	-1.10116	0.26066	-4.224	2.39e-05 ***
GroupPost-YA:Lexical_AspectStative	-0.53494	0.25146	-2.127	0.033395 *

Table 24: Output from cumulative link mixed modelling (ordinal regression) for all-groups AJT rating model

Table 24 shows that multiple variables had a significant effect on ratings in the AJT task. Firstly, the effect of viewpoint aspectual condition will be considered. Recall that the *Target_Aspect* variable was Helmert coded, with comparisons made in order of increasing difficulty as predicted by Feature Reassembly; therefore, if there are significant effects of any of the levels of *Target_Aspect* on rating, this may suggest a role played by reassembly in comprehension of viewpoint aspectual form-meaning mappings (although of course this model contains data from all three participant groups, including the L1 group, so conclusions relating to L2 development made using this dataset are necessarily limited).

When ratings in the perfective vs. combined imperfective were compared, there was a significant difference in ratings between the two conditions (z = 3.561, p = 0004). Figure 18 shows ratings in perfective vs. imperfective conditions per group.

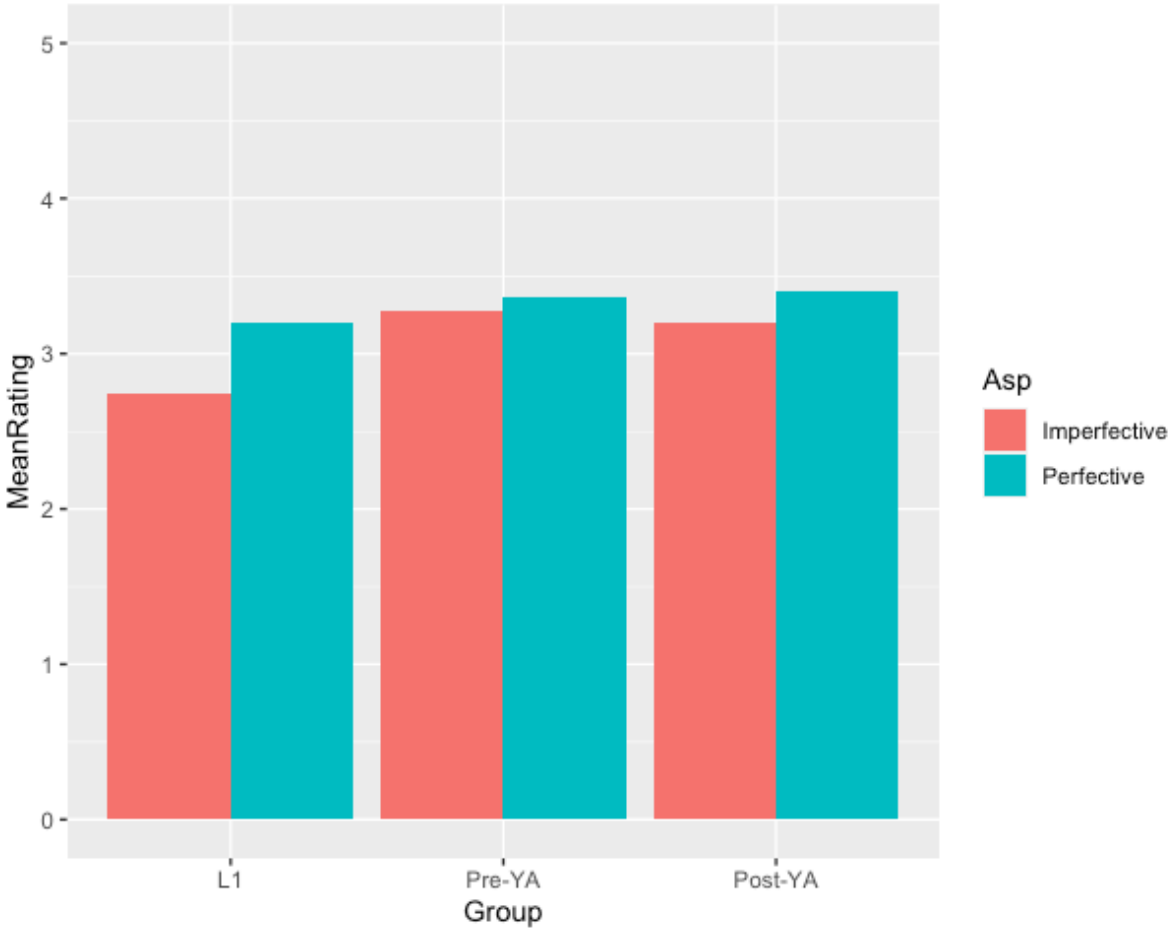


Figure 18: All-group ratings in the AJT in imperfective vs. perfective conditions.

It can be seen that perfective forms were rated more highly than imperfective forms, regardless of whether or not they were the target form for a given context. This difference appears to be particularly pronounced for the L1 group.

However, this finding is not overly informative without considering how the ratings varied according to *Form*, i.e. how the target form for a given aspectual condition was rated vs. the non-target form. Figure 19 displays how ratings of target vs. non-target forms varied per aspectual condition.

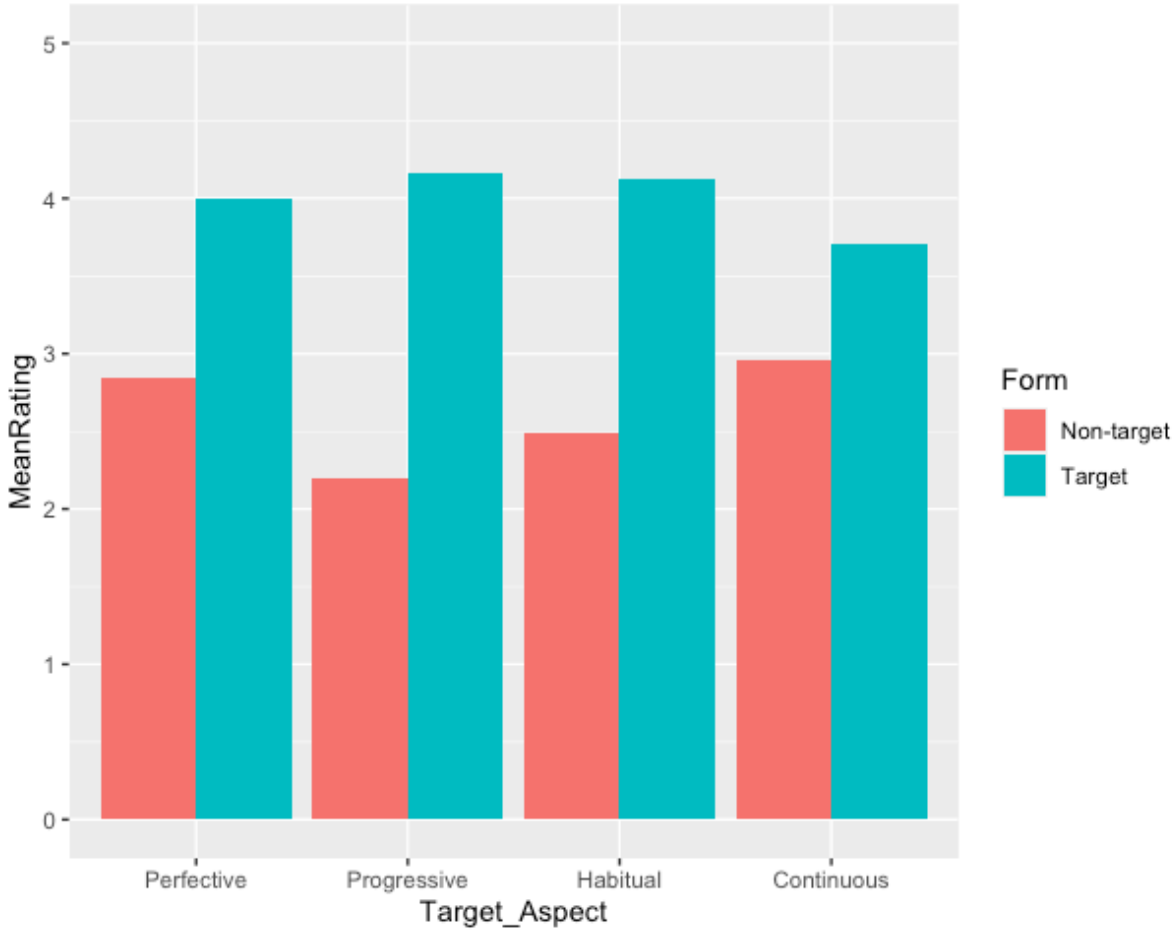


Figure 19: All-group ratings of target vs. non-target forms in the AJT by viewpoint aspectual condition.

In the all-groups ratings model, a significant interaction was found between *FormTarget* and *Target_Aspect2* i.e. the second level of *Target_Aspect* comparisons (z = 4.926, $p < .001$). As a reminder, this level of comparisons contrasted ratings in progressive conditions vs. those in combined habitual and continuous conditions. Looking at Figure 19, ratings of the target form (the Imperfect) in progressive and habitual conditions appear in fact to be relatively similar, but ratings of the Imperfect

target in continuous conditions are comparatively lower. It could therefore be inferred that the lower ratings of target continuous Imperfect are the source of the significant effect found here. This idea is supported by the fact that the model also contains a significant interaction between *FormTarget* and *TargetAspect3* ($z = 2.832$, $p = .004$), indicating that there is a significant difference between ratings of the target Imperfect in habitual vs. continuous conditions.

Although the above data on rating of the target form across aspectual conditions may point towards some differences in the extent to which participants accept the target form – which in itself might be informative of how far feature reassembly has progressed, for the L2 participants – the central point of interest for the ratings data is whether or not there is within-group differentiation between target and non-target forms, as previously discussed. The per-group trends presented in Figure 17 indicated a clear target vs. non-target differentiation for each group; these trends are substantiated not only in the data presented in Figure 19, but also in the significant main effect for *Form* in the ratings model ($z = 19.637$, $p < .001$). This provides strong support for the fact that, across the ratings data as a whole, target forms were consistently rated higher than non-target forms. There is not a significant interaction between *Group* and *Form* in the best-fitting rating model, which suggests that neither L2 group rates the target form significantly differently to the L1 reference category. Taken together with the evidence supporting a clear target/non-target differentiation, a picture begins to emerge from the AJT data wherein the acceptability ratings of the L2 French learners fairly closely resemble those of the L1 French speakers. This would suggest that learners' underlying L2 viewpoint aspectual representations are quite well-developed, in line with the fact that these are advanced learners with a substantial number of years of French exposure. Differences between the learner groups will be explored further in the following subsection.

Lastly, there are several other results of potential interest within the all-groups ratings model. The first of these is a main effect of *Semantic_connectedness* ($z = -2.241$, $p = .025$), the binary variable denoting the contextual congruence of a given AJT item. A summary of how ratings varied according to semantic connectedness is presented in Figure 20.

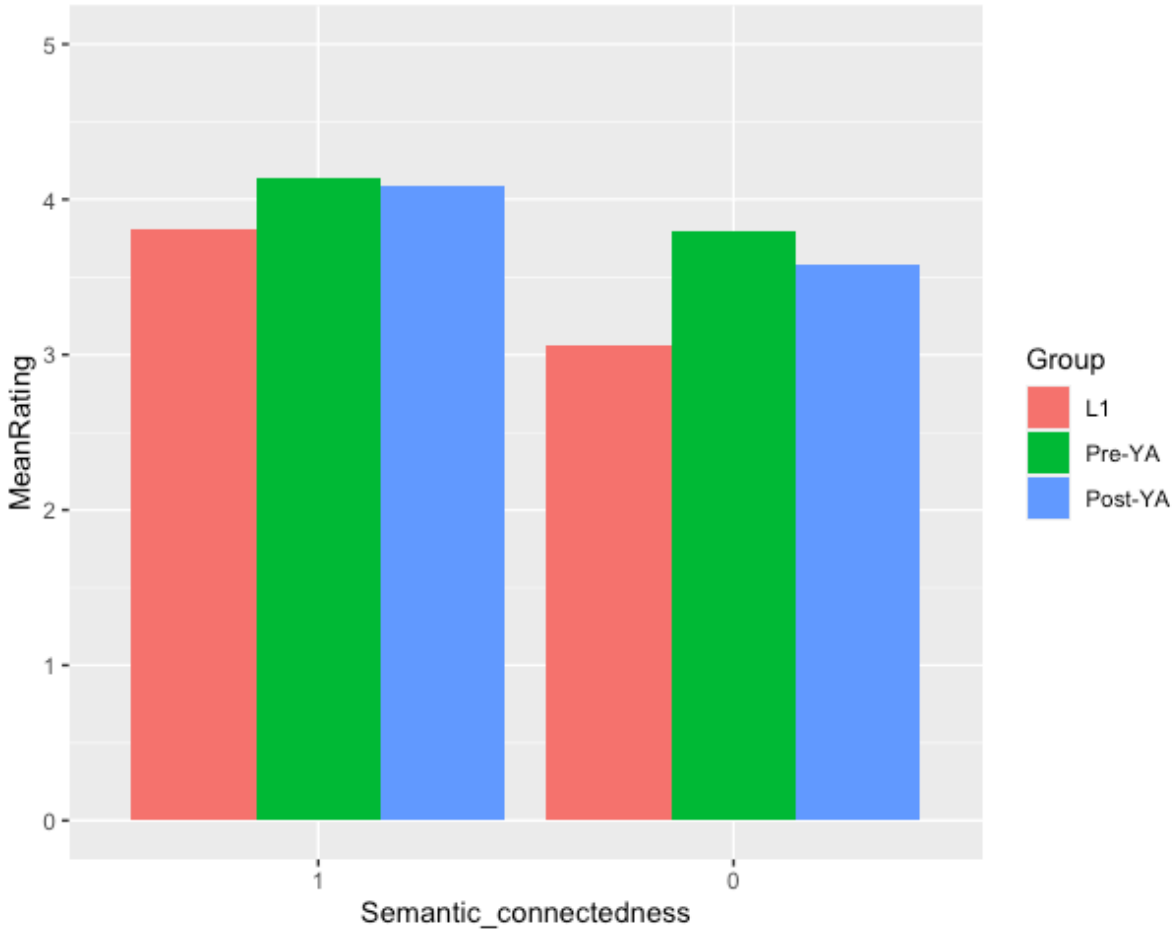


Figure 20: All-groups AJT ratings per contextual congruence condition (semantic connectedness).

As the data presented in Figure 20 show, all groups rated sentences with a contextual congruence score of 0 less highly than sentences with a contextual congruence score of 1: a difference that is statistically significant according to the model output (Table 24). Though there is no interaction between *Semantic_connectedness* and *Group* in the model, Figure 20 shows the largest difference in rating between the two congruence conditions for the L1 group, whilst the Pre-YA group shows the smallest difference. This provides some exploratory evidence that L1 speakers vs. L2 (instructed) learners may attend to different informational cues when completing an acceptability judgement task. This idea will be expanded upon later.

Finally, an effect of *Lexical_Aspect* was found on ratings, both as a main effect (z = 3.817, p = .0001), and as an interaction with both L2 groups (*GroupPreYA:LexicalAspectStative* : z = -4.224, p < .001; *GroupPostYA:LexicalAspectStative* : z = -2.127, p = .033). To give context to these results, a summary of how ratings varied according to the lexical aspect of the verbal predicate (eventive vs. stative) in the sentence to be rated is presented in Figure 21.

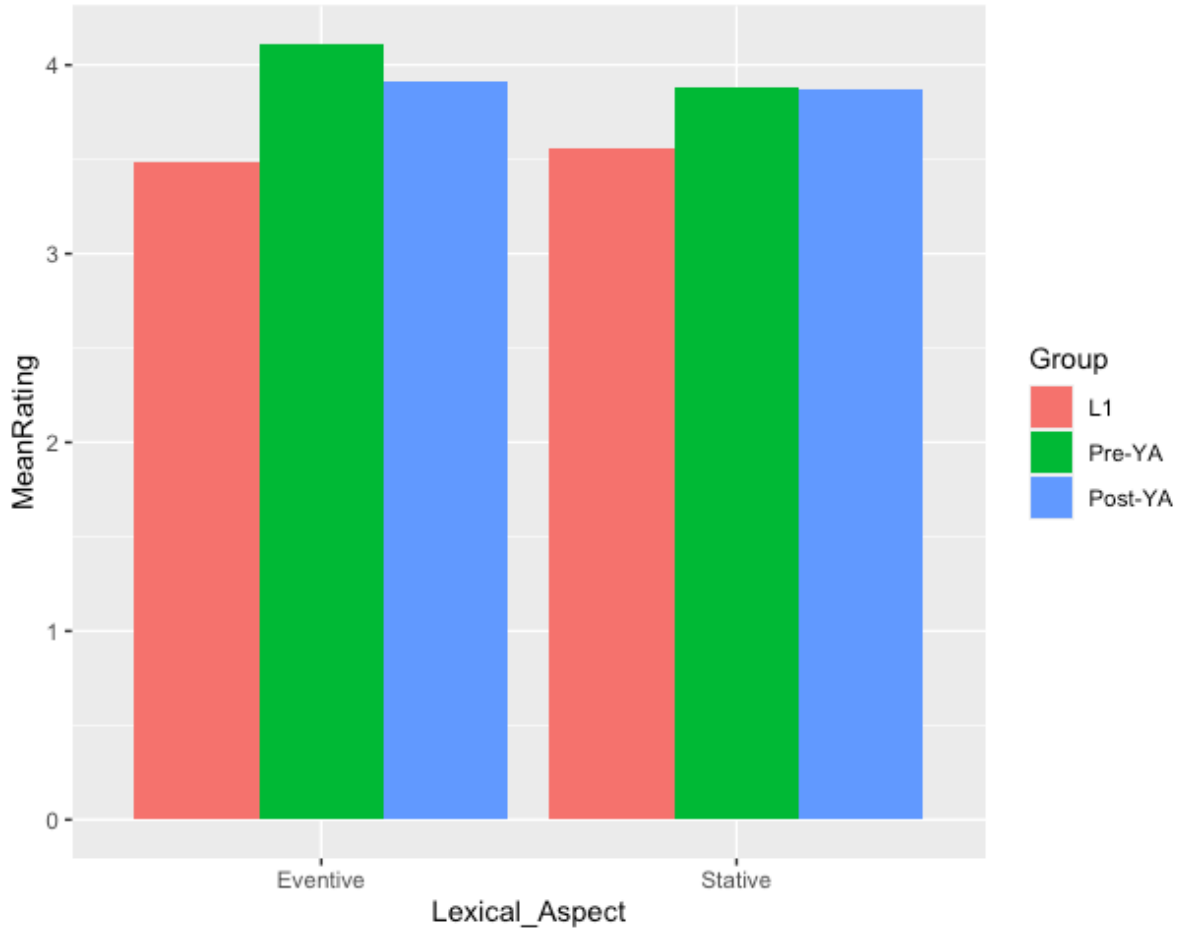


Figure 21: All-group AJT ratings per lexical aspect condition (eventive vs. stative).

Overall, the data in Figure 21 suggest that eventive predicates were overall rated more highly than stative predicates. Contrastingly to the *Semantic_connectedness* trends presented in Figure 20, where the L1 group appeared to be the most affected by differences in contextual congruence, differences in L1 ratings between stative and lexical predicates appear fairly minimal. Contrastingly, there is a larger difference between the two lexical aspectual conditions for the L2 groups, to which the significant interactions in Table 24 can be attributed. It is however worth noting at this point that the results pertaining to the effect of lexical aspect on rating should be interpreted with caution, given that the AJT items in each viewpoint aspectual condition could not always be equally divided across the two lexical aspect categories (see Chapter 4, Section 4.1.2 for more explanation on AJT design).

In summary, the analyses of the all-groups rating data highlight the influence of several variables on participants' rating of viewpoint aspectual forms in the AJT,

including viewpoint aspectual condition, contextual congruence, and lexical aspect of the verbal predicate. However, the key finding of these analyses is the significant difference between ratings of the target vs. non-target form. Coupled with the lack of interaction between *Group* and *Form* in the best-fitting model, this suggests not only that each group rates the target form significantly more highly than the non-target form, but that there is not a significant difference between each of the L2 groups and the L1 reference category in terms of target form ratings. Therefore, the all-groups rating analyses lend support to the idea that the L2 French viewpoint aspectual form-meaning mappings of both learner groups are at least reasonably well-established, in that they are able to differentiate between the target and non-target form in all viewpoint aspectual conditions in an acceptability judgement task. This is in contrast to the production data, where the L2 groups – particularly the Post-YA group - do not always behave in a similar manner to the L1 group.

In order to take a closer look at trends in the L2 data, and permit a clearer comparison between the Pre- and Post-YA learners, the following subsection will carry out analyses on AJT ratings using data from the L2 groups only.

L2 acceptability judgement ratings

As within-group differentiation between the target and non-target form has already been established through the all-groups analyses, the following L2 analyses will take a different central focus. Of particular interest is whether there are significant between-groups rating differences, as this has the potential to provide further insight into the state of each group's L2 aspectual representations. Additionally, significant within-group differences across aspectual conditions may be informative with regard to how successfully learners have reassembled a given form-meaning mapping, as alluded to in the previous section.

Before considering the results of the L2 rating regression analysis, however, it is useful to look at the patterns shown in the data. Figures 22 and 23 display ratings of target vs. non-target forms per aspectual condition in the AJT by the Pre-YA and Post-YA groups respectively.

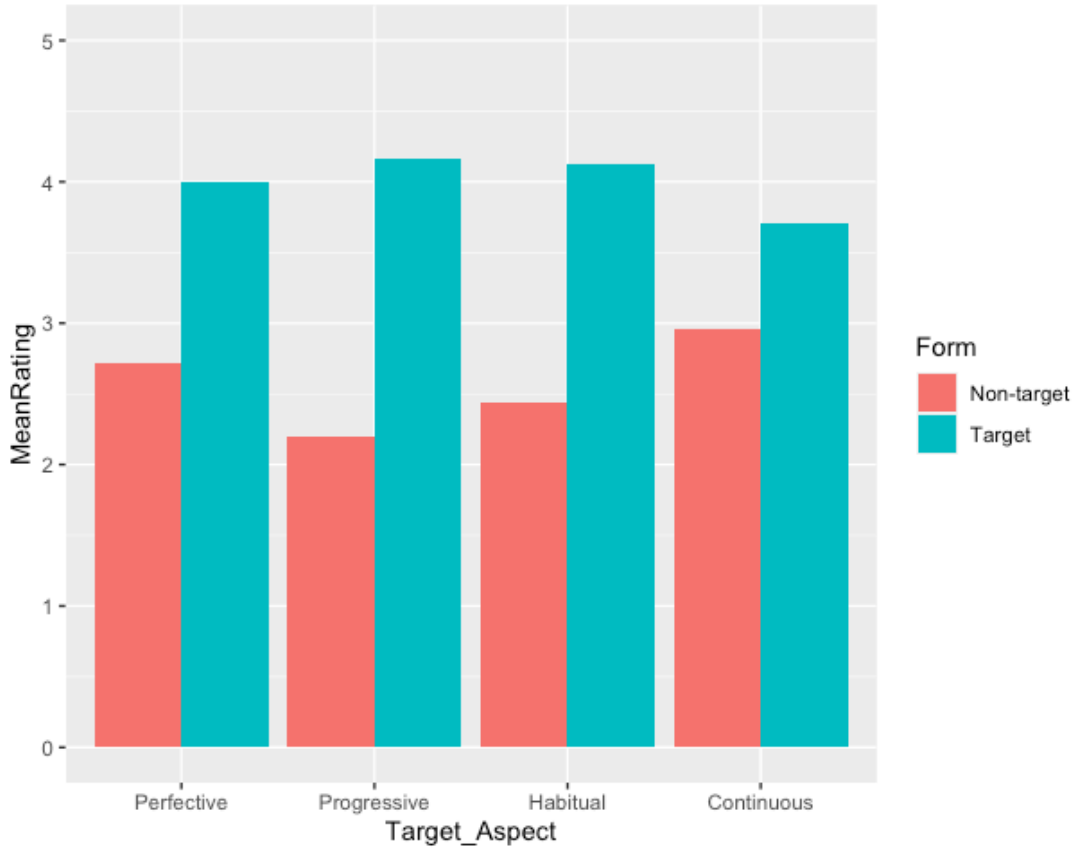


Figure 22: Pre-YA ratings of target vs. non-target forms per aspectual condition in the AJT.

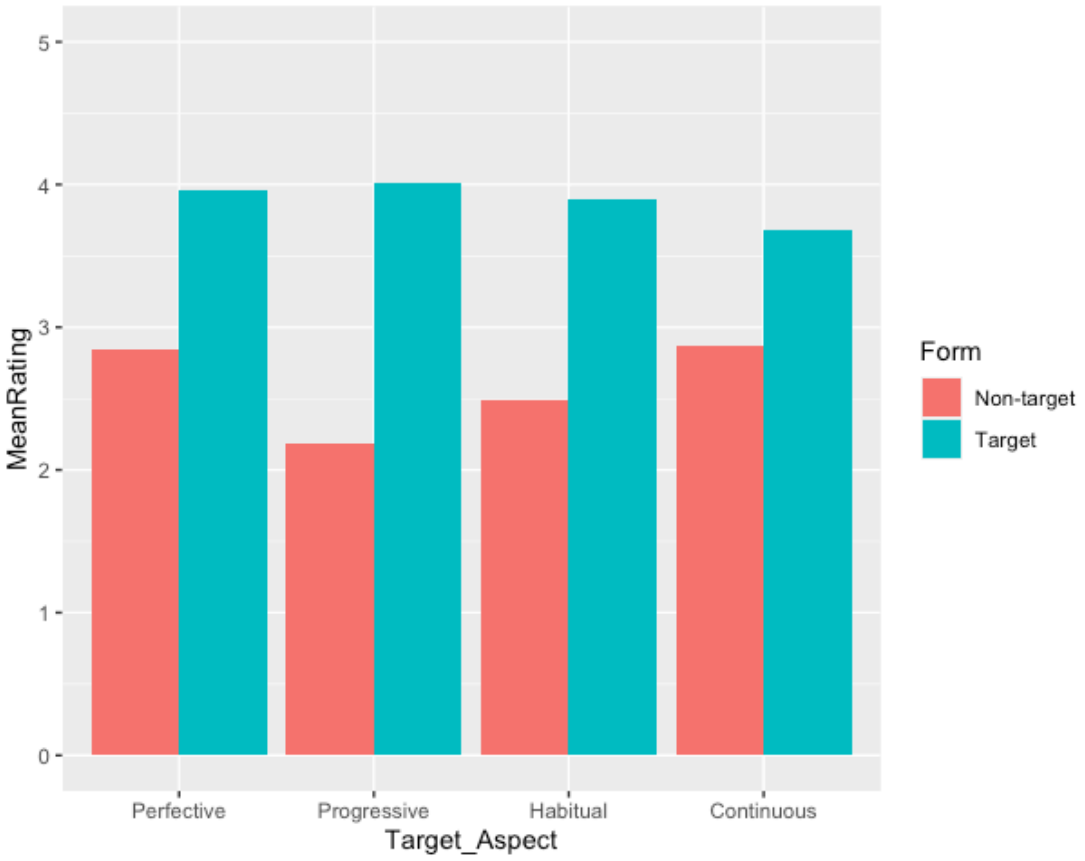


Figure 23: Post-YA ratings of target vs. non-target forms per aspectual condition in the AJT.

