The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study

Caroline Cordier

Doctor of Philosophy

School of Modern Languages, Newcastle University, UK

September 2013

Abstract

The present study is a longitudinal investigation of the presence, nature, and role of formulaic sequences (FS) in advanced English learners of French. The learners investigated are in their second year of an undergraduate degree in French at the onset of the study, and are tested before and after a seven-month stay in France. FS are defined psycholinguistically as multiword units which present a processing advantage for a given speaker, either because they are stored whole in his/her mental lexicon (Wray 2002) or because they are highly automatised.

The construct of FS is particularly relevant to investigate key linguistic issues such as the dynamism of linguistic representations, their idiosyncratic nature as well as the relationship between the lexicon and grammar. FS have been shown to be frequent in the oral productions of native speakers. They also play an important role in first language acquisition as well as in the initial stages of instructed second language (L2) acquisition. However, very little is known about their presence and role in advanced L2 learners, as most studies dealing with them have not adopted a psycholinguistic approach and have focused on L2 learners' knowledge and use of idioms and idiomatic expressions.

Conversely, this study seeks to evaluate and characterise the presence of psycholinguistically-defined FS in advanced learners as well as examine their longitudinal development in relation to the development of the learners' fluency and lexical diversity. It seeks to determine whether FS use can be said to play a role in the development of fluency and lexical diversity and if it does, describe the underlying mechanisms that account for this role.

Data from five learners performing five oral tasks (an interview, a story retell and 3 discussion tasks), repeated before and after their stay in France, was elicited and transcribed. FS were identified through the hierarchical application of a range of criteria aiming to capture the holistic nature of the sequences. The necessary criterion used for identification was fluent pronunciation of the sequence, and additional criteria were applied such as irregularity, holistic mapping of form to meaning or holistic status of the sequence in the input. Fluency was operationalised through 4 measures (phonation-time ratio, speaking rate, mean length of runs and articulation rate) and lexical diversity was measured using D.

The results show that psycholinguistically-defined FS represent about 27% of the language of advanced learners of French. The typology of the identified sequences shows that they are mostly grammatically regular but that despite the advanced level of the participants, some present non-nativelike characteristics. Individual differences in the learners' repertoires of FS as well as task effects are also found.

Between time 1 and time 2, across the group of 5 subjects, there is a general and statistically significant increase in FS use, fluency and lexical diversity. Significant correlations are found between FS use, fluency and lexical diversity. The qualitative analysis suggests that FS use plays a role in increasing fluency by allowing longer speech runs, contributing to the reduction of pausing time as well as the speeding up of the articulation rate. At the internal level of processing mechanisms, the results suggest that FS play a facilitating role not only in the formulation stage of speech production but also in the conceptualisation and articulation stages. Significant correlations are also found between FS use and lexical diversity, which suggests that FS, by lightening the processing burden and freeing some attentional resources, might facilitate the acquisition of new vocabulary.

The analysis of the development of the learners across all variables shows a single developmental path with similar processes of automatisation but with different rates of acquisition, as the learners vary in how efficient they are at proceduralising their language. Because of this, it is suggested that the year abroad is more likely to be beneficial for a given subject if their language has already reached a certain level of automatisation pre-time abroad.

Acknowledgements

I am very grateful to the Arts and Humanities Research Council for funding my doctoral research.

I cannot thank enough my main supervisor, Professor Florence Myles for her invaluable help, guidance and feedback, as well as for her unwavering support and faith in me.

I would also like to thank my second supervisor Professor David Howard, particularly for his help with statistics but also for his insightful suggestions and enthusiasm for my project.

In addition to my supervisors, I would like to thank Dr Christophe Dos Santos and Dr Jalal Al-Tamimi for their invaluable help with Praat, Dr Annabelle David for all her input on lexical diversity and help with the CHILDES software, Dr Jo Lumley for all his help and Dr Clare Wright for being a great mentor.

On a personal note, I would like to thank all my friends and particularly Adeline, Cécilia, Julie, Kevin, Pascale and Sophie for their continuous support. Also, I want to thank Suzie for making me feel at home in her house. I am truly grateful to my family and particularly to my unbelievably patient and supportive husband Borrie for assisting me through every stage of this very long process.

Finally, my thoughts go to the five students who kindly accepted to be part of this research project. I would like to thank them for their availability, reliability and enthusiasm for the French language.

Contents

List of tables

List of graphs

List of figures

Chapter 1. Introduction

The present study is a longitudinal investigation of the presence, nature and role of formulaic sequences (FS) in advanced learners of French, before and after a year abroad and in particular how it links with the development of fluency. It aims to contribute to the wide and dynamic research field on formulaicity by adopting both a psycholinguistic and a second-language acquisition point of view.

The scope of the research into formulaicity, i.e. multiword units, is very broad and includes the study of a variety of wordstrings from idioms (*to be caught red handed*), proverbs (*the more the merrier*), turns of phrase (*will you marry me?*), partly fixed frames (NP be-TENSE *past* POSSESIVE *sell-by date*) to collocational associations (*pitch black*).

Wray (2008) makes an essential distinction between (1) 'speaker-external' and (2) speaker-internal approaches to formulaicity. Speaker-external approaches investigate the phenomenon of formulaicity in the language outside the speaker, i.e. either in the formal properties of strings (for example their irregular semantic or syntactic nature), in their frequency of occurrences in various corpora or in their pragmatic functions. Speaker-internal approaches, on the other hand, investigate sequences considered formulaic because they are psycholinguistic units for a given speaker i.e. they are retrieved with greater efficiency than other linguistic strings by this individual speaker.

This study adopts a speaker-internal approach to FS, which are defined as multiword semantic / functional units presenting a processing advantage for a given speaker, either because they are likely to be stored whole in their lexicon or because they are highly automatised. The construct investigated in this study will henceforth be referred to as either 'psycholinguistic' or 'speaker-internal' FS as opposed to sequences which are formulaic 'in the language' either because they belong to the language's repertoire of idiomatic expressions or because they occur frequently in corpora, but which are not necessarily formulaic in the mind of a given speaker. This second type of sequences, 'speaker-external' FS will also be referred to as either 'linguistic FS' because they are formulaic 'in the language' or 'idiomatic' FS in the sense that they are 'readily acceptable to native speakers as an ordinary and natural form of expression' (Pawley and Syder 1983)

The abundance of research into formulaicity is epitomised by the high number of terms used to refer to it (over forty terms according to Wray (2000, 2002)). This variety of

approaches and terms can make the study of formulacity quite confusing. In certain cases, the difference is only terminological as the different terms refer in effect to the same construct. The variation in terminology can also reflect, however, the difference in the focus adopted by different approaches. For example, the term 'chunk' is used in psycholinguistic research whereas 'clusters' is favoured in corpus-linguistics. What is more problematic though, is when the same term is used by various researchers to refer to constructs which, although they might overlap, are not exactly the same. This is the case of the term 'formulaic sequence', which has been widely adopted and used by various researchers and has become an 'umbrella term' (Weinert 2010) since Wray (2002)'s synthesis on formulaic language. As underlined by Wray (2012), the confusion in terminology is potentially problematic when some claims are made about formulaic sequences in general while the approach taken only deals with one aspect of formulaicity and claims should not be extended to all kinds of formulaic language. Despite the potentially confusing nature of the term 'formulaic sequence', it was nevertheless chosen to refer to the construct investigated in this study. This terminological choice was made for two reasons. Firstly, as will be described in the next chapter, the definition of FS given by Wray (2002) has a clear psycholinguistic focus, which fits the research agenda of this study. Secondly, the adoption of the term 'formulaic sequence' reflects this study's will to contribute to the wider debate on formulaicity, although it must be underlined that all the results obtained in this study only apply to formulaicity as it is defined in this study as well as to the type of language users it focuses on: advanced L2 learners.

The rationale for this study stems from the fact that most of the research into formulaicity in advanced L2 learners is characterised by its speaker-external focus. Most of these studies have shown that L2 learners do not know idiomatic sequences as well as native speakers, which is hardly surprising given the difference in input these two types of language users receive. This study aims to show, however, that the fact that advanced learners have been shown to use fewer speaker-external FS than native speakers should not be equated with the claim that they do not use FS defined speaker-internally or that formulaicity does not play a role in L2 development.

This thesis is structured as follows:

Chapter 2 deals with the presentation of the psycholinguistic, speaker-internal approach used in this study and aims to show the specificity of this approach compared to

traditional speaker-external approaches to FS. The chapter also aims to show how the construct of FS understood psycholinguistically is relevant for the investigation of central questions in both general linguistic theory and first and second language acquisition.

As the relationship between FS use and the development of fluency in advanced learners is one of the main aims of this study, chapter 3 and 4 are dedicated to fluency. Chapter 3 deals with fluency at the cognitive level. It aims to analyse how the impact of FS on the development of fluency can be conceptualised at the cognitive level of underlying processing mechanisms, by integrating the construct of FS into models of speech production and showing which stages of speech production can be facilitated by the use of FS.

Chapter 4 deals the effect of the acquisition of FS on the development of utterance fluency i.e. the temporal characteristics of speech. Firstly, it presents how the fluency variable can be operationalised in several measures. It focuses in detail on the notion of fluent run, which is of central importance in this study, not only in order to measure the learners' fluency but also because it is at the centre of the FS identification process in this study.

Chapters 5 and 6 focus on the methodology used in the present study. Chapter 5 focuses on the identification of FS. It reviews the identification methodology used in L1 acquisition as well as the early stages of L2 acquisition before establishing a method of identification of FS in the advanced learners of this study.

Chapter 6 presents the methodology used in the present study. It describes the study design, explains how the data collected was analysed and accounts for all the methodological choices made.

The results of this study are presented and discussed from chapter 7.

Chapter 7 presents and analyses the use of FS by the learners of this study.

Chapter 8 and 9 present and discuss the results related to the learners' longitudinal development in FS use, lexical diversity and fluency. Chapter 8 presents the longitudinal development in FS use and its relation to the development in lexical diversity. Chapter 9 presents the fluency results and the correlations between FS use and the various fluency measures.

Chapter 10 attempts to link learners' longitudinal development in FS use, lexical diversity and fluency to the modalities of their stay in France.

The conclusions are presented in chapter 11.

Chapter 2. Psycholinguistic approach to formulaic sequences (FS)

This chapter focuses on the psycholinguistic, speaker-internal approach to FS adopted in this study. It has three main goals: (1) describe traditional speaker-external approaches to FS in order to set the research background on formulaicity and understand how the psycholinguistic approach relates to these approaches, (2) present and define the psycholinguistic, speaker-internal approach adopted in this study and (3) explain the importance of the construct of psycholinguistic, speaker-internal FS to central questions of linguistic theory as well as of first (L1) and second language (L2) acquisition.

The first part of this chapter deals with the traditional speaker-external approaches to formulaicity as well as the key principles at their heart. It explains why researchers within these speaker-external approaches believe that formulaicity is not a peripheral pursuit of linguistic research and should be placed at the centre of language competence. It also explains why the speaker-external approach to FS is at the origin of the speaker-internal approach to FS, as some researchers such as Pawley and Syder (1983) believe that FS defined speaker-externally are not just a phenomenon in the language but are also psycholinguistic constructs which present a processing advantage for native speakers. Their assumption is put to the test by reviewing many studies which investigate whether FS defined learner-externally present a processing advantage for native speakers and L2 learners. The first part concludes that the psycholinguistic reality of speaker-external FS cannot be taken for granted, especially in the L2 context.

The second part of this chapter therefore argues for a clear distinction between speaker-external and speaker-internal approaches to FS and presents the way psycholinguistic FS are defined in this study. It also shows how the construct of psycholinguistic FS can still be related to the key notions at the heart of formulaicity defined speaker-externally such as strong relationship between words or restricted creativity.

The last section of the chapter aims to show the importance of the study of psycholinguistic FS to understand central issues in both linguistic theory and language acquisition. It presents the usage-based framework for the study of language and shows how the study of FS can be a lens through which key questions within that framework can be investigated. It then describes the role of psycholinguistic FS in L1 and the early stages of L2 acquisition. Finally, in the context of the prevalence of studies dealing with speaker-external FS in advanced L2 learners, it argues for the necessity of also adopting

a psycholinguistic, speaker-internal perspective in the study of FS in advanced L2 learners.

2.1 Traditional speaker-external approaches to formulaicity

Various ways of approaching the study of formulaic language can be distinguished. Firstly, one way of looking at FS, mainly adopted in corpus linguistics, is statistical and studies recurrent clusters of words in corpora. Another possible approach is formal and focuses on strings which display various characteristics of irregularity such as semantically irregular sequences like *pull someone's leg,* or grammatically irregular sequences like *by and large.* Other researchers (Nattinger and DeCarrico 1992) adopt a pragmatic and functional account of formulaic language and focus on the contexts in which formulaic strings such as *how do you do* are used in social interaction.

In spite of important conceptual differences between these various approaches, there are a number of key notions underpinning them, which will be briefly summarised in the following section.

2.1.1 *Key notions at the heart of formulaicity*

2.1.1.1 *Strong relationship between words*

One of the core ideas behind the notion of formulaicity is that certain words have an especially strong relationship with each other (Wray 2008) either because corpus linguistics shows that they co-occur at a high frequency or because they form either a semantic or a syntactic whole. For example, if one considers the semantically irregular phrase *pull someone's leg*, the meaning of the sum of its parts does not equal the meaning of the whole. As a result, the phrase only gets its meaning when considered in its entirety. In the same way, the syntactically irregular phrase *by and large* has to be considered holistically and cannot be broken down since it is composed of the coordination of constituents of a different grammatical nature (a preposition and an adjective). It is the strength of the relationship between the different components of the sequence that gives it its unity. The strong relationship between words can also be due to the frequency of their co-occurrence e.g. *salt and pepper*.

2.1.1.2 *Restricted exchangeability, freedom and creativity*

For Erman and Warren (2000), who call formulaic strings 'prefabs', the decisive criterion to establish whether a given sequence is a prefab is that of 'restricted

exchangeability'. In other words, a word string can be considered formulaic only if it has a measure of conventionalisation, that is, if 'one member of the prefab cannot be replaced by a synonymous word without causing a change of meaning or function and / or idiomaticity'. An example chosen by them to illustrate this idea is the prefab *I'm afraid,* traditionally used to soften a piece of bad news and which cannot be replaced by *I'm scared* or *I'm frightened*.

Related to this idea of relative 'restricted exchangeability' is the notion of restricted freedom. Indeed as underlined by Sinclair (1991: 121), all the evidence available from corpus linguistics points to language as rigid, in the sense that particular combinations of words are favoured in preference to alternative combinations which would be equally grammatical but are not used by native speakers (Pawley and Syder 1983). In other words, only a small portion of the total set of grammatical sentences are nativelike and idiomatic in form, in the sense that native speakers accept them as ordinary and natural forms of expression.

The fact that native speakers have preferred choices to express certain meanings is also related to the idea that language use is not as creative as language competence would allow it to be. In this respect, it can be said that native speakers do not exercise the creative potential of syntactic rules to anything like their full extent (Pawley and Syder 1983: 193). This relative lack of creativity was brought to light by corpus linguistics. According to Sinclair (1991: 108), 'by far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns.' The patterning of words and phrases manifests far less variability than could be predicted on the basis of grammar and lexicon alone (Perkins 1999: 55-56). In other words, speakers underutilise the resources of language (Corrigan et al. 2009).

2.1.1.3 *Formulaicity as a discrete category or a graded notion*

It is possible to look at formulaicity as a discrete category to which some multiword sequences belong and some do not. In such a way of proceeding, one defines FS as a construct presenting several characteristics and rejects all the sequences which do not possess these characteristics. For example, if, in corpus linguistics, one's main criterion for formulaicity is frequency of occurrence, one can distinguish between the clusters that are formulaic because they occur above a certain frequency threshold and those that are not formulaic because they occur below that threshold. Within a formal approach, one could decide to only consider formulaic sequences that are semantically opaque and

grammatically irregular and reject expressions which are regular and semantically transparent.

However, it is not easy to decide on clear criteria on which to base the distinction between what is formulaic and what is not formulaic. For example, if identification is based on frequency of occurrence, it seems rather arbitrary to come up with a given frequency threshold deciding on the formulaic status of a given cluster of words (for example deciding that a given cluster is formulaic because it occurs ten times and another is not because it only occurs nine times). Or in the case of an approach based on semantic transparency, researchers might disagree about the status of certain sequences such as *spill the beans*, which are more transparent than a sequence like *kick the bucket* but less transparent than *tell the truth.*

That is why, many researchers, instead of conceptualising formulaicity as a discrete category, prefer to conceive of it as a graded notion (Coulmas 1994). Indeed, whatever the approach adopted by researchers, many of them agree that it is difficult to establish robust boundaries between what is formulaic and what is not. Consequently, rather than being categorical about the formulaic or non-formulaic nature of a given sequence, it seems more useful to place it along a continuum from the least to the most formulaic. Depending on one's approach to formulaicity, continua can be based on different dimensions of variation such as frequency of occurrence or semantic transparency and some researchers have also suggested using multi-dimensional continua (Wulff, in press).

2.1.2 *Speaker-external FS at the centre of language competence*

As mentioned in the introduction, although formulaicity was neglected as a research topic for a long time, there is now a consensus that formulaicity is a ubiquitous feature of language and as such should be placed at the centre and not the periphery of linguistic competence.

For some linguists, the fact that particular combinations of words are favoured by native speakers in preference to alternative, equally grammatical combinations only has to do with the realm of language use or, in Chomsky's (1965) terminology, performance. Researchers working in such a theoretical framework equate linguistic competence with knowledge of a generative grammar. As a result, for them, the linguistic phenomenon of formulaicity does not affect the way in which language competence should be defined

since, as put forward by Pinker (1994: 90): ‘using prefabricated chunks of language is a peripheral pursuit that tells us nothing about real language processing’.

For other researchers, however, formulaicity cannot be treated as an epiphenomenon for several reasons. Firstly, according to Pawley and Syder (1983), relegating formulaicity at the periphery rests on an incomplete definition of language competence. They define the ‘puzzle of nativelike selection’ as the ability of the native speaker to convey his meaning by an expression that is not only grammatical but also nativelike. According to them, this ability to select a natural and idiomatic sentence from all its possible grammatically correct paraphrases is necessary to be accepted as a native speaker and is therefore part and parcel of linguistic competence: ‘if a language learner is to achieve nativelike control, then, he must learn not only a generative grammar as this term is usually understood […] In addition he needs to learn a means for knowing which of the well-formed sentences are nativelike — a way of distinguishing those usages that are normal or unmarked from those that are unnatural or highly-marked’ (1983: 194). Although Pawley and Syder, by their own admission, fail to give a precise definition of the notion of naturalness, their contribution is essential in their acknowledgment that the issue of nativelike competence goes beyond that of grammatical correctness and that grammatical competence must only be regarded as a subset of nativelike competence.

Moreover, some researchers put formulaicity at the centre of language competence because they consider that a formulaic sequence is more than a valid construct to describe how language patterns or a prerequisite to sounding native and idiomatic: it is also relevant to describe speakers’ mental representations. In other words, some linguists regard formulaic multiword strings not only as a linguistic reality, but also as a psycholinguistic one, in the sense that they consider them mental wholes processed as units. For example, according to Pawley and Syder (1983: 192), the speaker is able to retrieve formulaic multiword expressions ‘as wholes or as automatic chains from the long-term memory’. Similarly, Sinclair (1991) proposes that, at the heart of language is the ‘principle of idiom’ according to which language-users have available to them ‘a large number of semi-pre-constructed phrases that constitute single choices, even though they might appear to be analysable into segments’. In other words, according to Sinclair, formulaic multiword sequences found in corpora are also psycholinguistic units despite the fact that, in theory, they could be generated from their constituents.

How these researchers conceive of the nature of these psycholinguistic units is not entirely clear however. Pawley and Syder's expression 'retrieved as a whole' suggests holistic retrieval of the sequence from the speaker's lexicon, which probably implies that the sequence is a lexical unit stored holistically in the lexicon However, retrieved 'as automatic chains' suggests the idea of a processing advantage of the sequence without making the claim that it is necessarily a lexical unit. Pawley and Syder do not elaborate further on these issues however. Similarly, Sinclair's notion of 'single choice' is not entirely clear. The idea of a single lexical choice suggests that a certain formulation is preferred over alternative ones to express a given semantic content. In Sinclair's definition, the notion of 'single choice' applies to grammatically regular sequences which means that even multiword sequences which can be decomposed into shorter segments can be regarded as 'single choices'. The fact that Sinclair describes formulaic phrases as 'semi-pre-constructed' also shows that he considers that frames with gaps for inserted variable items are formulaic. This, in turn, brings about questions about the degree of lexical fixedness of formulaic sequences and if some kind of generation still takes place in the case of sequences allowing for the insertion of variable items.

Moreover, although both Pawley & Syder and Sinclair evoke the issue of the relationship between the external linguistic characteristics of a sequence and its psycholinguistic treatment by a given speaker, they remain unclear in their explanation of how formulaicity in the language relates to formulaicity as a psycholinguistic phenomenon, i.e. the holistic treatment of a particular wordstring by a particular individual. Firstly, it is not clear whether Sinclair considers that all the recurrent sequences in corpora are stored whole in the lexicon of native speakers. As for Pawley and Syder (1983: 208) they make a distinction between (1) 'memorized sequences' which the speaker or hearer is capable of consciously assembling or analysing but which on most occasions of use are recalled as wholes or as automatically chained strings and (2) 'lexicalized sentence stems' which they define as a 'piece of timeless knowledge shared by the members of a language community'. In other words, they distinguish between (1) sequences that are a psycholinguistic whole for a given individual and (2) sequences which belong to the repertoire of idiomatic phrases known by all the speakers of a given language i.e. which are part of the lexicons of all native speakers. According to them, not all 'memorized sequences' are 'lexicalized sentence stems'. In other words, some of the sequences which are recalled as automatic wholes

by a given native speaker are idiosyncratic and are not conventional sequences for other native speakers. However, they say nothing about the fact that 'lexicalized sentence stems' might not all be memorised by a given speaker. Although they do not state it clearly, they seem to imply that 'lexicalized sentence stems' are psycholinguistically real for all the native speakers of English. In other words, they seem to consider that a speaker-external FS is also necessarily speaker-internal. The numerous issues raised by Pawley and Syder as well as Sinclair epitomise the fact that extending the construct of FS to the psycholinguistic domain is far from straightforward and that further investigation is needed with regard to the relationship between formulaicity in the language and formulaicity for a given speaker.

The next section aims to further develop this issue by giving an overview of the wide range of psycholinguistic studies which have investigated how FS defined speaker-externally are processed by individual speakers, either native speakers or L2 learners.