Looking at both Figures 22 and 23, a clear target vs. non-target differentiation for each aspectual condition is apparent, as previously discussed. Within this, however, some cross-aspectual variation can be observed, which appears to show a similar pattern across both groups. Both groups show the highest rating of the target form and the lowest rating for the non-target form in the progressive condition; it follows that this is the condition where there is the largest contrast made between the target (Imperfect) and non-target (*passé composé*) forms. Ratings of the target form in both habitual and perfective conditions are relatively similar to in the progressive condition, although it can be seen that the Pre-YA group display a higher mean rating of the habitual Imperfect than the Post-YA group. Lastly, the continuous condition is that where there is the smallest differentiation in ratings of the target and non-target form: both L2 groups rate the non-target form (the *passé composé*) the most highly in this condition, while rating the target Imperfect the lowest on average when compared to the other imperfective conditions. These observations are interesting for two reasons. Firstly, the pattern of results seen for the imperfective mappings aligns well with Feature Reassembly predictions: the progressive, with its one-to-one mapping between English and French, is the condition in which learners differentiate most decisively between the target and non-target forms, whereas the continuous, with its complex reassembly task involving detangling of imperfective feature from the perfectivity-hosting simple past, shows the least clear differentiation. However, these results in themselves contrast strongly with those from the production data, where learners performed very accurately in the continuous condition and considerably less accurately in both the progressive and habitual conditions. Potential reasons for this disparity will be explored in Chapter 6.

When fitting the model for the L2 rating data using ordinal regression, the same principles were applied as for the all-groups rating model. An attempt was again made to fit random slopes for as many of the variables appearing in the best-fitting model as possible, but as before only a random slope for *Semantic_connectedness* could be fitted within a converging model. The formula of the best-fitting model is presented below, with model output in Table 25.

Formula: Rating ~ Target_Aspect + Form + Target_Aspect:Form +

Semantic_connectedness + (1 + Semantic_connectedness | Partic_ID) +

(1 + Semantic_connectedness | AJT_context)

Coefficients	**Estimate**	**Std. Error**	**z-value**	**p-value**
Target_Aspect1	0.2514	0.1700	1.479	0.139150
Target_Aspect2	-0.4957	0.2356	-2.104	0.035398 *
Target_Aspect3	-0.2012	0.1657	-1.214	0.224756
FormTarget	1.7524	0.1027	17.059	< 2e-16 ***
Semantic_connectedness0	-0.3913	0.1967	-1.989	0.046650 *
Target_Aspect1:FormTarget	-0.2174	0.1410	-1.542	0.123118
Target_Aspect2:FormTarget	0.6946	0.1857	3.740	0.000184 ***
Target_Aspect3:FormTarget	0.4813	0.1316	3.658	0.000254 ***

Table 25: Output of ordinal regression analyses for L2 AJT rating model.

Table 25 contains several significant effects that support the trends shown in Figures 22 and 23. Firstly, as in the all-groups model, there is a significant main effect of *FormTarget* ($z = 17.059$, $p < .0001$), cementing the within-group differentiation that has already been attested. There are also two significant interactions between *Target_Aspect* and *Form*. The first is between the second level of *Target_Aspect* comparisons (progressive vs. [habitual + continuous]) and ratings of the target Imperfect ($z = 3.740$, $p = .0002$), suggesting that the observed difference between Imperfect ratings in the progressive and the other imperfective conditions is a significant one. The second significant interaction is found between the third level of *Target_Aspect* comparisons (habitual vs. continuous) and ratings of the target form ($z = 3.658$, $p = .0003$). This confirms the pattern observed in Figures 22-23, indicating that ratings of the Imperfect in continuous conditions are significantly lower than those in habitual (and other imperfective) conditions. Lastly, *Semantic_connectedness* continues to be a significant main effect in the L2 ratings model, as in the all-groups model, indicating that all groups are influenced by how congruent a sentence item was with its preceding context when rating viewpoint aspectual forms.

Also notable is the absence of certain variables from the model. The fact that *Group* is not present in the best-fitting L2 ratings model indicates that there is no significant difference between the performance of the two L2 groups in the AJT, which could already be inferred by the very similar patterns in ratings in Figures 22 and 23. Additionally, *Lexical_Aspect* did not feature in the L2 ratings model, despite significant interactions with L2 groups in the all-ratings model. Lastly, the *Seen_First* variable, relating to the order in which a participant viewed a given item within its "context pair", was ultimately not a significant variable in either rating model. This suggests that participants' ratings were not overly influenced by the decision to present each context twice – with a different sentence to rate each time – as opposed to in a two-alternative forced choice structure.

In summary, the results of the L2 ratings analyses confirm that both L2 groups differentiate between target and non-target forms across all four viewpoint aspectual conditions, meaning that their form-meaning mappings are established. However, there are some significant differences in how highly the target form is rated (and some descriptive differences in how the non-target form is rated) across the aspectual conditions. Notably, ratings of the habitual Imperfect are significantly higher than the continuous Imperfect, and ratings of the progressive Imperfect are higher than both of these. These observations align with the degree of reassembly required for each mapping, with the mapping requiring the most reassembly (the continuous) also being the one in which there is the smallest difference in rating between the target Imperfect and non-target *passé composé.* However, they contrast with L2 performance in the production task, where the imperfective-continuous was one of the aspectual conditions in which learners performed most accurately. An analysis of accuracy in the AJT – presented in the following subsection – will permit a closer comparison between production and comprehension, with the aim of further elucidating what is happening in the viewpoint aspectual systems of the participants.

5.3.3 Acceptability judgement accuracy analysis

Overview

In addition to analysing participants' ratings in the AJT, their accuracy in the task was also explored. Accuracy was coded as a binary variable: participants were given a score of 1 (accurate) for every item where a sentence containing the non-target form was given a rating lower than 3, or a sentence with the target form was given rating higher than 3. Non-target sentences rated higher than 3 and target sentences rated lower than 3 were given a score of 0. Ratings of 3 exactly were also scored 0, as this corresponded to 'I do not know whether this sentence is appropriate' on the scale used.

Before exploring the mean accuracy scores of each group in the comprehension task, some consideration will be given to individual participant scores per group (Figure 24). These individual accuracy score patterns will then be considered alongside those for overall production (combined Cat Story and Conversation accuracy scores), presented in Figure 25.

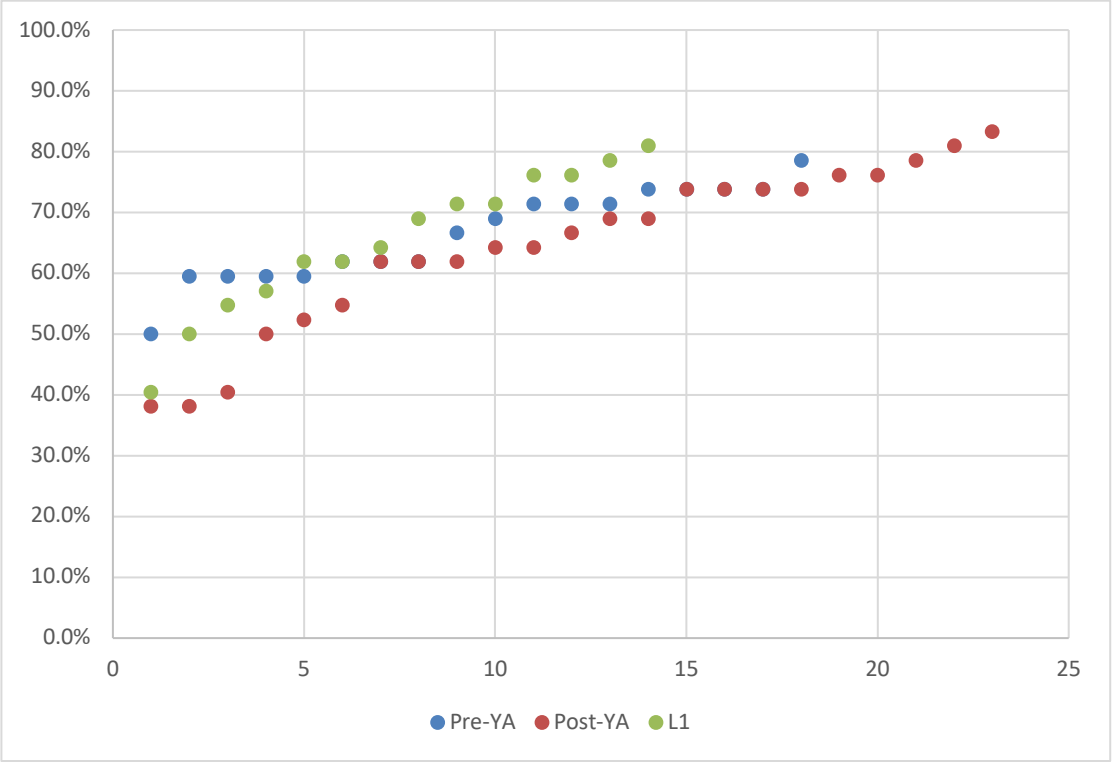


Figure 24: Individual participant scores in the AJT, per group.

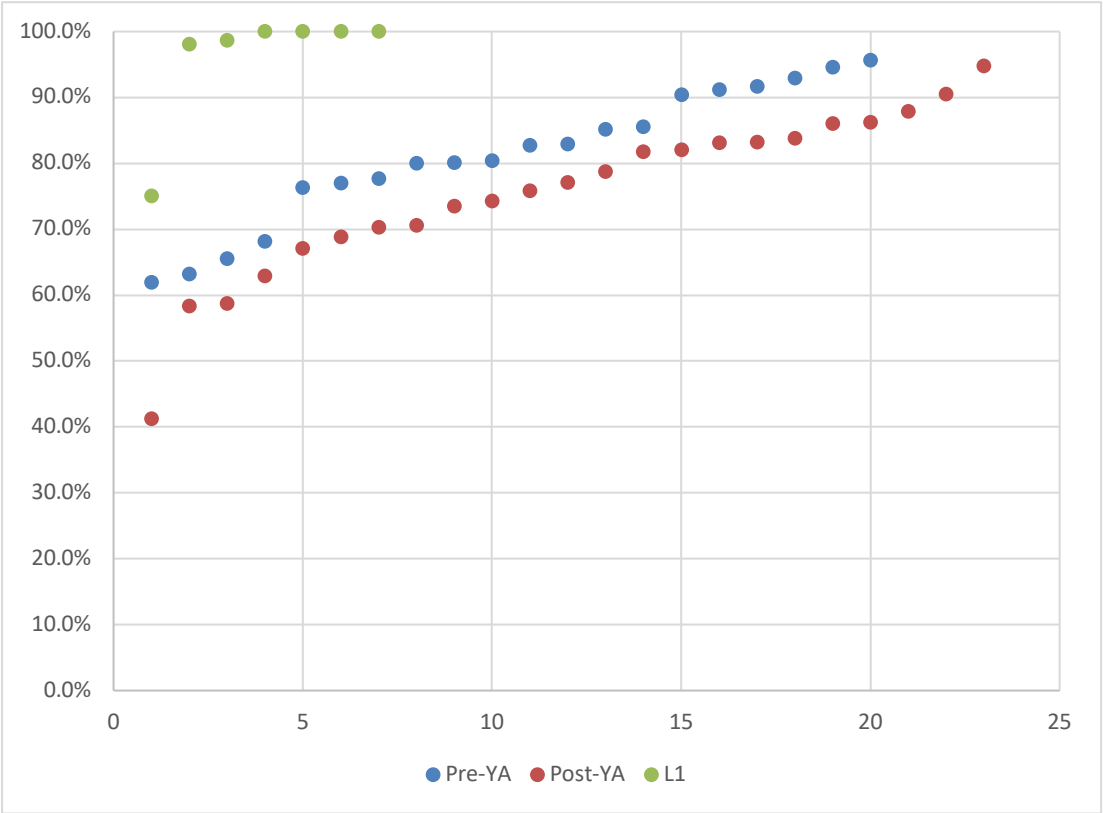


Figure 25: Individual participant scores in overall production, per group.

Considering first the individual accuracy data presented for the AJT in Figure 24, it can be seen that there is a wide array of scores ranging from scores in the 40% range to those in the 80% range. The Post-YA group arguably show the widest range of scores (although it is worth noting that this was also the group with the largest number of participants). When the comprehension scores are compared with the individual production scores in Figure 25, it can be seen that whilst there is a clear differentiation in overall accuracy patterns in the production data – where the L1 group perform at ceiling, and the Pre-YA are shown to be consistently higher-scoring than the Post-YA group – there is no such clear differentiation for the comprehension data. Though looking at the line gradients in Figure 24 may provide some evidence for overall lower scores from the Post-YA group, all three groups pattern closely together on the whole, suggesting a lack of notable between-groups accuracy differences.

It may also be insightful to consider participants' cross-task consistency, in order to assess how similarly participants performed across production vs. comprehension. Although it must be stressed that a score of, for example, 70% in overall production cannot necessarily be equated to a score of 70% in the AJT, it is

nonetheless of interest to explore whether the participants who performed the most accurately in production were the same participants who performed most accurately in comprehension. Table 26 contrasts the top ten highest scoring L2 participants in both production and comprehension. As previously, the IDs of participants who appear in the top ten for both production and comprehension are presented in bold and underlined font.

TOP SCORERS			
Production		*Comprehension*	
Participant ID	Score	Participant ID	Score
202	95.7%	411	83.3%
410	94.8%	**404**	81.0%
207	94.6%	**207**	78.6%
214	92.9%	421	78.6%
215	91.7%	405	76.2%
206	91.2%	415	76.2%
427	90.5%	223	73.8%
212	90.4%	201	73.8%
404	87.9%	**202**	73.8%
424	86.2%	**215**	73.8%

Table 26: Comparison of highest-scoring L2 participants in production vs. comprehension.

From the ten highest scoring participants in each task, it can be seen that three Pre-YA participants (202, 207, and 215) appear for both production and comprehension, whilst this is only the case for one Post-YA participant (404). This suggests not only that the Pre-YA group perform more accurately in some cases (notably production), but that they are also more consistent in their accuracy across tasks. With this said, it can also be observed in Table 26 that four of the top five comprehension scores in fact belong to the Post-YA group, illustrating the wide range of scores for this group previously illustrated in Figure 24.

Moving on to mean accuracy patterns, Figure 26 presents the overall accuracy scores for each group, divided by viewpoint aspectual condition. First considering the L2

groups, there are – as in the ratings data – some key differences in comparison with the production data. Whilst the latter showed a large discontinuity in accuracy between the habitual and progressive on the lower end and the continuous and the perfective on the higher end, no such contrast is visible in the AJT data. If anything, the inverse may be partly true, as learners appear to have the lowest mean accuracy scores in the continuous condition, and middling accuracy scores for the perfective, while scoring somewhat more highly in the habitual and progressive conditions (although accuracy across all four conditions is broadly more similar, in comparison with the rather polarised production accuracy data).

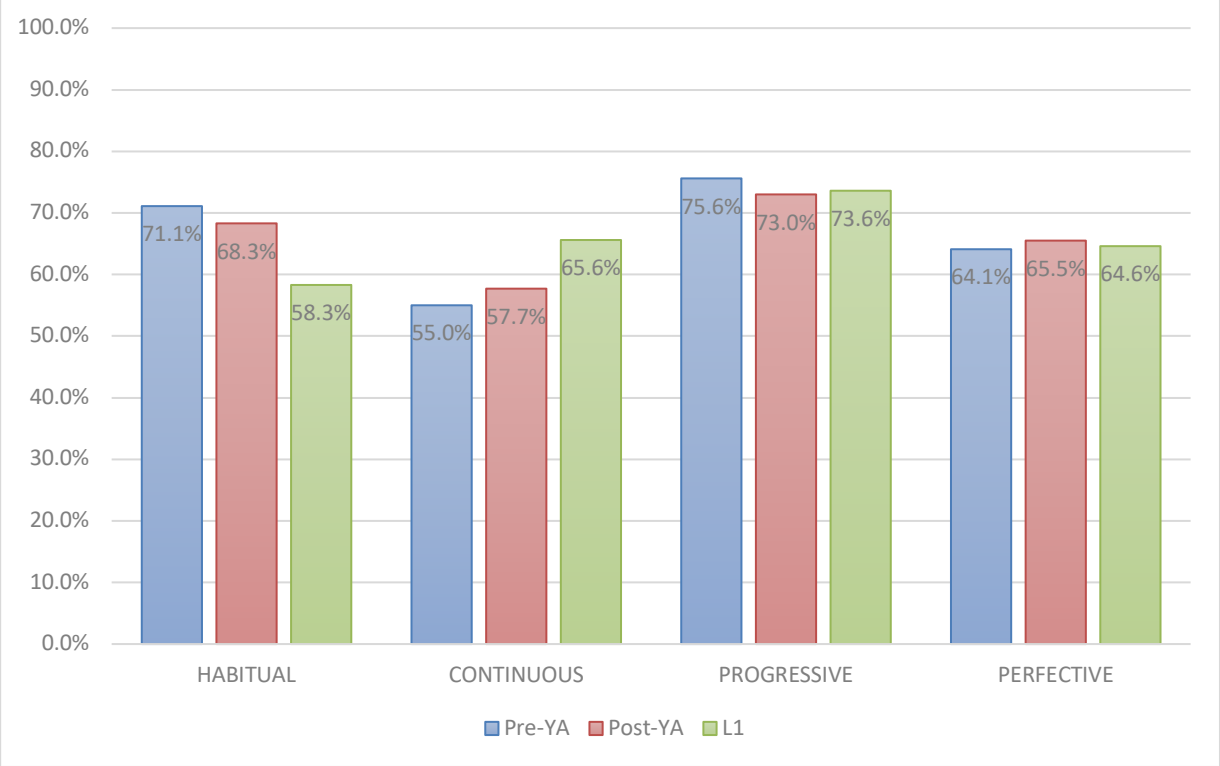


Figure 26: Mean AJT accuracy (%) per group and aspectual condition.

[SDs: Habitual: Pre-YA 0.45, Post-YA 0.47, L1 0.49; Continuous: Pre-YA 0.50, Post-YA 0.50, L1 0.48; Progressive: Pre-YA 0.43, Post-YA 0.45, L1 0.44; Perfective: Pre-YA 0.48, Post-YA 0.48, L1 0.48]

Considering the overall patterns for the L1 group presented in Figure 26, the results are also somewhat unexpected. It may have been anticipated that, as an L1 control, this group would perform at or close to ceiling, as is the case in the production data (although, as mentioned in Chapter 4, the L1 production data was by necessity

collected from a different cohort to the L1 comprehension data). However, it should be noted that the overall AJT scores for the L1 group were quite wide ranging, which is not fully reflected in Figure 26. This range may be related to the fact that the L1 sample was quite heterogenous in comparison to the L2 groups. Indeed, the L1 group diverged demographically from the L2 groups in two key ways – firstly in terms of age, and secondly in terms of occupation: whilst all L2 participants were languages students, the L1 group's occupations were unknown, and not all group members were necessarily educated to university level.

The implications of these demographic differences on linguistic performance will be explored throughout this section, but are included in Table 27 for reference, along with the mean score and range of scores for the AJT for each of the three groups out of a possible maximum score of 42.

	GROUP		
	Pre-YA (*n* = 18)	Post-YA (*n* = 23)	L1 (*n* = 14)
Age range (Mean, [SD])	19-21 (20.0, [0.53])	21-24 (21.7, [0.72])	24-59 (34.1, [10.16])
Level of education	University-level	University-level	Unknown
Occupation	Languages students	Languages students	Various unknown occupations
Mean AJT score [SD]	28.0 [3.32]	27.7 [5.30]	27.3 [5.30]
Range of scores (max. 42)	21.0-33.0	16.0-33.0	17.0-34.0

Table 27: Mean ([SD]) and range AJT scores per group, combined with key demographic information.

A particularly salient demographic difference of the L1 group compared to the L2 group is the markedly higher mean age of the former (there is also considerably more variance in age for the L1 group). In addition, whilst both L2 groups are (by definition) educated to university-level, the level of education of the L1 sample is unknown. As these variables have been indicated (e.g. Hulstijn 2011, Andringa et al 2012) to play a

role in participants' linguistic behaviour, it is arguably important to take the lower accuracy of the L1 group in the AJT task with a grain of salt, particularly given that they have already been shown to make targetlike contrasts between the target and non-target forms in all aspectual conditions. The role of age in particular may be especially key when we consider Schütze's (1996/2016**:** 13-14) affirmation that judgement tasks reflect '… the result of interactions between primary language faculties of the mind and general cognitive properties.' The different task- and sample-related factors affecting the L1 group's responses on the AJT will be explored more thoroughly in the following subsection, but for now it may be reasonable to posit that it is these aforementioned extralinguistic variables that have yielded the unexpected accuracy results for this group.

In an endeavour to shed further light on the patterns presented in the above overview, the following subsection will carry out regression analyses on the all-groups accuracy data, in order to explore the influence of variables – both related to and external from viewpoint aspect – on participant accuracy in this comprehension task.

Accuracy Modelling

The principal questions underpinning the comprehension accuracy analyses related to the variables of *Group, Target_Aspect,* and *Form.* Specifically of interest was whether there were significant differences in accuracy between the participant groups, and also whether there were significant within-groups difference according to either aspectual condition, or target vs. non-target condition. The presence of significant differences in either of these categories has the potential to inform on the viewpoint aspectual development of the L2 groups, as well as whether they differed significantly to the L1 group in their comprehension of viewpoint aspectual form-meaning mappings.

Given that the dependent variable, *Accuracy*, was binary in nature, logistic regression analyses using the *lme4* package in R were carried out, as for the production accuracy analyses. The model-fitting process was also similar: to start, a model containing interactions between all the key variables of interest was fitted, which also contained the 'background' variables (*Lexical_Aspect, Semantic_connectedness, Seen_First*) whose influence on accuracy was to be explored. The model was gradually simplified by removing variables that did not significantly contribute to fit (established

using likelihood ratio tests via the *drop1* function), with final model selection made via manual AIC comparisons. Random intercepts for *Partic_ID* and *AJT_context* were fitted as before, and an attempt was made to fit as many relevant random slopes as possible; ultimately, only a random slope for *Seen_First* could be accommodated within a converging model.

The formula for the best-fitting all-groups accuracy model can be seen below, with output presented in Table 28.

Formula: Accuracy ~ Group + Form + Group:Form +

Target_Aspect + Form + Target_Aspect:Form +

Seen_First +

(1 + Seen_First | Partic_ID) + (1 + Seen_First | AJT_context)

Fixed effects:	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	0.583952	0.209405	2.789	0.00529 **
GroupPre-YA	-0.331719	0.241521	-1.373	0.16951
GroupPost-YA	-0.329217	0.228219	-1.443	0.14915
FormTarget	-0.163729	0.181041	-0.904	0.36580
Target_Aspect1	-0.295807	0.185443	-1.595	0.11068
Target_Aspect2	0.397281	0.244152	1.627	0.10370
Target_Aspect3	0.027360	0.174599	0.157	0.87548
Seen_First0	0.257681	0.112297	2.295	0.02175 *
GroupPre-YA:FormTarget	0.967053	0.243275	3.975	7.03e-05 ***
GroupPost-YA:FormTarget	0.861384	0.230821	3.732	0.00019 ***
FormTarget:Target_Aspect1	0.481229	0.148637	3.238	0.00121 **
FormTarget:Target_Aspect2	-0.001083	0.193928	-0.006	0.99555
FormTarget:Target_Aspect3	0.110961	0.133142	0.833	0.40461

Table 28: Output from all-groups comprehension accuracy model (logistic mixed-effects regression)

Of the significant effects shown in Table 28, two relate to significant interactions between *Group* and *Form*. To provide visual context to these interactions, Figure 27 displays a graph of accuracy in target vs. non-target conditions, divided by group. Looking at Figure 27, it can be seen that, whilst there is not a large contrast between accuracy in target vs. non-target conditions for the L1 group, both L2 groups are more accurate when rating sentences containing the target form vs. the non-target form. Indeed, the difference in accuracy between target and non-target conditions is statistically significant for both L2 groups in the best-fitting accuracy model (*GroupPreYA:FormTarget: z* = 3.975, *p* < .001; *GroupPostYA:FormTarget: z* = 3.732, *p* = .0002).

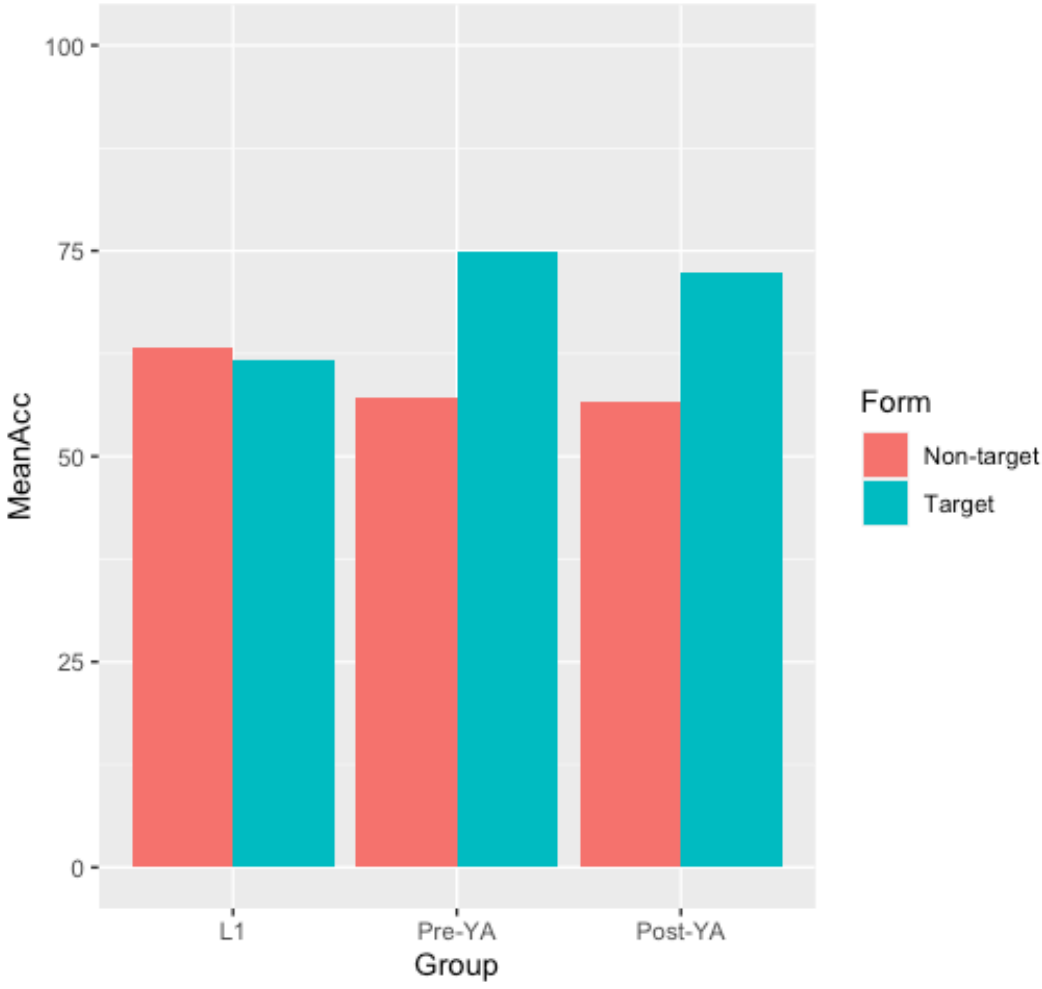


Figure 27: Graph of AJT accuracy per group in target vs. non-target conditions.

There is also a significant interaction (*z* = 3.238, *p* = .001) between accuracy in target conditions and the first level of *Target_Aspect* comparisons (perfective vs. combined imperfective). Figure 28 (below) displays a graphical representation of

accuracy in target vs. non-target conditions per aspectual conditions. It can be seen that, with regard to the target form, accuracy is highest in the progressive condition, with the perfective condition coming a close second. Contrastingly, accuracy in AJT items containing target Imperfect appears substantially lower, which may be the source of the significant contrast between perfective and combined imperfective accuracy. This contrast in accuracy between the progressive and continuous – also aligned with the trend seen in the ratings data – lends further support to the previous observations about the relative well-fittedness of the comprehension data to Feature Reassembly predictions, as well as its divergence from the production accuracy trends. Looking at the L2 accuracy data alone (in the next subsection) will permit a more decisive perspective on this from the lens of L2 development.

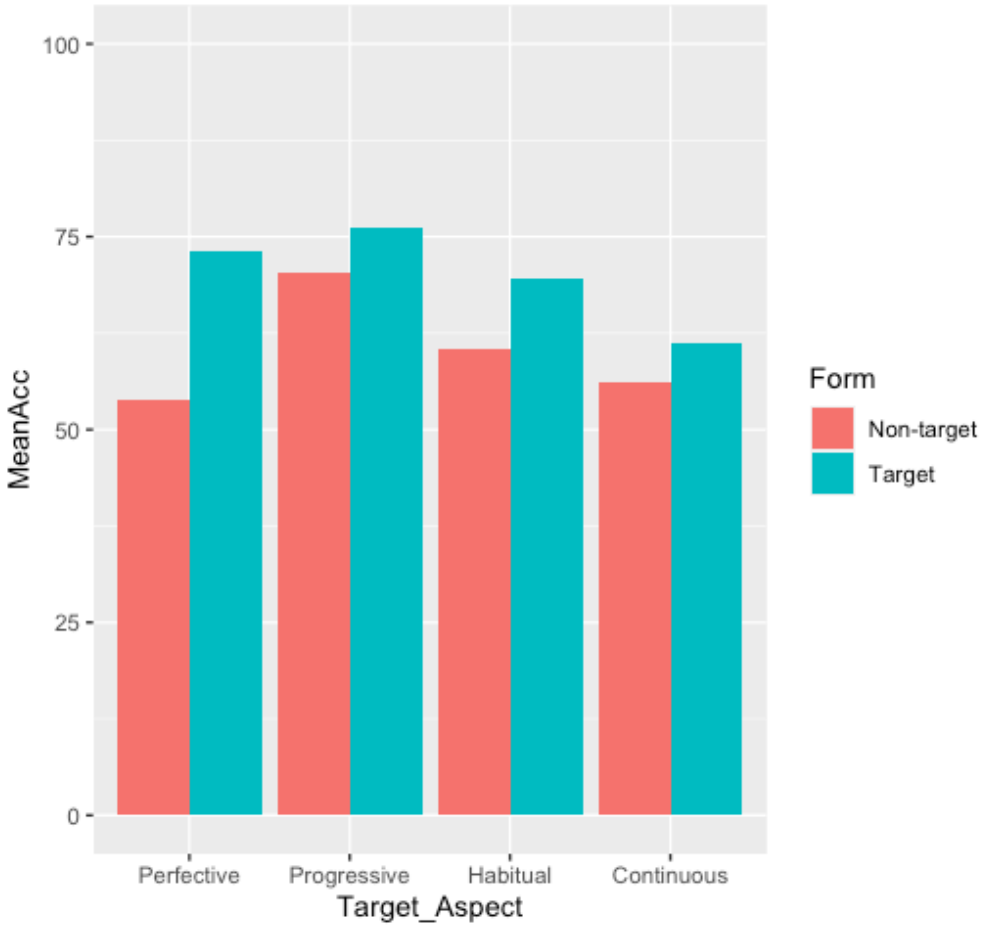


Figure 28: Graph of AJT accuracy in target vs. non-target conditions, per aspectual condition.

Finally, there is a significant main effect of *Seen_First* on accuracy ($z = 2.295$, $p = .02$). This is the binary variable indicating whether, for a given item, participants viewed this item first or second within its "context pair" (bearing in mind that each AJT context was presented twice, in a randomised order: once followed by a sentence containing the target form, and once followed by a sentence containing the non-target form). Figure 29 presents accuracy scores per group when *Seen_First* = 1 (i.e. the context was presented for the first time) vs. *Seen_First* = 0 (i.e. the context was presented for the second time).

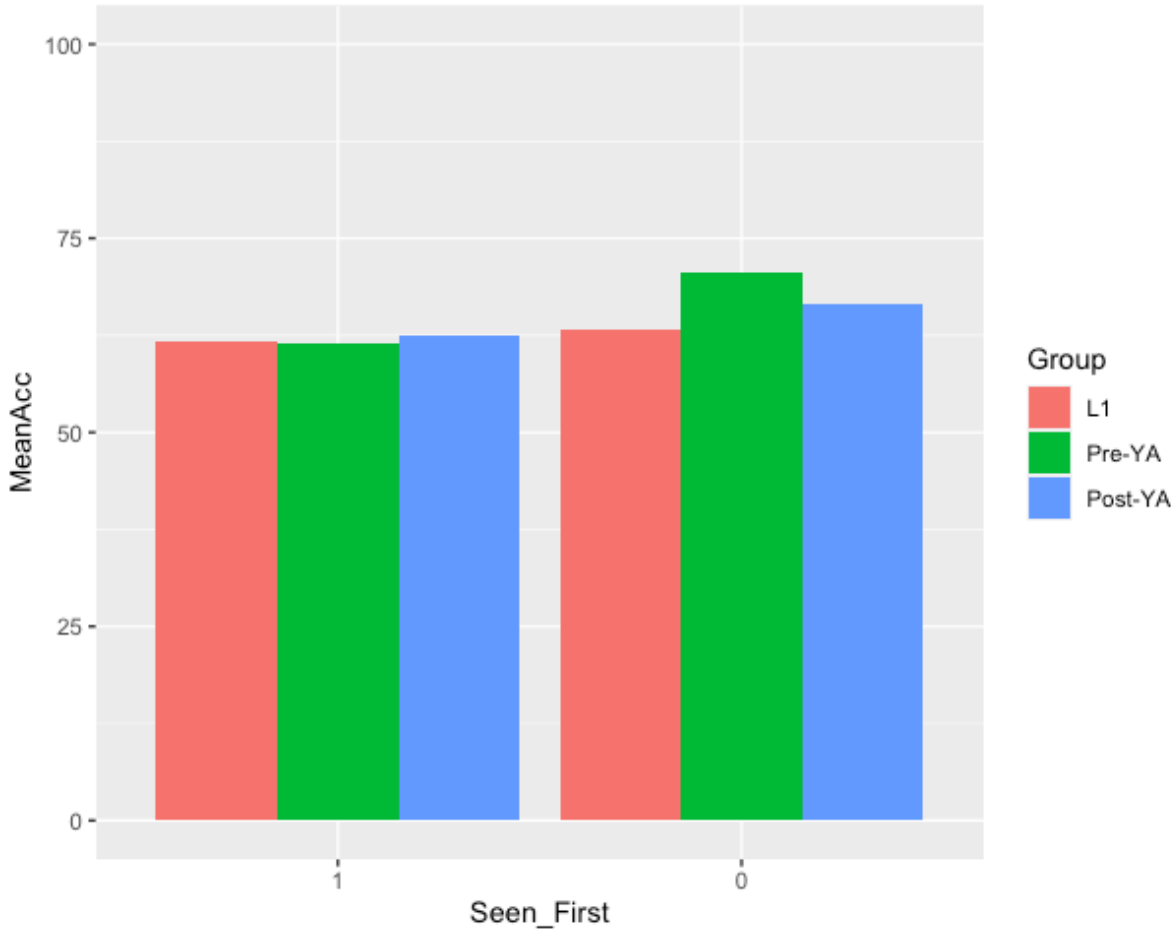


Figure 29: Graph of AJT accuracy according to values of 'Seen_First', per group.

As can be seen in Figure 29, accuracy for all groups remains very similar when viewing a context (and its associated item) for the first time. On seeing the context for the second time, however, it can be seen that accuracy increases for both L2 groups – particularly the Pre-YA group – whilst the L1 group do not appear to be especially

affected. It is at this point that the demographic differences between the L2 and L1 may be worth returning to.

Firstly, the higher accuracy seen for the L2 groups in the *Seen_First* = 0 condition could be said to indicate a "learning effect" for these groups: that is to say, on seeing the context for the second time, they realised that the focus of their acceptability rating should be the form of the verb, given that this was the only element that changed between the first and second presentation of the context. The fact that this pattern does not hold for the L1 group suggests that this group were perhaps less aware of this. This may be attributable to the fact that, whilst the L1 groups' occupations were unknown, the L2 groups were composed specifically of university-level languages students. This may have provided the L2 groups with a performance advantage in the AJT, given that, as languages students, they were likely more accustomed to focussing on verb form. Schütze (1996/2016) refers to a seminal work by Spencer (1973;87), who asserts that 'the linguist views language in a highly specialised way'. Though the members of the L2 groups were perhaps not fully-fledged linguists, they had nonetheless received many years of L2 instruction, sometimes for multiple languages (heightening their metalinguistic awareness), and many would have taken introductory linguistics courses as part of their degrees. All of these factors may have provided the L2 groups with a different perspective on the AJT in comparison with the L1 group, explaining not only the potential L2 "learning effect" but also the unexpectedly low accuracy of the L1 group. Although this variable does not contribute significantly to the fit of the accuracy model, the all-groups ratings analyses did reveal a significant main effect of *Semantic_connectedness*, which on further exploration appeared particularly pronounced for the L1 group. Taken together, the above findings suggest that L1 vs. L2 (instructed) learners of French may draw on different information when making viewpoint aspectual distinctions in a judgement task like the one used in this project.

There are also additional demographic factors that may have contributed to the lower-than-anticipated accuracy of the L1 group that might be borne in mind. As shown in Table 27, the mean age for the new L1 group was 34.1, making the average L1 participant over ten years older than the average L2 participant. Additionally, the L1 group were likely to be much more heterogeneous in terms of level of education: whilst all of the L2 group were currently pursuing an undergraduate degree, this was not

controlled for in the new L1 sample. Hulstijn (2011) highlights that levels of education are likely to have a particularly strong impact on how different participants perform on any given linguistic task, as education levels are likely to be (at least broadly) an indicator of participants' intellectual capacities, as well as their reading and writing abilities. Schütze (1996/2016) also acknowledges the importance of individual differences, or "organismic factors", to participants' performance in judgement tasks specifically, noting that 'a number of extragrammatical factors often implicated in acceptability… might be subject to inherent differences' (p.111). Among these "extragrammatical factors" Schütze cites working memory capacity, which, alongside other general cognitive capacities, has been shown to decline with age (e.g. Logie & Morris 2014). Therefore, the combination of their greater age with their more varied level of education may partly explain simultaneously the lower than unexpected performance by the L1 group in the AJT task, and the fact that the L2 groups appear to be less affected.

In short, not only the homogeneity of the L2 groups' *level* of education, but also their *subject* of education, may have provided them with an additional advantage. Contrastingly, it seems reasonable that the L1 group, coming from a diverse range of backgrounds, may not be as necessarily attuned to the linguistic content of the sentences they were rating, and instead may have focussed more on how semantically "good" or congruent the sentence was in light of the preceding context, which would explain why semantic connectedness appeared to more markedly affect this group in the ratings model. It is also worth mentioning here that when the original Spanish items for this task were conceptualised for the SPLLOC project (Mitchell et al 2008), the researchers tried to reduce as much as possible the use of temporal adverbials (such as yesterday, often, every day) in the items to rate in order to minimise priming; however, this may also have reduced the connectedness between contexts and items and further (disproportionately) affected the L1 group.

To summarise the findings emerging from the all-groups comprehension accuracy model, it was found that accuracy was significantly higher in perfective than combined imperfective conditions: this is likely attributable to the particularly low rates of accuracy in continuous conditions. Moreover, both L2 groups demonstrated significantly higher accuracy when rating target forms than rating non-target forms. This

may indicate a difficulty in rejecting non-target forms, suggesting possible difficulties in fully reassembling viewpoint aspectual form-meaning mappings in their L2 (despite indicating a clear target vs. non-target differentiation in the ratings data). These avenues of investigation will be explored more fully in the following subsection.

In terms of variables that were not present in the best-fitting accuracy model, the most notable absence is arguably that of a main effect of *Group*, indicating that there is not a substantial difference between the accuracy of the L2 vs. the L1 groups in the AJT. This could be attributable to a range of unintended demographic differences between the L2 and L1 groups, without which the data for the L1 comprehension group may have patterned significantly closer to that of the L1 production data. Also to consider is the potential fact that, as the L2 groups are relatively advanced, there may indeed not be significant between-groups differences in viewpoint aspectual comprehension. In this case, the non-ceiling scores across the board may be attributable to elements of the task design. The latter idea will be returned to and further expanded upon in Section 6.4.

L2 Accuracy Modelling

The principal aim of carrying out accuracy analyses specifically on the L2 data was to look more closely at differences between the L2 groups, as opposed to in comparison with the L1 group. This has the potential to inform on the role of time spent abroad on L2 viewpoint aspectual development. This can also be explored from a point of view of language exposure in the L2 model, via the *LEQFren* and *LEQEng* variables measuring learners' self-assessed degree of exposure to French and English, respectively. As in the L2 production model, *LEQFren* was reverse Helmert coded, meaning that comparisons between the exposure categories were made as follows:

1. Medium vs. Low
2. High vs. [Medium + Low]

LEQEng contained only two levels (Medium and Low), so was dummy coded with Medium as the reference category.

In additional to any exposure-related differences, any significant accuracy differences related to aspectual condition that appear in the L2 model could build upon the already observed patterns found in the all-groups model, providing potential further insight into the ease of acquisition of each of the form-meaning mappings and how closely this is aligned with Feature Reassembly predictions.

The approach to fitting a model for the L2 accuracy data was the same as that described for the all-groups data, again using logistic mixed-effects regression. The formula for the best-fitting model (containing the most maximal random effects structure possible) is presented below, with output in Table 29.

Formula: Accuracy ~ Group + LEQFren + Group:LEQFren +

Form + LEQFren + Form:LEQFren +

LEQFren + LEQEng + LEQFren:LEQEng +

Seen_First +

(1 + Seen_First | Partic_ID) + (1 + Seen_First | AJT_context)

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	0.203294	0.354170	0.574	0.56597
GroupPost-YA	0.062612	0.287787	0.218	0.82777
LEQFren1	-0.794938	0.353980	-2.246	0.02472 *
LEQFren2	0.031037	0.122861	0.253	0.80056
FormTarget	0.714582	0.116860	6.115	9.66e-10 ***
LEQEngHigh	-0.043202	0.240151	-0.180	0.85724
Seen_First0	0.300692	0.135090	2.226	0.02602 *
GroupPost-YA:LEQFren1	0.610883	0.288474	2.118	0.03421 *
LEQFren1:FormTarget	-0.365295	0.135586	-2.694	0.00706 **
LEQFren2:FormTarget	-0.109868	0.085614	-1.283	0.19939
LEQFren1:LEQEngHigh	0.893869	0.312086	2.864	0.00418 **
LEQFren2:LEQEngHigh	-0.001781	0.154756	-0.012	0.99082

Table 29: Output from L2 accuracy model (logistic mixed-effects regression).

It can be seen that there is no significant main effect of *Group* in the L2 accuracy model (p = 0.8); this is not overly surprising in view of the similar accuracy scores for the L2 groups shown in Figure 26, at the start of this subsection. However, there is a significant interaction (z = 2.118, p = .034) between *GroupPost-YA* and the first level of *LEQFren* comparisons, i.e. Medium vs. Low exposure. There is also a significant main effect of this first level of comparisons of *LEQFren* (z = -2.246, p = .025). Despite this, mean accuracy scores from both groups across the LEQFren categories are in fact quite stable, spanning a small range of 63.5-66.0%. To gain further clarity on the role of *LEQFren* on accuracy, it is perhaps more insightful to look at a more strongly significant interaction in the model: that between the first level of *LEQFren* comparisons and accuracy when rating the target form. This relationship is presented graphically in Figure 30, which displays accuracy in target vs. non-target conditions, sub-divided by *LEQFren* category.

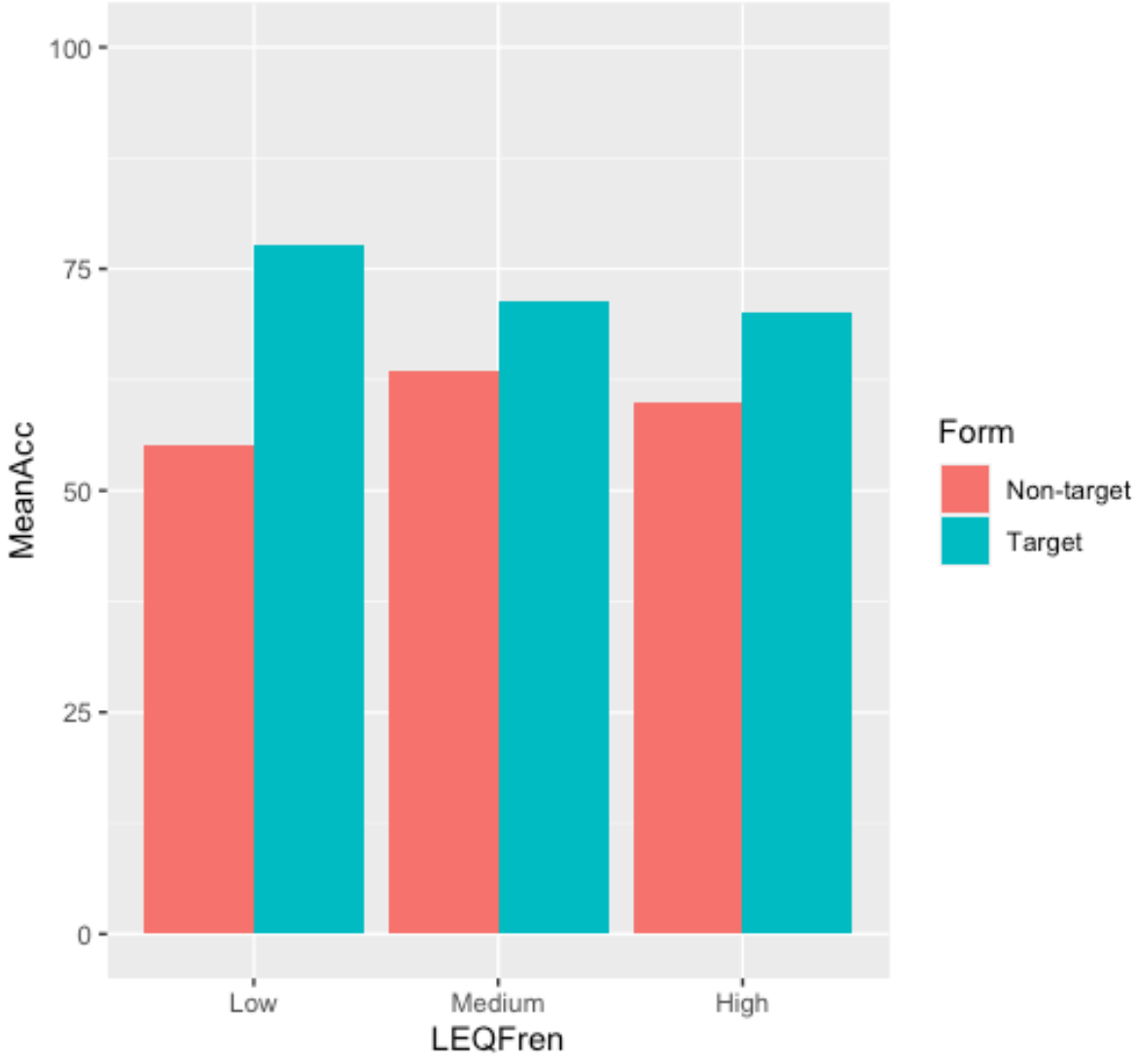


Figure 30: Graph of L2 AJT accuracy in target vs. non-target conditions, per LEQFren category.

In contrast to the fairly consistent levels of accuracy across *LEQFren* categories per group, some larger contrasts are shown in Figure 30, wherein the categories are divided by *Form*. As may have been anticipated, higher accuracy is observed in the *FormTarget* condition for each category of *LEQFren*; this is aligned with the highly significant main effect of *Form* ($z = 6.115$, $p < .001$), indicating that the L2 participants were indeed more accurate at accepting the target form then at rejecting the non-target form. This pattern was alluded to in the all-groups rating model, but its presence in the L2 model confirms that, while learners appear to have been relatively successful in forging new form-meaning mappings (such as linking the Imperfect to habituality and/or continuousness in their L2 French representations), they may have been less successful at severing the connection to supplementary "unhelpful" L1 mappings (e.g. the partial/total mapping of habituality/continuousness to the Simple Past in their L1 English representations). As for the relationship between *LEQFren* and accuracy in target conditions, it can be seen that, while this appears quite similar in the Medium vs. High exposure conditions, accuracy in the Low exposure condition is higher. This finding, which echoes a similar observation found in the L2 production accuracy analysis, initially appears incongruous – why would a lower exposure to French result in higher accuracy? However, it should be kept in mind that the majority of members of the Low group were the Pre-YA participants. In the discussion of the production analyses, it was pointed out that the Pre-YA group performed on the whole more accurately than the Post-YA group in production, thus explaining the elevated accuracy of the Low exposure group. This point is a little more difficult to make for the comprehension data, given that both L2 groups' performance was broadly similar (with the Post-YA group perhaps holding a slight upper hand, descriptively). Nonetheless, the significance of the trends depicted in Figure 30 indicate a potential advantage for the Pre-YA instructed group vs. the Post-YA group with regard to accurate rating of target forms in the AJT.

A final significant interaction pertaining to the exposure variables arises between *LEQEngHigh* and*LEQFren1* ($z = 2.864$, $p = .004$). Figure 31 displays how accuracy varies according to levels of both French and English exposure.

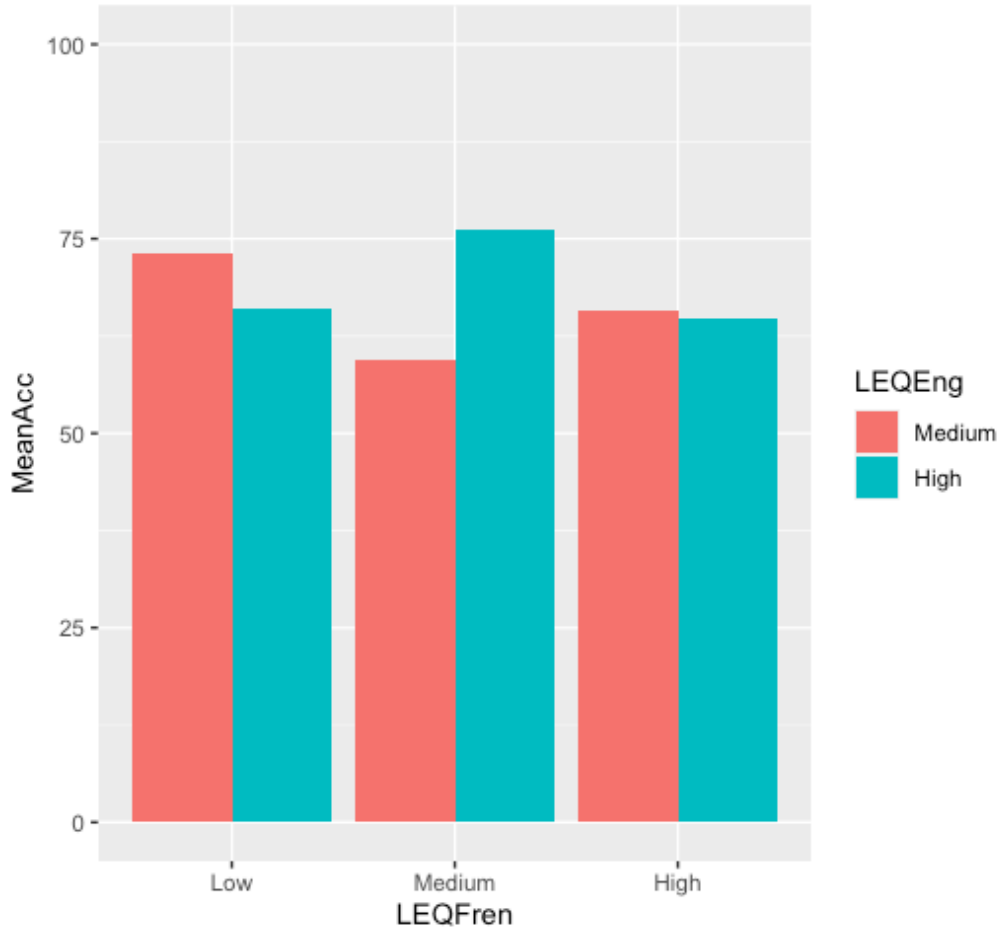


Figure 31: Graph of L2 AJT accuracy, divided by categories of LEQFren and LEQEng.