2.1.3 *Psycholinguistic studies on the processing of speaker-external FS*

Many studies using a range of psycholinguistic methodologies such as eye-tracking (Underwood, Schmitt and Galpin 2004), oral dictation (Schmitt, Grandage and Adolphs 2004), on-line grammaticality judgements (Jiang and Nekrasova 2007) and self-paced reading (Conklin and Schmitt 2008), have attempted to study the processing of some pre-identified speaker-external FS by both native speakers and L2 learners in order to investigate their psycholinguistic nature, i.e. whether they are also speaker-internal. In other words, these studies have aimed to examine whether some sequences which are formulaic 'in the language', either because of formal criteria or because of their highly frequent co-occurrence in corpora, can be equated with multiword lexical units stored whole in the lexicon or at least presenting a processing advantage over ordinary strings of words for individual speakers. These studies can be said to be working at the interface between the speaker-external and the speaker-internal approaches to formulaicity since they are testing whether speaker-external FS are speaker-internally valid.

2.1.3.1 *Psycholinguistic studies on idiom processing*

Studies on the nature of idiom processing

Many of the studies dealing with the processing of idiomatic sequences actually deal more precisely with the processing of one subtype, i.e. idioms. True idioms are defined

by Wray (2008: 10) as ‘a set of not all that frequent but particularly evocative multiword strings that express an idea metaphorically’, such as *kick the bucket, red herring*, and *raining cats and dogs*. Idiomatic FS constitute a larger category than idioms and can be defined as multiword sequences which are not only grammatically correct but are readily acceptable to native speakers as an ordinary and natural form of expression (Pawley and Syder 1983) such as *How are you doing?* or *this is out of the question*.

Despite the fact that idioms are only a subcategory of FS, idioms have been focused on a lot in psycholinguistic research because they are regarded as clear examples of FS as their semantic irregularity gives them unity. They are often conceived as ‘big words’ i.e. as multiword lexical units. According to some psycholinguistic experiments, however, acknowledging the fact that idioms are multiword units does not necessarily imply that they are processed like single lexical items or that when processing a multiword unit, no semantic or syntactic processing is taking place. For example, Cacciari and Tabossi (1988) do not see idioms as being encoded as separate entries in the mental lexicon. For them, the meaning of idioms is associated with particular configurations of words and is accessed when sufficient input has rendered the configuration recognisable (the configuration hypothesis). Peterson, Dell, Burgess and Eberhard (2001) tried to understand the processing differences between idioms such as *kick the bucket* and literal phrases such as *kick the ball*. They used both kinds of sentences, stripped of their final word, as primes for naming visually presented targets which were continuations with near-zero probability. By doing so, they meant to test whether syntactic expectations were present during the processing of predictable literal and idiomatic phrases. Their results for idioms showed that verb targets were named more slowly than noun targets, which points toward the existence of syntactic priming with idiomatic sequences. However, no conceptual priming was found for idioms, which suggests that by the final word of an idiomatic phrase, participants do not compute a literal interpretation of the phrase. According to Peterson et al., there therefore seems to be a dissociation between syntactic and semantic processing. Indeed, although the semantic processor terminates its analysis of the literal meaning of an idiomatic phrase once the figurative meaning has been retrieved, the termination of the semantic analysis does not result in subsequent elimination of syntactic processing. For the authors, these results point towards a modular view of language processing since the language comprehension system seems to be composed of a series of distinct processing levels which are autonomous from

each other. In terms of idiom processing, this means that although idioms present a processing advantage due to their semantic nature, they are still processed syntactically and cannot therefore be regarded as lexical units. Cutting and Bock (1997) experimentally elicited idiom blends (speech errors that accidentally combine two different idioms e.g. *the road to Chicago is as straight as a pancake*, which is a blend of *straight as an arrow* and *flat as a pancake*). Across all three experiments they carried out, idiom blends consistently involved structurally and semantically similar components. On the basis of these results, Cutting and Bock argue that idioms are not lexicalised chunks comparable to large single words but phrases with syntactic and semantic components. To sum up, according to the results of these various psycholinguistic experiments, the holistic semantic nature of idioms does not imply that their processing can be equated with the processing of individual lexical units as there is evidence of semantic and grammatical processing in the psycholinguistic treatment of idioms by language users.

Wray (2008: 30) calls into question the results reached by the above psycholinguistic experiments on the ground of their artificiality as, according to her, the gap between the artificial nature of these psycholinguistic experiments and what happens in real life casts some suspicion onto their approaches to testing. She emphasises that these types of experiments tap into a kind of processing that we can do in theory but that we do not do in normal circumstances. Indeed, if one considers the results of Cacciari and Tabossi (1988) for example, they only show that idioms are initially processed literally when there is an absence of contextual cues, which is unlikely to happen in real life when idioms are usually used in a clear pragmatic context.

Studies on the processing advantage of idioms over non-formulaic language

In any case, even if one agrees that idioms are not processed like single lexical items and that some semantic and syntactic processing is still taking place, the question of whether idioms present a processing advantage over non-formulaic strings of words can still be asked.

The question of the processing advantage of idioms has been raised since Swinney and Cutler (1979) compared the processing of idioms such as *break the ice* with that of matched literal phrases such as *break the cup* and found that native speakers processed the idioms more quickly than literal, non-formulaic phrases. Since then, this question has been studied through the use of sophisticated methodologies. For example,

Underwood et al. (2004) used eye-tracking to explore the processing of a series of ready-made expressions by a group of native speakers and a group of L2 learners who had to read short passages in which a series of idioms, proverbs, collocations, etc. had been included. What they regarded as critical data was the number and length of fixations by the participants on the last word of an idiomatic phrase versus on the same word in a non-formulaic sequence. The assumption behind this task was that, if some phrases possessed a processing advantage, the participants would be more likely to predict the final word of the sequence and as a result, there would be less need to fixate on this word. The results showed that terminal words in final position in formulaic sequences gained fewer fixations than the same words used in a non-formulaic context for both types of participants, although native speakers fixated the terminal words less often than the L2 learners. These results seem to show that both groups of participants were better at predicting the terminal words within a FS, which is consistent with the view that such FS present a processing advantage. In another and more recent eye-tracking study however, Siyanova-Chanturia, Conklin and Schmitt (2011) found that only the group of native speakers processed the idioms (*left a bad taste in my mouth*) faster than the matched control phrases *(the bad taste left in my mouth*). For the non-native speakers, the figurative phrases were read more slowly than the literal ones. Taken together, the above studies seem to support the view that idioms are processed more quickly than non-formulaic language by native speakers. However, the results of this research are mixed as to whether this processing advantage extends to L2 learners.

Several remarks need to be made regarding the above studies. First, it has to be borne in mind that any result finding some processing advantage for idioms cannot be taken as evidence that these sequences are units stored whole in the lexicon, although many authors e.g. Underwood et al. (2004) seem to make that equation. Moreover, the results obtained by these studies can only be applied to the limited set of sequences that were used as stimuli and that were selected from various dictionaries of idiomatic expressions. As a result, they cannot and should not be generalised, all the more so as the types of selected sequences represent a narrow conception of formulaic language as most of the sequences tested were idioms or even proverbs. As previously mentioned, idioms are usually chosen as stimuli as they are thought to be unambiguous and extreme instances of FS. But as pointed out by Wray (2012: 240), 'how safe can our assumptions be about what observations based on extreme types tell us about other kinds of examples'?

Indeed, idioms are only a subtype of idiomatic formulaic sequences and do not represent the whole range of them. Moreover, they tend to be not very frequent.

The use of idioms as stimuli in many experiments is particularly problematic when the group tested are L2 learners. Indeed, it is very likely that some of the idiomatic sequences used as experience stimuli were simply not known by the L2 learners. For example, in the study by Underwood and al. (2004) some of the stimuli chosen included *the straw that broke the camel's back* or *every cloud has a silver lining*, which were likely to be unknown by learners, even advanced ones. This is particularly problematic as the results of these studies would be undermined if a large proportion of the idiomatic sequences under scrutiny was indeed unknown by the L2 learners. Indeed, there will be no processing advantage if the idiom is not well known. It is even likely that there will be a processing disadvantage as the meaning of many idioms is not easily retrievable due to their lack of semantic transparency.

As a result, one can wonder whether a study design involving the processing of idioms is adequate in a second language acquisition context because of the likely gaps in the lexicon of L2 learners. In fact, Tabossi, Fanari and Wolf (2009) recently showed that rather than its idiomaticity or the transparency of its meaning, knowing an idiomatic expression is what determines the speed at which it is processed. Wray (2012: 242) also points out that knowing what is in an L2 learner's personal inventory of formulaic expressions is surely likely to be a crucial determinant of how they respond to stimuli, although she rightly remarks that such a requirement is difficult to accommodate adequately in a research design.

Using idioms as stimuli for psycholinguistic experiments therefore has several shortcomings. That is why other studies have focused instead on the processing of common, corpus-derived and mostly transparent idiomatic expressions.

2.1.3.2 *Psycholinguistic studies on idiomatic and corpus-derived FS*

Jiang and Nekrasova (2007) used two on-line grammaticality judgments to examine the effect of idiomaticity on reaction times in native English speakers and L2 learners. They did not use idioms as stimuli but only transparent and very common idiomatic expressions such as *take a look at* or *on the whole.* The responses on idiomatic phrases such as *on the other hand* or *at the same time* were compared with responses on non-formulaic phrases matched for word length and frequency such as *on the other bed* or *at*

the same building. They found shorter reaction times and fewer errors for idiomatic sequences, for both native speakers and L2 learners.

Bod (2001, 2001) compared the processing speed of frequent sentences such as *I like it* with lower-frequency control sentences such as *I keep it*. He found that the frequent sentences were processed more quickly than the less frequent ones. Tremblay and Baayen (2010) investigated the processing of four-word sequences (*in the middle of*) by native speakers and found that higher frequency of occurrence improved the participants' recalls of sequences.

Schmitt et al. (2004) tested the psycholinguistic validity of corpus-derived FS. They compiled a list of idiomatic sequences found frequently in different corpora and used them to create an oral dictation task. The bursts of dictation, in which the formulaic strings were included, were long enough to overload working memory, the authors' hypothesis being that if the formulaic sequences were available for use in the participants' memory, there would be a high likelihood for these sequences to be reproduced as part of the participants' response. An oral-response task was chosen over a written one to overcome the problem that correct reproduction does not mean that a cluster is holistically stored (it could have been produced via syntactic rules and lexical knowledge of the component words). Since the participants had to reproduce the dictation orally, time-pressure was put on them, which was supposed to lead to a preference for the quicker route of retrieval. Moreover, the manner of articulation (fluent or disfluent) could give a good idea of whether the sequence was processed holistically or not. The results showed that, even amongst native speakers, not all the clusters were reproduced in a manner which would suggest that they were holistically stored in the mind. This seems to show that the recurrent clusters are not a homogeneous set within the group of native speakers themselves. The L2 learners' scores only suggested holistic storage for a minority of the target sequences. Indeed, the vast majority of their productions was partially incorrect and / or disfluent. This shows that for them, the strings under scrutiny were not stored as whole units, which supports the observation that non-native speakers have difficulty with the mastery of idiomaticity; the highest level L2 learners, however, mirrored the native speakers' performance closely.

2.1.3.3 *Conclusion on studies on the processing of speaker-external FS*

To conclude this review of studies investigating the psycholinguistic nature of idiomatic and corpus-derived sequences (defined more or less narrowly depending on the studies), and whether or not they present a processing advantage, it can be said that the evidence provided by empirical research is mixed. Most of the evidence suggests that idioms present a processing advantage although this advantage does not seem to be shared by even proficient language learners. We have seen however, that the absence of processing advantage for language learners is hardly surprising given the unfamiliar nature of the material used as stimuli. With regard to studies using idiomatic word clusters or frequent multiword bundles / sentences as stimuli, the results are also mixed. Some point towards faster processing for both native speakers and language learners. It has also been shown, however, that not all corpus-derived clusters are psycholinguistically real and that even native speakers differ in the repertoire of sequences that present a processing advantage for them. Finally, it must be emphasised that although many studies assume that processing advantage implies holistic storage, studies dealing with the nature of the processing of idioms suggest that idioms cannot be regarded as longer lexical units. In other words, even if they present a processing advantage, it does not necessarily follow that they are stored whole in the lexicon and that they do not need to be processed semantically or syntactically.

In view of the above empirical evidence, it is necessary for the sake of methodological soundness to treat speaker-external FS and speaker-internal FS as two distinct constructs. The first one is derived from what is common in a large number of native speakers whereas the second one is internally derived as the set of sequences which present a processing advantage for a given speaker. There is an overlap of course, as the externally derived set is made up of the sum of what is automatised in a large number of speakers. However, although a large number of FS will have both internal and external validity, the set will nonetheless vary from speaker to speaker.

Although the above studies show that not all speaker-external FS present a processing advantage over non-formulaic strings of words, they also demonstrate that some do, at least for some language users. In other words, the notion of processing advantage exists and is an interesting phenomenon, which is worth investigation. Being clear about which approach one is adopting when investigating FS is crucially important, however.

2.1.4 *Importance of the distinction between speaker-external versus speaker-internal approach to formulaicity*

With regard to the last section, one can see why the distinction made by Wray (2008: 11), and introduced in chapter 1, between a speaker-external and a speaker-internal approach to formulaicity is of essential importance. Indeed, given the complex issue of the relationship between idiomatic sequences and the way they are dealt with psycholinguistically, it is crucial to disambiguate the relationship between, on the one hand, multiword sequences either recurrent in corpora or possessing a formal unity and, on the other hand, their psycholinguistic treatment by individual members of the language community.

Wray's clear distinction between speaker-external and speaker-internal is fundamental because it emphasises the fact that it is not because a sequence is a semantic unit or occurs in corpora at a high level of frequency that it will necessarily have a psycholinguistic reality for a particular individual. Conversely, a multiword sequence which presents a processing advantage for a given language user might not be a frequent bundle or in the case of an L2 learner, it could even be a sequence which is grammatically incorrect. This distinction between frequency in corpora or formal characteristics, on the one hand, and individual psycholinguistic processing, on the other, has the crucial correlate that what is formulaic for a given speaker might not be formulaic for another one and that there might be individual differences even amongst native speakers, who have a unique store of formulaic strings based on their own experience and language exposure: a *formulalect* or a *phrasalect* (Schmitt et al. 2004).

Stressing the distinction between the speaker-external and speaker-internal approach to formulaicity does not mean denying that there is a certain degree of overlap between the two approaches. Firstly, the formal characteristics of an idiomatic string can be a strong indication that the string in question is likely to have a psycholinguistic reality. Indeed, a string which is either semantically or syntactically irregular is unlikely to be generated by the grammar, e.g. *by and large*. Moreover, it is undeniable that the stock of idiomatic expressions stored in the memory of a native speaker is likely to resemble the one stored in the memory of another native speaker and many idiomatic strings found in corpora are likely to present a processing advantage for many native speakers. It is indeed reasonable to suppose, for example, that many sequences used in social routines are

used so frequently that they become automatised by all speakers of a given community, e.g. *how do you do* or *nice day today*.

However, it is clear that what is formulaic speaker-externally is not necessarily so speaker-internally and that therefore the overlap between the two should not be assumed without empirical evidence. As underlined by Wray (2012: 239) 'we, for the moment at least, cannot necessarily assume that there is a single phenomenon at the heart of [various researchers'] different activities'. Therefore, the present study strongly supports Wray's call for disambiguating the notion of FS as a linguistic versus psycholinguistic construct as it supports the position that without a clear awareness of the difference between the two constructs, researchers risk ending up 'not talking about precisely the same thing' (Wray 2012: 237) while thinking that they are. In other words, without a clear distinction between the two approaches, researchers might end up making claims about all types of FS when their results only apply to one type of FS and should not be generalised to other types. This is particularly true in the L2 context, where the input learners are exposed to is less rich and more variable, and where the automatisation processes have not necessarily been completed either.

In the light of the above disambiguation, the approach to formulaicity adopted in this study is psycholinguistic and speaker-internal. The next section is dedicated to the psycholinguistic definition of FS.

2.1.5 *Psycholinguistic definitions of FS*

The most widely used psycholinguistic definition of a 'formulaic sequence' (FS) is given by Wray (2002: 9) which defines it as 'a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.' Since then, Wray (2008: 12) has further refined this definition by developing the construct of Morpheme Equivalent Unit (MEU) which she defines as 'a word or word string, whether complete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any subparts it may have'.

Whichever of the two definitions is considered, the claim made by Wray is very strong in the sense that a FS is prefabricated because it is a multiword lexical unit retrieved whole from the lexicon without any grammatical processing taking place. However, it is

important to understand Wray's rationale in defining FS in such a way. Wray (2009: 29) comes back to the two above definitions and makes it clear that they are both 'stipulative', that is to say 'Definitions you start with to explore the fundamental nature of the defined phenomenon'. Wray underlines that these definitions are not meant to be working definitions, i.e. definitions 'that reliably identify examples' (ibid). In other words, they are not operational definitions in the sense that they cannot be directly operationalised in empirical studies. . Their purpose is to enable 'the laying out of a working space' (ibid). This is why, in the 2002 definition, the expression 'appears to be' is important. It draws attention to the fact that some multiword sequences possess some external characteristics which suggest that they are holistic at some internal level. As summarised by Wray (2009: 31-31), Morpheme Equivalence (Wray 2002: 265-9) is in essence the 'proposal that certain wordstrings take on characteristics associated with formulaicity (fluency of production, semantic and/or grammatical oddity, characteristic intonational contours, frequency of occurrence in text) because they have a dedicated entry in the mental lexicon'. Moreover, 'appears to be' acknowledges the absence of certainty and the exploratory purpose of her definition.

Wray is right to point out that her definitions are not operational definitions. Indeed, FS defined as lexical units are extremely difficult to investigate empirically as we have no direct access to speakers' internal linguistic representations. Some psycholinguistic experiments can indirectly tap into the nature of processing but their results might be questionable as these experiments are somewhat artificial as they are not based on natural language use. Secondly, the available evidence points toward the fact that idioms might still undergo semantic and grammatical processing, and it might therefore be wrong to regard them as multiword units stored whole in the lexicon. Moreover, with respect to the claim of holistic storage in the lexicon, both the 2002 and the 2008 definitions seem to contain a contradiction between the claim that there is no 'generation or analysis by the language grammar' or that it is 'processed like a morpheme' and the fact that a sequence can be 'discontinuous' and include 'gaps for inserted variable items'. Indeed, if the sequence is discontinuous, for example if it is a formulaic frame with slots for insertion of variable items, it is difficult to conceive that no grammatical processing is taking place at all.

For the above reasons, I will adopt a 'weaker' psycholinguistic definition of FS than that provided by Wray in the sense that it will focus on the processing advantage of FS

rather than their holistic storage. The definition of FS that will be used in the present study is the following:

A formulaic sequence (FS) is a multiword semantic / functional unit that presents a processing advantage for a given speaker, either because it might be stored whole in their lexicon or because it is highly automatised.

This definition is preferred to Wray's because it only claims processing advantage and not necessarily holistic lexical storage. Although it is not possible to reliably prove holistic storage, it is less methodologically problematic to demonstrate the faster and easier processing of certain sequences of words in relation to others. Moreover, as will be explained later in this chapter, this study adopts a usage-based theoretical framework according to which there is no clear separation between what is lexical and what is grammatical. FS defined as processing rather than lexical units fit that theoretical framework better.

2.1.6 *Psycholinguistic FS and core principles of formulaicity*

As a conclusion to the first section of this chapter, it is interesting to return to the core principles at the heart of formulaicity defined speaker-externally to see to what extent they can also apply to FS defined psycholinguistically.

It can be said that the notion of a strong relationship between the words of an idiomatic sequence also applies to psycholinguistic FS since it is the strength of the link between the different components of the sequence that allows it to be retrieved quickly. This notion will be further developed in the section about the process of chunking.

The idea of restricted creativity and freedom also applies to psycholinguistic FS as the notion of processing advantage might be linked to the fact that speakers may have a preferred sequence to express a given semantic content and do not exploit their creative linguistic power to the full. Because of the weight processing constraints place on spoken language, speakers are likely to rely on a set of sequences that can be recalled with minimal effort. In this respect, it is worth pointing out that the idea of 'preferential' processing can be interpreted in two ways that can potentially overlap. Firstly, a sequence might be preferentially processed because it is the sequence a given speaker tends to choose to express a given semantic sequence. Secondly, it is also preferentially processed because, as a result of being a speaker's preferred sequence, it presents a processing advantage for that speaker.

How does the construct of psycholinguistic FS relate to the notion of formulaicity as a discrete category or as a continuum? If one assumes that formulaic implies being stored holistically in the lexicon, a given sequence is either stored in the lexicon or not. So in this way, the definition of an FS as a unit stored holistically in the lexicon is in keeping with formulaicity envisaged as a discrete category, and it does not work well with the idea of formulaicity as a graded notion. However, if one considers that a formulaic sequence is an automatised processing unit, without making the claim that it is a lexical unit, the notion of formulaicity as a continuum is more relevant.

To sum up, although a conceptual distinction needs to be made between psycholinguistic FS and idiomatic FS, the core principles at the heart of formulaicity defined linguistically, i.e. strong relationship between words, restricted creativity and the issue of discrete category versus continuum also apply to formulaicity tackled from a psycholinguistic perspective.

The next section will focus on the importance of a psycholinguistic approach to FS by providing an account of their relevance both to broad linguistic issues and to language acquisition, particularly to second language acquisition, which is the primary focus of the present study.

2.2 Relevance of the psycholinguistic approach to FS to general linguistic inquiry

The study of FS can take its place in a usage-based theoretical framework such as Beckner et al. (2009) Bybee (2010) and Goldberg (2003).

2.2.1 *Brief overview of the usage-based theoretical framework*

The key idea of the usage-based theoretical framework is that each experience with language has an impact on cognitive representations which are built up through the application of general principles of human cognition to linguistic input (Bybee and McClelland 2005). In other words, the cognitive organisation of language is seen as based directly on one's experience with language. Language is described by a group of researchers such as Bybee, Croft, Ellis and Larsen-Freeman as a 'complex adaptive system' (Beckner et al., introduction to the special edition of *Language Learning* 2009) because speakers' behaviour is based on their past interactions. Consequently, the linguistic knowledge of a language user is not regarded as static as it changes slightly every time they process a new utterance (Ellis 2003). As described by Pierrehumbert (2001), the cognitive representations underlying language use are built up by the

categorisation of utterances into exemplars and exemplar clusters, based on their linguistic form as well as their meaning and the context in which they have been experienced. Because this categorisation is ongoing during language use, even adult grammars are not fixed but have the potential to change as experience changes (Beckner et al. 2009)

This usage-based theoretical framework is often referred to as constructivist as it assumes a construction grammar approach, under which language is conceived of as an inventory of constructions (Goldberg 2003). Constructions are defined as utterance templates which are direct form-meaning pairings that range from the very specific (words or idioms) to the more general (passive construction, ditransitive construction), and from very small units (words) to clause-level or even discourse-level units (Beckner et al. 2009). These constructions of various sizes and levels of abstraction all serve some communicative or socio-pragmatic function (Ambridge and Lieven 2011).