It might be expected that a higher level of French exposure and a lower level of English exposure may both have a beneficial influence on L2 French accuracy. Considering Figure 31, it can be seen that, when exposure to French is high, level of English exposure does not appear to impact on accuracy; however, when level of French exposure is Medium or Low, the level of exposure to English does have an impact. Specifically, when French exposure is low, higher accuracy is observed when exposure to English is also lower (note, no participants fell into the Low category for English), but when French exposure is moderate, a higher exposure to English is associated with higher accuracy. In short, the trends in Figure 31 do not present the clear picture that one might expect with regard to language exposure and performance on this acceptability judgement task. At this point, it is worth noting that participants who reported a high level of French exposure did not necessarily report a low level of English exposure, and vice versa (see Figures 8 and 9 for the full distribution of L2 participants

across LEQ categories), and so it is difficult to comment fully on the interaction between LEQ variables shown here.

Lastly, *Seen_First* continues to be a significant main effect ($z = 2.226$, $p = .03$) in the L2 accuracy model, as it was in the all-groups accuracy model. Figure 32 presents mean accuracy per group in each of the *Seen_First* conditions.

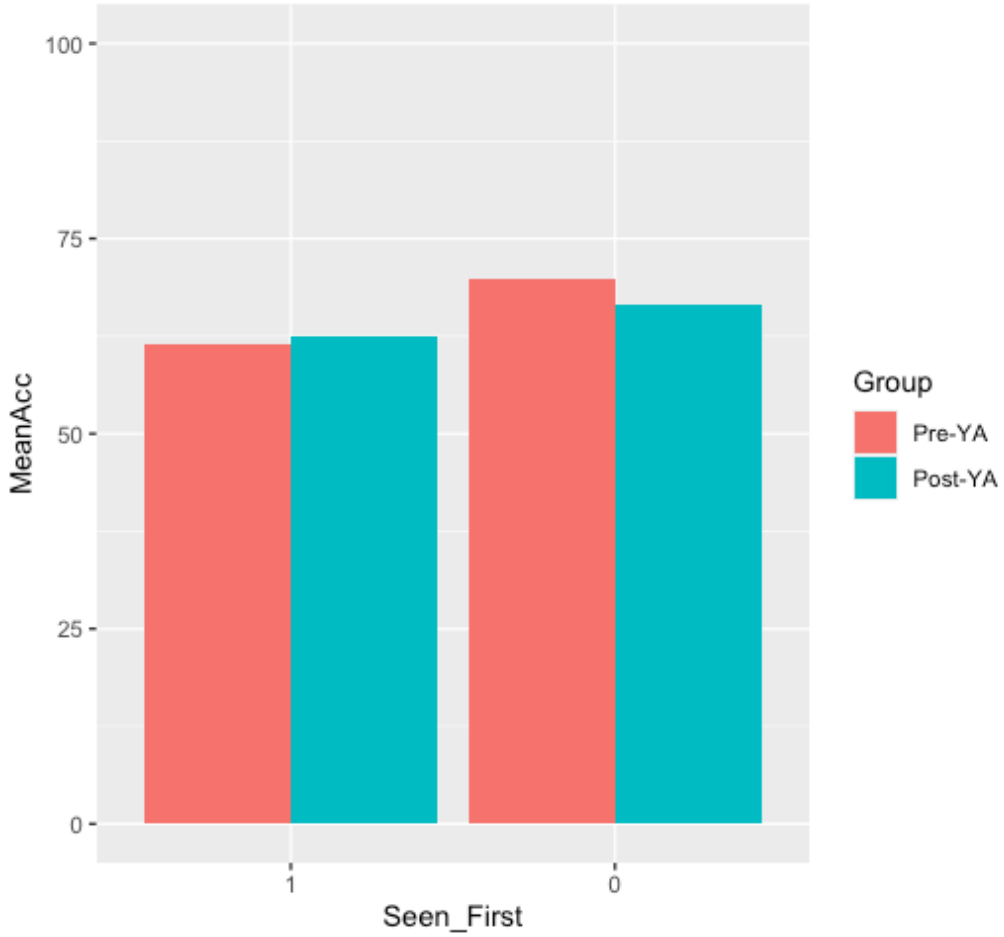


Figure 32: Graph of L2 AJT accuracy per group and Seen_First condition.

As can be seen, both groups perform more accurately upon seeing a context for the second time (*Seen_First* = 0). This provides further weight to the previous suggestion of a "learning effect" that is specific to the L2 groups, perhaps as a result of their background as languages students. This highlights the importance of maintaining awareness of participants' demographic backgrounds when designing tasks to gather linguistic data, particularly in view of the fact that universities are popular sources of participants for researchers. Suggestions for how this learning effect could have been minimised – or at least equalised across all three participant groups – will be made in Section 6.4.

In summary, the lack of a main effect of *Group* in the L2 accuracy model suggests that there is no significant difference in the Pre- and Post-YA group's comprehension of L2 French viewpoint aspectual form-meaning mappings. With regard to the potential impact of time spent abroad, the findings regarding language exposure are not especially clear, which may be linked to the fact that the year abroad setting was not synonymous with High French exposure and Low English exposure, but instead encompassed a wide range of levels of exposure to both languages (some of which inevitably will not have been captured via LEQ responses). There is some evidence that the Low French exposure group (composed primarily of Pre-YA participants) perform more accurately – particularly with regard to accepting target forms – but all learners show higher accuracy in target vs. non-target conditions. This may suggest that, even though both learner groups display an appropriate differentiation between target and non-target form in terms of their ratings (as explored earlier), their overall accuracy is higher when accepting target forms than when rejecting non-target forms. This may indicate a difficulty not in creating new form-meaning mappings in the L2, but in breaking ties with non-facilitative form-meaning mappings linked to the L1. Despite the absence of *Target_Aspect* in the L2 accuracy model specifically, the findings from the all-groups ratings model do point towards some influence of Feature Reassembly in participants' responses to the AJT. Finally, it is hoped that the discussion surrounding the unexpected behaviour of the L1 French group in the AJT highlights the methodological importance of considering extralinguistic factors (and their associated linguistic repercussions) during participant recruitment and data analysis.

The impact on L2 French viewpoint aspectual development of both feature reassembly (RQ1) and exposure to naturalistic input during the year abroad (RQ2) has been highlighted at multiple intervals throughout the above presentation of results. In the following chapter, the roles of both of these variables, both separately and in tandem, will be discussed in further detail, with reference to both the experimental and corpus data presented here.

6. Discussion and Conclusion

As a reminder, the three research questions explored in this thesis are as follows:

> **RQ1:** To what extent can predictions from Feature Reassembly accommodate the developmental path of the viewpoint aspectual system in L2 French for L1 English university-level learners?

> **RQ2**: What can studying the properties of naturalistic input tell us about the impact of programmes such as "the year abroad" on grammatical development, with a particular focus on the development of viewpoint aspect?

> **RQ3**: To what extent can an approach combining information from feature reassembly with information from the input explain the process of L2 (aspectual) development at advanced levels?

Some preliminary discussions surrounding RQ's 1 and 2 appeared in the previous chapter; these will be returned to and further developed(in Sections 6.1 and 6.2 respectively. These findings will be brought together as a basis to address RQ3 in Section 6.3, along with some conclusions. Finally the limitations of the present project, as well as suggestions for future research, will be discussed in Section 6.4.

6.1 Feature Reassembly and viewpoint aspectual development for L1 English learners of L2 French

A core reason proposed for the well-attested challenging nature of acquiring viewpoint aspect in an L2 for L1 English speakers lies in the differences in how English grammaticalises aspectual meanings, especially in comparison with Romance languages such as French or Spanish (Montrul & Slabakova 2002, 2003; Izquierdo & Collins 2008; McManus 2015; Domínguez et al 2017). For this reason, Feature Reassembly (Lardiere

2003, 2005, 2008, 2009) has become a popularly-espoused means to approach the L2 acquisition of viewpoint aspect, as it enables researchers to clearly visualise the details of the learning task – conceptualised in terms of remapping L1 meanings to the correct L2 forms – and to make predictions regarding the ease or difficulty of a particular form-meaning mapping based on the degree of reassembly that is required. As a reminder, the reassembly task for the acquisition of the L2 French viewpoint aspectual system by L1 English speakers (first shown in section 3.2.4) is displayed in Figure 33 below:

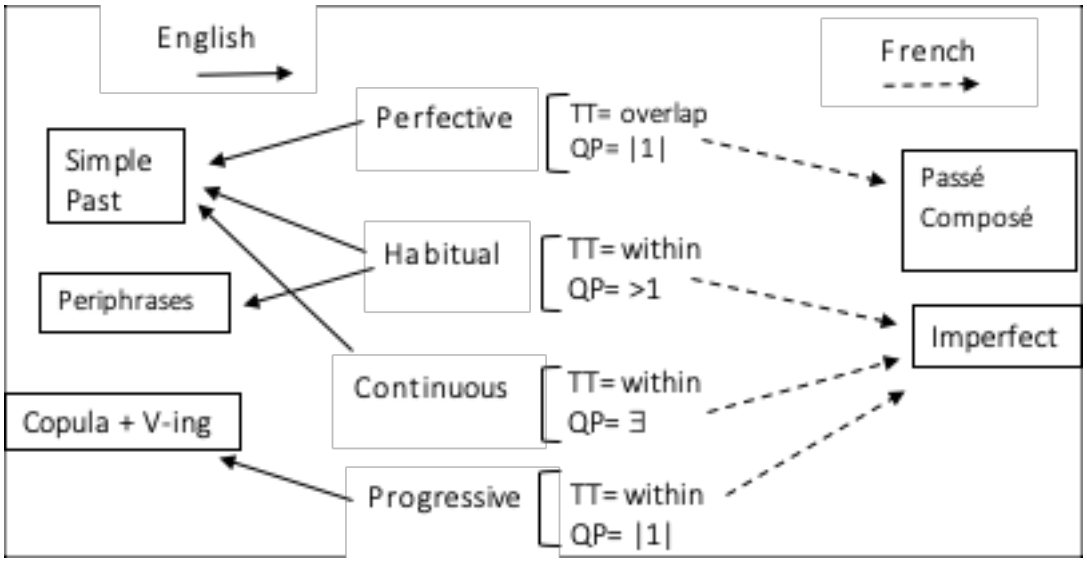


Figure 33: Diagram of feature reassembly task for viewpoint aspectual form-meaning mappings between L1 English and L2 French.

(adapted from Domínguez et al. 2017:7)

For the L1 English/L2 French pair, Feature Reassembly predicts that, while reassembling the perfective meaning from the Simple Past to the *passé composé* will be relatively straightforward, reassembly of the three imperfective meanings (habitual, continuous, and progressive) will be more challenging. This is linked to the fact that, whilst there is a one-to-one mapping between the Simple Past and the *passé composé* for perfectivity, imperfectivity is split across a range of forms in English, and learners must identify and disentangle these three meanings before reassembling them all onto the Imperfect in French. Moreover, some imperfective meanings are anticipated to be more challenging to reassemble than others: whilst the imperfective-progressive constitutes a one-to-one mapping and thus is expected to be relatively unproblematic,

both the habitual and continuous meanings can be expressed using the Simple Past, and so must be disentangled from the perfective meaning in order to be correctly reassembled. Of these two latter meanings, the continuous is predicted to be the most challenging to reassemble for learners, as it is the only imperfective meaning that is hosted solely on the Simple Past (habituality can also be expressed using periphrases such as *used to* + V or *would* + V). To summarise, then, Feature Reassembly would predict the following difficulty cline for L1 English learners acquiring the L2 French viewpoint aspectual system (in ascending order of difficulty): Perfective < Progressive < Habitual < Continuous. However, several studies of L2 viewpoint aspectual acquisition (e.g. Domínguez et al. 2011, 2017) have reported findings that do not fully align with feature reassembly predictions. As has already been discussed in Chapter 5, the findings presented in this project also do not completely align with feature reassembly. A summary of the results and their congruence with feature reassembly predictions are presented in Table 30.

	Difficulty/accuracy cline
FR predictions	**Least difficult** → **Most difficult** PERFECTIVE PROGRESSIVE HABITUAL CONTINUOUS
L2 production accuracy	**Most accurate** → **Least accurate** PERFECTIVE CONTINUOUS PROGRESSIVE HABITUAL
L2 comprehension accuracy	**Most accurate** → **Least accurate** PROGRESSIVE *[PERFECTIVE]* * HABITUAL PERFECTIVE CONTINUOUS

Table 30: Accuracy clines across viewpoint aspectual conditions for the L2 production and comprehension data in this project, in comparison with Feature Reassembly (FR) predictions.

** See below for further discussion on perfective comprehension accuracy.*

Table 30 presents the difficulty cline predicted by Feature Reassembly for the four form-meaning mappings of the French viewpoint aspectual system for L1 English speakers (shown in increasing order of difficulty), alongside the accuracy results of the L2 French learners in this study, in both production and comprehension (presented in

decreasing order of accuracy). If feature reassembly were the only factor influencing learners' aspectual development, we would expect all three clines to look the same, as learners would be the most accurate with the least difficult mappings according to Feature Reassembly and vice versa. However, as can be seen, this is not the case. Focussing first on production, we can see that although the perfective is the most accurate mapping (as Feature Reassembly would predict), learners' accuracy across the imperfective mappings does not line up with their respective degrees of reassembly. Most notably, the continuous, despite entailing the most complex reassembly task, is the most accurate imperfective meaning. In contrast, the progressive is not the most easily-acquired of the imperfective meanings, despite comprising a one-to-one mapping. Lastly, the habitual is the least accurate meaning although it is not associated with the most complex reassembly task. Taken together, the accuracy cline seen in the L2 production data suggests that, while Feature Reassembly can account for some of the findings – such as the perfective being more straightforward than the imperfective meanings – it appears only to be able to accommodate the full picture of these learners' L2 French viewpoint aspectual development to a certain extent.

A different picture again is seen when the L2 comprehension data is considered, where, at least regarding the imperfective meanings, the predictions of Feature Reassembly are upheld: learners perform with the greatest accuracy in the progressive condition and the lowest accuracy in the continuous condition, with the habitual condition falling in between these. With regard to the perfective, however, learners' low comprehension accuracy is unexpected: both from a Feature Reassembly perspective and additionally in light of learners' high production accuracy in this mapping. This said, it is worth referring back to Figure 26, where it can be seen that the accuracy differences between all four viewpoint aspectual mappings in the comprehension data were small: for instance, L2 accuracy in the habitual condition ranged from 68.3-71.1% on average, and in the perfective condition from 64.1-65.5% on average. With such small differences across aspectual conditions, it is possible that the presence in a specific condition of one or two "problematic" items in the AJT would have been sufficient to skew the order of the accuracy clines presented above.

On a return to the raw data, it is apparent that a few items in the perfective condition were indeed disproportionately poorly-answered, which may suggest an

influence of factors outside of the viewpoint aspectual contrast. In the two problematic contexts in question – PerfEv1 and PerfSta1 (see Appendix A.3 for full details of AJT items) – there is both a markedly higher overall acceptance of the (non-target) Imperfect (PerfEv1: 56.6%; PerfSta1: 86.8%) compared to other perfective items (mean IMP rating: 28.0% [11.23]), and a relatively lower acceptance of the (target) *passé composé* (PerfEv1: 62.2%; PerfSta1: 59.8%; mean PC rating in other perfective items: 81.0% [15.9]). These findings suggest that, despite best efforts to ensure every item yielded the correct aspectual interpretation, it is possible that another element of these two contexts skewed learners' ratings. There is also some evidence that these items were problematic for the control group: PerfSta1 yielded low degrees (23.5%) of Imperfect rejection (but was targetlike in PC acceptance), whilst PerfEv1 appeared problematic in terms of both Imperfect rejection (35.3%) and PC acceptance (47.1%). Based on these observations, L2 accuracy per context was recalculated excluding the two problematic perfective AJT items. This resulted in a reworked L2 perfective accuracy ranging from 71.7-73.0%, placing learners' perfective comprehension accuracy above the habitual and near-equivalent to the progressive (73.0-75.5% accuracy). This reworked accuracy cline (indicated by the square bracketed *[PERFECTIVE]* in Table 30) is more aligned with FR predictions, with the two one-to-one mappings constituting the most accurate aspectual conditions, as well as supporting the literature surrounding the relative ease of acquisition of perfective over imperfective meanings in general (e.g. Harley 1978, 1992; Bergström 1995; Montrul & Slabakova 2002, 2003).

Having addressed the unexpected discrepancy in the perfective condition for the comprehension data, the difference in learners' performance in production vs. comprehension for the imperfective meanings is nonetheless interesting to consider in light of RQ1. If Feature Reassembly were the only variable predicting learners' acquisition of viewpoint aspectual form-meaning mappings, one would probably expect the same accuracy clines to be present in both production and comprehension data. It is worth noting here that a similar discrepancy in L2 performance in comprehension vs. production data was also observed in Domínguez et al. (2017), whose L1 English learners of L2 Spanish displayed targetlike patterns of acceptance and rejection for the imperfective-habitual in a comprehension task, but were seen to overextend the perfective Preterite in habitual contexts in production data. A possible explanation for

the production vs. comprehension discrepancy in this project will be returned to later on.

Another unexpected finding from the production data in this project is the fact that the Post-YA group were overall slightly less accurate than the Pre-YA group in all aspectual conditions. This runs counter to expectations surrounding the supposed impact of time spent abroad: given that the Post-YA group were demonstrated to have overall been exposed to more French input from a wider range of sources (see Figures 8 and 9 in Section 5.1.1) during their year abroad, shouldn't they be *more* accurate? This perception of the positive impact of residence abroad is rooted in the findings of the study abroad literature, which have widely posited that students return from the year abroad with improved proficiency, at least in terms of oral skills such as fluency (e.g. Freed 1995, Segalowitz & Freed 2004; Kinginger 2008; Llanes & Muñoz 2009) (although, as discussed in Section 3.3, the picture in other areas of language development is less clear). Indeed, the Post-YA group did score slightly more highly than the Pre-YA group on the independent proficiency measure (see Table 2 in Section 4.1.1). The proficiency results also render the results of the production data surprising from a feature reassembly perspective, as the form-meaning mapping process has been suggested to develop as a function of proficiency (Montrul & Slabakova 2002, 2003; Salaberry 1999, 2002, 2003, 2005, 2008). Therefore, the fact that the L2 group, who are (at least descriptively) more proficient, display less evidence of target-like reassembly in production arguably represents another counter to the idea that feature reassembly is the sole factor underpinning learners' viewpoint aspectual development.

To explore these unexpected production accuracy results more closely, the full range of verbal forms used by learners in aspectual conditions was considered, and compared across the L2 groups. The visual summary of forms used in both the Cat Story and Conversation tasks (first presented in Section 5.1.2) is presented again here in Figures 34 and 35 as a reminder.

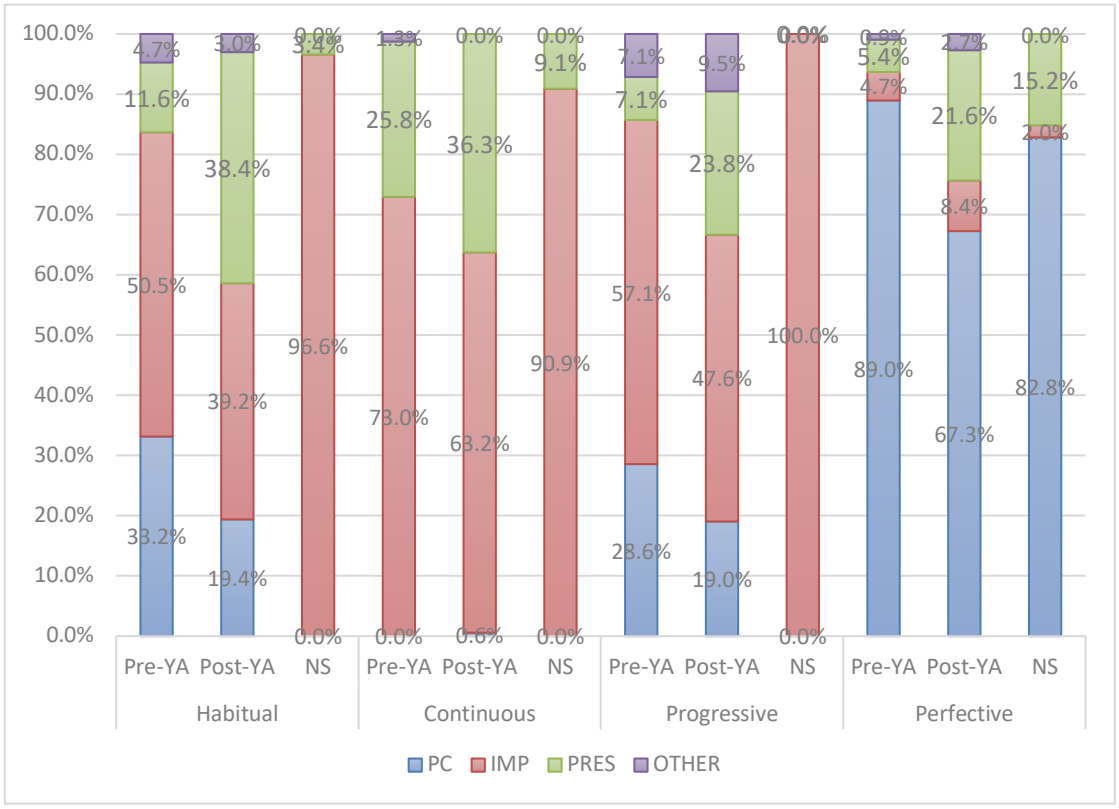


Figure 34: Proportions of forms used by aspectual condition by Pre- and Post-YA L2 groups and the control group (NS) in the Cat Story task.

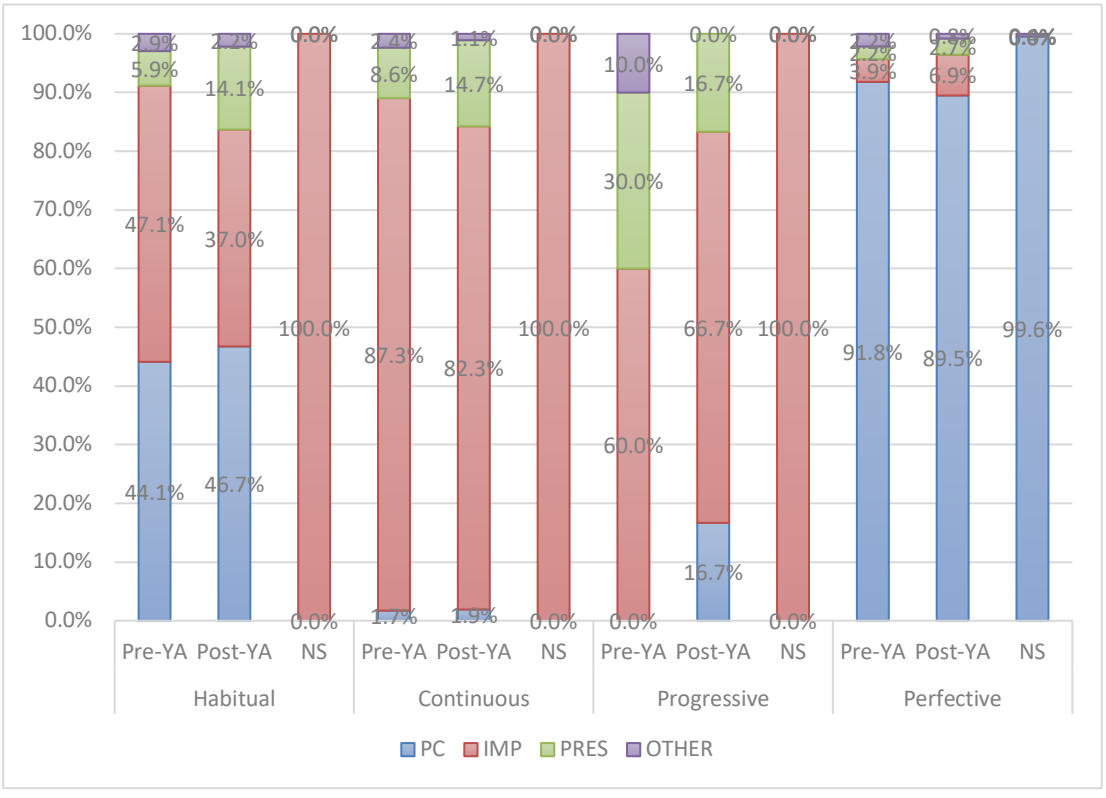


Figure 35: Proportion of forms used by aspectual condition by the Pre- and Post-YA L2 groups and the control group (NS) in the Conversation task.

The key finding observed in terms of learners' non-target production was the use of the Present by the Post-YA group. The results of the analyses conducted in Section 5.1.2 confirmed that the Post-YA group utilised the non-target Present in imperfective contexts in a way that significantly differentiated them from the control group in both production tasks, and that Post-YA Present usage was significantly impacted by aspectual condition in the Cat Story task, where they produced the Present significantly more in the habitual condition than the perfective condition. Importantly, no such significant results were seen in the forms used by the Pre-YA group. Taken together with the lower suppliance of the target form in every aspectual condition by the Post-YA group, the concomitant higher suppliance of the Present by this L2 group suggests that the Post-YA learners are relying on using the Present to express viewpoint aspectual meanings in a way that the Pre-YA learners are not.

What possible reasons can be advanced for this? Post-YA learners' use of the Present cannot be linked to influence from the L1, as it does not occur in cases where English would use the Present (indeed, all the form-meaning mappings of interest in this project encode past temporal reference in both English and French). In addition, though there is some evidence that both L2 groups utilise the perfective *passé composé* significantly more than the control group in the habitual condition (which is likely to indicate ongoing L1 influence), there is no evidence to suggest that the Post-YA group use the *passé composé* more than the Pre-YA group in either production task. Indeed, the results from the habitual condition in Figure 34 (Cat Story) suggest, at least descriptively, that the Post-YA group use less *passé composé* than the Pre-YA group, and also that *passé composé* usage for both groups in the continuous condition (where we might also expect to see L1 influence) is negligible. These patterns indicate the lower production accuracy of the Post-YA group cannot be attributed to some kind of "undoing" of the feature reassembly process, as learners are plainly not reverting to their L1 form-meaning mappings.