The construction grammar approach sits well with the construct of FS as it recognises the existence of multiword units since, as pointed out by Weinert (2010: 3), in such a model, there is 'no claim or expectation of maximal analycity and minimal representation'. In other words, within such an approach, multiword strings are allowed to have a direct memory representation. Conversely, psycholinguistic FS can be regarded as an ideal construct to investigate several issues of key relevance to the usage-based framework: the dynamism of linguistic representations, the individual differences in speakers' linguistic representations and the intertwinement between grammar and the lexicon.

2.2.2 *FS and the dynamism of linguistic representations*

According to researchers working in the usage-based framework, every linguistic experience affects speakers' abstract linguistic representations (Ellis 2003). The following section aims to describe the modalities of this impact.

2.2.2.1 *Frequency and the process of chunking*

According to Bybee and McClelland (2005), Bybee (2010), Ellis (2002) and many others, frequency of use impacts cognitive representations and one very general principle is that, as sequences of actions are performed repeatedly, they become more fluent and integrated. Indeed repeated sequences such as *in other words, take a break, to top it all, I don't know,* come to be processed as single units rather than morpheme-by-

morpheme. One effect of frequency, then, is that it reduces any tendency toward compositionality within the pattern (Bybee and McClelland 2005). Sequences become highly integrated, which renders them less likely to undergo analysis. This process is referred to as chunking.

Bybee (2010: 7) defines chunking as the cognitive process by which sequences of units that are used together cohere to form more complex units. Newell (1990: 7) defines a chunk as a unit of memory organisation, formed by bringing together a set of already formed elements (which, themselves, may be chunks) in memory and welding them together into a larger unit. As described by Ellis (2003), chunking appears to be a ubiquitous feature of human memory and has been at the core of short-term memory research since Miller (1956) first proposed the term. Although the number of items in short-term memory remains relatively static (estimated at 7 ± 2), the size of the items it handles can be increased through chunking, therefore increasing its information capacity. Patterned phone numbers are common examples of increased short-term memory thanks to the use of chunks as it is easier to remember 10 digits if they are grouped into 2-digit-numbers. In language, repeated sequences of words (or morphemes) are packaged together in cognition so that the sequence can be accessed as a single unit.

The status of a chunk in memory is thought to fall along a continuum; from words that have been experienced together only once and fairly recently, which will constitute a weak chunk whose internal parts are stronger than the whole, to more frequent chunks such as *lend a hand* which are easily accessible as wholes while still maintaining connections to their parts (Bybee 2010). Hay (2001, 2002) proposes that each instance of direct access of the complex unit strengthens that path of access and weakens the access through the component parts, at the same time weakening the relation with these parts and bringing on gradual loss of analysability. In other words, the stronger the sequence becomes as a unit, the less associated it will be to its component parts.

Ellis has applied the notion of chunking to the L2 acquisition process. According to him, associative learning is at the core of language acquisition. Sequences of words that are repeated across learning experiences become better remembered. Indeed ‘we are tuned to linguistic regularities: we process faster and more easily language which accords with the expectations that have come from our unconscious analysis of the serial probabilities in our lifelong history of input’ (Ellis 2002). For Ellis (2003), chunking is the mechanism of learning which underpins the acquisition and perception of formulaic

sequences. Chunking is a particularly powerful process in language as it operates at both concrete and abstract levels. Indeed, not only co-occurrences of tokens, but also co-occurrences of types, are registered by memory (Ellis 2003). In this second language acquisition context, MacWhinney (2008) adds another dimension to the notion of chunking as he sees it as not only a method for creating new long term memory units but also as a method for integrating phrases on line. He points out that in addition to retrieving constructions as chunks, learners must work out methods that produce new constructional chunks on the fly in real time. In other words, they develop fluency by repeated practice in making combinations. In this respect, chunks can be seen as production strategies in second language acquisition (Raupach 1984).

Chunking is not the only effect that can be observed with high frequency: phonological reduction can also occur. Bybee and McClelland (2005) point out that reductions are especially apparent in the process of grammaticalisation (e.g. the phonetic reduction of *going to gonna* as future marker) and in the development of discourse markers such as *I don't know* which undergo extreme reduction (*dunno*) when they come to serve as markers of interaction. Psycholinguistic FS can be interesting in this respect as they are potential loci for such phonetic reductions.

The very construct of FS defined psycholinguistically is the result of this chunking process. FS are therefore highly relevant within a usage-based theoretical framework as their investigation allows for the study of the 'close relationship between language use and language representation' (Weinert 2010: 1).

2.2.2.2 *FS and individual lexical differences*

As mentioned earlier, our cognitive representations of language are constantly affected by our experience with language. Because the input that speakers are exposed to is variable, the results will be representations that vary from one speaker to another.

Beckner et al. (2009) underline the idea that language exists both in individuals (as idiolect) and in the community of users (as communal language). They consider that language is emergent at these two distinctive but interdependent levels: an idiolect is emergent from an individual's language use through social interactions with other individuals in the communal language, whereas a communal language is emergent as the result of the interaction of the idiolects. Sociolinguistic studies have revealed the large degree of heterogeneity among idiolects (Weinreich, Labov and Herzog, 1968),

not only in language use but also in their internal organisation and representation (Dabrowska 1997)

As was mentioned earlier, this study's approach, by virtue of being psycholinguistic is also learner-internal. In other words, what is of interest is what is formulaic for a given learner, which is different from what is formulaic for another learner (although their repertoires of FS might share some common features) and also distinct from what is formulaic in the 'communal language'. As a result, the study of FS can be a lens through which individual differences in learners' idiolects can be investigated, i.e. through the investigation of learners' individual 'formulalects' (Schmitt et al. 2004).

2.2.2.3 *Language as a pastiche*

Another, very different, consequence of the idea that language representations are based on our linguistic experience is the notion that utterances are very often closely similar to previous utterances. The notion of similarity between utterances is one of the staple ideas of the theory of emergent grammar (Hopper 1998)**.** According to it, we say things that have been said before as 'real life discourse abounds in all sorts of repetitions' (Ibid: page number). This notion of language as pastiche is particularly interesting in the context of the study of formulaicity since, as Weinert (2010) points out, the ordinary, non-scientific definition of formulaic is 'lack of originality', the fact that something has been said or heard before.

2.2.2.4 *Nature of the relationship between lexicon and grammar*

As previously described, in the usage-based framework, cognitive representations are conceived of as abstractions resulting from one's cumulative experience with language (Bybee 2008). However, there is evidence that knowledge of specific instances of language use is not entirely lost in the abstraction process and especially with reinforcement through repetition, specific instances of constructions can have rich memory representations, called exemplars, which register details about linguistic experience. Consequently, it is not just idioms which, due to their unpredictable meaning, have direct representations. Many expressions which are regular in form and meaning such as *dark night* are also registered in memory because they are known to speakers as expressions they have experienced before (Bybee 2010: 28). In other words, once a category is formed or a generalisation is made, the speaker does not necessarily have to throw away the examples upon which the generalisation is based (Beckner et al.

2009). As a result, even completely predictable word combinations can be registered in memory.

According to Bybee (2008, 2010), this highly redundant, experience-based cognitive storage and access system for language leads to a conception of lexicon and grammar as highly intertwined rather than separate as it implies a continuum between lexis and grammar. In brief, language is seen as a highly redundant, experience-based cognitive storage and access system which contains words as well as phrases and sentences. Such a theory of redundant linguistic representations is in stark contrast with the generativist proposals of an abstract grammatical system with a redundancy-free lexicon (Chomsky, 1957, 1965).

Thinking of grammar and the lexicon as intertwined is not the only way of accounting for the existence of direct representations for multiword units. Other researchers (Peters 1983, Pawley and Syder 1983, Wray 2002, 2008) operationalise the direct representation of multiword units by arguing for a heteromorphic lexicon in which multiword items can be stored holistically. This model of the lexicon allows for the same lexical material to be stored in bundles of different sizes. Pawley and Syder (1983: 192) state that '[i]nsofar as many regular morpheme sequences are known both holistically (as lexicalised units) and analytically (as products of syntactic rules) it is necessary to specify these sequences at least twice in the grammar'. They make it clear that this duplication applies to lexicalised sentence stems of regular formation, as well as to idiomatic sequences other than those of irregular structure, e.g. *Will you marry me?* They argue that the duplication of a sequence of regular formation such as *Will you marry me?* is necessary in order to account for the fact that such a sequence has a dual status in the language. On the one hand, its potential occurrence and meaning is predicted by the productive rules of syntax and semantics. On the other hand, the 'dictionary entry' for the same sequence should note its status as a lexical item, because it's the one entry corresponding to a specific situation. In the same way, Wray (2008) argues for a heteromorphic lexicon, flexible enough to allow for the storage of both atomic items and whole phrases. In this model, the same linguistic material can be differently processed at different times, which permits multiple part-mappings of the same information in the lexicon. In this way, alongside an entire phrase that is stored whole with its associated meaning and phonological form, subparts may also be stored. This conception of the lexicon as heteromorphic does not follow the traditional compartementalisation of grammar into syntax (productive rules) vs. dictionary (fixed,

arbitrary usages). Indeed, according to such a lexical model, even regular multimorphemic sequences that could be produced according to the rules of grammar can have a separate entry in the lexicon.

In effect, both the usage-based conception and Wray's heteromorphic lexicon are similar in the sense that they allow multiword units with a grammatical structure to have mental representations. They just conceptualise this possibility differently, the first by postulating the absence of boundaries between syntax and the lexicon, the second by allowing phrases with a grammatical structure to be stored whole in the lexicon. As a result, the main difference, as pointed out by Weinert (2010), is that, in Wray's conception, the lexicon is kept separate from syntax, despite the fact that multiword units can be stored in it. According to Weinert, the advantage of the usage-based conception over Wray's heteromorphic lexicon is that it allows for a unified account of cognition, i.e. the existence of constructions with various levels of abstraction rather than a dual-mechanism account, i.e. the coexistence of productive grammatical rules and fixed lexical items.

2.3 Relevance of psycholinguistic FS to first and second language acquisition

As described in the preceding section, the construct of FS can be used to investigate key linguistic issues such as the effect of frequency of usage on speakers' mental representations, the individual differences in speakers' linguistic representations, the repetitive nature of language as well as the nature of the relationship between syntax and the lexicon. The construct of FS is also of paramount importance for the understanding of both L1 and L2 acquisition, in which FS have been shown to be entries into communication, production but also learning strategies (Yorio 1989, Weinert 1995).

2.3.1 *FS in L1 acquisition*

The role of FS in L1 acquisition has been studied extensively and there is a consensus that they constitute an important part of child language: 'That children do store and use complex strings before mastering their internal make-up is generally agreed' (Wray 2002: 105). As made clear by Wray's quote, in the context of L1 acquisition, FS need to be defined as unanalysed multiword units. Therefore, and as will be further developed in chapter 5, in the context of L1 acquisition, speaker-internal FS are conceptualised as unanalysed units, which makes their status as psycholinguistic units less controversial than in the context of adult native speakers.

As summarised by Wray (2002), FS are a set of starter utterances which give, at least to some children, entry into adult-like social interactions. In this respect, research on the early stages of L1 acquisition has focused on the difference in cognitive styles between children and on the resulting individual differences with regard to FS use. In her longitudinal study of 18 children, Nelson (1973, 1975) made a well-known distinction between referential and expressive children. Referential children tend to use language primarily to refer to things whereas expressive children mainly use language for interaction purposes. These two cognitive styles have consequences on the children's use of FS as the referential style seems to go hand in hand with an analytic approach to language whereas the expressive style favours the use of longer strings with an associated communicative function.

Some researchers such as Bates, Bretherton and Snyder (1988) have tended to characterise expressive children as slow language learners. Bates et al.'s negative view stems from the fact that their study suggests that acquisition of FS is a dead-end on the road to productive language use. Many researchers question the view of FS as a less advanced communication mode however. On the opposite, they see them as learning strategies and claim that their use by expressive children is evidence of a different rather than slower path than that used by referential children. Clark (1974) studied data from her son Adam and suggested that, in some cases, FS acted as primary data, allowing the child to carry out a gradual analysis of their internal structure, so that they eventually became productive. In the same way, for Peters (1983), it is possible to observe how FS are broken down by children and feed into their productive system.

Lieven and Pine (1992) and especially Pine and Lieven (1993), will be reported in detail as they are seminal studies which shaped today's research agenda as well as current methodologies.They revisited Nelson's distinction between referential and expressive children as, according to them, this dichotomy only captured differences in surface features of the children's utterances but said little about their underlying structure. That is why they adopted a coding scheme aiming to differentiate between unanalysed multiwords utterances and productive ones. They proceeded by treating multiword utterances as single units unless there was evidence that they represented instances of broader productive patterns. This method is now known as the trace-back method and has been used extensively since then. Their results suggest that the best defining feature of 'expressivity' is not the proportion of personal-social words in children's

vocabularies, but rather the proportion of frozen phrases acquired by the child in the first and second 50 words.

Moreover, Pine and Lieven (1993: 554) put forward the idea that the relationship between referential and expressive styles (Nelson 1975) may be best explained not so much in terms of 'noun preference' versus preference for social interaction as in terms of the direction from which different children break into structure, with some children constructing patterns by combining two or more items from their single-word vocabularies and others developing patterns by gaining productive control over slots in previously unanalysed phrases. This is why they emphasise that the distinction between analytical and holistic children must be dealt with carefully because according to them, there is analysis in both cases. The difference is only in the point from which analysis begins and the direction in which it progresses. Pine and Lieven warn that the distinction between analytical and holistic children has sometimes encouraged researchers to think in terms of different underlying systems but they stress that it makes more sense to view variation as the product of the same system operating in different ways or upon different kinds of input data.

Additionally, Pine and Lieven (1993) question the distinction between expressive and referential children on the ground that it has to be thought of more as a quantitative difference rather than a qualitative one. According to them, breaking down originally unanalysed phrases is a strategy used by all children, though to varying degrees. Indeed, they show that relatively few of the children's patterns appear to have been built up from two items already present in their single-word vocabularies. Instead, 66% of the grammatical patterns seem to have their roots in unanalysed phrases and even the least phrasal child had 4 out of 10 patterns which seemed to have developed in this way.

Finally and crucially, Pine and Lieven's study shows that, for all the children, rather than being a strategy restricted to the very earliest stages of language development, the acquisition of unanalysed phrases actually increases in importance as vocabulary development progresses. Indeed, the proportion of frozen phrases increased significantly between 50 and 100 words. They point out that this kind of process has also been reported in much older children (for example Clark's son was almost 3 years old), which suggests that it may continue to be useful until quite late in the language acquisition process.

Despite their influential nature, the above studies were still limited in terms of the quantity of data they were based on. Since then, as pointed out by Bannard and Lieven (2012), research in L1 acquisition has been characterised by a massive increase in the size of the data sets available for analysis. These very large samples of children's interactions with their caregivers have shown that children repeatedly encounter a great number of multiword units (Cameron-Faulkner, Lieven, and Tomasello 2003) and researchers working in this framework have argued that 'children have dedicated representations for word sequences that they frequently encounter' and that 'these sequences form the basis of their developing productive grammars' (Bannard and Lieven 2012: 4). The rationale behind this argument is that, for a child with an immature cognitive system, having a direct memory representation for frequently required sequences supports the efficient processing and production of language.

The traceback method has been further refined over the years and applied on very large data samples to confirm the above hypotheses. For example, Lieven et al. (2009: 313) examined the speech of four two-year-old children and recorded 30 hours of data for each child over several weeks. The traceback was performed on the final two hours of recordings for each child and the preceding 28 hours of recordings were searched for related utterances. Lieven et al. found that 25-40% of the children's target utterances were exact repetitions of strings already produced in the main corpus and a further 36-48% could be derived by just one operation (i.e. inserting one new item into a prefabricated frame).

To conclude on FS use in children, a growing body of literature suggests that FS are not just entries into basic communicative functions but are a core component of L1 acquisition. Far from being a peripheral phenomenon, they form the basis of the children's developing grammar. Moreover, researchers such as Bannard and Lieven (2012: 14) suggest that 'there is good reason to believe that these sequences that are the starting point of the system are not completely driven out, and that the most heavily entrenched sequences persist in the [children's language]'.

2.3.2 *FS in naturalistic child L2 acquisition*

In similar ways to L1 acquisition, there is a large body of evidence showing that FS are prominent in the early stages of child L2 naturalistic acquisition, (Itoh and Hatch 1978; Karniol 1990; Wong-Fillmore 1976) and that they are used extensively both as

communication and learning strategies. Wong-Fillmore (1976) is the most extensive study of child L2 acquisition in a naturalistic setting.

Wong-Fillmore adopted Jespersen (1924)'s distinction between formulas (fixed expressions in which neither stress, rhythm nor words can be altered and which are acquired whole) and productive speech in her analysis of the development of English as a second language in young Spanish-speaking children. However, she broadened the definition of a formula as 'language which functions wholly or partly as unanalysed, fixed or automatic units for the speaker (1976: 295). Thus her definition allowed for the notion of a formulaic frame, a formula which has been partly analysed, so that there is some substitutability in a grammatical slot within the formulaic construction.

She studied 5 Spanish speaking Mexican immigrant children over a nine-month period as they acquired English at kindergarten and school. One of the children, Nora, was later described by Wong-Fillmore as a 'spectacular language learner (1979: 221). Her remarkable success was linked to her use of FS and the way they fed into her productive rules. Wong-Fillmore showed how Nora used specific FS such as *I wanna play wi' dese* and progressively moved from them to more general patterns such as '*I wanna* + VP'. Similarly to L1 acquisition, the longitudinal study of FS in naturalistic child L2 acquisition can therefore give us an insight into the complex process of construction of the linguistic system.

Despite the important difference in contexts of acquisition, FS have also been shown to play a crucial role in the early stages of instructed L2 acquisition.

2.3.3 *FS in instructed L2 acquisition*

2.3.3.1 *FS in the early stages of instructed L2 acquisition*

FS occupy a prominent position in the UK teaching context, at least as far as secondary schools are concerned. This is shown clearly by Mitchell and Martin (1997)'s longitudinal study of French teaching and learning in two secondary schools in southern England. Their study documented the development in French of 60 11 to 13-year-old pupils over a two-year period and showed that, despite the end of the audio-lingual method, prefabricated phrases have maintained a significant place in contemporary classroom practice, often under the newer guise of 'exponents' for communicative functions: 'Our learners were explicitly taught a curriculum consisting very largely of unanalysed phrases. Typically these took the form of face-to-face questions and answers

exchanges, clustered around topics such as "the family", "hobbies", or "likes and dislikes", and for much of the time they were memorized and rehearsed unaltered' (1997: 23). Mitchell and Martin point out that despite the important amount of time devoted to 'communicative' activities, in practice, these activities primarily offered occasions for further recycling of learned chunks, and rarely put pupils under real pressure to analyse and restructure them for new communicative ends.

The teachers under scrutiny in this study were all well aware of the key role of the rote learning of FS in early classroom learning as is well illustrated by the following comment by one of the teachers: 'So I've just given them *je suis allé* (I went), we've already learned *au cinéma* (to the cinema), *à la piscine* (at the swimming pool), so now we've added a different chunk *je suis allé*, and we have put the two together…so yes we do deal a lot in chunks'. All the teachers viewed the memorisation and re-use of unanalysed chunks as a central part of effective classroom practice at this level and, according to them, it was the most basic means of developing pupils' knowledge of French, for the first 2 or 3 years at least. They saw this kind of learning as within the capability of (almost) all learners as clearly put by one of the teachers: 'I think that's filling the gap for children who don't understand grammar, so *je suis allé* becomes a chunk rather than a grammatical point'.

A crucial question raised by Mitchell and Martin's study is how the learners 'move on from regurgitation of learned, fixed phrases to a more creative and flexible control of the target language' (1997: 23). According to the teachers observed in the study, grammar explanation and conscious understanding played a key part in the eventual development of syntactic control over the chunks. However, in the researchers' opinion, the prime driving force behind the learners' unpacking of unanalysed FS was communicative need, e.g. the necessity, triggered by certain activities, to extend reference beyond the first and second person reference typically embodied in the questions and answer formulae.

Myles, Hooper and Mitchell (1998) directly investigated whether FS played a role in the development of the learners' emerging grammatical competence by tracking the development of several verbal FS: *j'aime* (I like)*, j'adore* (I love)*, j'habite* (I live). Their data clearly showed that the use of formulas facilitates entry into communication and speeds up production in the early stages of instructed acquisition. In fact, learners

could not initially rely on much else in order to hold the kind of 'conversations' required by the classroom context.

Moreover, and more importantly, they found that FS played a role in the development of the learners' grammatical competence and could in no way be dismissed as a peripheral phenomenon playing no part in this creative process. According to them, there was a continuum in the pupils observed, from those who never managed to break down the chunks to those who succeeded in doing so, a process they found to be clearly linked to the emergence of the subject pronoun system. Indeed, according to the researchers, as third person formulas entered the learners' repertoire, the segmentation process itself began. The pressure of communicative needs beyond the well-practiced classroom routines triggered the breakdown process. However, far from dropping the chunks from their interlanguage at this stage, the learners actively 'worked on' them, and fed them directly into the creative construction process. The learners' first step was to keep the chunk intact but add a lexical noun phrase to it in order to make reference clearer, tagging on a correct reference to the unchanged formula as clearly shown by examples such as *Richard j'aime le musée* ('Richard I love the museum' with the intended meaning 'Richard loves the museum'). Then chunks were further broken down and in the case of some learners, the appearance of the third-person pronoun could be seen like in examples such as *Euh j'ai adore...oh no Monique j'ai adore..no Monique elle adore la....regarder la télévision*. ('Erm I have love ...oh no Monique I have love, no Monique she loves the....watch television').

Myles, Mitchell and Hooper (1999) looked at the question / answer formulaic sequence *Comment t'appelles-tu? Je m'appelle* ('How + reflexive pronoun+ call you?' / 'I reflexive pronoun call' = 'What's your name?'/ 'My name is'). This particular sequence was chosen as it is taught in all classrooms. Moreover, since it is particularly complex structurally (it has wh-fronting, inversion as well as a reflexive pronoun), it is a clear example of unanalysed chunk as it is not expected to be produced creatively in the early stages of learning French. Like for the declarative verbal chunks of the previous study, the subjects were found to progress along a common general route: from the inappropriate and over-extended use of the chunk, through several progressive steps of breaking down to the correct third person form.