Instead, it could be suggested that the surge in non-target Present usage for the Post-YA group represents a kind of "placeholder", not dissimilarly to what is proposed in Prévost & White's (2000) Missing Surface Inflection Hypothesis (MSIH), wherein the higher cognitive demand associated with producing spontaneous speech results in optionality in L2 acquirers' use of correctly-inflected forms and the subsequent optional replacement of these with a simpler form. Prévost & White studied

the use of verbal inflections in adult learners of French and German, and observed that, even if inflections are not always supplied when required, inflected (finite) verbs are never produced in non-finite contexts where they are not required. The authors conclude that these systematic distributions provide evidence for the fact that learners do have the requisite abstract knowledge of tense and agreement in their L2, but that actually producing the inflection on the "surface" is more challenging, leading to learners sometimes resorting to non-finite forms. Although the Present is not non-finite, we suggest that its less complex morphology and computation may make it a valid candidate for learners to "fall back on" in their attempt to produce the more complex inflections of either the Imperfect or *passé composé*.

The ideas underpinning the MSIH can be extended to help explain some key observations in the data, notably the production vs. comprehension discrepancy: if 'overt reali[s]ation of morphology' (Prévost & White 2000:104) is an area of difficulty for L2 learners, this aligns with the fact that, whilst there are disparities between the Pre- and Post-YA groups in production, the two groups behaved near-identically (in terms of both overall accuracy and ratings of target/non-target forms) in the comprehension task (see Section 5.3.2). This supports the central proposal of the MSIH: that is, while production may be variable, learners' underlying representations (in this case, their knowledge of French viewpoint aspectual form-meaning mappings) are intact. Moreover, a core tenet of the MSIH is that the use of placeholder forms by learners occurs as a result of 'mapping problems' (Lardiere 2000) between forms and abstract (semantic) features (Prévost & White 2000: 127), so the proposal of the existence of "placeholder Present" in the Post-YA production data meshes very well with the complex reassembly task for L2 French viewpoint aspect that the learners in this project are in the middle of.

Although the Missing Surface Inflection Hypothesis is able to successfully accommodate some important findings, it is necessary to explore why this "falling back" onto the Present is observed in some conditions more than others. Firstly, the fact that more marked Present usage is seen in the Cat Story can be quite readily accounted for. Given that this task is more structured, it may have imposed greater cognitive demands on learners, who were obliged to produce viewpoint aspectual mappings in line with the specific contexts elicited rather than being able to circumvent areas of weakness

such as was possible in the more open-ended Conversation task. This cross-task contrast also highlights the methodological importance of collecting production data from a range of tasks that vary in terms of characteristics such as degree of structuredness in order to gain a representative picture of learner production, as has already been emphasised with regard to L2 aspectual research (Labeau 2005, 2011; Domínguez 2019).

Less completely understandable from a Feature Reassembly perspective is the extent to which we observe Present fallback across the four aspectual conditions. If it is being posited that the Post-YA learners are using the Present as a "placeholder" form in instances where it is too cognitively challenging to produce the target form, it would be logical to anticipate the highest rates of Present usage in the conditions where the most reassembly is required. However, this is not found to be entirely the case. Tables 31 and 32 display each L2 group's Present suppliance by aspectual condition in Cat Story (where it was most utilised) and in overall production respectively. Aspectual conditions are presented left to right in ascending order of production accuracy.

"Placeholder Present" suppliance: Cat Story				
	Habitual	Progressive	Continuous	Perfective
Pre-YA	11.6%	7.1%	25.8%	5.4%
Post-YA	38.4%	23.8%	36.3%	21.6%
% Diff.	+ 26.8%	+ 16.7%	+ 10.5%	+16.2%

Table 31: Present suppliance by L2 group and aspectual condition in the Cat Story task. "% Diff." shows the difference in suppliance between the two groups.

"Placeholder Present" suppliance: Overall production				
	Habitual	Progressive	Continuous	Perfective
Pre-YA	10.7%	16.7%	13.0%	3.5%
Post-YA	27.6%	20.5%	17.4%	9.3%
% Diff.	+16.9%	+ 3.8%	+ 4.4%	+ 5.8%

Table 32: Present suppliance by L2 group and aspectual condition in overall production. "% Diff." shows the difference in suppliance between the two groups.

As can be seen, Tables 31 and 32 show a positive percentage difference (% Diff.) between the Present suppliance of the Pre- and Post-YA groups in every aspectual condition, both in Cat Story and overall. This patterns with previous discussion of the characteristic use of non-target Present across the board by the Post-YA group, as does the fact that there is greater Present fallback in the more structured and cognitively-demanding Cat Story task (in line with the Missing Surface Inflection Hypothesis). However, it can also be seen that the amount by which the Post-YA group produce *more* Present than the Pre-YA group does not align exactly with the degree of reassembly for a given mapping in every case. For example, Table 32 shows a relatively small difference in increased Post-YA Present usage between the imperfective-continuous and -progressive in overall production, despite the fact that the latter has a much more straightforward reassembly task than the former. This divergence from feature reassembly predictions is all the more apparent in Cat Story (Table 31), where is observed a larger increase in Post-YA fallback Present in progressive over continuous contexts. This may be partially explained by the fact that there is (as mentioned) increased Present usage by the Post-YA group in every aspectual condition, and that - at least in Cat Story - incidence of the Present in progressive contexts was much lower than in continuous contexts to start with. This reasoning may also explain the seemingly relatively high increase in Present suppliance for the perfective context, which was found to be the most accurate viewpoint aspectual mapping in production overall.

Beyond this, however, it is noted that patterns of Present usage do approximately align with overall accuracy for the two extremes of the spectrum: that is to say, there is the greatest percentage increase of Present fallback for the habitual mapping, which is the least accurate aspectual condition, and the continuous has among the lowest rates of increased Present use and is also the most accurate of the imperfective mappings. This arguably supports the suggestion that the Post-YA learners are utilising the Present as a placeholder, particularly for aspectual mappings that already appear challenging to reassemble. It is also worth re-emphasising the differential frequencies of the form-meaning mappings here: for example, the continuous, which shows very little Present fallback, is very frequent in both learner production and the corpus sample, whilst the habitual, with its higher incidence of Present fallback, is much less frequent.

Collectively, exploring the extent to which Post-YA learners utilise the Present as a "placeholder" form provides some support for the idea that learners' differential production accuracies across viewpoint aspectual form-meaning mappings is governed by Feature Reassembly – but again, only to a certain extent. In addition, a satisfactory explanation has not yet been found for why the Post-YA group produce less of the target form and more of the Present in every aspectual condition, which is a striking and unexpected finding from the L2 production data. Given that the Post-YA learners have experienced a change in input in their transition from university to their year abroad, it may be the case that this is a potential source of these unexpected findings. It therefore seems worth turning to the input and exploring the suggestion (Wallington 2017, following the argumentation of Cho & Slabakova 2014) that the frequency of occurrence of each of the French aspectual form-meaning mappings may also play a role in their ease of reassembly and consequently their acquisition. The following section will return to the analysis of the frequency and distribution of these mappings in L2 French production and in L1 French input, with a view to further exploring the suggestion advanced in Wallington (2017) and also to addressing RQ2.

6.2 Exploring naturalistic input and the impact of the year abroad on grammatical development

In this section, the second research question will be addressed in two stages. Firstly, the question of what can be known about the input learners have received prior to and during the year abroad, and how this may affect the frequencies and proportions with which they produce viewpoint aspectual form-meaning mappings, will be returned to (Section 6.2.1). Secondly, the findings of this frequency-distributional analysis are applied to the year abroad context (Section 6.2.2). Commentary is provided on the areas of the L2 which may be most likely to benefit from this kind of extended residence abroad, and some possible reasons for this are suggested that centre on the characteristics of the input encountered in naturalistic vs. instructed settings.

6.2.1 Examining "year abroad input" and its implications for L2 frequency-distributional patterns

In order to be able to answer the question of whether studying naturalistic input can inform us on the effectiveness of programmes such as the year abroad on language development, it is first necessary to assess a key related question: that is, what is different about the naturalistic input learners are exposed to during the time they spend abroad, compared with the instructed input they are exposed to in the language classroom? Of course, responding to this question is extremely challenging, due in main to two particular reasons. First is the well-established methodological challenge of measuring linguistic input as a whole, and second is the fact that there is no one-size-fits-all label for what constitutes "year abroad input", in view of the wide-ranging heterogeneity of year abroad experiences (Coleman 2009; Klapper & Rees 2012). These are important caveats that should be borne in mind throughout the following discussion.

In light of the above, it is acknowledged to be beyond the scope of this project to provide a detailed analysis of the precise quantities and contents of the input provided to learners in either the year abroad or pre-year abroad settings. However, the responses collected from the Language Engagement Questionnaire (LEQ, first presented in Section 5.1.1) enabled the painting of a general picture of two important elements of the input – the extent to which learners were exposed to the target L2 (French) and their L1 (English) – in each setting. To briefly summarise the findings from the LEQ, it was found that the Pre-YA group (instructed setting) were predominantly categorised as receiving Low French exposure, with just a few attaining Medium exposure levels; contrastingly, all respondents reported High exposure to English. For the Post-YA group - who responded based on their experiences *during* the year abroad - a different story was revealed: the majority of respondents fell into either the High or (to a slightly lesser extent) Medium exposure categories for French, and participants were divided near-equally between the High and Medium exposure categories for English. However, it should be acknowledged that just over a quarter of the Post-YA group were categorised as Low French exposure, once again emphasising the heterogeneity of the year abroad experience.

Another important reality of the year abroad experience that is reflected in the LEQ data is that, irrespective of their level of exposure to French, learners maintained a medium-to-high level of exposure to English during their year abroad. This continued contact with the L1 is echoed in other research that has investigated learner interactions in the study/residence abroad setting. For example, McManus (2019) examined the social networks of L1 English UK university students during a year abroad in France, and found that, though French was used in contexts such as the workplace and for some organised face-to-face socialising, students also reported consistent use of English, particularly in virtual contact with friends and family. Overall, this L1 use formed a more prevalent part of learners' overall social interactions than their L2 French, and additionally remained at a constant level of use over the duration of the year abroad (McManus 2019). As mentioned in the previous chapter, it is possible that this continuous exposure to English even while in a naturalistic French environment may have the potential to impact learners' L2 development, particularly given that one of the central tenets of Feature Reassembly pertains to overcoming L1 influence. It would be interesting for future research to explore the linguistic development in a naturalistic setting of participants with a range of levels of exposure to their L1 (in particular, those with little to no L1 exposure). However, this would admittedly be challenging to achieve within the study abroad research paradigm, as the finite duration of these programmes means that participants are naturally extremely motivated to maintain L1-using connections to the life that they will ultimately return to.

Continued levels of English exposure notwithstanding, on consideration of the overall picture provided by the LEQ data it can generally be concluded that, during the year abroad, students were exposed a higher quantity of French input from a wider range of sources. This corresponds with Rehner & Mougeon's (2003 in J-S. Yang 2016:67) description of the year abroad setting as an opportunity for 'intensive, regular, contextualised L2-use opportunities in situ'. Contrastingly, the French exposure of the instructed Pre-YA group was limited to just a few contexts of use and so was necessarily lower. Overall, the results from the LEQ permit the drawing of an (exploratory) conclusion that there is indeed a difference in the input experienced by learners in the year abroad vs. instructed setting, specifically pertaining to the amount and diversity of exposure to the L2.

Is this difference in input experienced by the Post-YA group reflected in how learners produce L2 French viewpoint aspectual mappings? To answer this question, data presented in section 5.2.2 is referred to. Firstly, it is apparent that learners are sensitive to the distribution of viewpoint aspectual form-meaning mappings in French, as it was observed that the relative frequencies with which learners produce each of the mappings (Progressive < Habitual < Continuous < Perfective) parallels that in the L1 French corpus data. This was found to be the case particularly in the Conversation task, whose structure most closely resembles natural conversational French. Section 5.2.2 also explored evidence in support of the idea that the differential frequencies of the viewpoint aspectual mappings can partly explain the discontinuity in accuracy seen in the L2 production data. To summarise, the perfective and continuous mappings, which are most frequent in the L1 (and L2) data, are the mappings which learners produce with most accuracy, whilst the habitual and progressive mappings are markedly less frequent across all datasets and also less accurate in L2 production. These findings align with McManus's (2013) suggestion that learners will be more proficient in aspectual mappings that they are exposed to more frequently in the input, as well as Cho & Slabakova's (2014:160) observation that, conversely, 'feature reassembly may be slow to occur or may not occur at all if the relevant feature is rare or contradictory in the linguistic input.'

In view of the seemingly positive correlation between the frequency of a given aspectual mapping in naturalistic data and its accuracy in L2 production, it may initially appear redundant to state that the frequency of occurrence of a form-meaning mapping in the L1 French data is also reflected in how readily it is produced by L2 learners. However, it is important to highlight that a higher frequency of use does not always equate to a greater task-specific accuracy for the learners in this project. For example, in the Cat Story task, whose structure might have been expected to influence the frequencies with which learners produced each mapping, it was instead found that learner production arguably more closely reflected the frequency cline seen in the corpus data, as opposed to the proportions of aspectual contexts actually elicited by the task. In particular, Table 20 shows that learners produce higher numbers of tokens of the (naturally frequently-occurring) continuous mapping then the (less naturally frequently-occurring) habitual mapping, despite the fact that the latter was specifically elicited in the first section of Cat Story (see Section 4.1.2 for task details). Turning attention to the less structured Conversation task (Table 22), it appears that the Post-YA

learners are particularly sensitive to the high frequency of the continuous mapping in naturalistic French, as they produce a disproportionately large number of these tokens (43.0 on average per participant, constituting over 60% of their total viewpoint aspectual tokens) in comparison to both the Pre-YA and L1 production groups (21.3 and 32.3 average continuous tokens per participant respectively, constituting around 50% of total aspectual tokens in both cases). Some potential differences between naturalistic and instructed input in terms of the respective frequencies of viewpoint aspectual mappings will be explored at a later point, but for now it is suggested that the above findings provide support not only for the idea that learners are sensitive to viewpoint aspectual frequencies in the input, but that the Post-YA group in particular appear to be "overapplying" some of these natural frequencies in their own production. Again, this may be linked to the fact that, unlike the Pre-YA group, the Post-YA group have been exposed to an extended period of naturalistic input (where the aforementioned frequency-distributional patterns are likely to be particularly evident and abundant).

The continuous mapping seems at this point to be a recurring source of interest in the L2 data, not least for its contrasting performance in production vs. comprehension: to recap, it was the most accurate imperfective mapping in production (and also extremely frequent, particularly for the Post-YA group), yet also the condition for which target Imperfect was rated the lowest (and non-target *passé composé* the highest) in the acceptability judgement task. One possible explanation for this is that although learners produce the continuous mapping as a whole very freely and accurately, they may only do so in a narrow range of verbal contexts. This would also explain how, when required to rate a wider range of continuous meanings in the acceptability judgement task, learners were less accurate and more likely to be influenced by their L1 mappings (as indicated by the relatively high *passé composé* ratings observed in the continuous contexts).

To evaluate whether this idea about learners' performance with the continuous was upheld in the data, a small-scale type-token ratio (TTR) analysis was undertaken. TTR analysis (Templin 1957) divides the total number of tokens – for example, the number of instances when the continuous mapping was produced – by the total number of individual verb types seen across the tokens. The numerical output ranges from 0-1, with a higher value indicating a greater diversity of verbs across which the mapping was used. For the purpose of this small-scale analysis, TTR values were

calculated for all four viewpoint aspectual mappings, across 5 Conversation files in each of the L2 groups and the L1 control group. The results of the TTR analysis are presented in Table 33.

	Viewpoint aspectual mapping							
	Habitual		**Continuous**		**Progressive**		**Perfective**	
	Total tokens	Mean TTR [SD]	Total tokens	Mean TTR [SD]	Total tokens	Mean TTR [SD]	Total tokens	Mean TTR [SD]
Pre-YA (*n* = 5)	4	0.83 [0.24]	111	0.31 [0.16]	5	1.0 [0.00]	99	0.78 [0.13]
Post-YA (*n* = 5)	12	0.95 [0.10]	174	0.19 [0.08]	4	1.0 [0.00]	91	0.75 [0.15]
L1 (*n* = 5)	8	0.92 [0.17]	166	0.33 [0.09]	14	0.90 [0.22]	171	0.69 [0.08]

Table 33: Summary of type-token ratio (TTR) analysis by group and aspectual mapping.

Table 33 presents the mean type-token ratio per aspectual mapping and group, as well as the total number of tokens for each mapping. It can be seen that the total number of tokens for the habitual and progressive group samples are very low, which is not surprising given the previous findings as to the relative infrequency of these mappings. Consequently, although both L2 groups display a high TTR in these conditions, this should be taken with a pinch of salt given the small number of tokens of these mappings overall. It is perhaps more useful to compare TTRs between the continuous and perfective mappings, given that each of these contains a substantial number of tokens per group.

Looking first at the perfective, it can be seen that both L2 groups produce a similar number of tokens (99 Pre-YA vs. 91 Post-YA), and that these tokens are spread over a fairly similar number of different verbs, as indicated by the similar TTR values for each group (0.78 Pre-YA vs. 0.75 Post-YA). Moreover, the mean value of the TTR indicates that lexical diversity in perfective contexts is relatively high. When compared to the TTR of the L1 group, we see that that the lexical diversity of the L2 groups is broadly similar to that of the controls: the TTR of the L1 group is actually slightly lower, but this should be balanced against the fact that the L1 group also produce close to twice the number of perfective tokens compared with the L2 groups.

Moving on to the continuous, it can be seen that, even for L1 French speakers, mean TTR is lower: despite producing a comparable number of tokens in each condition (166 continuous vs. 171 perfective), the continuous TTR for the control group is 0.33, compared with 0.69 in the perfective condition. This suggests that the continuous mapping naturally co-occurs with a smaller number of distinct verbs than the perfective. This makes sense when we consider that the existential component of continuousness is inherently not semantically compatible with eventive verbs (see Section 3.2.1 for more information), which necessarily excludes many verbs from being paired with this meaning. However, there is nonetheless a contrast between the two learner groups for the continuous mapping: whilst the mean TTR of the Pre-YA group (0.31) is similar to that of the L1 group, the TTR of the Post-YA group is lower (0.19). Moreover, this lower TTR cannot definitively be attributed to the simple fact of the Post-YA group producing more continuous tokens: although this is the case when compared with the Pre-YA group (111 tokens vs. 174 Post-YA), the L1 group produce a similar number of tokens to the Post-YA group and yet maintain a higher TTR.

Taken together, the findings of this small-scale type-token ratio analysis suggest that, despite behaving comparably to the L1 controls in conditions such as the perfective, the Post-YA group in particular do indeed appear to utilise the continuous across a smaller range of distinct verbs[3] when compared with L1 French speakers. Therefore, the hypothesis on learners' continuous performance can be said to be substantiated with regard to the Post-YA group - although the same cannot be said for the Pre-YA group, who appear to be quite targetlike in the diversity of verbal contexts over which they use the continuous mapping. Nonetheless, the lower continuous lexical diversity for the Post-YA group, coupled with their disproportionate spike in production of continuous tokens – and not forgetting their lower accuracy in production overall - could arguably be said to perhaps reflect a phase in which this group's L2 French viewpoint aspectual development has stalled. Referring back to previous discussion on

[3] The most frequent verb used in the continuous mapping by Post-YA learners was by a large margin *être* (to be); though this was also frequent in the Pre-YA sample, occurring 43 times out of 111 total continuous tokens, it was markedly more prominent in the Post-YA sample, occurring 142 times out of a total of 174 continuous tokens.

learners' use of "placeholder Present" in a manner reminiscent of the Missing Surface Inflection Hypothesis – indicating a cognitive demand that surpasses their ability to produce the target form – this more reductive use of the Imperfect that focusses heavily on the continuous mapping and a small array of verbs also arguably supports the idea of a temporary plateau in aspectual development. Given that learners' sensitivity to the frequency and distribution of viewpoint aspectual mappings in the input has already been established, the fact that the Post-YA group have experienced a transition from instructed to naturalistic input that the Pre-YA group have not may be at least somewhat responsible for the former group's divergent behaviour in production. This key argument will be returned to in Section 6.3.

In summary, this section has discussed the results of the frequency-distributional analysis reported in Section 5.2.2. It was highlighted that the frequency with which a given aspectual mapping occurs in L1 French has the potential to also impact its frequency in L2 production, sometimes to the extent of overriding the requirements of the task being carried out. Evidence is also noted in support of a relationship between the frequency of a given mapping and the accuracy with which it is used in learner production. Finally, exploratory evidence is considered regarding learners "overextending" the frequency patterns observed in L1 data – such as the large number of continuous tokens produced by the Post-YA learners, over a diminished variety of verbs – in a manner which ultimately represents a divergence from the nativelike patterns of frequency and lexical diversity adhered to by the Pre-YA group. Crucially, the fact that this less targetlike behaviour is present only in the Post-YA group, who are the only L2 group to have undergone a transition between an instructed and a naturalistic setting, suggests that learners are sensitive not only to input but to *changes* in input. This is an important concept, which will be addressed more fully in Section 6.3. Before doing so, however, it is necessary to consider how the findings presented in this section can be applied to an evaluation of the linguistic benefits of programmes such as the year abroad.

6.2.2 Applying findings from frequency-distributional analysis to the year abroad context: what kind of benefits for language development?

Given that the previous section established that the characteristics of the input received by learners during their year abroad differed in comparison with that received by learners in instructed settings, it is important to consider whether this different "input type" may affect L2 development, and in which ways. This will permit a consideration of one of the most central questions of existing study abroad research: that is, what is the effect of programmes such as the year abroad on language development? Unsurprisingly, this question has already been the subject of a fairly substantial amount of research attention (see Kinginger 2009, J-S. Yang 2016, Borràs & Llanes 2019 for several reviews) - even if, as Llanes (2011) notes, the study abroad context is comparatively under-researched in relation to other language learning settings such as classroom or immersion settings. However, the most dominant findings from L2 research in the study/residence abroad domain have typically pertained to development of more global language skills, especially oral fluency and overall proficiency (e.g. Carroll 1967; Segalowitz & Freed 2004; Llanes & Muñoz 2009; Mora & Valls-Ferrer 2012; Muñoz & Llanes 2014). Though this has contributed to and provided support for the public perception that study abroad programmes have a significant positive impact on L2 language development (Freed 1995, Kinginger 2008), it should be noted that there has been less focus on the development of specific grammatical properties in the study/residence abroad setting, and a lack of consensus among such research as to whether time spent abroad is beneficial to grammatical development or not (e.g. Collentine 2004; Isabelli-García 2010; McManus & Mitchell 2015 vs. Howard 2001, 2005b, 2006; Edmonds & Gudmestad 2018).

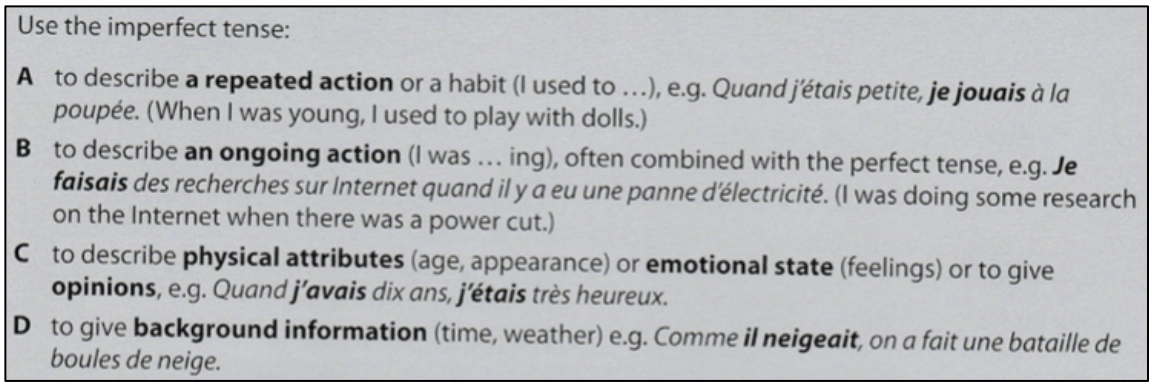
Indeed, the findings of this project do not provide clear evidence in favour of a positive impact of the year abroad on grammatical development (with specific regard to the development of viewpoint aspect): Post-YA learners are shown to behave near-identically to their Pre-YA counterparts in overall comprehension of aspectual form-meaning mappings, and to perform slightly less accurately in their production. What possible reasons could this be attributed to? The previous section presented evidence to the fact that learners are sensitive to the frequency and distribution of viewpoint aspectual mappings in the input, and suggested that the Post-YA group were in some way responding to the change in input experienced during their year abroad. To further

explore how this may have affected their viewpoint aspectual production, the traits of the input in both naturalistic and instructed settings will now be discussed in further detail from the perspective of viewpoint aspect.

Section 5.2.1, presented a breakdown of the L1 French corpus data, and explored the extent to which evidence for the four viewpoint aspectual form-meaning mappings was present in naturalistic French input. The analysis revealed that forms expressing viewpoint aspect as a whole were not especially naturally abundant, occurring close to 17% less often than verbal forms such as the Present and constituting just 3.8% of the entire 10,500+ word corpus sample. In addition, and as has been touched on in the previous section, not every mapping is equally represented within viewpoint aspect: notably, the perfective (53%) and continuous (41%) combined represented 94% of all viewpoint aspect tokens in the sample, with the habitual (5%) and the progressive (1%) markedly less prevalent. Therefore, if evidence for the totality of viewpoint aspect is not particularly frequent in naturalistic French input, evidence for a given mapping – particularly the imperfective-habitual and -progressive – is even less so.

The differing natural frequencies of the viewpoint aspectual mappings have already been observed to correlate with L2 production, with learners performing most accurately in the continuous and perfective conditions and least accurately in the progressive and habitual conditions. If learners' accuracy is indeed impacted by frequency, it is important to assess whether such frequency patterns would be present across both instructed and naturalistic settings, albeit acknowledging that we cannot be definitive about the precise quantities or characteristics of the input in either setting. Nonetheless, it might be supposed that it is much more commonplace in an instructed setting for the input to be "reinforced": such as to, in this case, highlight all three possible meanings of the Imperfect equally. This represents a contrast to the naturalistic setting where the three imperfective mappings would occur in their natural, "unbalanced" frequencies. In addition, L2 input provided in an instructed setting such as the language classroom may often be accompanied with glossed examples or explanations in the L1 to ensure learner comprehension, particularly in cases where one form in the L2 encompasses multiple meanings in the L1 (as is the case for the French Imperfect). Figures 36 and 37, sourced from two different grammar textbooks intended

for intermediate-to-advanced L2 students of French (Jacob & Schofield 2008; Thacker & D'Angelo 2013), indicate the kind of L1 English glossing or explanation that may accompany examples of the French viewpoint mappings, which could just as easily constitute part of spoken input from language teachers.



Use the imperfect tense:

A to describe **a repeated action** or a habit (I used to …), e.g. *Quand j'étais petite,* ***je jouais*** *à la poupée.* (When I was young, I used to play with dolls.)

B to describe **an ongoing action** (I was … ing), often combined with the perfect tense, e.g. ***Je faisais*** *des recherches sur Internet quand il y a eu une panne d'électricité.* (I was doing some research on the Internet when there was a power cut.)

C to describe **physical attributes** (age, appearance) or **emotional state** (feelings) or to give **opinions**, e.g. *Quand* ***j'avais*** *dix ans,* ***j'étais*** *très heureux.*

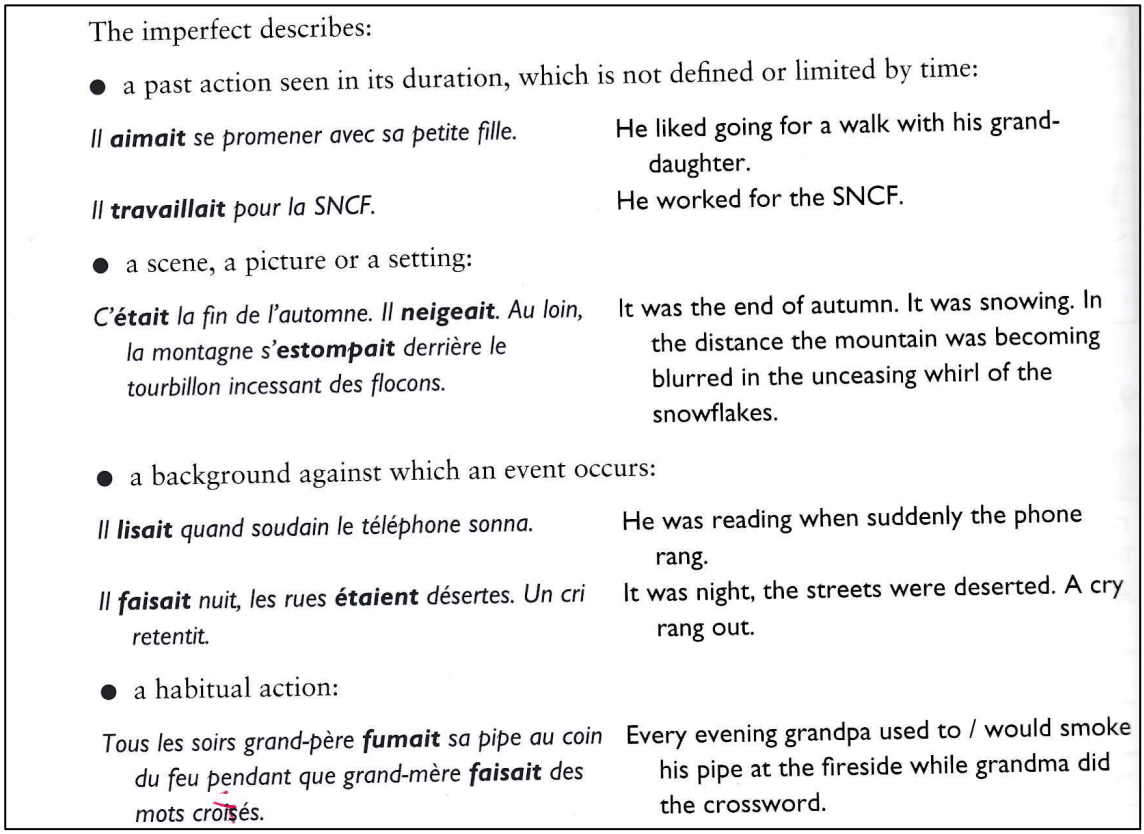
D to give **background information** (time, weather) e.g. *Comme* ***il neigeait****, on a fait une bataille de boules de neige.*

Figure 36: Example of explicit explanation linking form to meaning for the imperfective form-meaning mappings.

Sourced from Jacob & Schofield (2008:42)[4]

[4] *Heinemann A Level French Grammar Practice,* text © Pearson Education Limited 2008.

Reproduced in accordance with the provisions of the Copyright, Designs and Patents Act 1988.



The imperfect describes:

- a past action seen in its duration, which is not defined or limited by time:

*Il **aimait** se promener avec sa petite fille.*	He liked going for a walk with his grand-daughter.
*Il **travaillait** pour la SNCF.*	He worked for the SNCF.

- a scene, a picture or a setting:

*C'**était** la fin de l'automne. Il **neigeait**. Au loin, la montagne s'**estompait** derrière le tourbillon incessant des flocons.*	It was the end of autumn. It was snowing. In the distance the mountain was becoming blurred in the unceasing whirl of the snowflakes.

- a background against which an event occurs:

*Il **lisait** quand soudain le téléphone sonna.*	He was reading when suddenly the phone rang.
*Il **faisait** nuit, les rues **étaient** désertes. Un cri retentit.*	It was night, the streets were deserted. A cry rang out.

- a habitual action:

*Tous les soirs grand-père **fumait** sa pipe au coin du feu pendant que grand-mère **faisait** des mots croisés.*	Every evening grandpa used to / would smoke his pipe at the fireside while grandma did the crossword.

Figure 37: Example of glossed examples of French imperfective form-meaning mappings. Sourced from Thacker & d'Angelo (2013:218)[5]

When the acquisition of viewpoint aspectual forms is considered from a Feature Reassembly perspective, the kind of L1 support shown in Figures 36 and 37 has the benefit of making the link between form and meaning very explicit for learners – which, as mentioned, may be particularly useful when one L2 form (such as the Imperfect) encodes multiple meanings that are expressed across a range of different forms in the L1. Contrastingly, in a naturalistic setting, the semantic cue for the use of a particular form may not be as readily apparent. Gabriele (2009) posits that acquiring the semantic component of aspectual forms (or any grammatical property with a complex semantic

[5] *Essential French Grammar* © 2013 Mike Thacker & Casimir d'Angelo. Reproduced with permission of the Licensor through PLSclear.

association) may in fact be particularly input-sensitive, as the target form must arise in a semantic context which is sufficiently transparent for its meaning to be identified in order for the appropriate form-meaning mapping to be subsequently integrated into the L2 grammar. This is not always the case given that, as Gabriele (2009:397) notes, semantic cues can often be very subtle – and, arguably, are certainly likely to be less direct in general in naturalistic over instructed settings. With this in mind, it seems clearer not only why the Post-YA group performed less accurately across every aspectual condition, but also why the form-meaning mappings where learners continued to perform the most accurately were those that occurred most frequently – and thus had their form-meaning mapping most often reinforced – in the input. Though it should again be stressed that, in the absence of a systematic and detailed analysis of the input to which learners are exposed in instructed and naturalistic settings, only exploratory suggestions can be made about the advantages and disadvantages of either setting for L2 viewpoint aspectual acquisition, the above suggestions nonetheless pattern with the findings of this project and also evidence the essential nature of form-meaning mapping in aspectual acquisition. To build on this, any future research which is able to undertake a more detailed quantitative comparative analysis of the input provided to L2 acquirers in both instructed and naturalistic settings – particularly for an "input-sensitive" property such as viewpoint aspect – would undoubtedly be extremely valuable.

Despite the suggestion that the naturalistic input to which learners are exposed during their year abroad is perhaps not the most facilitative of the form-meaning mapping process (and thus of viewpoint aspectual development), it is worth highlighting that some of the key beneficial findings of the study/residence abroad literature are in fact substantiated in this project. Not only did the Post-YA group score more highly on the overall language proficiency test (mean scores: Pre-YA 64.7% [SD 10.5], Post-YA 67.2% [SD = 12.7]), but they were also shown to produce more average tokens per participant (overall and per aspectual condition) in the Conversation production task than the Pre-YA group (see Table 22). If this is taken as a marker of the Post-YA learners being able to speak at comparatively greater length in an informal conversational setting, these findings align with the widely-reported observations of increased overall oral proficiency and fluency following a stay abroad (e.g. Carroll 1967; Freed 1995;

Llanes & Muñoz 2009), and demonstrate a key source of the linguistic benefit of programmes such as the year abroad component in UK undergraduate foreign language degrees. As for the fact that the impact on grammatical gains is less decisive (both in this project and elsewhere), it could be said that this state of affairs gains clarity when viewed from the perspective of approaches such as Feature Reassembly, which forefront the role of L1 influence and help to illustrate why some grammatical properties may be more easily acquired than others depending on the L1/L2 pairing in question. It is therefore important when assessing the role of study/residence abroad on "grammatical development" to consider that the variable findings reported by studies investigating a wide range of L2s and grammatical properties may be partially attributed to the degree of reassembly between the L1 and the L2 for the property in question, and not solely attributable to the period abroad itself.

To summarise, this section began by demonstrating that, at least on an essential level, learners are generally exposed to different input while on their year abroad compared to prior to it. Although the notion of "year abroad input" is a misnomer due to the inherent diversity of year abroad experiences, it was found that students were generally exposed to more French input from a wider range of sources during their year abroad, whereas pre-year abroad students' French exposure was limited to a small number of contexts of use and so was lower overall. It was also highlighted that all students reported medium-to-high levels of English exposure during the year abroad – irrespective of their level of French exposure – which patterns with previous findings regarding high incidences of sustained L1 interactions during this kind of programme.

The frequency-distributional analysis presented in the previous chapter was subsequently returned to, with a discussion of the observation that, whilst both learner groups demonstrated a sensitivity in their own production (both in terms of frequency and accuracy) to the relative frequencies of the four viewpoint aspectual form-meaning mappings in L1 French, the Post-YA group showed some unusual behaviour which appeared to diverge from the nativelike distributional patterns to which the Pre-YA group generally adhered. This, taken alongside the lower production accuracy of the Post-YA group in every aspectual condition, led to the proposal that the unexpected behaviour of the Post-YA group may be attributed to their experience transitioning

between an instructed and naturalistic input type. This also ties in with the findings presented in section 6.1, where the Post-YA group were shown to produce a distinctive "placeholder Present" form that was not seen in Pre-YA production. The section closed by considering ways in which the naturalistic and instructed input types might vary, with reference to the L1 corpus sample for the former and to examples of textbook input for the latter. A tentative conclusion was advanced that instructed input may yield some potential benefits over naturalistic input for grammatical properties (such as viewpoint aspect) which involve a complex reassembly process, as all mappings can be equally (artificially) reinforced and the link between meaning and form made explicit with help from glossed examples and L1 explanation, in contrast to naturalistic input where mappings are present in their natural "unbalanced" frequencies and semantic cues may be less clear. However, it was also highlighted that the Post-YA learners in this project were shown to be more proficient overall in the independent proficiency measure and could also speak at greater length in the conversational production task, testifying to the well-established notion that time spent abroad facilitates oral and global proficiency.

The following section of this thesis brings together the findings discussed in Sections 6.1 and 6.2 (which respectively considered the roles of feature reassembly and of the input on L2 viewpoint aspectual development), with a view to exploring the extent to which an approach integrating information from both of these perspectives may be used to explain not only the development of viewpoint aspect but L2 development more generally.

6.3 Integrating generative predictions into a study of the input: observations from L2 French viewpoint aspectual development

Section 6.1 considered the predictions instantiated by Feature Reassembly with regard to the four form-meaning mappings of the French viewpoint aspectual system for L1 English speakers, and the extent to which they aligned with the production and comprehension results of the L2 French learners in this project. It was concluded that, while certain predictions (such as the relative difficulty of acquisition of the imperfective over the perfective) were supported in the data, Feature Reassembly could

only partially accommodate learners' differential rates of accuracy across the imperfective mappings, particularly in the case of the production data. Not only did learners' production accuracy not align with the degree of reassembly required in every case, but behaviour was observed that was unrelated to any relevant L1 aspectual mapping: specifically, the distinctive use of the Present by the Post-YA group. Given that both L2 groups behaved near-equivalently in the comprehension task, it was suggested that the non-target Present use in viewpoint aspectual contexts by the Post-YA group represented a kind of placeholder form (similarly to Prévost & White's (2000) Missing Surface Inflection Hypothesis), the presence of which may in itself indicate some kind of difficulty with the reassembly process that is exacerbated by the cognitive demands of spontaneous speech production. However, the low incidence of the *passé composé* in the habitual and continuous contexts indicates that learners are not reverting to their L1 form-meaning mappings. A key question that therefore arises is how else these unexpected results can be accounted for.

As the ease of acquisition of a given form-meaning mapping has been suggested, in various guises (e.g. Miller & Schmitt 2012; Gil & Marsden 2013; McManus 2013; Cho & Slabakova 2014; Wallington 2017), to be correlated with the frequency with which the form and meaning co-occur in the input, it was of interest to explore the implications of this from a viewpoint aspectual perspective. Given the methodological challenges of measuring input, this project studies two learner groups expected to have experienced different "types" of French input (instructed vs. naturalistic), which was hoped to aid in making any role of frequency, and input in general, more apparent. Section 6.2 discussed evidence in support of the fact that learners are indeed sensitive to frequency-distributional patterns of viewpoint aspectual mappings in the input, and may be influenced by this in ways that subvert Feature Reassembly predictions or task requirements.

Moreover, it was observed that the frequency of occurrence of a mapping contributed to learners' production accuracy in that mapping, with the highest accuracy scores observed in the two most frequent mapping conditions and vice versa. This provides empirical evidence for the exploratory suggestions advanced in Wallington (2017:54), in which was proposed the following interplay between Feature Reassembly and frequency to account for learners' differential aspectual accuracies: 'IMP-Habitual is not as frequent and requires reassembly; IMP-Progressive has a one-to-one form-

meaning mapping but is very infrequent; and lastly IMP-Continuous, despite entailing a feature reassembly task [relatively] comparable to the habitual, is considerably more frequent that both other Imperfect interpretations'. Not only does the production accuracy cline of Wallington (2017) (Habitual < Progressive < Continuous < Perfective) match with the current project, but the exploratory suggestions made regarding the relative frequencies of the viewpoint aspectual mappings (which were based on the frequency of each mapping in learner production) were substantiated in the L1 French production and corpus data analysis that formed a cornerstone of the present project. This aligns with a perspective wherein Feature Reassembly does contribute to the relative ease or difficulty of acquisition of a given viewpoint aspectual form-meaning mapping by L1 English learners of L2 French, but that the effect of Feature Reassembly is mediated by the frequency with which that form and meaning co-occur in the input. Taken together, these collective findings provide growing evidence that, with regard to L2 French viewpoint aspectual development, the combination of Feature Reassembly predictions and the relative frequency of the target mapping in the input has greater explanatory power than either variable alone.

It should however be noted that, interestingly, the effect of frequency is most apparent in production, and less so in comprehension, where the accuracy cline of the imperfective mappings was in fact fully aligned with Feature Reassembly predictions. The imperfective mapping that showed the greatest accuracy contrast between production and comprehension was the continuous, which was the most accurate imperfective mapping in production but the least (in terms of both accuracy and targetlike ratings contrasts) in comprehension. The results of the small-scale type-token ratio analysis confirmed suspicions that, although learners were very accurate with the continuous mapping in production, Post-YA learners in particular only produced it with a small selection of verbs, and all learners were less accurate when challenged to consider the acceptability of continuous Imperfect in a more diverse range of verbal contexts. These findings go some way in explaining the production vs. comprehension discrepancy observed in the data, and also highlight the importance of considering the diversity of verbal contexts over which mappings are used when evaluating constructs such as L2 accuracy in production. Additionally, the fact that the accuracy cline predicted by Feature Reassembly is present in the L2 comprehension data provides

further support for the validity of Feature Reassembly in considering the L2 French viewpoint aspectual learning task. With this said, further research to tease apart the respective contributions of frequency, lexical diversity, reassembly, and task type in accounting for production vs. comprehension differences (as were also seen in Domínguez et al. 2017) would be greatly valuable.

Another unexpected finding explored in the previous sections was the fact that the Post-YA group were found to be less accurate in production in every aspectual condition. This was surprising, not only as the Post-YA group may have been assumed to be at a more advanced stage in their L2 French development (as supported by their higher score in the overall proficiency measure), but also because they had just returned from an extended residence in a French-speaking country, which is widely perceived to be facilitative of language development. When the relative frequencies with which the Post-YA learners produced viewpoint aspectual mappings were explored, some evidence was found of divergence from naturalistic patterns - such as the large spike in continuous tokens in their Conversation data – which may arguably have resulted from an "overextension" of naturalistic patterns such as the high natural frequency of the continuous. Moreover, the Post-YA group also showed a lower TTR in the continuous condition in comparison to both the Pre-YA and L1 groups, suggesting that their increased use of the continuous mapping does not correspond with an ability to extend this use over a wider range of appropriate verbal contexts. Consequently, the impression is not that the Post-YA group are using the continuous mapping in a more targetlike way – if anything, the opposite may be true.

Given that the Pre-YA group show comparatively more convergence to naturalistic frequency-distributional patterns, and also have a higher production accuracy without significant use of placeholder Present, it is posited that these observations conspire to suggest that the unusual production behaviour of the Post-YA group is in fact indicative of their transition between instructed and naturalistic input types. This suggestion is grounded in the fact that if – as has been demonstrated – learners are sensitive to patterns in the input, it seems logical to extend this to a sensitivity to *changes* in input. Extrapolating from this, it may be proposed that the feature reassembly process is itself also sensitive to the input and input changes, in line with the suggestions of researchers such as Cho & Slabakova (2014). Such a supposition

does not appear too outlandish if one considers that the restructuring component that follows Full Transfer (Schwartz & Sprouse 1994, 1996), the starting point of Feature Reassembly (Lardiere 2003, 2005, 2008, 2009), occurs in response to a failure of the transferred L1 mappings to analyse the L2 input. Following this logic, if, for example, an L2 learner of French moved from an instructed environment (such as the foreign-language classroom) to a naturalistic environment (such as an extended stay in a French-speaking country), it is likely that the French input they received would also change. This chapter has explored the ways in which instructed input may differ from naturalistic input, including the observation from the LEQ data that learners during a year abroad receive on average more exposure to French from a more diverse range of sources than learners prior to the year abroad. Contrastingly, in instructed settings, the input may be artificially "reinforced", such as by giving equal emphasis to form-meaning mappings that do not naturally occur with equal frequency or by using the L1 to make the link between meaning and form more explicit. This exploratory analysis of the two input settings arguably presents some support for the concept that university learners who undertake a year abroad are exposed to a different type of input during this period. Returning to the notion of restructuring, it therefore could be said that this process (that is itself integral to feature reassembly) may be impacted as a result of the change in input setting, depending on how well the current mappings in the learner's L2 grammar are able to accommodate this different type of input.

In view of the above, it is proposed that the changes in input experienced by a (UK) university student of French who participates in a year abroad may elicit a feature reassembly "plateau" in response to these input transitions. The existence of such a plateau is arguably visible in the Post-YA use of placeholder Present, which reflects the fact that that learners are not regressing in their reassembly of aspectual form-meaning mappings (as there is no evidence of increased L1 influence in production), but instead are merely adjusting and responding to the changes in input type. This is arguably more demanding on learners' cognitive resources than remaining in an environment with no change in input type, and so there is an increased incidence of learners "falling back" on the simpler Present form in production. At the same time, feature reassembly appears to have paused, as indicated by the slightly lower production accuracy of the Post-YA group and their somewhat "reductive" use of the Imperfect which is disproportionately centred on the continuous mapping (itself expressed across only a narrow range of verbs). Crucially, though, the fact that comprehension accuracy remained stable across

the L2 groups suggests that the impact of the feature reassembly plateau is not permanent and may be overcome, as learners' underlying aspectual representations do not appear to be affected. It is acknowledged that the differences observed between the learner groups are subtle - which may be partly attributable to sample size – but nonetheless, these subtle differences also are in accordance with the understanding of Feature Reassembly as an incremental, gradual process.

Indeed, the incremental nature of Feature Reassembly meshes well with models such as C. Yang's (2002) Variational Model, which emphasises that, while innate knowledge is necessary to constrain acquisition, 'statistical learning seems most naturally suited to modelling the gradualness of language development' (C. Yang 2002:24). Moreover, the optionality in both production and comprehension seen from the learners in this project aligns not only with the incremental nature of Feature Reassembly, but also with the mechanism central to the Variational Model, wherein the selection of a target grammar is probabilistically linked to its ability to parse the incoming input (C. Yang 2006, 2018).

The high degree of compatibility between Feature Reassembly and input-centred models such as the Variational Model is encouraging to observe, particularly in light of the recent increased demand for generative L2A research to focus on input in a manner that extends beyond the notion of "the poverty of the stimulus" (e.g. Rankin & Unsworth 2016; Yang & Montrul 2017). Embracing approaches to language acquisition which factor in the role of input alongside Universal Grammar (e.g. Westergaard 2009, 2014; Lidz & Gagliardi 2015; Pearl 2021) has arguably become of an even greater importance for generativists in view of the ongoing and lively contemporary debate surrounding what Universal Grammar actually consists of (see e.g. Biberauer 2019 for one perspective). Drawing on Chomsky's (2005) model of the "three factors" underpinning the language faculty, C. Yang (2010:1160) summarises the current state of the field, stating that, while Universal Grammar *is* essential, equally essential is not to 'ask[…] for too much' from it: particularly given that one of the other two "factors" (alongside general cognitive factors) is linguistic experience. If Universal Grammar is indeed less richly-structured than was originally posited under Chomsky's (1955/1975, 1978, 1981a) "poverty of the stimulus" formulation of generative language acquisition, then it is arguably all the more essential to direct more of our attention towards the

stimulus itself. As the analysis and discussion in this chapter endeavours to demonstrate, this can be achieved without relegating UG to a non-essential role.

6.4 Conclusions, limitations, and suggestions for future research

In conclusion, I hope to have illustrated that not only is it theoretically desirable for generative and input-based approaches to come together, it also makes good sense in view of L2 data such as that presented in this project. This project joins voices with those calling for a greater integration of the input into generatively-grounded studies, and specifically demonstrates that combining frequency-distributional information from L1 and L2 French input with Feature Reassembly, a model of second language acquisition that reposes on the seminal generative concepts of Full Transfer and Full Access, permits a more comprehensive account of the processes and trajectories underlying L2 viewpoint aspectual development. I posit that this kind of integrated approach is particularly applicable to the later stages (i.e. beyond Full Transfer) of later-acquired properties such as viewpoint aspect, and would also like to advocate for a similar consideration of the interplay between reassembly and frequency in the study of later-stage L2 development of other grammatical properties – particularly those which bring together different permutations of frequency of occurrence and degree of reassembly. For example, it may be interesting to explore the interplay of frequency and reassembly in the acquisition of the French subjunctive, as an area of the grammar that not only has a lower frequency and more restricted distribution than viewpoint aspect, but that also straddles the boundary between morphology and discourse and consequently would have a markedly different reassembly task. It would also be of great interest to explore the reassembly/frequency interface of one given property, but using a range of L2s: for example; L2 French viewpoint aspect could be explored for speakers of L1 Spanish, English, and German. Of course, exploring L1 influence on the same L2 property for learners of a range of L1 backgrounds is not a novel idea, but integrating this with an analysis of how this property varies in terms of frequency and distribution within each L1 could provide rich cross-linguistic detail to the ideas sketched out in this project.

At this point, it is essential to acknowledge the limitations of the current project. One limitation that has been previously mentioned is the inconsistency

between the L1 production and comprehension groups. If this project were to be replicated, it would be beneficial not only to have the same L1 participants complete both the production and comprehension tasks, but also to have a larger L1 sample, in view of the unexpected heterogeneity of the results reported in this project. This heterogeneity may admittedly be linked to another potential shortcoming in the design of this project: specifically, with regard to the acceptability judgement task. A future replication of this project could consider utilising a two alternative forced choice design for the AJT component and could certainly incorporate some distractor items, thus minimising the "learning effect" observed in the L2 comprehension data. Care should also be taken to control for as many demographic variables as reasonably practicable (e.g. age, level of education, linguist vs. non-linguist) when recruiting participants, to ensure that the collected data reflects participants' true linguistic competence and is not obscured by non-linguistic factors. Additionally, it is acknowledged that, in comparison with the size of the "corpora" created from the experimental production data, the CFPP2000 corpus utilised in this project is arguably on the smaller side. Though it would have been a challenge to incorporate analysis of a much larger corpus within this already rather methodologically wide-ranging project, future research building off of this work should consider analysing a larger sample of corpus data, to ensure that the sample contains sufficient data to be truly representative.

Finally, problems and restrictions on participant recruitment prevented this project from having a longitudinal design, wherein the same participants were followed before, during, and after their year abroad. Though the current cross-sectional design still provided insight, a longitudinal design would eliminate any interfering factors in the results linked to individual differences between the learner groups. A future study could endeavour to capitalise on the variable input types experienced by university languages students who undertake an extended residence abroad, such as by longitudinally analysing a cohort throughout their undergraduate degrees and assessing their linguistic development, varying L1 and L2 exposure, and responses to the different input "settings" inherent to the structure of their degree programmes. In particular, a post-test administered at the end of students' final year – a year after returning from the year abroad – would be especially enlightening for those wishing to explore the extent to which the response to the instructed/naturalistic input transition may endure.

References

AITCHISON, J. (1976) *The Articulate Mammal* (1st edition). New York: Routledge

ALLEN, H.W. (2010) Language-learning motivation during short-term study abroad: An activity theory perspective. *Foreign Language Annals,* 43(1). pp.27-49

ANDERSEN, R.W. & SHIRAI, Y. (1994) Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition,* 16. pp.135-156

ANDERSEN, R.W. (1988). Models, processes, principles, and strategies: Second Language Acquisition in and out of the classroom. *Issues and Developments in English and Applied Linguistics (IDEAL)*, *3*. pp. 111–38.