According to Myles et al., this development path made evident the interaction between the two chunks *Comment t'appelles-tu?* and *il s'appelle*. The interaction between these

two is a first indication that learners do not simply discard chunks but compare them. They are clearly instrumental in the learning process: as a linguistic database to work on and which is used as a springboard for creative construction. Consequently, similarly to what happens in L1 acquisition, the acquisition of FS is not a dead-end to L2 language learning. On the contrary, in Myles et al. (1998, 1999), it is clear that the learners who were able to memorise FS successfully and who were still working on them by the end of the study were also the learners who were earliest to engage in creative construction and who progressed farthest along the developmental continuum during the course of the 2-year study. Conversely, the learners who did not memorise FS stayed stuck in a pre-grammatical verbless stage.

2.3.3.2 *FS in more advanced stages of L2 acquisition*

We know very little about the role of FS, defined psycholinguistically, in advanced L2 learners. This is because, contrary to the research focusing on beginner learners, most of the research focusing on advanced learners investigates the acquisition of idiomatic sequences rather than psycholinguistic FS, and defines formulaicity in a learner-external way.

For a long time, there was virtually no work on the acquisition of idiomatic strings by L2 learners since research focused mainly on the acquisition of syntax. Punctual remarks on the use of idiomatic sequences by L2 learners can only be found here and there in studies dealing with other aspects of linguistic development (Schumann, 1978; Shapira 1978). However, the realisation, thanks to corpus linguistics, of the prevalence of idiomatic strings in the language of native speakers led to an interest in research into the L2 acquisition of idiomaticity (Bolander, 1989; Yorio, 1989; Biskup, 1992; Farghal and Obiedat 95; Foster, 2001). Studies seem to show that idiomaticity is problematic for L2 learners, even at advanced levels. For example, Forsberg (2009) investigated differences in the distribution of speaker-external formulaic sequences used by advanced, very-advanced learners and native speakers of French. She adopted a typology of formulaic strings based on Erman and Warren (2000), distinguishing between lexical, grammatical and discourse formulaic sequences. She found a significant difference between the advanced learners and the other two groups. Indeed, advanced learners overused discourse formulaic sequences and did not use as many lexical formulaic sequences as the two other groups.

However, one must be careful when drawing implications from the finding that that native and non-native speakers (even advanced learners) differ significantly in their use of idiomatic sequences. Firstly, Forsberg (2009) showed that with abundant native input, very advanced learners who have spent a long time living in France can end up resembling native speakers closely, in both the number and types of formulaic sequences they use. Moreover, the studies mentioned above define L2 formulaicity in a learner-external way and focus on idiomaticity. As a result, they only tell us about idiomatic strings and not about L2 learners' FS meant as psycholinguistic wholes.

However, the fact that L2 learners do not use many idiomatic strings does not necessarily imply that they do not have a repertoire of psycholinguistic FS. For example, Raupach (1984)[1] showed that L2 learners used various seemingly holistically recalled units as speech production strategies to perform various functions (fillers, speech organisers). The psycholinguistic FS used by L2 learners, however, might not necessarily be as idiomatic as those used by native speakers. Therefore adopting a psycholinguistic learner-internal approach to formulaicity is essential in the context of L2 acquisition, because it allows for the consideration of sequences that could be overlooked by studies focusing solely on idiomaticity.

According to Wray (2002, 2008), the relevance of FS in L2 acquisition is likely to be limited because the L2 teaching context encourages a lot of grammatical analysis and an atomic approach to language, which is prone to the breaking down of multiword units. However this assumption can be questioned on several fronts. Firstly, as previously shown, FS are omnipresent in the early stages of instructed L2 acquisition. In a usage-based framework of language acquisition, it is plausible to assume that initially unanalysed sequences that were learnt holistically such as *comment t'appelles-tu?* remain as a memory representation even after they have been analysed by the grammar. Moreover, if one assumes the possibility of chunking (Ellis 2002) or in Peters (1983)'s terms 'fusion', some L2 FS could be phrases that were initially generated but have become fused through frequency of use. Indeed the L2 learning context is prone to practice and repetition and as a result, the high frequency of use of certain sequences such as *je pense que* (I think that) might lead to their becoming a single processing unit. For all the above reasons, the presence of psycholinguistic FS in advanced learners should not be discarded and needs to be further characterised and understood.

[1] This study will be reviewed in detail in chapter 5.

Consequently, the following research questions must be asked about the use of psycholinguistic FS in advanced learners.

- How far do advanced learners of French use FS?
- Does the type of task used have an effect on FS use?
- What types of FS do they use?
- Are there individual differences between subjects in terms of quantitative and qualitative FS use?
- What are the implications of FS use in L2 learners for our understanding of language and the status of FS within it?

Moreover, it must be noted that advanced learners are more likely than beginner learners to be exposed to native input either through various media or by spending a year abroad. Indeed, the study abroad context is likely to provide them with more naturalistic input than an instruction setting at home (Foster 2009, see 4.3.3). In such a propitious learning context, they might acquire more vocabulary, including new FS. Indeed, FS can be regarded as a subset of vocabulary as they are automatised sequences containing at least two lexical items. In other words, FS can be regarded as hybrid constructs in the sense that, on the one hand, they are made of lexical items and on the other hand, these lexical items need to be automatised as a unit in order to be formulaic. In this respect, FS are based both on the learners' competence and performance as they are composed of lexical items but also depend on procedural knowledge. Because of their lexical dimension, their development needs to be examined in relation to that of general lexical diversity in order to investigate whether it can be placed within a general trend of acquisition of new vocabulary.

Finally, if advanced L2 learners use sequences which reduce the cognitive processing load, the effect of these sequences on their speech must be examined in order to further our understanding of the development of their fluency. In the existing literature, the development of L2 fluency is usually explained by proceduralisation of existing grammatical knowledge (Towell et al. 1996). However, the development of fluency could also be considered and explained from a memory-based lexical perspective (Kormos 2006). In this respect, investigating the effect of the use of FS on the development of fluency could be enlightening. However, as will be shown in chapter 4, studies focusing on this issue are few and far between. Before reviewing these studies, it is necessary to examine, in chapter 3, how the role of FS can be conceptualised at the

cognitive level of underlying processing mechanisms, by showing which stages of speech production can be facilitated by the use of FS.

Chapter 3. The impact of FS on cognitive fluency: integrating FS in models of speech production

As mentioned in chapter 2, one of the roles FS might play in the context of L2 acquisition is to contribute to our understanding of the development of fluent speech in advanced learners. In order to understand this potential role, this chapter aims to analyse how it can be conceptualised at the cognitive level of underlying processing mechanisms, by showing which stages of speech production can be facilitated by the use of FS.

The first section of this chapter is dedicated to the definition of fluency, with a particular focus on Segalowitz (2010)'s distinction between fluency at the cognitive level (the speaker's ability to efficiently mobilise and integrate the underlying cognitive processes responsible for producing utterances) and fluency at the level of utterances (i.e. the temporal variables of speech). The second section is dedicated to a concise description of stages of speech production (Levelt 1989, Levelt 1999) in order to understand, in the third section, how FS can be integrated into a model of speech production and at which level of speech production they can have a facilitative effect. Finally, the last section shows how various theories of automatisation can be adapted to FS in order to account for their acquisition and their contribution to the development of fluency.

3.1 Defining fluency: broad versus narrow approach

Fluency is a very common term. It is used frequently in the realm of language teaching as well as in everyday life. However, many researchers agree on the difficulty of defining fluency, emphasising that it is a 'complex phenomenon that encompasses a multitude of linguistic, psycholinguistic and sociolinguistic features' (Freed et al. 2004: 279). Following Lennon (1990), one can say that there are two main ways of understanding the construct of fluency. Firstly, Lennon distinguishes a broad meaning of the term which corresponds to overall linguistic proficiency. For example, when people say that a person is very fluent, they mean that the person can use the language effectively usually in reference to that person speaking a foreign language. Secondly, fluency can be approached more narrowly usually in the context of foreign language teaching and research. In this sense, it only corresponds to one of the components of general proficiency, namely oral fluidity. This narrow interpretation of the construct of fluency is what Segalowitz (2010) calls 'utterance fluency' since it refers to the actual

physical properties of the utterance. It is this narrow definition of fluency i.e. oral fluidity which is considered in the present study.

3.1.1 *Tension between utterance fluency and other aspects of proficiency*

When discussing L2 proficiency, Lennon (1990) makes an important distinction between skill and knowledge pointing out that although utterance fluency is one of the components of general linguistic proficiency, it is different in type from the other components. Indeed, while components of linguistic proficiency such as lexical range, syntactic complexity and use of idioms can be assigned to linguistic competence, fluency belongs to the realm of performance. As Lennon (1990: 391) puts it, 'there is presumably no fluency "store"'. However, it is very difficult to entirely separate fluency from the other components of linguistic proficiency, since as pointed out by Chambers (1997), the level of knowledge in any of the linguistic areas affects fluency. For example, it makes no sense to conceive of a fluent speaker with absolutely no knowledge of vocabulary.

This tension between utterance fluency and other aspects of proficiency is of essential importance for this study. Indeed, one of its aims is to examine whether the knowledge of FS (i.e. one aspect of general linguistic proficiency) has an effect on utterance fluency. The construct of FS is particularly interesting in the context of the relationship and interaction between competence and performance. Indeed the construct of a FS encompasses both competence and performance. In order for a FS to be used by a given subject it must be part of their lexical and / or grammatical competence but the issue of the automaticity of FS retrieval belongs to the realm of performance.

3.1.2 *Utterance fluency and cognitive fluency*

The study of the temporal variables of fluency is important because it can provide a window into cognitive fluency. Indeed the study of utterance fluency enables researchers to gather valuable empirical evidence on psycholinguistic mechanisms at stake in speech production since processes of language production themselves are not directly accessible to observation (Chambers 1997). The cognitive focus of several fluency researchers is noticeable in their definitions of fluency. For example, according to Rehbein (1987: 104) 'fluency means that the activities of planning and uttering can be executed nearly simultaneously by the speaker of the language'. In the same way, Lennon (2000: 26) points out that 'a working definition of fluency might be the rapid,

smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing'. In other words, both these definitions focus on the link between the underlying cognitive system ('planning', 'thought or communicative intention') and the external characteristics of speech. Segalowitz (2010) calls this focus on the working of underlying cognitive mechanisms 'cognitive fluency' which he defines as the speaker's ability to efficiently mobilise and integrate the underlying cognitive processes responsible for producing utterances. According to him, it can be said that the cognitive system exhibits fluency in how efficiently and fluidly it is able to carry out this mobilisation and integration of processes. Similarly, Lennon defines (1990: 391) fluency as 'an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently'. It is interesting to notice that this single definition encompasses the three constructs distinguished by Segalowitz (2010) namely perceived fluency ('an impression on the listener's part'), cognitive fluency ('psycholinguistic processes of speech planning') and utterance fluency ('speech production').

However, despite the cognitive focus of the above definitions and their underlying assumption that cognitive fluency underpins utterance fluency, which is its visible manifestation, very little research has been carried out in order to explore the links between utterance fluency and cognitive fluency. One of the aims of the present study is to make a contribution in this area by focusing on the relationship between the use of FS at the cognitive level and the external characteristics of speech at the level of utterance fluency. In other words, this research project seeks to investigate how FS use contributes to more effective cognitive fluency by observing the effect of FS use on utterance fluency. In order to understand how FS may contribute to the speeding up of speech production, it is necessary, in the first place, to give a brief account of how speech production works.

3.2 Model of L1 speech production and Levelt's blueprint of the speaker (1989, 1999)

Although the present study focuses on L2 acquisition, the most comprehensive work on models of speech production has been carried out on native speaker use. In the literature, there is a consensus that language production has four important components:

- Conceptualisation i.e. the planning of what one wants to say

- Formulation, which includes the grammatical, lexical and phonological encoding of the message
- Articulation, in other words, the production of speech sounds
- Self-monitoring, which involves checking the correctness and appropriateness of the produced output.

It is agreed that, for L1 speech production at least, planning the message requires attention, whereas formulation and articulation are automatic, which makes speech generally smooth and fast.

Although there exist other theories of L1 speech production such as Spreading Activation (Dell 1986), for the purpose of the present study, I will solely present Levelt's model of speech production, which was first developed in 1989 and updated in 1999. This preference for Levelt's model can be accounted for by several reasons. Firstly, Levelt's model is the most widely used theoretical framework in L1 and L2 production research. It has been used, for example, by studies dealing with the development of fluency in L2 learners of French such as Towell et al. (1996) or Towell and Dewaele (2005). Secondly, Levelt's model has been tested more thoroughly than other models such as Dell's and therefore has a much firmer empirical basis (Kormos 2006). Moreover, compared to the other existing theories of speech production, Levelt's model is more detailed at every level of the speech production process. This level of detail is crucial for the present study. Indeed, as mentioned in the introduction of this chapter, one of the goals of this study is to understand how the construct of FS can be integrated into a model of speech production. Therefore Levelt's model allows for conceptualising in detail how FS can have a facilitative effect on speech production at each stage of the process. Finally, as will be explained below, contrary to Dell's model which is a frame-slot model in which generative rules build a frame with slots to be filled in by insertion rules, Levelt's model is lexically-driven, which means that words activate syntactic building procedures. It is therefore more adapted to the focus of the present study since FS can be regarded as a special type of lexical items.

3.2.1 *Processing components and knowledge stores*

As illustrated by Figure 1, the 1989 model is made of different processing components (represented by boxes) and several knowledge stores (represented by circles and ellipses). A distinction is made between declarative knowledge (such as conceptual and

lexical knowledge) contained in knowledge stores and procedural knowledge for the processing components. In the 1999 version of the model, there are three knowledge stores. The first one contains the speaker's knowledge of the external and internal world. It comprises the discourse model, which is 'a speaker's record of what he believes to be shared knowledge about the content of the discourse as it evolved (1989: 114), the model of the address i.e. knowledge about the present context of interaction and the ongoing discourse, and encyclopaedic knowledge i.e. information about the world. The second store is the mental lexicon which is the repository of lexical entries which are composed of lemmas (that contain syntactic information) and morpho-phonological codes (that carry information about the morpho-phonological form of the lexical entry). Finally, the last knowledge store is the syllabary which contains gestural scores that is chunks of automatised movements used to produce the syllables of a given language.

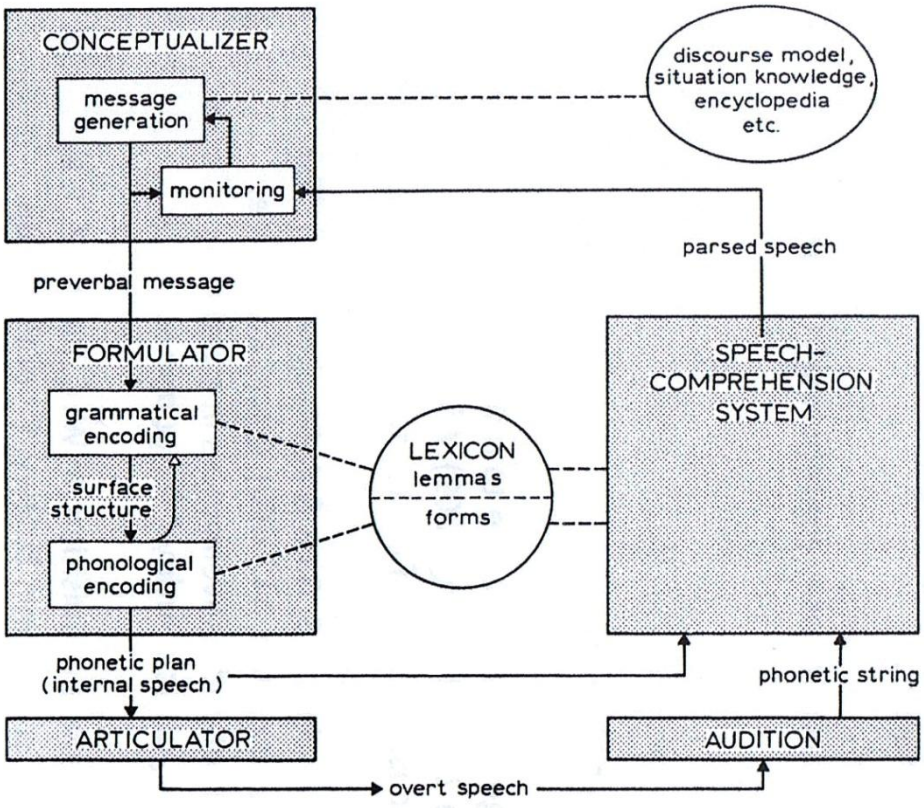


Figure 1: Levelt's model of speech production

In Levelt's model, speech production is conceived of as modular that is to say that the processing components work autonomously from each other. In other words, these processing components are specialists in the particular functions that they have to execute (1989: 14). Another very important characteristic of the model is that it is incremental, which means that the next processing component in the general flow of

information can start working on the still incomplete output of the current processor. This implies that a processing component can be triggered into action by a fragment of its characteristic input. This incremental working is necessary in order to account for the high speed of speech production.

As shown by Figure 1, the processing components are the following: (1) the conceptualizer in which conceptual planning takes place, (2) the formulator where grammatical and morphophonological encoding happen and where the preverbal message is converted into a speech plan, (3) the articulator which converts the speech plan into actual speech.

3.2.2 *Phases of speech production*

- **Conceptual preparation**

Despite the few differences between the two versions of Levelt's model, phases of speech production are conceptualised in the same way in both**.** Let us take a more precise look at how they are described in the 1999 version. The first phase consists of conceptual preparation in which the message is generated through two core processes: macroplanning and microplanning. Macroplanning deals with the elaboration of the communicative intention. Once the communicative intention has been decided on, microplanning can take place. Roughly speaking, in microplanning, speakers decide on the perspective they need to take in conveying the message. Decisions on what constitutes old and new information are made. In addition, the message is given propositional content that is to say the argument structure of the message is specified: thematic roles are assigned, referents are specified. The outcome of this conceptual preparation is called the preverbal plan. In the 1989 version of the model, both macroplanning and microplanning are conceived of as language independent. In the 1999 version of the model however, although the preverbal plan is not linguistic, it is thought of as linguistically accessible, that is to say that it contains all the necessary information to convert meaning into language. In other words, the terminal elements of the preverbal plan must be lexical concepts. According to Levelt (1999), conceptual preparation is not language-independent because languages differ in their range of lexical concepts and in the conceptual information that is obligatorily expressed. For example, in some languages like French and English, it is obligatory to mark tense, contrary to languages such as Chinese. However, it must be pointed out that despite Levelt's arguments, no consensus has yet been reached on whether conceptualisation is or is not language-independent (Kormos 2006).

The preverbal plan is the input used by the next processing phase i.e. grammatical encoding.

- **Grammatical and morpho-phonological encoding**

In the process of grammatical encoding, information from the speaker's mental lexicon is retrieved. The mental lexicon consists of lexical entries which are made up of lemmas, which contain syntactic information about the lexical entry, and lexemes, which carry information about the morpho-phonological codes of the lexical entry.

The first step of grammatical encoding is lemma selection that is to say that the lemma whose meaning best matches the semantic information carried by the preverbal plan is retrieved. In this respect, grammatical encoding can be said to be a lexically driven process because Levelt assumes that when a lemma is selected, its syntactic properties become available for further syntactic construction. The output of grammatical encoding is the surface structure.

The surface structure is further processed in the course of morpho-phonological encoding. As soon as a lemma is selected, the speaker gets access to its morphological and phonological composition. It is during this process that the word is syllabified in its syntactic context. The final result of phonological encoding is the phonological score of the utterance with its syllabified words and intonation patterns.

- **Phonetic encoding and articulation**

In the course of phonetic encoding, each of the syllables in the phonological score triggers an articulatory gesture. It is at this stage that speakers resort to their syllabary. The outcome of phonetic encoding is the articulatory score. Finally, the execution of the articulatory score by the laryngeal and supra-laryngeal apparatus produces the end product of the whole process that is overt speech.

- **Monitoring / Self-perception**

When we speak, we monitor our own output, which involves the same comprehension system that we use to listen to others. Thanks to the process of self-monitoring, we can stop and correct ourselves.

An over-detailed description of all the processes involved in speech production is not the aim of the present study. Therefore some of the components of the model especially those dealing with articulation and monitoring have been mentioned only briefly. However, an overview of the different phases of speech production is necessary in order to frame the discussion on how FS can be integrated into the processes of speech production.

The aim of this chapter is to examine the effect that FS have on speech production and the claim that FS present a processing advantage because they allow the bypassing of encoding processes. In order to examine in detail how this might happen, a detailed

description of the structure of the lexicon is needed, as well as the presentation of the precise working of formulation procedures.

3.2.3 *Structure of the lexicon in Levelt's model*

In Levelt's model, the lexicon has a very important role because it is an essential mediator between conceptualisation on the one hand and grammatical and phonological encoding on the other. Levelt (1989: 182) defines a speaker's mental lexicon as 'a repository of declarative knowledge about the words of his language'. In the 1989 model, lexical knowledge is composed of four kinds of features: semantic, syntactic, morphological and phonological.

In the 1999 version of the model, conceptual and semantic information is stored at a different level from where grammatical information is stored (lemma level) and where morpho-phonological forms are stored (lexeme level). Conceptual knowledge and semantic knowledge are seen as interdependent and stored at an integrated level. Concepts are viewed as undivided wholes which are activated in their entirety. They can be either lexical (when they can be expressed by a single word) or non-lexical (when they have to be encoded by multiple words). Lexical access involves the activation of concepts, which further spread activation to lemmas, which contain syntactic information about the lexical entry but no information on meaning. A lemma's syntactic information specifies the item's syntactic category, its assignment of grammatical functions, and a set of diacritic feature variables. The example chosen by Levelt to illustrate this point is the lemma *give*. It is a verb, which works with a subject, an object and an indirect object. Moreover, a lemma relates to specific morpho-phonological form information: it can be said to 'point' to a form entry. This entry contains several word forms (or lexemes). For example, for the lemma *give*, there are various word forms such as *gives*, *gave* or *given*. These forms can be distinguished only by assigning value to several features (tense, mood, aspect, person, number) of the lemma.

3.2.4 *Grammatical encoding*

Since in Levelt's model, grammatical encoding is lexically driven, it implies that the encoding operations are largely controlled by the grammatical properties of the lemmas retrieved. Once the lemmas have been retrieved, dedicated syntactic procedures set up the appropriate grammatical frame. Levelt compares grammatical encoding to a process of solving simultaneous equations because the emerging syntactic structure should

simultaneously satisfy all the constraints required by the various lemmas. The output of grammatical encoding is the surface structure. Levelt (1989) points out that the surface structure is the type of representation that forms the hinge between grammatical encoding and phonological encoding. So in other words, surface structures are the input to phonological encoding in the course of which word forms are accessed and prosodic patterns are generated. The lowest-level or terminal nodes in a surface structure are lemmas with their lexical pointers indexed for diacritic features. A lemma's lexical pointer indicates a form entry where the corresponding word-form information is stored. There can be diacritic features for each lexical pointer. In other words, the pointer to a form entry may be indexed with various features that will affect the word form retrieved. That is why the surface structure should indicate case, number, person, tense, aspect, definiteness and whatever other features are to be morphologically or phonologically marked on the word form pointed to. However, inflections themselves are not yet specified. The pointers are only abstract instructions for inflectional procedures to be run in the second phase of encoding i.e. morpho-phonological encoding.