ANDERSEN, R. (1993). Four operating principles and input distributions as explanations for underdeveloped and mature morphological systems. IN: HYLTENSTAM, K. & VIRBORG, A. (Eds.), *Progression and regression in language*. Cambridge: Cambridge University Press. pp. 309–339

ANDERSEN, R.W. & SHIRAI, Y. (1996) Primacy of aspect in first and second language acquisition: the pidgin/creole connection IN: RITCHIE, W.C. & BHATIA, T.K. (eds.) *Handbook of second language acquisition.* San Diego, CA: Academic Press. pp.527-570

ANDRINGA, S. & OLSTHOORN, N. & VAN BEUNINGEN, C. & SCHOONEN, R. & HULSTIJN, J. (2012) Determinants of success in native and non-native listening

comprehension: an individual differences approach. *Language Learning* 62(Suppl. 2). pp.49-78

ARCHE, M.J. (2006) Individuals in time, Tense, aspect and the individual/stage distinction. Amsterdam: John Benjamins

ARCHE, M.J. (2014) The construction of viewpoint aspect: the imperfect revisited. *Natural Language and Linguistic Theory*, 32(43). pp.791-831

AYOUN, D. (2013) *The Second Language Acquisition of French Tense, Aspect, Mood and Modality*. Amsterdam: John Benjamins

BAAYEN, R.H. & DAVIDSON, D.J. & BATES, D.M. (2008) Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language,* 59(4). pp.390-412. doi: 10.1016/j.jml.2007.12.005

BAKER, M. (2008) The macroparameter in a microparametric world IN: BIBERAUER, T. (ed.) *The limits of syntactic variation.* Amsterdam: John Benjamins. pp.351-74

BAKER, C.L. & MCCARTHY, J. (eds.) (1981) *The logical problem of language acquisition*. Cambridge, MA: MIT Press

BARDOVI-HARLIG, K & COMAJOAN-COLOMÉ, L. (2020) The Aspect Hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Studies in Second Language Acquisition.* pp.1-31. doi:10.1017/S0272263120000194

BARR, D. & LEVY, R. & SCHEEPERS, C. & TILY, H. (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language,* 68. pp.255-278

BARTNING, I. (1997) L'apprenant dit avancé et son acquisition d'une langue étrangère. *Acquisition et interaction en langue étrangère,* 9. pp.9-50

BARTNING, I. & SCHYLTER, S. (2004) Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French language studies,* 14. pp.281-299

BATES, D. & MAECHLER, M. & BOLKER, B. & WALKER, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

BERGSTRÖM, A. (1995) The expression of past temporal reference by English- speakers of French. Unpublished PhD dissertation, Pennsylvania State University

BERTINETTO, P.M. (1997) Il domino tempo-aspecttuale: Demarcazioni, intersezioni, constrasti. Torino: Rosenberg & Sellier

BERTINETTO, P.M. (2001) On a frequent misunderstanding in the temporal- aspectual domain: 'The Perfective = Telic Confusion' IN: CECCHETTO, C. & CHIECRCHIA, G.

& GUASTI, M.T. (eds.) *Semantic Interfaces: Reference, Anaphora and Aspect.* Stanford: CLSI Publications. pp.177-210

BIBERAUER, T. (2019) Factors 2 and 3: Towards a principled approach. *Catalan Journal of Linguistics*. pp.45-88. doi: https://doi.org/10.5565/rev/catjl.219

BLEY-VROMAN, R. (1989) What is the logical problem of foreign language learning? IN: GASS, S.M. & SCHACHTER, J. (eds.) *Linguistic perspectives on second language acquisition.* Cambridge, UK: Cambridge University Press. pp.1-68

BLEY-VROMAN, R. (1990) The logical problem of foreign language learning. *Linguistic Analysis,* 20. pp.3-49

BLOCK, D. (2007) *Second language identities.* New York, NY: Continuum

BLOM, E. & PARADIS, J. & DUNCAN, T.S. (2012) Effects of input properties, vocabulary size, and L1 on the development of third person singular *-s* in child L2 English. *Language Learning,* 62. pp.965-994

BORER, H. (1984) *Parametric Syntax: Case Studies in Semitic and Romance Languages.* Dordrecht: Foris

BORRÀS, J. & LLANES, À. (2019) Re-examining the impact of study abroad on L2 development: a critical overview. *The Language Learning Journal*. doi: 10.1080/09571736.2019.1642941

BRECHT, R. & DAVIDSON, D. & GINSBERG, R. (1995) Predictors of foreign language gain during study abroad IN: FREED, B. (ed.) *Second language acquisition in a study abroad context.* Amsterdam: John Benjamins. pp.37-66

BRANCA-ROSOFF, S. & FLEURY, S. & LEFEUVRE, F. & PIRES M., 2012, *Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*. http://cfpp2000.univ-paris3.fr/CFPP2000.pdf

BROOKS, R. & WATERS, J. (2010) Social networks and educational mobility: the experiences of UK students. *Globalisation, Societies, and Education,* 8(1). pp.143-57

BYBEE, J. (2008) Usage-based grammar and second language acquisition. In: Robinson, P. & Ellis, N.C. (eds.) *Handbook of cognitive linguistics and second language acquisition.* New York: Routledge. pp. 216–36

BYBEE, J. & DAHL, Ö. (1989) The creation of tense and aspect systems in the languages of the world. *Studies in Language*, 13. pp.51-103

BYBEE, J. & PERKINS, R. & PAGLIUCA, W. (1994) *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: The University of Chicago Press

CARROLL, J. (1967) Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1. pp.131-151

CARROLL, S. (1996) Parameter-setting in second language acquisition: explanans and explanandum. *Brain and Behavioral Sciences,* 19. pp.720-1

CARROLL, S. (2001) *Input and evidence: the raw material of second language acquisition.* Amsterdam: John Benjamins

CATTANI, A. & ABBOTT-SMITH, K. & FARAG, R. & KROTT, A. & ARRECKX, F. & DENNIS, I. & FLOCCIA, C. (2014) How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language and Communication Disorders,* 49. pp.649-71

CHUNG, S. & TIMBERLAKE, T. (1985) Tense, aspect, and mood IN: SHOPEN, T. (ed.) *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon* (vol. 3). Cambridge: Cambridge University Press. pp.202-258

CHO, J. & SLABAKOVA, R. (2014) Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research*, 30(2). pp.159-190

CHOMSKY, N. (1955/1975). *The logical structure of linguistic theory*. New York: Plenum Press; University of Chicago Press

CHOMSKY, N. (1959) A review of B.F. Skinner's *Verbal Behavior. Language,* 35. pp.26-58

CHOMSKY, N. (1965) *Aspects of the theory of syntax.* Cambridge, MA: MIT Press

CHOMSKY, N. (1967). The formal nature of language IN: LENNEBERG, E.H. (ed.) *Biological foundations of language.* New York: Wiley and Sons. pp. 397–442.

CHOMSKY, N. (1978). A naturalistic approach to language and cognition. *Cognition and Brain Theory*, 4. pp. 3–22

CHOMSKY, N. (1980) (A) *Rules and representations.* Oxford: Blackwell

CHOMSKY, N. (1980) (B) On binding. *Linguistic Inquiry,* 11(1). pp.1-46

CHOMSKY, N. (1981) (A) *Lectures on government and binding.* Dordrecht: Foris

CHOMSKY, N. (1981) (B) Principles and parameters in syntactic theory IN: HORNSTEIN, N. & LIGHTFOOT, D. (eds.) *Explanation in linguistics: the logical problem of language acquisition.* London: Longman. pp.32-75

CHOMSKY, N. (1986) *Knowledge of language: its nature, origin and use.* New York: Praeger

CHOMSKY, N. (1995) *The minimalist program.* Cambridge, MA: MIT Press

CHOMSKY, N. (2001) Derivation by phase IN: KENSTOWICZ, M. (ed.) *Ken Hale: A life in language.* Cambridge, MA: MIT Press. pp.1-52

CHRISTENSEN, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

COLEMAN, J. (1997) Residence abroad within language study. *Language Teaching,* 30. pp.1-20

COLEMAN, J. (2015). Social circles during residence abroad: What students do, and who with. IN: MITCHELL, R. & TRACY-VENTURA, N. & MCMANUS, K. (Eds.), *Social interaction, identity and language learning during residence abroad*. EUROSLA Monographs Series. Amsterdam: The European Second Language Association. pp.33-52

COLLENTINE, J. (2004) The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition,* 26. pp.227-248

COLLENTINE, J. (2009) Study abroad research: Findings, implications, and future directions IN: DOUGHTY, C. & LONG, M. (eds.) *Handbook of language teaching.* Malden, MA: Blackwell. pp.218-233

COLLENTINE, J. & FREED, B. (2004) Introduction: Learning context and its effects on second language acquisition. *Studies in Second Language Acquisition,* 26. pp.153-72

COLLENTINE, J. (2009) Study abroad research: Findings, implications, and future directions IN: DOUGHTY, C. & LONG, M. (eds.) *Handbook of language teaching.* Malden, MA: Blackwell. pp.218-233

COLLINS, L. & TROFIMOVICH, P. & WHITE, J. & HORST, M. & CARDOSO, W. (2009) Some input on the easy/difficult grammar question. *Modern Language Journal*, 93. 336-53

COMRIE, B. (1976) Aspect: An Introduction to the Study of Verbal Aspect and Related Problems. Cambridge: Cambridge University Press

COPPIETERS, R. (1987) Competence differences between native and non-native speakers. *Language*, 63. pp.544-573

CUNNINGS, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382.

DE BOT, K. (2015) Moving where? A reaction to Slabakova et al. (2014). *Applied Linguistics,* 36. pp.261-4

DEMIRDACHE, H. & URIBE-ETXEBARRIA, M. (2000) The primitives of temporal relations IN: MARTIN, R. & MICHAELS, D. & URIAGEREKA, J. (eds.) *Step by step: Essays on minimalist syntax in honor of Howard Lasnik.* Cambridge: MIT Press. pp.157-186

DEKEYSER, R. (1991) Foreign language development following a semester abroad IN: FREED, B. (ed.) *Foreign language acquisition and the classroom.* Lexington, MA: D.C. Heath. pp.104-119

DEKEYSER, R. (2010) Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals,* 43(1). pp.80-92

DÍAZ-CAMPOS, M. (2004) Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition,* 26. pp.249-273

DILLON, B. & WAGERS, M.W. (2019). Approaching gradience in acceptability with the tools of signal detection theory. OSF Preprints.

DOMÍNGUEZ, L. (2019) A 'mixed methods' approach for investigating aspect in a second language: Evidence from the SPLLOC project. *Dutch Journal of Applied Linguistics*, *8*(1). pp.41-66.

DOMÍNGUEZ, L. & ARCHE, M.J. & MYLES, F. (2011) Testing the Predictions of the Feature Reassembly Hypothesis: Evidence from the L2 acquisition of Spanish Aspect Morphology IN: DANIS, N. & MESH, K. & SUNG, H. (eds.) *BUCLD 35: Proceedings of the 35th annual Boston University Conference on Language Development* (Vol. 1). Sommerville: Cascadilla Press. pp.183-196

DOMÍNGUEZ, L. & ARCHE, M.J. & MYLES, F. (2017) Spanish Imperfect revisited: exploring L1 influence in the reassembly of imperfective features onto new L2 forms. *Second Language Research*, online. Available from: https://doi.org/10.1177/0267658317701991

DUDEN, D. (1995) *Die Grammatik* (vol. 5). Mannheim, Germany: Dudenverlag

DUFON, M.A. & CHURCHILL, CHURCHILL, E. (eds.) *Language learners in study abroad programs.* Clevedon: Multilingual Matters

DURRELL, M. (2006) Using German: A Guide to Contemporary Usage. Cambridge: Cambridge University Press

EDMONDS, A. & GUDMESTAD, A. (2018) Gender marking in written L2 French: Before, during and after residence abroad. *Study Abroad Research in Second Language Acquisition and International Education,* 1(3). pp.58-83

ELLIS, N. (2002) Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition,* 24. pp.143–88

ENÇ, M. (1987) Anchoring conditions for tense. *Linguistic Inquiry,* 18. pp.633-657

FELIX, S. (1986) *Cognition and language growth.* Dordrecht: Foris

FREED, B. (1995) What makes us think that students who study abroad become fluent? IN: FREED, B. (ed.) Second language acquisition in a study abroad context. Amsterdam: John Benjamins. pp.123-48

FREED, B. (1998) An overview of issues and research in language learning in a study abroad setting. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 4. pp.31-60

GABRIELE, A. (2009) Transfer and transition in the SLA of aspect. *Studies in Second Language Acquisition,* 31. pp.371-402

GABRIELE, A. & MARTOHARDJONO, G. & MCCLURE, W. (2003) Why *swimming* is just as difficult as *dying* for Japanese learners of English. *ZAS Papers in Linguistics,* 29. pp.85-103

GABRIELE, A. & MARTOHARDJONO, G. & MCCLURE, W. (2005) Evaluating the Role of the L1 in the L2 Acquisition of Aspect: A Study of Japanese Learners of English IN: COHEN, J. & MCALISTER, T. & ROLSTAD, K. & MACSWAN, J. (eds.) *ISB4: Proceeding of the 4th International Symposium on Bilingualism.* Somerville: Cascadilla Proceedings. pp.808-826

GATHERCOLE, V.C.M. (2002) Command of the mass/count distinction in bilingual and monolingual children: An English morphosyntactic distinction IN: OLLER, D.K. & EILIERS, R.E. (eds.) *Language and literacy in bilingual children.* Clevedon: Multilingual Matters. pp.175-206

GIL, K.H. & MARSDEN, H. (2013) Existential quantifiers in second language acquisition: A feature reassembly account. *Linguistic Approaches to Bilingualism,* 3. pp.117-19

GOLONKA, E. (2006) Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. *Modern Language Journal*, 90. pp.496-505

GOODLUCK, H. (1991) *Language acquisition: a linguistic introduction.* Oxford: Blackwell

GOSSELIN, L. (1996) Sémantique de la temporalité en français. Louvain-la-Neuve: Duculot

GREGG, K.R. (1996) The logical and developmental problems of second language acquisition IN: RITCHIE, W.C. & BHATIA, T.K. (eds.) *Handbook of second language acquisition.* San Diego, CA: Academic Press. pp.49-81

GUNTERMANN, G. (1995) The Peace Corps experience: Language learning in training and in the field IN: FREED, B. (ed.) *Second language acquisition in a study abroad context.* Amsterdam: John Benjamins. pp.149-170

GUO, Y. (2020) From a simple to a complex aspectual system: Feature reassembly in L2 acquisition of Chinese imperfective markers by English speakers. *Second Language Research*. pp.1-28. doi: 10.1177/0267658320911433

HAHNE, A. (2001) What's different in second-language processing? Evidence from event-related brain potentials. *Journal of psycholinguistic research*, *30*(3), pp.251-266.

HARLEY, B. (1978) An analysis of the verb system used by young learners of French. *Interlanguage Studies Bulletin,* 3(1). pp.35-79

HARLEY, B. (1992) Patterns of second language development in French immersion. *Journal of French Language Studies,* 2. pp.159-183

HAWKINS, R. & CHAN, C.Y.H. (1997) The partial availability of Universal Grammar in second language acquisition: The 'failed functional features' hypothesis. *Second Language Research,* 13(3). pp.187-226

HOPP, H. (2006) Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), pp.369-397.

HORNSTEIN, N. & LIGHTFOOT, D. (1981) *Explanation in linguistics: the logical problem of language acquisition.* London: Longman

HOWARD, M. (2001) The effects of study abroad on the L2 learner's structural skills: Evidence from advanced learners of French. *EUROSLA Yearbook,* 1. pp.123- 141

HOWARD, M. (2005) (A) Les contextes prototypiques et marqués de l'emploi de l'imparfait par l'apprenant du français langue étrangère IN: LABEAU, E. & LARRIVÉE, P. (eds.) *Nouveaux développements de l'imparfait.* Amsterdam: Rodopi.

HOWARD, M. (2005) (B) On the role of context in the development of learner language: Insights from study abroad research. *ITL International Journal of Applied Linguistics,* 148. pp.1-20

HOWARD, M. (2006) The expression of number and person through verb morphology in French interlanguage. *International Review of Applied Linguistics,* 44(1). pp.1-22

HULSTIJN, J. H. (2011). Language proficiency in native and non-native speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8. pp. 229–249. DOI: 10.1080/15434303.2011.565844

HWANG, S.H. & LARDIERE, D. (2013) Plural-marking in L2 Korean: A feature-based approach. *Second Language Research,* 29(1). pp.-57-86

HYAMS, N. (1986) *Language acquisition and the theory of parameters.* Dordrecht: Reidel

IMBS, P. (1960) *L'emploi des temps verbaux en français moderne.* Paris: Klincksieck

ISABELLI-GARCÍA, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study Abroad and at home. *Foreign Language Annals,* 43(2). pp.289–303

IZQUIERDO, J. & COLLINS, L. (2008) The facilitative role of L1 influence in tense-aspect marking: A comparison of Hispanophone and Anglophone learners of French. *The Modern Language Journal,* 92(3). pp.350-68

JACOB, S. & SCHOFIELD, J. (2008) *Heinemann A Level French Grammar Practice.* Heinemann. p.42

JUAN-GURAU, M. & PÉREZ-VIDAL, C. (2007) The effect of context and contact on oral performance in students who go on a stay abroad. *Vigo International Journal of Applied Linguistics,* 4. pp.117-134

JUDY, T. (2011) L1/L2 parametric directionality matters: More on the null subject parameter in L2 acquisition. *EUROSLA Yearbook,* 11(1). pp.165-90

KANNO, K. (1997) The acquisition of null and overt pronominals in Japanese by English speakers. *Second Language Research,* 13. pp.265-287

KIHLSTEDT, M. (1998) La référence au passé dans le dialogue: Étude de l'acquisition de la temporalité chez les apprenants dits avancés de français (Cahiers de la Recherche 6). Stockholm: G. Engwall & J. Nystedt

KIHLSTEDT, M. (2002) Reference to past events in dialogue: The acquisition of tense and aspect by advanced learners of French IN: SALABERRY, R. & SHIRAI, Y. *The L2 Acquisition of Tense-Aspect Morphology.* Amsterdam; Philadelphia: John Benjamins. pp.323-61

KINGINGER, C. (2007) Research on language education through programmes of study and residence abroad: Recent publications. *International Journal of Bilingual Education and Bilingualism,* 10. pp.104-12

KINGINGER, C. (2008) Language learning in study abroad: Case studies of Americans in France. *The Modern Language Journal,* 92. pp.1–124

KINGINGER, C. (2009). *Language learning and study abroad: A critical reading of research.* Basingstoke: Palgrave Macmillan.

KLAPPER, J. & REES, J. (2012) University residence abroad for foreign language students: Analysing the linguistic benefits. The Language Learning Journal, 40. pp.335-358

KLEIN, E. & MARTOHARDJONO, G. (1999) Investigating second language grammars: some conceptual and methodological issues in generative SLA research IN: KLEIN, E. & MARTOHARDJONO, G. (eds.) *The development of second language grammars: a generative perspective.* Amsterdam: John Benjamins. pp.3-34

KLEIN, W. (1994) *Time in language.* London: Routledge

LABEAU, E. (2005) *Beyond the Aspect Hypothesis: Tense-Aspect Development in Advanced L2 French.* Oxford: Peter Lang

LABEAU, E. (2009) Le PS: cher disparu de la rubrique nécrologique? *Journal of French Language Studies*, 19(1). pp.61-86

LABEAU, E. (2011) An Imperfect Mastery: The Acquisition of the Functions of *Imparfait* by Anglophone Learners IN: LABEAU, E. & MYLES, F. (eds.) *The Advanced Learner Variety: The Case of French.* Peter Lang

LARDIERE, D. (1998) (A) Case and tense in a 'fossilized' steady state. *Second Language Research,* 14(1). pp.1-26

LARDIERE, D. (1998) (B) Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research,* 14(4). Pp.359-375

LARDIERE, D. (2000) Mapping Features to Forms in Second Language Acquisition IN: ARCHIBALD, J. (ed.) *Second Language Acquisition and Linguistic Theory.* Oxford: Blackwell. pp.102-129

LARDIERE, D. (2003) Second Language Knowledge of [±Past] vs. [±Finite] IN: LICERAS, J.M. & ZOBL, H. & GOODLUCK, H. (eds.) Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002). Sommerville: Cascadilla Proceedings. pp.176-189

LARDIERE, D. (2005) On morphological competence IN: DEKYDTSPOTTER, L & SPROUSE, R.A. & LILJESTRAND, A. (eds.) *Proceedings of the 7th Generative Approaches to*

Second Language Acquisition Conference (GASLA 2004). Sommerville: Cascadilla Proceedings Project. pp.178-192

LARDIERE, D. (2008) Feature assembly in second language acquisition IN: LICERAS, J.M. & ZOBL, H. & GOODLUCK, H. (eds.) *The Role of Features in Second Language Acquisition.* Mahwah, NJ: Erlbaum. pp.106-140

LARDIERE, D. (2009) Some thoughts on contrastive analysis of features in second language acquisition. *Second Language Research*, 25. pp.173-227

LASNIK, H. & LIDZ, J.L. (2016) The Argument from the Poverty of the Stimulus IN: ROBERTS, I. (ed.) *The Oxford Handbook of Universal Grammar.* Oxford: OUP.

LEGATE, J.A. & YANG, C.D. (2002) Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review,* 19. pp.151-62

LENNON, P. (1990) Investigating fluency in EFL: A quantitative approach. *Language Learning,* 40. pp.387-417

LICERAS, J. & ZOBL, H. & GOODLUCK, H. (eds.) (2008) *The role of formal features in second language acquisition.* London; New York: Lawrence Erlbaum Associates

LIDZ, J. & GAGLIARDI, A. (2015) How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics*, 1. pp.333-53. doi:10.1146/annurev-linguist-030514-125236

LINCK, J.A. & CUNNINGS, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185-207

LLANES, À. (2011) The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism,* 8(3). pp.189-215

LLANES, À & MUÑOZ, C. (2009) A short stay abroad: does it make a difference? *System,* 37(3). pp.353-365

LOGIE, R.H. & MORRIS, R.G. (eds.) (2014) *Working memory and ageing*. Psychology Press

LONG, M.H. (2000) Focus on form in task-based language teaching. *Language policy and pedagogy: Essays in honor of A. Ronald Walton.* pp.179-192

LONG, M. & SATO, C.J. (1984) Methodological issues in interlanguage studies: an interactionist perspective IN: DAVIES, A. & CRIPER, C. & HOWART, A.P.R. (eds.) *Interlanguage.* Edinburgh: Edinburgh University Press. pp.253-280

MAI, Z. & YUAN, B. (2016) Uneven reassembly of tense, telicity and discourse features in L2 acquisition of the Chinese *shì… de* cleft construction by adult English speakers. *Second Language Research,* 32(2). pp.247-276

MCMANUS, K. (2011) The development of aspect in a second language. PhD thesis, Newcastle University

MCMANUS, K. (2013) Prototypical influence in second language acquisition: What now for the Aspect Hypothesis. *International Review of Applied Linguistics*, 51. pp.299-322

MCMANUS, K. (2015) L1/L2 Differences in the Acquisition of Form-Meaning Pairings: A Comparison of English and German Learners of French. *The Canadian Modern Language Review/La revue canadienne des langues vivantes,* 71(2). pp.155-181

MCMANUS, K. (2019) Relationships between social networks and language development during study abroad. *Language, Culture and Curriculum*, 32(3). pp. 270-284, DOI: 10.1080/07908318.2019.1661683

MCMANUS, K. & MITCHELL, R. & TRACY-VENTURA, N. (2014) Understanding insertion and integration in a study abroad context: the case of English-speaking sojourners in France. *Révue française de linguistique appliquée,* XIX-2. pp.97-116

MCMANUS, K. & MITCHELL, R. (2015) Subjunctive use and development in L2 French: a longitudinal study. *Language, Interaction, and Acquisition,* 6(1). pp.42-73

MILLER, K.L. & SCHMITT, C. (2012) Variable Input and the Acquisition of Plural Morphology. *Language Acquisition*, 19(3). pp.223-261

MITCHELL, R. (2015) The development of social relations during residence abroad. *Innovation in Language Learning and Teaching,* 9(1). pp.22-33

MITCHELL, R. & DOMÍNGUEZ, L. & ARCHE, M. J. & MYLES, F. & MARSDEN, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. *EUROSLA Yearbook 8*(1), pp. 287-304.

MITCHELL, R. & MCMANUS, K. & TRACY-VENTURA, N. (2015) Placement type and language learning during residence abroad IN: Mitchell, R. & Tracy-Ventura, N. & McManus, K. (eds.) Social Interaction, Identity and Language Learning During Residence Abroad. EuroSLA Monographs, 4. pp.115-138

MITCHELL, R. & TRACY-VENTURA, N. & MCMANUS, K. (2017) Anglophone students abroad: Identity, Social Relationships and Language Learning. Abingdon; New York: Routledge

MONTRUL, S. & SLABAKOVA, R. (2002) The L2 acquisition of morphosyntactic and semantic properties of the aspectual tenses Preterite and Imperfect IN: PÉREZ-LEROUX, A.T. & MUÑOZ LICERAS, J. (eds.) *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection.* Dordrecht: Kluwer Academic Press. pp.113- 149

MONTRUL, S. & SLABAKOVA, R. (2003) Competence similarities between native and near-native speakers: An investigation of the Preterite-Imperfect contrast in Spanish. *Studies in Second Language Acquisition,* 25. pp.351-398

MORA, J. & VALLS-FERRER, M. (2012) Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly,* 46(4). pp.610–641

MUÑOZ, C. & LLANES, À. (2014) Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal*, 98(1). pp.432–449

O'BRIEN, I. & SEGALOWITZ, N. & FREED, B. & COLLENTINE, J. (2006) Phonological memory and lexical, narrative, and grammatical skills in second-language oral production by adult learners. *Applied Psycholinguistics*, 27. pp.377-402

O'BRIEN, I. & SEGALOWITZ, N. & FREED, B. & COLLENTINE, J. (2007) Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition,* 29. pp.557-81

PARADIS, J. (2010) Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning,* 60. pp.651-80

PÉREZ-LEROUX, A.T. & GLASS, W.R. (1999) Null anaphora in Spanish second language acquisition: Probabilisitic vs. generative approaches. *Second Language Research,* 15. pp.220-249

PHINNEY, M. (1987) The pro-drop parameter in second language acquisition IN: ROEPER, T. & WILLIAMS, E. (eds.) *Parameter setting.* Kluwer: Dordrecht

PINKER, S. (1984) *Learnability and cognition: the acquisition of argument structure.* Cambridge, MA: MIT Press

PINKER, S. (1994) *The Language Instinct* (1st edition)*.* New York: William Morrow and Co

PULLUM, G. & SCHOLZ, B. (2002) Empirical assessment of stimulus poverty arguments. *The Linguistic Review,* 19. pp.8-50

R CORE TEAM (2018) *R: a language and environment for statistical computing* [Software]. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org

RANKIN,T. & UNSWORTH, S. (2016) Beyond poverty: Engaging with input in generative SLA. *Second Language Research,* 32(4). pp.563-572. doi: 10.1177/0267658316648732

REHNER, K. & MOUGEON, R. (2003) The effect of educational input on the development of sociolinguistic competence by French immersion students: The case of expressions of consequence in spoken French. *Journal of Educational Thought,* 37(3). pp.259-81

REICHENBACH, H. (1947) *Elements of symbolic logic.* New York: The Free Press

ROTHMAN, J. & SLABAKOVA, R. (2018) The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition,* 40. pp.417-442

RINGBÖM, H.A. (2007) Cross-linguistic Similarity in Foreign Language Learning. Clevedon: Multilingual Matters

ROBERTS, L. & LISZKA, S.A. (2013) Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research,* 29(4). pp.413-439

RYAN, J. & LAFFORD, B. (1992) Acquisition of lexical meaning in a study abroad environment: Ser and estar and the Granada experience. *Hispania,* 75(3). pp.714- 722

SALABERRY, M.R. (1999) The development of past tense verbal morphology in classroom L2 Spanish. *Applied Linguistics,* 20. pp.151-178

SALABERRY, M.R. (2000) The acquisition of English past tense in an instructional setting. *System,* 28. pp.135-52

SALABERRY, M.R. (2002) Tense and aspect in the selection of Spanish past tense verbal morphology IN: SALABERRY, M.R. & SHIRAI, Y. (eds.) *Tense-Aspect Morphology in L2 Acquisition*. Amsterdam; Philadelphia: John Benjamins. pp.397- 415

SALABERRY, M.R. (2003) Tense and aspect in verbal morphology. *Hispania,* 3. pp.559-573

SALABERRY, M.R. (2005) Evidence for transfer of knowledge of aspect from L2 Spanish to L3 Portuguese IN: AYOUN, D. & SALABERRY, M.R. (eds.) *Tense and Aspect in Romance Languages.* Amsterdam; Philadelphia: John Benjamins. pp.179- 210

SALABERRY, M.R. (2008) Marking Past Tense in Second Language Acquisition: A Theoretical Model. London: Continuum

SANZ, C. & MORALES-FRONT, A. (2018) Introduction: Issues in study abroad research and practice IN: SANZ, C. & MORALES-FRONT, A. (eds.) *The Routledge Handbook of Study Abroad Research and Practice.* New York; London: Routledge. pp.1-16

SCHÜTZE, C.T. (1996) *The empirical base of linguistics: Grammaticality judgments and linguistic methodology* (1st ed.). University of Chicago Press

SCHÜTZE, C.T. (2016) *The empirical base of linguistics: Grammaticality judgments and linguistic methodology* (2nd ed.). Language Science Press.