The purpose of morpho-phonological encoding is to build a phonetic plan for each lemma and for the utterance as whole. In modular models such as Levelt's, morpho-phonological segments have their own abstract representations, which are stored as a group of features in memory. During the phase of phonological encoding, word forms are accessed and prosodic patterns are generated. The outcome of phonological encoding is the phonetic plan, which is not overt speech yet but which is the internal representation of how the planned utterance should be articulated.

3.3 Models of L2 speech production

Existing models of L2 speech production are heavily based on Levelt's model. Indeed, researchers who have adapted Levelt's model to L2 production such as De Bot (1992) and Kormos (2006) have endeavoured to modify it as little as possible. To account for his choice, De Bot emphasises the firm empirical basis of Levelt's model, which is based on several decades of psycholinguistic research and a wealth of empirical data through experimental research and the observation of speech errors. Moreover, he points out that, since L1 and L2 speech production processes bear more common points than differences, a single model to describe both monolingual and bilingual speakers is a better solution than two separate models. Similarly, Kormos' model is not significantly different from Levelt's model constructed for monolingual speakers. Kormos justifies her stance by quoting Abutalebi et al. (2001) whose meta-analysis of

existing research suggests that neither the extent of brain activation nor the regions involved in processing in L1 and L2 are different for bilinguals who learned the L2 early in their lives and for highly proficient speakers with extensive L2 exposure. According to her, this evidence supports a common model of speech production for monolingual and bilingual speakers. So, in the same way as Levelt's model, Kormos's model is modular, that it is to say it is composed of separate processing components: the conceptualizer, the formulator and the articulator. Moreover, as Levelt's model, her model is incremental as encoding procedures in a module can be triggered by only a fragment of the characteristic input of this module. The incremental nature of the model entails that for L2 learners above a certain level of proficiency, parallel processing is possible.

Consequently, as stages of speech production are conceptualised in the same way in models of L2 speech production, this study also adopts Levelt's model in order to conceptualise the effect that FS use has on fluency.

Another reason for not developing a dedicated model of L2 speech production is that it would involve tackling a series of highly complex issues surrounding the question of bilingual lexical encoding which are beyond the scope of this thesis. Indeed, as pointed out by Kormos (2006), although much research is devoted to the bilingual lexicon, there is still a lot of disagreement about its structure, the information it contains, and how access to lexical items is conceptualised. As summarised by Kormos (2006), there is now a considerable amount of evidence that the conceptual system sends activation to both L1 and L2 lexical items. The selection of the right language is assumed to be achieved by the existence of a language cue. However, researchers disagree on whether this language cue situates itself at the conceptual level (La Heij 2005) or at the level of lemma representation (Poulisse and Bongaerts). Moreover, there is a considerable amount of disagreement amongst researchers in the way they conceptualise the organisation of the bilingual lexicon. For researchers such as Paradis (2000) and Pavlenko (1999), semantic representations are stored in the lexicon but not conceptual ones because for them, concepts are independent of word meanings. For other researchers such as Levelt et al. (1999) however, conceptual knowledge and semantic knowledge are interdependent and should be represented at the same level. Concepts can be lexical, i.e. expressed by one word, or nonlexical, which means that they can only be encoded by multiple words. However, for them, this semantic/ conceptual level is represented outside the lexicon, which only contains lemmas and lexemes and no

semantic information. For Kormos, the lexicon contains everything: concepts, semantic information and word forms. These different ways of conceptualising the lexicon and the organisation of concepts, word meanings and word forms respectively have in turn consequences for the way researchers conceptualise lexical access in the L2 and the extent to which they see it as a process mediated by the L1.

A thorough description of these various and conflicting theories of the bilingual lexicon and lexical access would be necessary to account for the status of FS in the bilingual lexicon as well as the way they are accessed. Such a description would be interesting in order to conceptualise cross-linguistic influence for FS and precisely understand at which level(s) i.e. conceptual/semantic/lemma this influence takes place. However, although relevant and on the agenda for future research, such a complex description is beyond the scope of the present study, which does not deal directly with the issue of cross-linguistic influence with respect to FS but with their role in stages of speech production. This is why, the next section will conceptualise the integration of FS within Levelt's model of speech production.

3.4 Integrating FS into models of speech production

3.4.1 *Levelt's position on idioms*

Levelt (1989: 186-7, 1999: 95) briefly acknowledges the presence of phrases and idioms in the mental lexicon of speakers. He assumes that they might be stored in the lexicon in the same way as single words. According to him, idiomatic collocations are entries in the mental lexicon and each entry consists of one or more items. For example, the idiom entry for *kick the bucket* has one lexical entry which contains two items: the infinitive form and the past tense form. According to him, idioms, like words, have their characteristic conceptual conditions. If such a condition is met in the message, the idiom will be accessed. As a result, it is possible for certain concepts to map directly onto phrases: 'Some lexical concepts or rather "idiom concepts" map onto idioms of one kind or another. Idioms are encoded by going from a single concept to a complex idiom lemma with its own syntactic properties' (Levelt 1999: 95). For example, some idioms do not allow passivisation (Jackendoff 1997).

Although Levelt, mentions that the amount of idiom and collocation in the mental lexicon is probably of the same order of magnitude as the number of words, he does not deal with this question any further. Since for him, they work in the same way as single lexical items, they do not require any special treatment. Levelt's relative disinterest for

the question of multiword sequences is representative of the fact that, as underlined by Kormos (2006: pXX), existing models in the psycholinguistic literature focus on speech production as a creative process, in the course of which utterances are constructed word by word using rules of syntax and phonology. As a result, these models mainly deal with creative language processing and do not consider FS and the role these could play in the processes involved in speech production. It is in order to bridge this gap in the literature that Kormos makes an attempt to integrate formulaic language into models of speech production and the development of fluency. She discusses how FS are stored and retrieved and how theories of automaticity can account for learning formulas.

3.4.2 *Kormos' integration of FS in theories of speech production*

According to Kormos (2006), whatever model of speech production we consider, the only place where formulaic knowledge can be stored is the lexicon. Her conception of the lexicon can therefore be equated to Wray's heteromorphic lexicon (in which the same lexical material can be stored in bundles of different sizes). As a result, whatever theories of speech production assume about how words are accessed also applies to FS. As far as retrieval is concerned, she assumes that FS are retrieved like other lexical items in the mental lexicon. Kormos first adopts a psycholinguistic definition of FS as sequences of words or phrases retrieved from memory as one unit (2006: XX). She mentions (2006: 45) that these holistically-retrieved sequences can be of different types such as idioms, multiword phrases and collocations. According to her, these sequences are stored whole in the lexicon and then retrieved as single units. Accounting for their holistic retrieval is therefore not problematic since they are retrieved in the same way as normal words.

She points out, however, that it is more difficult to account for the retrieval of longer sequences such as whole sentences which express a pragmatic function such as apologising. Her solution to this problem is to assume that, given the fact that in any model of speech production concepts activate lexical items, chunking or the creation of larger units takes place at the conceptual level. Many functions can be conceptualised as one unit and these conceptual units send activation to pre-assembled lexical items. Kormos adopts La Heij's (2005) view of complex selection and simple access, and hypothesises that most pragmatic functions are probably conceptualised as one unit and include specifications concerning the level of formality, style and so forth and that these conceptual units send activation to pre-assembled lexical units. This can account for the

fact that a native speaker will retrieve the phrase ‘I regret to tell you’ as one unit from the lexicon rather than individually accessing the words that constitute the phrase. Kormos is very brief about the role FS play in the syntactic encoding of the message. She states that as regards the syntax of FS, in modular models, formulas can also point to various types of syntactic information just as other lexical items can, and this information is used in syntactic encoding.

Kormos’ contribution is essential because she is one of the first researchers to acknowledge the gap between research on FS and research on speech production. However, as described above, her work is based on a heteromorphic conception of the lexicon. But if chunking is to be thought of as a process rather than a result (McWhinney 2008), it is not necessary to postulate the holistic storage of FS in the lexicon to investigate how they contribute to the speeding up of speech production. The difference between conceptualising FS as lexical units forming part of a heteromorphic lexicon, or as units which have been automatised through repeated activation, is that the effect they have on the processes of speech production can be conceptualised either as the entire bypassing of some phases of speech production (lexical units) or as their increased speed and efficiency (automatic units).

If one argues that FS present a processing advantage in the sense that they allow encoding or some part of it to be either bypassed or speeded up, then one needs to describe precisely which encoding phase is bypassed or speeded up and how this takes place. However, it can be argued that there might be various ways in which processing is shortcut depending on what type of FS is involved. In Kormos’ model, how FS might affect the process of speech production is not very detailed. Indeed, her account of how encoding is bypassed is very brief and she says hardly anything on the role of FS in syntactic encoding. She only mentions that FS point to syntactic information in the same way as other lexical items. This might be due to the fact that her definition of formulaic sequences is rather vague and imprecise. Although she points out that there are different types of FS such as idioms, multiword phrases and collocations, she does not deal with the fact that they might affect speech production in different ways. As a result, it is difficult to clearly understand from her account what precise role FS might play in speech production.

3.4.3 *FS and the bypassing or speeding up of phases of speech production*

If one integrates the construct of FS into a model of speech production, one needs to understand how FS use facilitates fluent speech. Facilitation of speech production can be interpreted either as bypassing (if one postulates the possibility for multiword sequences to be stored whole in the lexicon) or as facilitation (if one postulates the putting together of some sequences is facilitated by the fact there are memory traces for them). To understand how this processing advantage takes place, one needs to consider different types of FS at different levels of abstraction since different cases of bypassing can be envisaged depending on the type of FS.

- Speeding up grammatical encoding

The speeding-up of grammatical encoding can be envisaged in the case of formulaic frames which are partly-fixed and are composed of some fixed parts and some slots that have to be filled. Since the syntactic frame is already either entirely or partially built, some of the syntactic procedures do not need to take place. Let us take the example of the frame (subject) *be* (tense) *sorry to keep* (tense) *someone waiting*. In this frame, the grammatical structure is already built. As a result, the only operations that need to be performed are choosing the tense and person for the first verb *be*, choosing between the present and past form of the infinitive *keep* and to fill in the *someone* slot.

- Bypassing grammatical and morpho-phonological encoding entirely and speeding up conceptual preparation

It can also be envisaged that some FS allow for the bypassing of not only the phase of grammatical encoding but also that of morpho-phonological encoding. This global bypassing of the whole of the encoding phase could be allowed by FS which are either automatised or lexically stored whole in a fixed form. For example, it can be envisaged that the whole interrogative sentence *How are you?* has a direct memory representation and can be recalled without needing to be encoded. To go back to the example (subject) *be* (tense) *sorry to keep* (tense) *someone waiting*, it is very plausible that a doctor will have automatised one precise realisation of this formulaic frame (such as *I'm sorry to keep you waiting*!) that he will produce automatically each time he is late and welcomes a patient. Such a fully automatic retrieval and full bypass of encoding processes can be hypothesised in the case of the most common pragmatic functions. As described above by Kormos, in the case of such sequences, it can be assumed that, since concepts in any

model of speech production activate lexical items, chunking or the creation of larger units takes place at the conceptual level. Kormos's hypothesis seems viable as it seems highly plausible that many pragmatic functions can be conceptualised as one unit. If the bypassing of encoding processes is accounted for in such a way, though, it consequently means that the conceptual phase of speech production is also speeded up as the content of the message is extremely common and predictable. As a result, formulaicity situates itself both at the conceptual and linguistic level as the multiword linguistic chunk corresponds to a conceptual chunk. This is interesting in the context of the notion of formulaicity conceived of as 'lack of originality' (chapter 2): in the case of common pragmatic functions, the prefabricated nature of the linguistic units can be regarded as a direct consequence of the common nature of the message at the conceptual level.

It makes sense to assume that this entire bypassing of encoding only happens for the most frequently used morpho-phonological forms. In this respect, the same verb could be stored lexically for certain persons, most likely first person singular (since it is used very frequently) but not for other persons.

- Speech facilitation by FS in the L2 context

In the L2 context, it is interesting to see how the use of FS can solve potential problems with which L2 learners might be confronted in the phases of both grammatical and morphophonological encoding. For example, at the level of lemma selection, if one considers the use of a formulaic grammatical frame such as *be sorry to keep someone waiting*, the learner does not have to hesitate about which lemma to retrieve for the verb 'keep' because of the prefabricated nature of the frame. If the learner does not know such a frame, then it is likely that, trying to express the same semantic content, they might hesitate between several verbs, which would slow down production.

At the level of morpho-phonological encoding and lexeme selection, the automatisation of multiword sequences can suppress problems potentially faced by the learner at that level of encoding. For example, if a learner manages to automatise an entire FS such as *j'aime le football* (I like football), it avoids problems both at the lemma and at the lexeme level. At the lemma level, all the lexical pointers are already indexed for diacritic features so the learner does not need to select the correct features (first person singular form of the pronoun and the verb, definite article, masculine gender). As a result, the learner does not have to wonder whether they should use the definite article or not or whether the noun *football* is feminine or masculine. At the lexeme level,

encoding problems might also be avoided thanks to the use of FS. Indeed, it is possible that the learner might not encounter problems at the level of the surface structure (i.e. is able index the lexical pointers with the right features (such as masculine form of the definite article) but is not able to retrieve the appropriate form for it at the level of morpho-phonological encoding. However, if a FS is automatised, all the potential problems with retrieving the correct lexeme are also avoided and production can run smoothly. With respect to morpho-phonological encoding, it is plausible to assume that some forms of the same verbal lexeme are much more likely to present a processing advantage than others because they are called for much more frequently by the use of FS. For example, it makes sense that for the pronominal verbal lemma '*s'appeler'*, the form je m'appelle can be retrieved automatically because of the frequent use of the FS '*je m'appelle* + name' whereas *nous nous appelons* might be retrieved with difficulty.

Integrating FS into the phases of speech production shows how they can facilitate fluent speech. FS therefore need to take their place in an account of the development of speech production in L2 learners.

3.5 FS and the development of fluency in speech production

Except in the case of very advanced L2 learners, an important difference between L1 and L2 production is the slower nature of L2 speech production. There are of course individual differences in speed of speech production and some speakers are slower than others in both their mother tongue and second language (Towell 2002). However, on a general basis, several studies have shown that speech rate and the mean length of runs are lower in L2 than in L1, even in the case of advanced university learners (Towell, Hawkins and Bazergui 1996). One of the most important reasons why L2 speech is slower than L1 speech is the degree of automaticity with which L1 and L2 speech is produced. Indeed, whereas lexical, syntactic, morphological and phonological encoding is mostly automatic in L1 production, these mechanisms are only partially automatic in L2 production, even in the case of advanced learners. Due to this lower level of automaticity, processes of L2 production do not run in parallel as efficiently as in L1, which slows speech down. However, some advanced learners have been shown to be able to reach high levels of fluency (Towell et al. 1996). Therefore, the question of what role automaticity plays in L2 speech production and how it develops (automatisation) in the course of L2 learning is of great significance to SLA research. However, as pointed out by De Bot (1992) and Towell et al. (1996), Levelt's model is a steady-state model, not a language-learning model. In other words, it deals only with language production

by mature native speakers and has nothing to say about where this knowledge comes from, how it develops, or how it is given procedural form. This is why Towell et al. point out that, in the context of L2 acquisition, Levelt's model needs to be complemented by a theory of automatisation (i.e. the development of automaticity).

The preceding section has described how FS can speed up production. The acquisition of FS can therefore be thought of as one of the phenomena responsible for the development of fluent speech production. Consequently, one needs to account for the development of FS and to find a theory of automatisation able to explain it. In spite of its importance, according to Kormos (2006), the question of automaticity in L2 speech production has long been neglected. She points out the existence of a gap in research, with on the one hand, research on automaticity and on the other hand, research on models of speech production but virtually no research trying to link the two fields. Before taking a look at how some researchers have attempted to bridge this gap, automaticity will be defined and different theories aiming to account for its development will be presented.

3.5.1 *Definitions of automaticity*

As explained by Segalowitz (2003), automaticity has been discussed in different ways in the research literature and different researchers associate different characteristics with it. The characteristic most frequently associated with it is high speed of processing. However, Segalowitz points out that not all fast processing is necessarily automatic and it is too simplistic to equate automatic with fast. Indeed, for many researchers, automaticity does not refer only to a quantitative change in performance. It also entails a qualitative change, resulting from restructuring of the underlying mechanisms involved in carrying out the performance. In this respect, the fact that there is a qualitative change and it is not only a simple speed-up process is shown by the evolution of learners' performance (Segalowitz and Segalowitz 1993; Segalowitz, Segalowitz and Wood 1998). At the beginning of the practice session, learners' performance is characterised by high variability and low efficiency, then processing becomes more efficient and less prone to errors and variability of performance decreases, which the authors interpret as a sign that a qualitative change has taken place in tasks components. However, Segalowitz (2003) still wonders whether the automatic / non automatic distinction should be viewed as a strictly dichotomous qualitative difference or as end points of a continuum stretching from automatic to non-automatic.

Automaticity is also associated with load-independent and effortless processing (Newell 1990, Schmidt 1992). The construct of psycholinguistic FS relates well to these characteristics. Firstly, with regard to automatic units as load-independent, the section about chunking showed that chunking allows for the processing of much longer sequences. FS can then been seen as load-independent as although they might be composed of several items, they are psycholinguistically light and considerable reduce the processing burden. Finally, the idea of effortless processing relates well to the definition of FS adopted in this study i.e. a multiword semantic / functional unit which is processed more easily than other strings of words.

3.5.2 *Theories of automatisation*

Kormos distinguishes between two main groups of theories of automatisation: rule-based and item-based approaches. The first one sees automatisation as the conversion of declarative knowledge into procedural rules which become automatised with practice, whereas the second one views automatisation as the single-step access of a memorised item becoming faster and more efficient.

3.5.2.1 *Rule-based approach*

One of the most widely used theories within the rule-based approach is Anderson's (1983) ACT (adaptive control of thought) and ACT-R (1995). Anderson's work is built on the assumption that all knowledge is initially declarative, that is to say consciously held, explicit and which requires attention. It assumes that skill acquisition involves a transition from declarative knowledge to procedural knowledge (knowledge evident in the person's behaviour but which the person does not need to be consciously aware of e.g. L1 grammatical knowledge). In this model, automaticity describes an end point in the acquisition of a skill.

For behaviour which requires rapid performance, such as the production of speech, conversion into procedural knowledge is essential for two reasons. First, working memory has a limited capacity and cannot deal with a too great quantity of information. Since procedural knowledge does not necessarily require the attention of the speaker, it can be processed by working memory in larger units without exhausting its capacity. Moreover, declarative and procedural knowledge are stored and accessed in different ways. Declarative knowledge is retrieved by interpretive mechanisms. As a result, though flexible, it is slow. On the opposite procedural knowledge of language consists

of units known as productions and takes the form of condition/action pairs which are very fast to access.

Anderson argues that there are three stages in the conversion from declarative into procedural knowledge. The first stage is the cognitive stage in which declarative knowledge is accessed slowly through interpretive mechanisms. The second stage is the associative stage in which behaviour depends partly on declarative and partly on procedural knowledge. The final stage is the autonomous stage which is fully procedural. In this final autonomous stage, knowledge is accessed very rapidly but cannot be easily modified by the speaker.

According to Anderson's theory, five learning mechanisms contribute to the development of automatic performance: composition, proceduralisation, generalisation, discrimination and strengthening. More precisely, composition or the creation of macro-productions is a chunking process, that is, a psychological process of transforming items into larger units in order to help making it manageable for working memory (see chapter 2). Proceduralisation is the removal of declarative knowledge. Once a production has become automatic, the initial declarative knowledge underlying it is often not retrievable anymore. A very important aspect of this theory is that automatisation entails not only a quantitative change but also a qualitative change in the nature of processing. This qualitative change can come about with the help of three tuning processes: generalisation, discrimination and strengthening. Generalisation widens the scope of the application of the declarative knowledge necessary for production and ensures that production rules are applied in all the contexts where they are appropriate. Discrimination narrows the scope of application of rules so that they are applied only in the appropriate contexts. Finally, strengthening allows for the weakening of poorer rules and the strengthening of better rules: it increases the likelihood with which a particular production procedure is selected.

To conclude, crucial to rule-based approaches is the view that automatisation comes from the transformation of factual knowledge into production rules. By contrast, item-based approaches do not resort to rules to explain the automatisation of knowledge.

3.5.3 *Item-based approach*

Instance theory developed by Logan (1988) exemplifies the item-based approach to automatisation and addresses the issue that not all learning involves the conversion from

declarative to procedural learning. For Logan, automatic processing equals memory retrieval. In other words, the use of an algorithm is substituted by a single-step retrieval of the solution from memory. With practice, associations between problems and the memory traces of their solutions become stronger, and consequently retrieval speeds up. According to this theory, there is a competition between rule-based processing and memory-retrieval: the speed of the two different processes determines which one will be applied.

Other theories also attribute the development of automaticity to memory. According to MacKay's (1982) Strength Theory, connections between response and stimulus become stronger with practice. Strength Theory is compatible with a connectionist approach to language acquisition. Indeed, according to the connectionist theoretical framework, language learning takes place through gradual strengthening of the associations between co-occurring elements of language. In this view, learning involves observing the distributional characteristics of the language input, and extracting from them rule-like, but not rule-governed, regularities.

3.5.4 *The automatisation of FS*

The theories of automatisation briefly outlined above were developed to account for the acquisition of complex skills in general. How can they be applied to language and more specifically to the acquisition of FS?

According to Peters (1983), Wray (2002) and Kormos (2006), the automatisation of FS can be envisaged in two different ways: a top-down way or a bottom-up way. Firstly, from a top-down perspective, FS can be conceived as sequences memorised as a whole, that is to say acquired holistically in the first place. For example, a learner could be taught sequences such as 'nice to meet you' or 'how do you do' as holistic units and told to use these sequences when they meet someone for the first time. These FS acquired holistically from the start are best accounted for by theories of automatisation such as Instance Theory and Strength Theory, which stress the importance of memory and lexical retrieval. With regard to Instance theory, FS acquired in such a top-down way can be conceptualised as being retrieved by memory in the first place. Strength theory can account for the progressively stronger memory link between the different components of the sequence. At the beginning of the acquisition process, links between words that form a particular phrase are weak, therefore learners create variable phrases. In the course of learning, connections between words are strengthened and the FS can

be retrieved with increased efficiency. Strength theory can also account for the stronger and stronger link between the conceptual stimulus (e.g. say one's name) and the retrieval of the multiword FS (e.g. *je m'appelle* + name).

The automatisation of FS can also be conceived of in a bottom-up way, that is to say as phrases or clauses that were first assembled with the help of syntactic and phonological rules but have become automatic with practice and frequency of use. This is a process that Peters calls fusion and that is also described as chunking. The bottom-up automatisation of FS can be accounted for by Anderson's theory if one equates it with the fact that the encoding processes become proceduralised for given multiword sequences.

In terms of instance theory, the bottom-up acquisition of FS can be seen as a competition between, on the one hand, the encoding procedures that assemble larger linguistic units with the application of syntactic rules and, on the other, the retrieval of memorised units. With experience and practice, the speed of memory retrieval exceeds that of rule-based processing, and formulaic expressions are accessed in memory as one unit.

Kormos refers to Oppenheim (2000) as supportive evidence for both strength theory and chunking theory as described by Anderson. Oppenheim (2000) investigated the speech of 6 non-native speakers who were required to give the same speech twice. The results showed that the participants tended to use partially overlapping sequences. They often varied the sequences by adding new elements or combining them into larger units. She claimed that strength theory and theories of chunking can explain how the participants applied the recurrent phrases in a variable manner. Indeed strength theory can account for the variability of sequences because of the different levels of connections that can exist between words. The fact that students create larger units from smaller ones can be explained by chunking theory since it predicts that lexical items that often occur together tend to form chunks and when the conceptual specifications call on them, they are retrieved as one unit.

The difference between Anderson's theory and instance theory is that the first one puts the emphasis on the greater efficiency of encoding processes whereas the second one emphasises the conversion into lexical retrieval. In effect though, it would be inaccurate to oppose the two theories by saying that Anderson's theory only postulates a quantitative change whereas for instance theory the change is qualitative. Indeed, as

explained above, for Anderson, proceduralisation also entails a qualitative change. In my opinion, the difference between the two approaches is better summarised in terms of how they conceptualise chunking. Instance theory puts the emphasis on the end result that is to say on the psycholinguistic unit as a chunk; whereas Anderson's theory emphasises the process of chunking itself, in the same way as MacWhinney (2008). In other words, the qualitative change is less in the psycholinguistic nature of the unit than in the nature of the process itself. One might prefer one theory over another depending on one's view of FS. If one conceives of FS as multiword units stored whole in the lexicon (Wray 2002, Kormos 2006) Instance theory seems more suited, whereas a theory of proceduralisation such as Anderson's is more adapted to a view of FS as units presenting a processing advantage.

To sum up, this chapter dealt with cognitive fluency, that is to say with the internal mechanisms underlying utterance fluency (Segalowitz 2010). A brief description of Levelt's model of speech production was provided in order to understand the various levels of processing which might be affected by FS. Particular attention was dedicated to the formulation stage in order to investigate how FS might facilitate different encoding mechanisms. Finally, in order to account for the development of fluency in L2 learners, and the role that FS might play within this development, various theories of automatisation were presented. The only way to try and further the understanding of how the use of FS might affect cognitive fluency, however, is through the examination of the relationship between FS use and utterance fluency, that is, the temporal variables underpinning fluency, which is the object of the next chapter.

Chapter 4. Fluency: measures, development and relationship to formulaic sequences

As explained in chapter 3, in order to understand how the use of FS might affect cognitive fluency (i.e. the mechanisms underlying speech production), one needs to examine the relationship between FS use and utterance fluency i.e. the temporal variables of speech. Utterance fluency and its relationship to FS is therefore the main focus of this chapter. The first part is dedicated to the measurement of utterance fluency. It presents measures which have been used in previous studies and deals with the complex issue of the treatment of pauses and other hesitation phenomena interrupting a fluent run. The notion of fluent run is of central importance to this study not only for the measurement of the learners' fluency but also because, as will be shown in chapter 5 which deals with the identification of FS, fluent pronunciation (i.e. uninterrupted by a pause or any other hesitation phenomenon) is the most important criterion used for identifying FS in this study. In the second part of this chapter, drawing on Towell et al. (1996), various fluency measures are interpreted in order to show how utterance fluency can provide an insight into cognitive fluency. After reviewing studies dealing specifically with the longitudinal development of fluency as well as the effect of the study-abroad context on fluency and lexical diversity, the last part of this chapter focuses on the few exploratory studies that have investigated the link between FS use and fluency in L2 learners.

4.1 Measuring utterance fluency

4.1.1 *Fluency measures*

Empirical research on fluency generally involves the elicitation of a speech corpus through various tasks and the analysis of its temporal aspects. Even within the framework of the narrow definition of fluency (see chapter 3), there is no consensus on the best way to apprehend it and the most important elements to emphasise in its definition. Various measures have been used in the literature to try and measure the fluency of an utterance. The most common ones are speech rate, articulation rate, phonation-time ratio and mean length of runs.

Speaking rate or speech rate (SR) can be defined as the total number of syllables produced in a given speech sample divided by the amount of time required to produce the sample (including pause time). It is therefore usually expressed in number of syllables per second. Empirical studies have shown that L2 learners' speech rate

generally increases over time (Freed 1995) and that it tends to correlate with judges' perceptions of fluency: i.e. the higher it is, the more fluent the speaker appears to be to the listener (Riggenbach 1991).

Articulation rate (AR) is not to be confounded with speech rate since it refers to the total number of syllables produced in a given speech sample divided by the amount of time required to produce the sample but excluding pause time. Like speaking rate, it is expressed in number of syllables per second and is logically higher than speaking rate since pause time is not included.

Phonation-time ratio (PTR) is the percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample. It is obtained by multiplying the total speaking time (excluding pauses) by 100 and then dividing it by the total task time (i.e. speaking time + pause time).

Mean length of runs (MLR) is the average number of syllables produced in fluent runs i.e. utterances pronounced between pauses of a certain length, generally between 0.2 and 0.3 second (see further down debate on how to deal with pauses and on which cut-off point to choose).

Other measures dealing with pausing time, such as the average number of pauses per minute or mean length of pauses are also used in the literature, though less frequently.

4.1.2 *The study of pauses and other types of hesitation phenomena*

If one reviews all the fluency measures detailed in the list above, it becomes quickly apparent that they depend heavily on how pauses are identified and measured. In other words, different ways of dealing with pauses can dramatically affect the results obtained since all the measures are dependent on the value of pause time. This is why Wood (2010) points out that pause phenomena are the most informative elements of the study of fluency. They are also very complex however, as will be shown in this section.

The treatment of pauses in terms of both identification and measurement in the present study is of paramount importance. Firstly, as has just been mentioned, the fluency measures carried out to investigate the learners' fluency development depend directly on the way pauses are dealt with. However, the treatment of pauses is also essential for a more important reason: in the present study, for a multiword sequence to be considered as potentially formulaic, it must be pronounced fluently i.e. not to be

interrupted by any pause or other hesitation phenomena. In the literature on fluency, there is no consensus on what exactly constitutes a pause and different studies have taken different methodological approaches in their identification of pauses. When dealing with the identification of pauses, a first distinction must be made between silent (unfilled) and filled pauses.

4.1.2.1 *Silent Pauses*

Physically, a silent pause is an interruption of the normal speech flow, meaning that there is no acoustic signal. It is an observable segment, since it shows no significant amplitude in the visual representation of speech. Silent pauses may be the result of physical restraints, such as inspiration, swallowing, any laryngo-phonatory reflex, or a silent expiration (Zellner 1994: 44), and/or may occur as the result of psycholinguistic or cognitive processes.

Now that technological advances have made pause measurement much more precise and easier than when stop watches were used, there is a considerable amount of debate about the cut-off point to be used when studying silent pauses. In other words, what is the minimum length of a pause in order to be regarded as a pause? According to Goldman Eisler (1968), the auditory threshold is around 0.20 to 0.25 seconds. This suggests that pauses shorter than this threshold can hardly be perceived and are, therefore, likely to have no meaning for fluency perception. A lower cut-off point of 0.1 second has been put forward by Griffiths (1991). However, as highlighted by Towell et al. (1996: 91), if the cut-off point is too low, the analyst may be confused by displays in which an apparent pause is in fact the stop phase of geminated plosives or other normal phenomena. On the other hand, if the cut-off point is too high, one runs the risk of omitting significant amounts of pause time.

With regard to the cut-off point issue, one has to consider not only the minimum cut-off point but also the maximum one. There can be occurrences of a total breakdown in speech where the speaker has entirely lost the thread of their message and are at a loss with a given utterance. Such fluency breakdowns are not representative of a speaker's typical speech and might have to be discarded. Moreover, some long pauses can simply the sign that a speaker has nothing more to say on a given topic and should not be considered as signs of disfluency. It is with such issues in mind that some researchers suggest discarding pauses which last over a certain length of time. For example, Riggenbach (1991) recommends discarding any pause time over 3 seconds.

4.1.2.2 *Filled pauses*

Riggenbach (1991: 426-428) defines filled pauses as 'voiced fillers which do not normally contribute additional lexical information'. She further distinguishes between:

1. Non lexical fillers such as *uh, uhm,* etc., which are not recognised as words as they contain no semantic information.
2. Sound stretches that are vowel elongations of 0.3 seconds or more. Sound stretches have been recognised by researchers as a pause phenomenon in a psycholinguistic sense: Zellner (1994), Freed (1995, 2000), Pawley & Syder (2000), and Raupach (1984) call this phenomenon drawls, syllable lengthening, or slowdowns. A problem with sound stretches is that they are quite difficult to measure, as a normal syllable/sound length has initially to be defined in order for an extended sound to be identified.
3. lexical fillers such as *you know* or *I mean.* These types of fillers are recognised as words but in context contribute little semantic information. They are problematic methodologically because they can either have a literal or a 'filler' meaning. In quantitative corpus linguistics, these two different uses cannot be distinguished for each occurrence. That is why, according to Dahlman (2010), the a priori categorisation of lexical fillers as pauses is problematic.

4.1.2.3 *Repetitions and false starts*

Repetitions and false starts can be seen as breaks in the flow of speech as even when they are not accompanied by a pause, they break the semantic and / or syntactic unity of an utterance. Like pauses, they are evidence of online processing taking place and they can also be regarded as time buyers for speech planning.

Looking at the list of various possible types of pauses and disfluencies, it is clear that a fluent run can be interrupted by a whole range of different phenomena. As a result, the definition of a fluent run is directly dependent on the type of phenomena one regards as an interruption of fluency. Chapter 5 will describe in detail which phenomena have been considered as interrupting a fluent run in the present study.

4.1.3 *Utterance fluency: a complex phenomenon to study*

As summarised by Segalowitz (2010), despite several decades of work, researchers have not agreed on universally applicable, objective measures of utterance fluency. As

underlined by Kormos (2006), one of the problems of fluency research has to do with the fact that researchers have not been consistent in the way that they have operationalised oral performance variables, for example in the definition of what constitutes a pause. Segalowitz (2010) points out that speech elicitation techniques have varied from study to study, with the consequence that the results obtained were not comparable. For example, certain studies used narratives whereas others used interviews. It therefore makes little sense to compare studies which have used such different tasks.

These methodological discrepancies between studies on utterance fluency are likely to be inevitable and are not necessarily problematic as such. However, because of the inconsistency in the operationalisation of fluency variables, it is of paramount importance, when looking at any results, to know exactly how they were obtained and especially how pauses were defined and measured. This is why the way in which fluency measures were calculated in this study will be described precisely in the next chapter so that the results obtained can be interpreted accordingly.

4.2 Utterance fluency as an insight into 'cognitive fluency' (Segalowitz 2010)

As explained in the preceding chapter, the study of the temporal variables of speech, i.e. utterance fluency, is important because it can provide indirect access to the underlying cognitive processes of language production. Thus it can be said that utterance fluency is a reflection of cognitive fluency, i.e. the speaker's ability to efficiently mobilise and integrate the underlying cognitive processes responsible for producing utterances. As Grosjean (1980: 51) puts it, 'studying temporal variables for their own sake is important but it is even more valuable to integrate the results obtained into a model (or models) of production, perception and language acquisition'. However, despite the cognitive focus of many definitions of fluency (Lennon 2000, Rehbein 1987), very little research has been carried out in order to explore the links between utterance fluency and cognitive fluency. One important question to consider is the cognitive meaning of the various measures of fluency described above. What do these various measures precisely tap into? How can the link between temporal variables and underlying cognitive mechanisms be conceptualised?

This is one of the questions asked by Towell et al. (1996), a study which stands out from other studies on fluency because the authors attempted to account for their findings by relating them both to Levelt's (1989, 1999) model of speech production and to theories of automaticity such as Anderson's (1983, 1995). Towell et al. (1996)

resorted to Anderson's theoretical framework because, according to them, becoming fluent in oral second language production requires the conversion of declarative knowledge into procedural knowledge (see chapter 3 for a more precise description of Anderson's model).

In order to examine proceduralisation of language production, they focused on several temporal variables: speaking rate (SR), phonation / time ratio (PTR) i.e. the percentage of time spent speaking, articulation rate (AR) and mean length of runs (MLR). Although they point out that the relationship between these temporal variables and underlying cognitive mechanisms is not straightforward, they suggest that temporal variables can provide evidence about the involvement of the various processors of the Levelt model (see chapter 3) in the development of on-line speech production in L2 learners. According to them, SR is a very general measure as it encompasses the working of the whole model i.e. conceptualising, formulating and articulating the message (see chapter 3 for a more precise description of phases of speech production). That is why Towell et al. assume that an increase in SR can be taken as evidence that the processes of proceduralisation have taken place somewhere within the model. However, because of the general nature of SR, it cannot indicate precisely in which speech production processor the change has happened. MLR is *a priori* a more precise measure than SR as increase in MLR (i.e. the fact that speakers are able to produce longer utterances uninterrupted by pauses) suggests that proceduralisation in the formulator has increased since the speaker is able to 'formulate' longer runs of speech.

However, according to the authors, increase in MLR alone is not enough to be interpreted as proceduralisation in the formulator since it could be argued that increased MLR might be due to greater planning time. Greater planning time may manifest itself in decreased phonation / time ratio either because there has been an increase in the average length of pauses or because there has been a greater number of pauses. Consequently, Towell et al. argue that the pattern which needs to be found to indicate greater proceduralised knowledge in the formulator is: increased mean length of runs with, at the same time, no increase in average length of pause and stability or increase in phonation / time ratio. If there is an increase in the mean length of runs with no parallel increase in pausing time, the improvements observed in the overall measure of the speaking rate can mostly be accounted for by changes in MLR. In terms of the Levelt model, this suggests that the changes in fluency measures, rather than being attributable

to changes in the conceptualizer or the articulator, are due to changes in the formulator, and therefore to the fact that linguistic knowledge has been proceduralised.

To sum up, Towell et al. (1996) is an essential study because of its attempt to relate fluency measures to the internal mechanisms underlying speech production. More details on this study will be given in the next section, which not only reviews studies investigating the link between FS use and L2 fluency but also studies on the development of L2 fluency.

4.3 Studies on the development of L2 fluency, the role of the study abroad context and the link between FS and fluency

There is a very large body of studies dealing with the development of L2 fluency. For reasons of space, it is impossible to review them all. In this review, I will therefore focus exclusively on the studies which are the most relevant in relation to the specific characteristics of the present study: it focuses on advanced learners of French, it has a pre-post year abroad longitudinal design and it specifically deals with the development of FS and its effect on fluency and lexical diversity. Therefore, only a specific subset of the very dense corpus of studies on fluency will be reviewed: those dealing with the longitudinal development of fluency in learners of French, the role of the study abroad context on the development of fluency and lexical diversity, and the link between FS and fluency.

4.3.1 *Towell et al. (1996): Key longitudinal study on the development of L2 fluency*

Towell et al. (1996) is highly relevant for the present study for 2 main reasons: it is also a longitudinal study of undergraduate learners of French and, like the present study, it attempts to relate the temporal variables of fluency with underlying cognitive mechanisms.

The subjects of Towell et al. were recorded before and after a year abroad, of which six months were spent in France. The task used was the story-retelling of a film in a recording booth with no interlocutor present. Although the authors admit that it is not the most natural way of collecting data, they chose it because it allowed for the task to be easily replicable from one time to another. The subjects were also asked to retell the story in their L1 at another time.

As explained in the preceding section, Towell et al. stands out from other studies on fluency because the authors attempted to account for their findings by relating them

both to Levelt's (1989) model of speech production and to theories of automaticity as previously described. Before their study, only Raupach (1987) and Möhle and Raupach (1987) had placed the temporal variables of fluency in the context of Anderson's theories and no one had attempted to link developmental processes to Levelt's model of speech production. Towell et al. (1996) resorted to Anderson's theoretical framework to account for the development of fluency in L2 acquisition. According to them, becoming fluent in oral second language production requires the conversion of declarative knowledge into procedural knowledge. They argue that proceduralisation of linguistic knowledge takes place in the formulator module of Levelt's model because it is the place where syntactic, morphological and phonological rules are applied and where conversion from declarative knowledge to procedural application of rules can take place.

The results of their study showed a statistically significant increase in SR and MLR, though SR and MLR at T2 were found to be inferior to SR and MLR in the L1. According to the authors, this shows that knowledge which underlies L2 processing is in some way different from the knowledge which underlies the processing of L1. Moreover, the dispersion of performance fell over time. In other words, there was a tendency for the students who scored highest at T1 to be those who, relatively speaking, show the least improvement at time 2 and conversely. The authors concluded that as exposure and practice increases so does fluency, but not, however, to levels comparable to those achieved in the native language. At the same time, since there was no fall in phonation-time ration and no increase in average length of pause, Towell et al. concluded from the quantitative evidence that the improvements observed in the overall measure of SR were mostly accounted for by changes in MLR. In terms of the Levelt model, this suggested, according to them, that the major changes in fluency are attributable to changes in the formulator, and therefore to the way linguistic knowledge is stored as procedural knowledge.

In order to find out what exactly has been proceduralised by the learners, Towell et al. conducted a qualitative analysis of the speech of two subjects in order to test the hypothesis that the learners have increased their MLR by proceduralising more knowledge. Subject 1 was chosen by the authors because they had made the greatest improvement in fluency from a low starting point while subject 2 was selected because, although they were already very fluent at time 1, their MLR increased significantly at time 2. From the qualitative analysis, the authors concluded that the causes of subject 1's improvement were the proceduralisation of the ability to use (1) 'sentence builders'

defined by Nattinger and DeCarrico (1992) as lexical phrases that provide the framework for whole sentences and contain open slots such as *not only X but also Y* and (2) the improved use of 'institutionalized expressions' defined by Nattinger and DeCarrico (1992) as formula for social interaction, such as *have a nice day*, which are usually invariable and function as separate utterances.

For subject 2, the authors attributed the cause of increased fluency to the use of more complex syntax, as subject 2 used subordination not just as a means of providing additional information but also as a means of highlighting information. Moreover, at time 2, there were far fewer internal hesitations, which, in Anderson's terms, suggests that the subject had established various productions i.e. condition/action pairs which are very fast to access in an autonomous formulator (see chapter 3 for more details on Anderson's theory).

From both the quantitative and qualitative analysis, Towell et al. (1996) concluded that the increase in fluency is not the result of a quantitative reduction in the amount of pausing that subjects do, nor in the increase in the speed with which they articulate what they say. Rather there is an increase in the length and complexity of the linguistic units which are uttered between pauses. According to them, what has changed in the learners is the rapidity with which syntactic and discourse knowledge can be accessed for on-line speech production. According to them, the component of the model in which proceduralisation takes place is the formulator, which is centrally concerned with assigning grammatical representations to messages and passing these grammatically-encoded messages to the articulator. A core feature of development in advanced L2 learners therefore seems to be the conversion of linguistic knowledge already acquired into rapidly-usable on-line 'productions' (Anderson 1983, 1995).

Towell et al. is a very significant study because of its attempt to link quantitative measures of speech (utterance fluency) to underlying cognitive mechanisms (cognitive fluency) by both integrating fluency into a model of speech production and accounting for its development thanks to theories of automatisation. The empirical design chosen, i.e. looking for a certain pattern in the combination of fluency measures (increased speaking rate and mean length of runs combined with no increase in average length of pause and no decrease in phonation time ratio) seems like a sophisticated and valid way of showing evidence of proceduralisation in the formulator.

The study has a few limitations, however. Firstly, the only task used to elicit data was a story retell. Although it is a way of ensuring that data is elicited in a standardised way (which makes results comparable across participants), it might not be the best way of eliciting spontaneous oral data, all the more so as the recording took place in a recording booth.

Another problematic aspect of this study has to do with the way the authors account for the type of linguistic knowledge that has been proceduralised by the two learners studied for the qualitative analysis. Towell et al. draw a clear opposition between subject 1 for whom the increase in fluency is attributed to the increased use of lexical phrases (defined by Nattinger and DeCarico (1992) as collocations that have been assigned pragmatic functions) and subject 2, for whom the increase in fluency is explained by proceduralisation of already existing grammatical knowledge. However, it is possible to interpret the differences between the 2 learners in a less clear-cut way than Towell et al. Indeed, it can be argued that the type of proceduralisation that has happened for learner 2 is of a more lexical nature than what is reported by the authors. In other words, what they interpret as proceduralisation of syntactic competence might in fact be the automatisation of a restricted number of syntactic formulaic frames. Towell et al.'s interpretation is based on their restricted definition of formulaic sequences which they equate with lexical phrases. However, if one adopts a broader view of FS, i.e. one that recognises automatised syntactic frames as formulaic, the role of FS use in their learners' increase in fluency might be assessed as more important than thought by Towell et al.

4.3.2 *Role of the study-abroad context on fluency development: Freed, Segalowitz and Dewey (2004)*

Freed et al. (2004) is interesting in the context of the present study for two main reasons. Firstly, like Towell et al. (1996), it also deals with the acquisition of fluency in advanced learners of French. Moreover, it examines the role of the context of learning on the development of second language fluency as the authors wanted to put to the test the widely held idea that the study abroad context is highly favourable to the development of fluency. More precisely, their study aimed to 'identify the characteristics of learner language that presumably have most contributed to the development of popular perceptions regarding superior language gain in the SA context' (2004: 277).

The authors compared the acquisition of various dimensions of fluency by 28 students of French studying in three different learning contexts: formal language classrooms in an at home institution (AH), an intensive summer immersion program (IM) and a study abroad setting (SA). The students participated in oral interviews and provided information regarding their language use and interactions.