SCHWARTZ, B.D. (1998) The second language instinct. *Lingua,* 106. pp.133-160

SCHWARTZ, B.D. & SPROUSE, R.A. (1994) Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage IN: HOEKSTRA, T. & SCHWARTZ, B.D. (eds.) *Language Acquisition Studies in Generative Grammar.* Amsterdam: John Benjamins. pp.317-368

SCHWARTZ, B.D. & SPROUSE, R.A. (1996) L2 cognitive states and the full transfer/full access model. *Second Language Research,* 12. pp.40-72

SCHWARTZ, B.D. & SPROUSE, R.A. (2013) Generative approaches and the poverty of the stimulus IN: YOUNG-SCHOLTEN, M. & HERSCHENSOHN, J.R. (eds.) *The Cambridge Handbook of Second Language Acquisition.* Cambridge: CUP. pp.137-159

SEGALOWITZ, N. & FREED, B. (2004) Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26. pp.173-199

SLABAKOVA, R. (2001) *Telicity in the second language* (vol. 26). Amsterdam: John Benjamins Publishing

SLABAKOVA, R. (2008) Meaning in the Second Language. Berlin: Mouton de Gruyter

SLABAKOVA, R. (2016) *Second Language Acquisition.* Oxford: Oxford University Press

SLABAKOVA, R. (2018) Back to our roots. *Linguistic Approaches to Bilingualism,* 8(6). pp.776-781

SLABAKOVA, R. & MONTRUL, S. (2002) Aspectual tenses in Spanish L2 acquisition: a UG perspective IN: SMITH, Y. (ed.) *The parameter of aspect.* Dordrecht: Kluwer Academic Press

SMITH, C.S. (1991) *The Parameter of Aspect* (1st edition). Dordrecht: Kluwer Academic Publishers

SMITH, C.S. (1997) *The Parameter of Aspect* (2nd edition). Dordrecht: Kluwer Academic Publishers

SMITH, C.S. (2006) The pragmatics and semantics of temporal meaning IN: DENIS, P. & MCCREADY, E. & PALMER, A. & REESE, B. (eds.) *Proceedings of the 2004 Texas Linguistic Society Conference.* Somerville, MA: Cascadilla Press. pp.92-106

SORACE, A. (2011) Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1. pp.1-33

SORACE, A. & FILIACI, F. (2006) Anaphora resolution in near-native speakers of Italian. *Second Language Research,* 22. pp.339-68

SPENCER, N.J. (1973) Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research,* 2(2). pp.83–98

SPINNER, P. (2013) The second language acquisition of number and gender in Swahili: A feature reassembly approach. *Second Language Research,* 29. pp.455-79

STOWELL, T.A. (1993) Syntax of tense. Manuscript, University of California, Los Angeles

STOWELL, T.A. (1996) The phrase-structure of tense IN: ROORYCK, J. & ZARING, L. (eds.) *Phrase structure and the lexicon [studies in natural language and linguistic theories].* Dordrecht: Kluwer. pp.277-91

STOWELL, T.A. (2007) The syntactic expression of tense. *Lingua,* 117. pp.437-63

TAGLIAMONTE, S. & LAWRENCE, H. (2000) “I used to dance, but I don’t dance now”: The habitual past in English. *Journal of English Linguistics*, 28. pp.324-353

TEMPLIN, M.C. (1957) Certain language skills in children: Their development and interrelationships. Westport, CT: Greenwood

THACKER, M. & D’ANGELO, C. (2013) *Essential French Grammar.* London; New York: Routledge. p.218

THOMAS, E.M. & WILLIAMS, N. & JONES, L.A. & DAVIES, S. & BINKS, H. (2014) Acquiring complex structures under minority language conditions: Bilingual acquisition of plural morphology in Welsh. *Bilingualism: Language and Cognition*, 17. pp.478-94

THORDARDOTTIR, E. (2011) The relationship between bilingual exposure and vocabulary development. *The International Journal of Bilingualism,* 15. pp.426-45

THORDARDOTTIR, E. (2015) The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology,* 17. pp.97-114

TOKOWICZ, N. & MICHAEL, E. & KROLL, J. (2004) The roles of study-abroad experience and working-memory capacity in the types of errors made using translation. *Bilingualism: Language and Cognition,* 7(3). pp.255-72

TOMASELLO, M. (2003) *Constructing a Language: A Usage-based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press

TRACY-VENTURA, N. & CUESTA MEDINA, J. (2018). Can native-speaker corpora help explain L2 acquisition of tense and aspect? A study of the "input". *International Journal of Learner Corpus Research,* 4, 277-300

TREMBLAY, A. (2011) Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acqusition,* 33. pp.339-72

TREMBLAY, A & GARRISON, M.D. (2010) Cloze tests: A tool for proficiency assessment in research on L2 French IN: PRIOR, M.T. & WATANABE, Y. & LEE, S-K. (eds.) *Selected proceedings of the Selected Language Research Forum 2008.* Somerville, MA: Cascadilla Press. pp.73-88

UNSWORTH, S. (2013) Assessing the role of current and *cumulative* exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition,* 16. pp.86-110

UNSWORTH, S. (2014) Comparing the role of input in bilingual acquisition across domains IN: GRÜTER, T. & PARADIS, J. (eds.) *Input and experience in bilingual development.* Amsterdam: John Benjamins. pp.181-201

VAINIKKA, A. & YOUNG-SCHOLTEN, M. (1994) Direct access to X'-theory: Evidence from Korean and Turkish adults learning German IN: HOEKSTRA, T. & SCHWARTZ, B.D. (eds.) *Language acquisition studies in generative grammar.* Amsterdam: John Benjamins. pp.265-316

VAINIKKA, A. & YOUNG-SCHOLTEN, M. (1996) Gradual development of L2 phrase structure. *Second Language Research,* 12(1). pp.7-39

VERKUYL, H.J. & VET, C. & BORILLO, A. & BRAS, M. & LE DRAOULEC, A. & MOLENDJIK, A. & VIEU, L. (2005) Tense and aspect in sentences IN: CORBLIN, F. & DE SWART, H. (eds.) *Handbook of French Semantics.* Stanford: CSLI Publications

WALLINGTON, A. (2017) *Exploring the reassembly of imperfective features in advanced learners of L2 French: aspectual development during 'the year abroad'.* Unpublished MA dissertation, University of Southampton

WESTERGAARD, M. (2009) *The acquisition of word order: Micro-cues, information structure and economy*. Linguistik Aktuell / Linguistics Today 145. Amsterdam: John Benjamins

WESTERGAARD, M. (2014) Linguistic variation and micro-cues in first language acquisition. *Linguistic Variation,* 14. pp.26–45

WHITE, L. (1989) *Universal Grammar and second language acquisition.* Amsterdam; Philadelphia: John Benjamins

WHITE, L. (2003) *Second language acquisition and Universal Grammar.* Cambridge: Cambridge University Press

WILMET, M. (2003) *Grammaire critique du français.* Brussels: Duculot

YANG, C. (2002) *Knowledge and Learning in Natural Language.* Oxford: Oxford University Press

YANG, C. (2006) *The infinite gift: How children learn and unlearn the languages of the world*. New York: Scribner

YANG, C. (2018) A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism,* 8(6). pp.665-706

YANG, C. & MONTRUL, S. (2017) Learning datives: the Tolerance Principle in monolingual and bilingual acquisition. *Second Language Research*, 33(1). pp.119-144. doi: 10.1177/0267658316673686

ZAGONA, K. (1990) Times as temporal argument structure. Manuscript, University of Washington, Seattle

Appendix A - Data collection materials

A.1 Sample pictures from “Cat Story” production task

(Source: langsnap.soton.ac.uk/tasks/html, accessed 20/05/19)

“Tous les matins étaient pareils…”

“Mais il est arrive un jour…”

A.2 Language Engagement Questionnaire (LEQ) – List of items rated

Participants were asked to give each of the following exposure contexts one of the following ratings: 'Every day'; 'Several times a week'; 'A few times a week'; 'A couple of times a month'; 'Rarely'; 'Never'. The questionnaire was completed separately for both English and French.

*How often do you do the following in **FRENCH/ENGLISH?***

- Watch TV
- Watch films
- Browse the internet (e.g. read news etc.)
- Use social networking sites (e.g. Facebook/Twitter)
- Read emails
- Write emails
- Listen to music
- Listen to talk radio
- Listen to lectures
- Participate in seminars/language classes
- Read literature (e.g. fiction, poetry, short stories)
- Read academic texts
- Read newspapers
- Read magazines
- Read text messages
- Write text messages

- Write reports (e.g. work, academic)

- Write for leisure (e.g. journal)

- Use instant messaging

- Have short phone conversations (<5 minutes)

- Have long phone conversations (>5 minutes)

- Teach a class

- Engage in service encounters

- Engage in small talk

- Engage in long casual conversations

- Participate in organised social activities (e.g. clubs, church, sports etc.)

A.3 Acceptability judgement task (AJT) items

Aspectual condition/ Lexical aspect/ Target Form **(Item name) [Contextual congruence ("semantic connectedness") (0/1)]**	**Context (Eng, Fr)**	**Sentences to Rate (Target sentence *italicised*)**
Habitual / Eventive / Imperfect (HabEv1) [0]	Pierre's building company has shut down. It's a pity because his company was involved in a reconstruction programme that worked in war zones whenever necessary. La compagnie de construction de Pierre a dû fermer. C'est dommage vu que cette compagnie faisait partie d'un programme de reconstruction qui travaillait dans des zones de guerre en cas de nécessité.	*La compagnie construisait des hôpitaux dans les zones de conflit.* [The company built-IMP hospitals in conflict zones.] La compagnie a construit des hôpitaux dans les zones de conflit. [The company built-PC hospitals in conflict zones.]
Habitual / Eventive / Imperfect (HabEv2) [0]	Jean says that he has fond memories of his childhood, especially when he went on picnics with his grandparents. Jean dit qu'il a de bons souvenirs de son enfance, surtout de quand il faisait des pique-niques avec ses grands-parents.	Jean a mangé au parc. [Jean ate-PC in the park.] *Jean mangeait au parc.* [Jean ate-IMP in the park.]
Habitual / Eventive / Imperfect (HabEv3) [1]	I was always a bit lazy when I was in secondary school, and it was always difficult for me to wake up early on school days. J'étais toujours un peu paresseux quand j'étais au collège, et je le trouvais toujours difficile de me réveiller tôt pendant la semaine.	*J'arrivais en classe en retard.* [I arrived-IMP late to class.] Je suis arrivé(e) en classe en retard. [I arrived-PC late to class.]

Habitual / Stative / Imperfect (HabSta1) [1]	When Anne-Marie was a child she had a very close friend, Amélie, and she liked to spend a lot of time at her house after school. Quand Anne-Marie était petite elle avait une amie proche qui s'appelait Amélie, et elle aimait passer beaucoup de temps chez elle après l'école.	Anne-Marie est souvent allée chez Amélie à la sortie du collège. [Anne-Marie often went-PC to Amélie's house after school.] *Anne-Marie allait souvent chez Amélie à la sortie du collège.* [Anne-Marie often went-IMP to Amélie's house after school.]
Habitual / Stative / Imperfect (HabSta2) [0]	When my brother Sam was in secondary school he did not do very well in his classes whenever he was going out with a girl. Quand mon frère Sam était au collège il ne réussissait pas très bien ses cours quand il sortait avec une fille.	*Sam avait besoin d'aide avec ses devoirs quand il avait une copine.* [Sam needed-IMP help with his homework when he had a girlfriend]. Sam a eu besoin d'aide avec ses devoirs quand il avait une copine. [Sam needed-PC help with his homework when he had a girlfriend.]
Habitual / Stative / Imperfect (HabSta3) [1]	Martine has moved to different flat in a much quieter part of town. Before, she was too close to a train station and couldn't sleep well at all. Martine a déménagé dans un nouvel appartement dans un quartier beaucoup plus calme. Avant, elle habitait trop près d'une gare et elle ne dormait pas bien du tout.	Martine a entendu les trains le matin. [Martine heard-PC trains in the morning.] *Martine entendait les trains le matin.* [Martine heard-IMP trains in the morning.]

Aspectual condition / Lexical aspect / Target form **(Item name)** **[Contextual congruence (0/1)]**	**Context (Eng/Fr)**	**Sentences to rate**
Perfective / Eventive / Passé composé (PerfEv1) [1]	My friend Pippa is very caring. She prefers to spend her holidays volunteering and helping others in less fortunate parts of the world. For example, this Christmas she was in Haiti working to build an orphanage. Mon amie Pippa is très gentille. Elle préfère passer ses vacances à faire du bénévolat et à aider ceux qui vivent dans des pays défavorisés. Par exemple, à Noël elle était en Haïti où elle a travaillé sur la construction d'un orphelinat.	Pippa construisait un orphelinat. [Pippa built-IMP an orphanage.] *Pippa a construit un orphelinat.* [Pippa built-PC an orphanage.]
Perfective / Eventive /Passé composé (PerfEv2) [1]	My mum is such a book worm. She reads whenever she gets a chance. This past Christmas, I gave her the last Harry Potter book and on Boxing Day she was threatening to give the ending away. Ma mère est une lectrice avide ; elle lit à tout moment possible. À Noël je lui ai donné le dernier tome de Harry Potter et le lendemain elle menaçait déjà de nous révéler la fin.	*Ma mère a lu le dernier livre de Harry Potter.* [My mum read-PC the last Harry Potter book.] Ma mère lisait le dernier livre de Harry Potter. [My mum read-IMP the last Harry Potter book.]
Perfective / Eventive / Passé composé (PerfEv3) [1]	It was so warm and nice that Jean decided to go out for a walk during his break and have lunch outdoors. Il faisait si beau et si chaud que Jean a décidé d'aller se promener pendant sa pause et de déjeuner dehors.	Jean mangeait au parc. [Jean ate-IMP in the park.] *Jean a mangé au parc.* [Jean ate-PC in the park.]

Perfective / Eventive / Passé composé (PerfEv4) [0]	My brother is 18 and has never had a girlfriend. But this morning my mum found a handbag in the car, which forced my brother to explain what he did last night. Mon frère a dix-huit ans et il n'a jamais eu de copine. Mais ce matin ma mère a trouvé un sac à main dans la voiture, ce qui a obligé mon frère à expliquer ce qu'il a fait hier soir.	*Mon frère est sorti avec sa copine.* [My brother went out-PC with his girlfriend.] Mon frère sortait avec sa copine. [My brother went out-IMP with his girlfriend.]
Perfective / Eventive / Passé composé (PerfEv5) [0]	I woke up very late and I missed the bus to school. So, I had to phone my mum and ask her to take me to school. Je me suis réveillée très tard et j'ai raté le bus, alors j'ai dû téléphoner à ma mère pour lui demander de m'emmener à l'école.	*Je suis arrivé(e) en classe en retard.* [I arrived-PC late to class.] J'arrivais en classe en retard. [I arrived-IMP late to class.]
Perfective / Stative / Passé composé (PerfSta1) [1]	Rachel's grandma is normally very healthy. However, last winter she caught a cold that became very complicated and she ended up in hospital for a month. D'habitude, la grand-mère de Rachel est en très bonne santé. L'hiver dernier, cependant, elle a attrapé un rhume qui s'est compliqué et elle a fini par passer un mois à l'hôpital.	Sa grand-mère était très malade. [Her grandmother was-IMP very ill.] *Sa grand-mère a été très malade.* [Her grandmother was-PC very ill.]
Perfective / Stative / Passé composé (PerfSta2) [0]	My mum told me yesterday morning that my friend Sam had phoned to cancel our revision session for that afternoon. Later, I found out that he had got the class notes from somebody else. Hier matin ma mère m'a dit que mon ami Sam avait téléphoné pour annuler notre séance de révisions qui était prévu pour l'après-midi. J'ai découvert plus tard que quelqu'un d'autre lui	*Sam n'a pas eu besoin d'aide avec ses devoirs.* [Sam didn't need-PC help with his homework.] Sam n'avait pas besoin d'aide avec ses devoirs. [Sam didn't need-IMP help with his homework.]

	avait déjà donné une copie de ce qu'on avait noté en classe.	
Perfective / Stative / Passé composé (PerfSta3) [1]	Last night, Martine got very scared when she was in bed. Around 2 am there was a loud car crash in her street and it woke her up. Hier soir, Martine a eu très peur alors qu'elle était au lit. Vers 2h du matin un grand bruit venant d'un accident de voiture dans la rue l'a réveillée.	Martine entendait un bruit. [Martine heard-IMP a noise.] *Martine a entendu un bruit.* [Martine heard-PC a noise.]

Aspectual condition / Lexical aspect / Target form **(Item name) [Contextual congruence (0/1)]**	**Context (Eng/Fr)**	**Sentences to rate**
Continuous / Stative / Imperfect (ContSta1) [0] (ContSta2) [1]	Last weekend I spent some time with my neighbour Jean. He has been having lots of problems with his new puppy Olivier. Le week-end dernier j'ai passé du temps avec mon voisin, Jean. En ce moment son nouveau chiot, Olivier, lui cause beaucoup de soucis.	Quand j'ai rendu visite à Jean, son chien a paru très fatigué. [When I visited Jean, his dog seemed-PC very tired.] *Quand j'ai rendu visite à Jean, son chien paraissait très fatigué.* [When I visited Jean, his dog seemed-IMP very tired.]
(ContSta3) [1] (ContSta4) [1]	My husband and I have moved to the south of France looking for some sun. Although we liked Scotland, we were a bit tired of the cold weather. Mon mari et moi avons déménagé au sud de la France à la recherche du soleil. Bien que nous ayons bien aimé vivre en Écosse, nous avions marre du temps froid.	*En Écosse, il faisait très froid.* [In Scotland, it was-IMP very cold.] En Écosse, il a fait très froid. [In Scotland, it was-PC very cold.]
	We had plans to go to a Chinese restaurant last Saturday after watching the new Bond movie. On our way to the restaurant the bus broke down so we arrived very late. Nous avions prévu d'aller à un restaurant chinois samedi dernier après avoir vu le dernier film James Bond. Le bus est tombé en panne en route pour le restaurant, alors nous y sommes arrivés très tard.	Quand nous sommes arrivés, le restaurant a été fermé. [When we arrived, the restaurant was-PC closed.] *Quand nous sommes arrivés, le restaurant était fermé.* [When we arrived, the restaurant was-IMP closed.]
	Guillaume has been a bit depressed lately: his girlfriend has left him and he is not doing well in his classes. Last weekend we	*Guillaume se sentait très triste.* [Guillaume felt-IMP very sad.]

	ran into him on our way to the sports centre. Guillaume est au creux de la vague ces temps-ci: sa copine vient de le quitter et il ne réussit pas très bien ses cours. Le week-end dernier nous l'avons croisé en route pour le centre sportif.	Guillaume s'est senti très triste. [Guillaume felt-PC very sad.]

Aspectual condition / Lexical aspect / Target form **(Item name) [Contextual congruence (0/1)]**	**Context (Eng/Fr)**	**Sentences to rate**
Progressive / Eventive / Imperfect (ProgEv1) [1] (ProgEv2) [0] (ProgEv3) [1] (ProgEv4) [1]	We went to the teachers' room to look for Mademoiselle Dupont, the new French language assistant, but she wasn't there. Instead, Ms Robinson the English teacher was there, working on our final exam. Nous sommes allés à la salle des profs à la recherche de Mlle. Dupont, la nouvelle assistante de langue française, mais elle n'était pas là. Pourtant, Mme. Robinson (la professeure d'anglais), qui travaillait sur notre examen final, s'y trouvait.	La professeure d'anglais a préparé l'examen final. [The English teacher prepared-PC the final exam.] *La professeure d'anglais préparait l'examen final.* [The English teacher was preparing-IMP the final exam.]
	I have just come back from visiting my cousin Oscar. He had just come back from school and was keeping himself	*Oscar lisait un livre.* [Oscar was reading-IMP a book.] Oscar a lu un livre. [Oscar read-PC a book.]

	occupied until dinner time. Je viens de rendre visite à mon cousin Oscar. Il venait de rentrer de l'école et s'occupait avant l'heure du dîner.	
	My sister was invited to a concert but she got there late. When she finally arrived, the pianist had already started playing. On avait invité ma sœur à un concert mais elle est arrivée en retard. Quand elle y est enfin entrée, le pianiste avait déjà commencé à jouer.	Le pianiste a joué du piano quand ma sœur est arrivée. [The pianist played-PC the piano when my sister arrived.] *Le pianiste jouait du piano quand ma sœur est arrivée.* [The pianist was playing-IMP the piano when my sister arrived.]
	Ségolène has just broken up with her boyfriend and she is not her usual happy self. She hasn't been going out much and I haven't seen her in a while. Ségolène vient de rompre avec son copain et elle ne semble pas être dans son assiette. Elle ne sort plus trop et ça fait un moment que je ne la vois pas.	*Quand Ségolène sortait avec Thibault, elle était heureuse*. [When Ségolène was going out-IMP with Thibault, she was happy.] Quand Ségolène est sortie avec Thibault, elle était heureuse. [When Ségolène went out-PC with Thibault, she was happy.]

Appendix B : Individual scores per task

B.1 Individual Scores: Pre-YA (*n* = 20)

	PRODUCTION			COMPREHENSION
Partic. ID	Cat Story	Conversation	Overall	AJT
201	90.0%	81.3%	85.2%	73.8%
202	100.0%	94.4%	95.7%	73.8%
203	43.9%	94.0%	82.7%	61.9%
204	50.0%	88.9%	80.1%	59.5%
205	64.7%	77.6%	76.3%	69.0%
206	94.0%	90.0%	91.2%	59.5%
207	95.5%	94.4%	94.6%	78.6%
209	79.4%	85.8%	80.0%	71.4%
210	97.0%	80.1%	85.6%	71.4%
212	91.0%	89.2%	90.4%	66.7%
213	66.7%	90.4%	82.9%	
214	91.7%	91.0%	92.9%	61.9%
215	80.7%	90.0%	91.7%	73.8%
218	56.5%	96.7%	77.7%	59.5%
220	63.3%	92.5%	80.4%	59.5%
221	34.7%	74.5%	65.5%	50.0%
223	45.8%	79.8%	68.2%	73.8%
224	29.4%	84.6%	63.2%	71.4%
226	56.4%	93.3%	77.0%	
229	28.7%	70.8%	61.9%	61.9%

B.2 Individual Scores: Post-YA (*n* = 23)

	PRODUCTION			COMPREHENSION
Partic. ID	Cat Story	Conversation	Overall	AJT
402	57.5%	79.0%	77.1%	52.4%
403	68.5%	88.0%	82.1%	38.1%
404	100.0%	84.5%	87.9%	83.3%
405	58.0%	75.3%	73.5%	76.2%
406	71.3%	87.6%	81.8%	73.8%
407	42.3%	84.5%	70.6%	66.7%
408	94.5%	84.8%	86.0%	73.8%
409	47.1%	73.7%	58.4%	64.3%
410	91.7%	96.5%	94.8%	54.8%
411	25.2%	83.0%	74.3%	81.0%
412	29.1%	67.1%	58.7%	69.0%
413	53.3%	88.4%	70.3%	61.9%
415	32.5%	85.3%	67.1%	76.2%
416	4.7%	64.6%	41.3%	64.3%
417	57.7%	94.6%	83.1%	40.5%
418	66.7%	69.6%	68.8%	73.8%
419	80.0%	79.4%	83.2%	69.0%
421	59.7%	87.6%	83.8%	78.6%
422	0.0%	96.1%	62.9%	38.1%
423	53.7%	80.6%	78.8%	50.0%
424	80.8%	90.5%	86.2%	61.9%
425	41.7%	82.6%	75.8%	61.9%
427	90.2%	90.5%	90.5%	73.8%

B.3 Individual Scores: L1 Production (*n* = 7)

	PRODUCTION		
Partic. ID	Cat Story	Conversation	Overall
603	94.1%	100.0%	98.7%
604	100.0%	100.0%	100.0%
605	*45.9% **	97.7%	75.1%
606	100.0%	100.0%	100.0%
607	100.0%	100.0%	100.0%
608	95.2%	100.0%	98.1%
611	100.0%	100.0%	100.0%

* Partly recounted Cat Story narrative in the Present

B.4 Individual Scores: L1 Comprehension (*n* = 14)

COMPREHENSION	
Partic. ID	AJT
901	54.8%
902	78.6%
903	71.4%
904	50.0%
905	81.0%
906	40.5%
907	61.9%
908	71.4%
910	57.1%
911	61.9%
912	64.3%
914	69.0%
915	76.2%
916	76.2%