The researchers' main finding was that the students in the SA context demonstrated significant gains in several features of oral fluidity compared to those in the AH context. However, when compared to the AH group, they made fewer gains than the IM group. This could be explained by the fact that, surprisingly, students in the SA context reported using more English than French in out-of class contact. The authors point out that such a finding is consistent with other reports concerning the paucity of L2 use by students in an SA setting (Frank, 1997; Wilkinson, 1998). Freed et al. concluded that 'it is not the context per se that promotes various types of learning but rather the nature of the interactions, the quality of the experiences , and the efforts made to use the L2 that render one context superior to another with respect to language gain' (2004: 298).

Although Freed et al. (2004) confirms that fluency tends to develop with a period of residence abroad, it really questions the role of the learning context and does not take for granted the common conception that study abroad is necessarily the best learning context, for example compared to an immersion context. Another useful contribution of this study is the elaboration of a language contact profile questionnaire to be filled by the participants in order to get a precise idea of how much L2 they have used during the learning period under scrutiny. This questionnaire will be described in more detail in the methodology section, as the questionnaire used for the present study was based on it.

An important shortcoming of Freed et al's study, however, is the fact that only two-minute segments of speech were analysed before and after the stay abroad, which is unlikely to be enough to be representative of the speech of the learners under scrutiny. Moreover, the data was not analysed qualitatively. Therefore, the increase in fluency, though witnessed, was not accounted for and contrary to Towell et al. (1996) the authors did not make any suggestions about which changes in the learners' language might be at the source of their increase in fluency.

4.3.3 *Role of the study-abroad context on lexical development: Foster (2009)*

Foster (2009) examines the role of the context of learning from another perspective as it investigates its effect on learners' lexical development. Her study is relevant to the present study for several reasons. Firstly, as mentioned at the end of Chapter 2, FS have a lexical dimension as they are composed of lexical items automatised as units. Therefore, their development needs to be studied within the more general context of lexical development. Moreover, Foster's study also focuses on the development of FS, although, as explained below, she defines them learner-externally.

In Foster's opinion, the L2 instructed learning context is not a rich enough environment to allow for the building of native-like primings for words. In Hoey's (2005) terms, classroom-acquired vocabulary lacks the necessary primings which result from repeated encounters with words in different contexts, different situations and different speakers. On the contrary, for learners living in the target language environment, the situation is different: 'language is encountered in a multitude of contexts allowing for at least the possibility that learners will be able to build more authentic networks of L2 word associations' (Foster 2009: 93). That is why, according to Foster, it is not surprising that that the learning context is a significant variable in the acquisition of L2 lexical knowledge.

Foster contrasted two comparable groups of intermediate learners of English studying in 2 different learning contexts: 40 participants in a 'study abroad' (SA) environment in London from a wide variety of mother tongue backgrounds and 60 participants in an 'at-home' environment (AH) in Tehran, all native speakers of Farsi. Her study builds on several studies which point to lexical organisation, especially of the formulaic language kind, as the main area of benefit for study-abroad learners (Marriott 1995, Siegal 1995).

The results showed that the context of learning indeed influences the learners' lexical development. Although the SA abroad learners were not found to be better than the AH learners in terms of grammatical accuracy, in terms of lexical diversity, they were found to be closer to the native speakers than to the AH learners. Like the native speakers and unlike the AH participants, the SA learners were more likely to use a delexicalised verb such as 'have a drink' instead of 'drink' and to use the existential construction 'there is/are'. They also used more colloquialisms, which were virtually non-existent in the language of the AH learners.

Foster's approach to FS (that she calls lexical phrases) is clearly learner-external as she used a mix of several pragmatic and corpus-based criteria in order to identify them: she considered a sequence of words formulaic if it was judged to be associated with a specific situation and / or register, or to be commonly used to convey an idea, or to have been encountered by the speaker before. She then checked the sequences for occurrences in the BNC. Examples of identified lexical phrases included: *in the background, blissfully unaware* or *a day off*. Foster points out that only nativelike sequences were looked for, although she acknowledges that it cannot be excluded that the learners may also have used idiosyncratic prefabricated sequences.

Foster found that lexical phrases were much more frequent in the SA corpus (mean total of 10.6 lexical phrases per participant) than in the AH corpus (5), though less frequent than in the NS corpus (21.1). Foster also found that greater use of FS could not really be associated with greater fluency, as only one measure suggested that the SA learners were less likely to pause in the middle of a clause than the AH learners. This shows that they were better able, to some degree, to plan their utterances as whole clauses. However, since the development of fluency was only a secondary aim of Foster's study, the learners' fluency was not measured in much detail since utterances were only coded for repair and breakdown, and Foster does not give any detail on how the coding was carried out. As a result, her statement that there is no link between FS use and fluency cannot be taken for granted. Moreover, the fact that her definition of FS is learner-external might explain the absence of association between FS use and increased fluency: the learners might have learnt new multiword sequences without necessarily automatising them.

Forster's study shows that FS defined learner-externally develop hand in hand with lexical diversity but as her approach is learner-external, it does not say anything about the link between the development of lexical diversity and FS defined learner-internally like in the present study. This link therefore remains to be investigated more thoroughly. Looking at the development of FS use in relation to that of lexical diversity is a way of investigating whether a potential increase in the quantity of FS is due to the fact that new lexical types of FS are acquired or if it can simply be accounted for by the fact that more lexical items already present in the learners' vocabulary have been automatised as units. Several configurations are possible: Lexical diversity might increase but not the quantity of FS because the new vocabulary acquired by the learners does not contain automatic sequences. Another possibility would be that the quantity of FS increases but

lexical diversity does not, which would mean that what has developed in the learners is the proceduralisation of lexical knowledge but not lexical knowledge itself. A final possibility could be that the use of FS increases at the same time as lexical diversity increases. In other words, FS use may increase not only because more pre-existing lexical knowledge has been automatised as multiword sequences but also because FS made of new lexical knowledge have been acquired.

4.3.4 *Studies dealing with the link between FS use and fluency*

According to Wood (2010), within the current body of knowledge about fluency, a gap exists between the work focusing on temporal measures of fluency and the work dealing with the use and role of FS. Indeed, very few studies have investigated fluency development and use of FS by attempting to link development in the temporal aspects of learners' speech with their use of FS. They are reviewed below.

- **Raupach (1984)**

Raupach (1984) found evidence of formulaic constructions contributing to fluency, particularly modifiers and rhetorical organisers. More details on his study will be given in the next section, as the methodology he used fits a psycholinguistic approach to FS and inspired, to a certain extent, the identification method used in this study.

- **Ejzenberg (2000)**

Ejzenberg (2000) examined the role of formulaic speech in oral fluency. Drawing from Krashen and Scarcella (1978)'s distinction between routines and patterns, she defined formulaic speech as prefabricated speech produced as strings of words either in fixed or semi-fixed sequences. This definition suggests that Ejzenberg assumes that routines and patterns are also a psycholinguistic reality, as reflected by her method for identifying them. Although she uses learner-external criteria such as 'semantic characteristics', she also uses intra-learner frequency (i.e. frequency of a given sequence within the same learner) including frequency of non-standard sequences, which is a criterion characteristic of a learner-internal approach. As a result, her approach to FS somehow lacks clarity with regard to whether it is learner-internal or learner-external.

In her study, fluency was not measured quantitatively but qualitatively as it was assessed by 4 judges according to three qualitative criteria: rate of speech, ease of expression and continuity. As a result, the study only deals with the link between FS use and perceived fluency, not utterance fluency. Ejzenberg's finding that speech is faster and more fluent if the speaker engages in formulaic production therefore needs to be qualified as it only applies to perceived fluency. Moreover, even at a qualitative level, the 3 subcategories that the judges had to rate i.e. rate of speech, ease of expression and continuity are confusing as they all overlap to a certain extent and it seems unlikely that they could be judged independently.

Ejzenberg's qualitative analysis draws an opposition between the highly fluent and less fluent speakers in relation to their use of FS. On the one hand, compared to the less fluent speakers, the most fluent speakers used more formulaic speech and fillers in the monologues while apparently searching for direction or planning ahead. Moreover, they were able to embed the FS contextually within their discourse. In other words, they used preassembled sequences in ways that fitted the preceding and following discourse segment. Fluent speakers also added to the sequences to form longer stretches of speech. In contrast, low-fluency speakers tended to insert FS inappropriately in the context of their discourse, failed to retrieve them in grammatically accurate ways and also merged incompatible sequences. For example, one learner's disfluency episode was caused by her merging *it's gonna be fun* and *we're gonna have fun* before choosing to rephrase: 'I'm sure we're gonna be fun...we....we're gonna to.....we're going to have (breath) a great time'. The low-fluency speakers were also found to be more bound to have

disfluency episodes if the variable part of the semi-fixed sequences fell within the sequence rather than at the end. For example, one speaker was unable to break: *I'd like to*+ infinitive, in order to insert *you* when they were trying to say *I'd like you to + infinitive*.

To sum up, Ejzenberg's study showed that FS had an ambivalent effect on fluency according to the speakers' level of proficiency. Indeed, although FS had a beneficial effect on the fluency of the more proficient speakers, they were found to have a detrimental effect on the fluency of non-proficient speakers who had not automatised the retrieval of sequences. In other words, her study suggests that the way in which L2 learners use FS affects how fluent they are perceived to be.

- **Wood (2009)**

Wood (2009) is a case study of the spoken discourse of a Japanese learner of English, which analyses in detail the nature and functions of FS in fluency gain. More precisely, the study examines the learner's use of FS before and after a six-week fluency development workshop. The fluency workshop was composed of four stages during which the learners were taught how to use fillers and sequences with a discourse function and were also given time to automatise the learnt sequences. The learner of the case study was asked to produce narratives spontaneously in a language lab on topics of personal relevance before and after the six-week fluency workshop.

To study the learner's gain in fluency, Wood used two types of temporal variables: speaking rate (SR) and mean length of runs (MLR). The cut-off point used for pauses was 0.3 seconds. To identify FS, Wood used 5 criteria, based largely on Wray (2002) and Wray and Namba (2003). (1) phonological coherence, (2) the taxonomy used by Nattinger and DeCarrico (1992),[2] (3) greater length and complexity than other output, (4) semantic irregularity and (5) syntactic irregularity. Moreover, idiosyncratic and non-canonical FS were accepted in order to fit the L2 context. Wood therefore used both learner-internal (1 and 3) and learner-external criteria (2,4 and 5) for the identification process, which suggests that, like Ejzenberg, he assumes that sequences defined

[2] Nattinger and DeCarrico (1992) distinguish between:

- Syntactic strings
- Collocations: strings of specific lexical items that co-occur with a mutual expectancy greater than chance
- Lexical phrases: collocations that have been assigned pragmatic functions

externally are psycholinguistically real . With regard to the precise way in which the identification criteria were applied, Wood stresses that they were used in a 'holistic manner' (2009: 48) that is to say that 'no particular criterion or combination of criteria were deemed as essential for a word combination to be marked as formulaic, and judgements were made based on one, several or all of these' (2009: 46-47).

The results showed strong gains in fluency measures from the first sample to the second. MLR increased by 26.3% and SR by 13.8%, which showed that the speaker was more readily able to produce faster speech and longer fluent runs. The quantity of FS used was counted in syllables and the mean length of FS went from 3.17 syllables in sample 1 to 4.46 syllables in sample 2. According to Wood, this increase of 40% in the mean length of FS no doubt plays a role in the overall increase of the length of runs. Although the speaker did not use proportionally more FS in the second than in the first sample, they were generally longer and more complex. Indeed, in the first sample, they were mainly fillers such as *I think* and *you know* whereas in the second sample, the FS used were of a wider range of functions and types such as FS used for temporal marking such as *in the daytime, two years ago*. More complex FS such as *the interesting thing is that* seemed to have been acquired in the fluency workshop undertaken by the subject between time 1 and time 2.

	Sample 1	Sample 2
Number of FS	18	52
Number of FS from fluency workshop	2	18
Number of syllables	530	760
%syllables from FS	11.3	12.5

Table 1: FS in Wood (2009)

Though acknowledging that the results from such one brief case study cannot be generalised, Wood points out that they still suggest that increased use of FS has a positive effect on fluency development. An interesting result of this study was that what mattered most was less the proportion of FS used, which did not change very much, than their length and complexity.

Another shortcoming of this study is the fact that, similarly to Ejzenberg (2000), Wood does not make a distinction between a learner-internal and a learner-external approach to FS as he uses identification criteria both learner-internal and learner-external, which he applies in a holistic manner i.e. no one criterion is necessary for the judgement of formulaicity, nor are all the criteria required to apply to all cases. This identification

method might have led him to identify some sequences which do not belong to the same construct as some might be only learner-internal (identified on the basis of learner-internal criteria only), some only learner-external (identified on the basis of learner-external criteria only), and others both learner-external and psycholinguistically valid for the learners under scrutiny in his study (identified by both learner-internal and learner-external criteria).

- **Wood (2010)**

This exploratory longitudinal study focuses on the possible effect of a growing repertoire of FS on fluency measures and is one of the first attempts to explore in depth the connection between the use of FS in adult English L2 learners and speech fluency development.

The speech samples were collected 6 times over the course of a 6-month period from 11 study abroad participants who studied at an intensive ESL program at a Canadian university. They all lived in homestay situations which provided a naturalistic environment with rich and sustained opportunities for English input and communication. Participants had three different L1s: Japanese, Chinese and Spanish.

The participants were asked to retell the story of three short animated silent films of similar length and narrative complexity. Each film was viewed twice at three months' interval: the first film was viewed the first and fourth month, the second film the second and fifth month, and the third film the third and sixth month. The nature of the task (narrative retell) was chosen for several reasons. First it was thought to produce clearer results than tasks such as interviews due to its more constrained nature. Moreover, it was considered the safest way to ensure equality of opportunity to speak for all participants and avoid the influence of the presence of an interlocutor (which might cause wide differences in the nature of the speech produced by participants). Finally, according to Wood, the narrative retell procedure provides optimal opportunities to produce discourse exhibiting clause chaining (Pawley and Syder 2000) and use of FS.

The corpus was analysed for FS using the same 5 criteria[3] as those used in Wood (2009) but they were applied by three expert native speaker judges and not the researcher alone. The judges not only examined the transcripts, they also listened to the sound files.

[3] (1) phonological coherence and reduction, (2) the taxonomy of Nattinger and DeCarrico (1992), (3) greater complexity than other output, (4) semantic irregularity and (5) syntactic irregularity.

Sequences had to be identified by at least two out of the three judges in order to be considered formulaic. Like in Wood (2009), the judges were asked to apply the criteria in a holistic manner i.e. a sequence could be considered formulaic based on one, several or all of the criteria.

A combination of quantitative and qualitative data analysis methods was used since the speech samples were subjected to a discourse analysis after the fluency gain had been established.

Following Towell et al. (1996), Wood interpreted increase in fluency as evidence of increased proceduralisation in the formulation stage of speech production. However, contrary to Towell et al. who mainly focused on the proceduralisation of syntactic knowledge, the key linguistic element Wood focused on with respect to automatisation is the use of FS, including all multiword units such as collocations and lexical phrases.

Wood's hypothesis was that FS would appear more frequently in the longer runs between pauses. The temporal variables calculated were phonation-time ratio (PTR), speaking rate (SR), articulation rate (AR) and mean length of runs (MLR). The cut-off point chosen for the identification of pauses was 0.3s. Filled pauses were transcribed but not timed. Prosodic lengthening or drawling of syllables was not marked.

A formula / run ratio (FRR) was also calculated that is to say the measure of the proportion of formulas to the number of runs in order to provide an indication of how the number of formulas per run may have changed over the time of the study. Wood points out that the FRR is a contribution to the methodology of studies of fluency and formulaic language because it bridges the gap between quantitative analysis of temporal aspects of speech and discourse analysis of how formulas are used in speech.

Although a repeated-measures one-way ANOVA was used to determine the statistical significance of changes in the temporal variables of speech over time, Wood underlines the fact that the small cohort of participants combined with the large variances among participants and scores, makes it very difficult to determine statistical significance. Still, the quantitative results which emerged from the whole group data on FRR were that increased formula use went hand in hand with improvement on other temporal variables, including MLR. However, this correlation was not statistically significant. Moreover, there were important individual differences in the performance of participants. Some participants, such as one of the Japanese males and one of the Chinese males, showed

development in both fluency and automatisation of FS. However, in the case of other participants, increases in fluency measures did not relate to the use of formulas. For example, in the case of one of the Japanese females, the FRR was found to decline over time. As a result, in her case, the increase in her fluency profile cannot be due to the automatisation of formulas. An additional problem was that, in the case of some of the participants (e.g. the Chinese female or one of the Spanish males), there were no clear fluency gains over time. Wood points out that the absence of a clear fluency gain in the case of some participants may have had to do with the nature of the task. Indeed, some participants attempted to describe the film in detail and ended up in a situation in which they were challenged by the task. For example, in the case of one of the Chinese males, there was loss of momentum in fluency development in the last half of the study.

For the qualitative analysis, speech samples in which increased use of FS paralleled the changes in temporal measures for the same film prompt were examined in order to identify exemplars of the facilitating role of formulas in increased fluency over time. Wood points out that, interestingly, in no case did a participant attempted to express exactly the same idea twice. The qualitative analysis enabled Wood to distinguish five broad categories of FS use which contributed to increased length of runs in later retellings:

1. Use of self-talk and fillers: *I don't know, I think, I guess*
2. Repetition of FS in a run: *came back, really happy, in his room, to the house*
3. Use of multiple FS to extend a run: *went upstairs to her house, came back to the house*
4. Use of FS as rhetorical devices: *that is the end of the story, at the beginning*
5. Reliance on one FS or filler repeatedly: *and then, and next*

These varied ways of using FS prompted Wood to conclude that FS play a role in facilitating fluent speech over time. He remarks, however, that the complex and shifting nature of the development of fluency tracked in his study highlighted issues around the analysis of spontaneous spoken language data. Indeed, according to him, the development of spoken language is by no means linear, nor readily captured and examined in empirical research.

Although, according to Wood, the results of the study are evidence for a link between FS use and fluency development, the results remain inconclusive. Indeed, the figures only work in the case of all the results taken together and in the case of 3 out 11

participants. However, the problematic aspect of this study is less to show a link between FS and fluency than to show a development of the participants' utterance fluency over time. As acknowledged by Wood himself, this shortcoming is probably due to the undesired effects of the story retells used for elicitation. Moreover, it must be pointed out that the fact that three different films were used for the story retells prevented the results from being comparable, though of course, there would have been an equally unwelcome learning effect if the same story had been used for the 6 retells. Another likely explanation for the lack of development in the participants' fluency is the fact that data was collected every month, which makes any dramatic improvement between times of data collection unlikely. Still, even when comparing the retell of one film with the later retell of the same film 3 months later, the increase in fluency was not clear for some of the learners.

Moreover, despite the resort to three independent native speaker judges, the identification method used suffers from the same problem as Wood (2009) (see last paragraph on Wood 2009) as it does not distinguish clearly between FS defined learner-internally and FS defined learner-externally.

Despite these shortcomings and the inconclusive nature of the results, this study is important in its attempt to bridge the gap between research on fluency and research on formulaicity. In this respect, the measure formula / run ratio is an important contribution as is the qualitative analysis of the different modalities in which the use of FS can facilitate fluent speech.

4.4 Conclusion and additional research questions

This chapter presented how the complex construct of utterance fluency has been operationalised in the literature in terms of the various temporal variables used to measure it as well as the treatment of pauses and hesitation phenomena. Moreover, in order to relate utterance fluency to cognitive fluency, it described how the various measures of fluency can be interpreted. Chapter 6 in the methodology section will describe how utterance fluency was operationalised in this study.

This chapter also reviewed the studies considered relevant to the research goals of the present study because of the way they investigated either fluency, the role of the context-abroad on the development of fluency and lexical diversity or the relationship between FS use and fluency. The review of the studies dealing with the link between FS use and fluency development has shown that in most of these studies (Ejzenberg 2000,

Wood 2009, 2010), the psycholinguistic advantage of sequences defined learner-externally is assumed by the researchers. At the core of the present study, however, is the belief that FS defined learner-internally and FS defined learner-externally are different constructs and should not be equated, especially in the context of L2 acquisition.

More research is therefore needed on the longitudinal development of learner-internal FS (clearly defined) and its relationship with both lexical diversity and fluency. Given the literature outlined above, the best context to examine that development seems to be that of the year abroad. As a result, in addition to the research questions outlined in chapter 2, the present study seeks to answer the following research questions:

- Longitudinal development of FS use, lexical diversity and correlation between the two
    - How does the use of FS by advanced learners of French develop during a seven months' stay in France?
    - How does their lexical diversity develop?
    - How does FS use develop in relation to lexical diversity? Is there a correlation between the two?
- Longitudinal development of fluency and relationship between FS use and fluency
    - How does the fluency of advanced learners of French develop during a seven months' stay in France?
    - Does the development of FS use by L2 learners play a role in the development fluency during the year abroad? → Quantitative analysis
    - If it does, how does it take place? → Qualitative analysis
    - Which underlying mechanisms can explain this role?
- Is there a relationship between the learners' engagement with the French language during their stay abroad and their development in FS use, lexical diversity and fluency?

Before the above research questions can be answered, FS need to be reliably identified in advanced L2 learners. Therefore, the next chapter deals with the issue of identifying psycholinguistic FS and puts forward an identification method for learner-internal FS in advanced L2 learners.

Chapter 5. Identifying formulaic sequences in advanced L2 learners

An obvious requirement to answer the research questions outlined at the end of chapters 2 and 4 is the reliable identification of FS in the corpus of L2 learners' productions collected for the present study. However, as underlined by Wray (2009: 28), this is no simple task: 'Researching formulaic language has many challenges but probably the single most persistent and unsettling one is knowing whether or not you have identified all and only the right material in your analyses'. In other words, in relation to the identification process, the researcher is faced with two opposite risks: that of not identifying all the right material and that identifying too much material.

When tackling the problematic issue of identification, it is of paramount importance to be clear about which construct exactly one seeks to identify since the definition used for FS has a crucial impact on the relevant criteria to be used in order to reliably identify them (For a discussion of the various definitions of the term FS, see chapter 1 section 1.1.2.). In the present study, a formulaic sequence is defined psycholinguistically as a multiword semantic / functional unit that presents a processing advantage for a given speaker, either because it is stored whole in their lexicon or because it is highly automatised. This is why, although a lot has been written in the literature about the issue of identification, only the studies dealing with the identification of FS from a psycholinguistic perspective will be reviewed in this section.

5.1 The identification of FS in L1 acquisition

Although the present study deals with L2 learners, the literature on the identification of FS in L1 acquisition is of particular interest and relevance because it clearly tackles formulaicity from a psycholinguistic learner-internal perspective. It has been well-documented that FS are prevalent in child language (Clark 1974, Peters 1983) and that children reproduce strings of words holistically before they have fully analysed their constituents. Thus, in the L1 literature, formulaic sequences are synonymous with unanalysed sequences recalled holistically.

Peters (1983) made an essential contribution to the identification process by pointing out six characteristics of unanalysed holistic units in children's speech:

1. The utterance is an idiosyncratic chunk that the child uses repeatedly and in exactly the same form

2. The construction of the utterance is unrelated to any productive pattern in the child's current speech
3. The utterance is somewhat inappropriate in some of the contexts in which it is used
4. The utterance coheres phonologically
5. The usage of the expression is situationally dependent for the child
6. The expression is a community-wide formula

In other words, according to Peters, a formulaic utterance in L1 acquisition stands out from productive utterances for several reasons: its idiosyncratic and frequent nature, its sophisticated structure compared to other productive utterances produced by the child, its frequent inappropriate use, its phonological coherence, its use in connection to a specific situation and the fact that it has more than likely been picked up by the child in the linguistic input around them. Peters remarks that these six characteristics need not be present at the same time for a sequence to be considered a formulaic unit (1976: 312). This is obvious since, as pointed out by Hickey (1993: 31), some of these characteristics appear to be mutually exclusive. For example, a given utterance cannot be idiosyncratic at the same time as being a community-wide formula. Peters' method therefore lacks precision as she does not go beyond stating that an utterance needs not meet all the criteria in order to classify as formulaic and she leaves it unclear whether some criteria should be considered more important than others. However, the elaboration of these criteria is an important contribution at it emphasises the specific characteristics of FS in comparison with the rest of L1 learners' utterances.

Another key methodological contribution to FS identification in L1 acquisition is that of Hickey's (1993). We will come back to her method of identification in more detail in section 5.3 since the method of identification used in the present study is partly drawn from it.

5.2 Identifying FS in the early stages of instructed L2 acquisition

Following Weinert (1995), Myles et al. (1998, 1999) adapted Peters' criteria to instructed L2 acquisition in order to identify unanalysed chunks of language that are used by beginner learners as communicative crutches when their grammatical competence is not developed enough to enable them to create sentences productively. In

a similar way to L1 acquisition, the crucial criterion for the identification of unanalysed formulaic chunks used by beginner learners is the fact that they are clearly beyond the learners' generative capacity. This can be noticed when there is an obvious discrepancy between, on the one hand, complex chunks that are uttered in a fluent fashion e.g. *comment t'appelles-tu?* (what's your name?) and, on the other hand, simple utterances generated from scratch that are uttered non-fluently e.g. *le nom?* (the name?). Moreover, because these formulaic chunks are used without being fully analysed by the grammar, their meaning tends to be over-extended by beginner L2 learners resulting in their frequently being used inappropriately e.g. **comment t'appelles-tu le garçon?* (**what's your name the boy?* instead of *what is the boy's name*?). (Examples taken from Myles et al. 1999.)

Despite these identification criteria however, it can still sometimes be problematic to discriminate between what is productive from what is not, especially in the case of shorter sequences. For example, if it is easy to identify *comment t'appelles-tu?* (what's your name?) as formulaic, it is more difficult to determine whether sequences such as *il/elle a* (he/she has) are unanalysed or productive sequences for certain learners. On the whole though, and despite remaining identification issues, it can be said to sum up that, in L1 acquisition as well as in the early stages of L2 acquisition, the identification of FS can be carried out relatively reliably because of all the characteristics that make FS clearly stand out from productive utterances.

5.3 Issues in identifying FS in more advanced learners

5.3.1 *The particularity of advanced learners*

It is easy to understand why adopting a psycholinguistic approach to FS in L1 acquisition and the early stages of natural or instructed L2 acquisition makes sense. Indeed in all these acquisitional contexts, the holistic processing of unanalysed multiword sequences is a key mode of operation and therefore FS use by these types of learners cannot be ignored.

In both L1 acquisition and the early stages of L2 acquisition, the crucial element that renders the process of FS identification easier is the gap between the learners' simple productive utterances and their seemingly grammatically-sophisticated non-analyzed formulaic productions. In the case of both L1 learners and beginner L2 learners, FS are retrieved holistically because the learners do not have the ability to break them down grammatically. From the researcher's point of view, the positive consequence of this

gap between formulaic productions and generated productions is that formulaic productions stand out from productive utterances because of their relative complexity as well as frequent inappropriate use and are therefore possible to identify.

In the case of more advanced learners, the discrepancy between competence and performance cannot be apprehended in the same way since advanced learners' grammatical competence can allow them to productively create complex grammatical sentences, and as a consequence formulaic productions do not stand out as clearly from productions generated from scratch. In other words, an underlying processing difference between generated utterances and utterances retrieved holistically is unlikely to be as obvious from differences in performance. This is the case for two reasons. Firstly, if need be, these learners would be able to analyse grammatically the FS they process holistically. In other words, in the case of advanced learners, holistic processing is a processing shortcut strategy and is not constrained by an under-developed grammatical competence like it is for L1 learners or L2 beginner learners. Moreover, because advanced L2 learners' grammatical competence is well developed, the fact that they produce fluent and sophisticated runs is no guarantee that these runs are processing units. They could also be utterances produced from scratch and uttered fluently. As a result, although some of the identification criteria used for L1 learners and L2 beginner learners might sometimes be useful in the identification process for more advanced learners, they are not straightforwardly applicable and need to be adapted in the context of advanced learners.

Very little in the literature has been written about identifying psycholinguistically-defined FS in advanced learners. Firstly, and as was previously mentioned (see chapter 1 section 1.3.2. and chapter 2 section 2.2.3.2.), many studies dealing with FS in advanced learners tackle the construct of FS from a learner-external idiomatic perspective (Yorio 1989, Forsberg 2009). In other words, what these studies investigate is idiomaticity in L2 learners. As a result, the identification criteria used in such studies are not concerned with the holistic processing of the sequences. Moreover, as was previously pointed out (chapter 1 section 1.1.2.), many researchers choose not to focus on the construct of psycholinguistic FS as they consider it impossible to be investigated empirically. Indeed, since they define FS psycholinguistically as units stored whole in the lexicon and that there is no possibility of direct access to speakers' mental linguistic representations, they deduce that psycholinguistic FS are not a pursuable object of linguistic inquiry. This is precisely because of this methodological difficulty that, in the

psycholinguistic framework of the present study, it was chosen to define FS more broadly as units presenting a processing advantage rather than more narrowly as lexical units. As rightly underlined by Lin (2010), a processing unit is a broader unit than a lexical unit (which Lin calls 'storage unit'). As a result, by defining FS as processing units, the preferential processing of some given units can be investigated, without making the claim that these units are necessarily lexical units stored whole in the lexicon, while recognising the possibility that some of them undoubtedly are. In other words, for the sake of methodological validity, the present study only claims that some sequences present a quantitative difference in the way they are processed, without making the claim that this preferential processing has to do with a qualitative difference in the nature of these sequences, though recognising that it might still be the case. As a result, this way of defining FS as semantic / functional units presenting a processing advantage makes their identification methodologically feasible and a method of identification will be described later in this section.

Finally, the most important reason why the identification of psycholinguistic FS has not received more attention is due to the fact that most studies are ambiguous with respect to the perspective on formulaicity that they are adopting. In other words, they tackle the construct of FS globally by approaching it both speaker-externally and speaker-internally (Wray 2008). In other words, in this type of study, although the psycholinguistic perspective is acknowledged as being important in the way FS are defined, it does not percolate into rigorous psycholinguistic operationalisation. Consequently, although the construct of a FS is defined as psycholinguistic, the actual way it is measured in most studies relies on speaker-external criteria such as input frequency or idiomaticity, which says nothing about the psycholinguistic status of these FS in particular speakers.

5.3.2 *Wray's diagnostic approach to identification*

There is a consensus across these very inclusive studies that several criteria need to be applied for reliable identification. Although there are of course minor differences between studies, the identification method resorted to is epitomised by the method described by Wray (2008) entitled 'the diagnostic approach'. The aim of this criterion-based approach is to establish reliable justifications for researchers' intuitive judgments of formulaicity. Wray (2008: 116-121) describes eleven diagnostic criteria on which to base intuitive judgements about formulaicity. These criteria encompass all the different

criteria that have been used to identify FS across various approaches to formulaicity (formal, pragmatic, statistical etc...) and for various types of speakers (from native speakers to L1 and L2 learners).

These criteria include:

- Grammatical irregularity, since the ungrammaticality found in some FS comes about because the word string is first processed holistically, and then, as a result, is isolated from changes in the language over time e.g. *if I were you*
- Lack of semantic transparency when an FS means something as a whole that is different from the combined meaning of its parts e.g. *kick the bucket*
- Specific pragmatic function when the FS is associated with a specific situation such as *happy birthday!*
- Idiosyncratic use by the speaker when the FS is the expression most commonly used by the speaker when conveying a given idea e.g. overuse of *don't get me wrong*
- Specific phonological characteristics used to demarcate the FS from the rest of speech e.g. when the sequence is pronounced fluently and with a specific intonation contour e.g. *you're joking?*
- Inappropriate use e.g. *excuse me* in a context where *I'm sorry* would be appropriate
- Unusual sophistication compared to the rest of the speaker's standard productions e.g. *what time is it?* versust*Time?*
- Performative function e.g. *I pronounce you man and wife*

When adopting an exclusively psycholinguistic approach as is the case in the present study, Wray's diagnostic approach can be considered problematic because there is a very high risk that it might lead to the identification of sequences that are very different from one another in nature. For example, if one takes the case of an idiom such as *kick the bucket*, it is clearly semantically irregular and for this reason could be considered formulaic from a speaker-external perspective. However, it could be used hesitantly by a L2 learner, which would show that the speaker is constructing his utterance on line, in which case it could not be considered a processing unit. On the other hand a sequence of words spoken fluently and with a coherent intonation contour could be identified as formulaic on prosodic grounds but might not be grammatically or semantically irregular. For example, *j'aime le foot* (I like football) can be seen as a processing unit because it

has been learnt and retrieved holistically by an L2 learner although it is a perfectly regular sequence. In other words, when using a set of criteria, one runs the risk of ending up with a very heterogeneous set of sequences which have very little in common with one another such as *passer l'arme à gauche* and *j'aime le foot*: some might be idioms which are also processing units for a given speaker; some might be idioms in the external linguistic environment but which have no psycholinguistic reality for a given speaker; and some might be processing units for learners, whatever their status is in the target language (some might be idiomatic but others might not display any feature of idiomaticity and some might not even be grammatically correct in the case of language learners). Wray is well aware of this issue and rightly underlines that not all of the criteria are applicable to all examples and that a subset of criteria needs to be chosen in order to answer specific research agendas and suit the type of data studied (for example native speakers or L2 learners) . However, most studies in the literature suffer from this ambiguity in the identification method they adopt (Wood 2010).

To sum up, although the need for using several criteria goes without saying, it is however insufficient to ensure methodologically-sound identification. Indeed, although Wray underlines that a subset of criteria needs to be chosen in order to suit certain types of data, she does not tackle directly the issue of the relative importance of some criteria in relation to others. More precisely, from the moment the researcher has defined from which perspective they are looking at the construct of FS, they need to decide on the hierarchical importance of some of the criteria in relation to others within the set of identification criteria they are using as will be explained in more detail in the next paragraph.

5.3.3 *A hierarchical approach to identification*

There is a consensus that not all criteria need to be present for a sequence to be considered formulaic (Wray 2008, Wood 2010). Indeed, it does not make sense to expect that the potential characteristics of a FS should all be present for each sequence because different types of FS exist and they will not all display the same characteristics, especially if the type of speaker under scrutiny is different. For example, it is clear that some features such as unusual complexity and inappropriate use are much more appropriate to the L1 or L2 acquisition contexts rather than to native speakers. However, the fact that not all criteria need to be present for a sequence to be considered formulaic does not imply that some of these criteria do not necessarily need to be present. In fact,

what is of paramount importance, in order to ensure coherence between definition and identification, is that the criteria which are considered defining criteria are necessarily fulfilled.

An essential fact to understand is that a heavier weight given to one criterion rather than another might drastically affect the corpus of identified FS ultimately obtained by the researcher. For example, if one includes a psycholinguistic dimension in one's definition of FS, then an identification criterion showing evidence of preferential processing (like phonological coherence, see 5.4.1.) cannot just be an optional criterion, it has to be a necessary one. This implies that a sequence might display some other characteristics of formulaicity such as semantic opacity, but it should not be regarded as formulaic if it does not fulfill the phonological criterion. In other words, the definition of FS adopted by the researcher must determine not only which identification criteria must be used but also whether some are more important than others in the sense that the identification of a given sequence must be rejected if a necessary criterion is not present notwithstanding the fact that the other remaining criteria have been fulfilled.

The essential idea of the relative importance of some criteria over others is to be found in the L1 acquisition literature and is that of Hickey's (1993). Hickey roughly reused the identification criteria set by Peters (1983) but set them in a 'preference rule system' (Hickey 1993: 31), previously developed by Jackendoff (1983). A preference rule system 'distinguishes between conditions which are necessary, conditions which are graded i.e. the more something is true, the more secure is the judgement- and typicality conditions which apply typically but are subject to exceptions' (1993: 31). Moreover, another important characteristic of the preferential rule system is that in this system, 'there is no subset of rules that is both necessary and sufficient, since the necessary conditions alone are too unselective' (1993: 31). In other words, the fact that a criterion is necessary does not make it sufficient, which means that, although it is necessary, it needs to be completed by additional criteria. Applying this preference rule system to Peters' existing criteria and adding a few additional ones, Hickey comes up with the following 'conditions for formula identification' in L1 acquisition (1993: 32).

Condition 1 (Necessary and graded): the utterance is at least two-morphemes long

Condition 2 (Necessary): phonological coherence

Conditions 3 to 9: all typical and graded

- Individual elements of an utterance not used concurrently in the same form separately
- Grammatical sophistication compared to standard utterances
- Community-wide formula occurring frequently in the parents' speech
- Idiosyncratic
- Used repeatedly in the same form
- Situationally-dependent
- Used inappropriately

Whatever context of identification one deals with, if, as is the case in the present study, the process of identification is carried out gradually i.e. by applying one criterion after the other, adopting a hierarchical method of identification has an important methodological consequence. Indeed, if some criteria are necessary and others are only typical, this difference in relative importance has to be reflected in the order in which identification is carried out. In other words, if, during the identification process, the researcher proceeds gradually by eliminating all the sequences that do not fulfil one criterion, thereby establishing narrower and narrower subsets of candidate FS, then the most important criterion must be the first to be applied in order for the right FS material to be identified. This is because the order in which criteria are applied importantly affects the nature of the sequences identified. For example, if one first applies some formal criteria (e.g. semantic or grammatical irregularity) to identify a subset of candidate FS, then some processing units that are not irregular will not be identified as they will not be retained in the first subset of potential FS candidates. For example, one won't identify the processing unit *J'aime le foot* in an L2 learner. This is not problematic as long as holistic processing is not the key element of the definition used for FS. For example, if the main research agenda is to identify semantically or grammatically irregular sequences, then not identifying *J'aime le foot* is not problematic. However, missing the identification of such a sequence in the present study would be an issue because holistic processing is at the core of the definition of FS which is adopted. On the other hand, if one chooses to apply first prosodic criteria of phonological coherence such as fluent pronunciation, intonation contour or phonetic reductions to isolate a subset of sequences, then some idiomatic sequences which were not pronounced fluently will not be part of the corpus of identified candidate FS. For example, the sequence '*it is raining* / pause / *cats and* / pause / *dogs*', hesitantly produced by a speaker, would not be identified as formulaic since it does not present a

processing advantage as it is clearly not stored and retrieved as a whole in this particular learner. In the same way, this is not an issue as long as idiomaticity is not at the centre of the definition of FS. To sum up, when the researcher proceeds by progressively reducing the subset of candidate FS, a hierarchical approach to identification criteria has consequences on the order in which criteria must be applied. More precisely, if holistic psycholinguistic processing is at the centre of the definition of FS, then identification criteria showing ease of processing must be applied first.

5.4 Elaboration of valid FS identification criteria for the present study

As has been underlined many times, the approach to formulaicity adopted in the present study is psycholinguistic and the key aspect of the definition of FS adopted is the fact that FS are units presenting a processing advantage. As a result, the question to answer is the following: which characteristics have to be fulfilled to show that a given sequence of words seems to present a processing advantage?

5.4.1 *Necessary criterion: phonological coherence*

Although we have pointed out that the criteria established for the identification of FS in the L1 and beginner L2 acquisition literature are not suitable in the context of advanced learners, it does not mean that some of the criteria established for L1 and beginner L2 learners are not relevant to more advanced learners. In this respect, within a psycholinguistic framework, the criterion of phonological coherence, mentioned in numerous L1 and initial L2 acquisition studies, can also be considered of essential importance for advanced learners.

Indeed, although there is no direct way of looking at the underlying mechanisms of language processing, what can be done is investigating the various characteristics of a sequence of words which show ease of processing. The various characteristics showing ease of processing evoked in the literature can be summarised under the term phonological coherence and concern either the temporal aspect of speech (such as fluent pronunciation and acceleration of the articulation rate) or the phonetic aspects of speech (such as coherent intonation contour and phonetic reductions). The key idea behind using the criterion of phonological coherence for FS identification is that external temporal and phonetic characteristics of multiword sequences reflect their underlying cognitive nature.

As pointed out by Dahlmann (2009), apart from fluent pronunciation, most of the other aspects, for example intonation, are very difficult to precisely measure in practice. This is why, when these features have been applied at all for the identification of holistic units, they have been used only in rather small data sets (e.g. Lin & Adolphs, 2009), or as a guidance for intuitive judgements (e.g. Plunkett, 1993; Wray & Namba, 2003) rather than systematically.

With this in mind, in the present study, the global criterion of 'phonological coherence' will only be operationalised through the study of fluent pronunciation (see next chapter) and the remaining characteristics of phonological coherence such as intonation and phonetic reductions will not be studied systematically or measured scientifically. Rather, some aspects, such as phonetic reductions and acceleration of the articulation rate, will occasionally be used as reinforcing factors in the identification process as it is considered that the absence of pauses or other types of interruptions of the speech flow provide a sufficient starting point for the study of phonological coherence, as the next section will show.

5.4.2 *Fluent pronunciation*

As underlined by Temple (2005), the study of the temporal organisation of speech, given that the act of speaking occurs "on-line" can lead to an understanding of the nature of language processing. This is why, in the present study, I will consider than an utterance has to be pronounced fluently in order to be evidence of a processing advantage.

5.4.2.1 *Use of fluent pronunciation in studies adopting a learner-external approach to formulaicity*

Fluent pronunciation has been used as a validation measure in the FS identification process by both Erman (2007) and Dahlmann (2009).

- Erman (2007)

Erman (2007) aimed to investigate (1) whether pausing occurs more in non-prefabricated structures than in prefabricated ones and (2) whether the degree of variability in prefabricated structures is reflected in pausing, i.e. whether structures in which some members can be chosen from a variable set generate pauses of longer durations compared to lexically more fixed structures. To illustrate the second research

question, Erman wanted to compare the length of pauses within, on the one hand, fixed prefabs such as 'to take advantage of' (in which only *advantage* can fill the slot to express this idea) and, on the other hand, prefabs allowing more variability such as *get the credit* (in which *credit* could be replaced by a restricted number of variants expressing the same semantic content such as *acclaim* or *honour*).

In order to answer the above research questions, Erman looked at pause distribution and pause duration in two native speaker spoken corpora of 30 000 words each, in which prefabs had already been identified thanks to the criterion of 'restricted exchangeability'.

The results showed that pausing was significantly more frequent and longer in non-prefabricated compared to prefabricated structures. Erman interpreted these results as showing that prefabs are retrieved as units involving only a minimum of cognitive effort.

In answer to the second question, Erman did not find any difference in pausing depending on degree of fixedness in prefabs. She suggests that all prefabs, regardless of degree of fixedness, seem to demand little cognitive effort compared to non-prefabricated structures. According to Erman, this could be evidence that although the linguistic system allows variation, individual speakers' idiolects may not: given a specific context and situation, they may make preferred choices and use a prefab which is fixed for them. This gives support to Sinclair (1991)'s idiom principle, according to which strings that would appear to be variable in the language nevertheless constitute single choices for individual speakers.

One major limitation of Erman (2007) is the fact that pauses were not measured accurately, relying on transcripts marking short versus long pauses without a clear dividing line, which is methodologically problematic as pointed out by Dahlmann (2009).

- Dahlmann (2009)

The purpose of Dahlmann's study is to work on reliable identification of FS and towards an inventory of FS, which she calls multiword units (MWUs), in spoken discourse in two oral corpora, a native speaker corpus and a learner corpus. She discusses the integration of pause phenomena as a reinforcing identification criterion completing statistically-based criteria. In other words, she first identified candidate FS in the corpora using frequency of occurrences of clusters as a primary criterion. She then looked at how pauses patterned in and around the pre-identified sequences and

what these pause patterns meant in terms of confirming or rejecting identification. The conclusion of her study is that pauses appear to be an informative asset, both with regard to the boundaries of FS and their internal structure. In other words, they are particularly useful in order to determine the start and end point of a sequence (boundaries) and whether it has internal coherence (i.e. when the multiword sequence is not interrupted by a pause).One of the strengths of Dahlmann's study is her highly informed and precise method of dealing with pauses and other hesitation phenomena. She carefully thought about all the instances of pauses or hesitation phenomena that are likely to correspond to acts of speech planning such as silent and filled pauses, syllable lengthening and various repair phenomena. Moreover, she considered pauses as short as 0.2.seconds, one of the lowest cut-off points used in the literature so far.

In the end, due to the extremely time-consuming nature of the work entailed, Dahlmann examined patterns of pauses in relation to only a few FS: *I think*, *I don't know why*, *you know what I mean* and *at the end of the day*. However, her contribution is essential as she provided a very detailed account of pause placement within and around FS and investigated differences in pausing patterns between native speakers and language learners. Her results showed that pauses within the sequences under scrutiny were extremely rare, for native speakers as well as language learners. Her study also showed how the study of pauses could be used to reject frequent sequences as FS. For example, in the language learner corpus she studied, there were 40 instances of the sequence *I think I*. However, nearly half of these instances contained an internal pause after *think* which showed that *I think I* could not qualify as a phonologically coherent sequence. Finally, she found that the most obvious difference when comparing the pause placement patterns in native and learner speech was that native speakers tended to embed FS more in fluent speech than language learners. On the other hand, there was a neat tendency for learner FS to be more clearly delineated by pauses and appear much more prominently as separate entities with clearly defined borders.

Both Erman (2007) and Dahlmann (2009) are particularly relevant for the present study because they deal with the phonological characteristics of formulaic language in language production, whereas, as previously reviewed (see chapter 2), most studies with a psycholinguistic objective deal with the processing of FS by native speakers and language learners, but from the point of view of the reception of FS. By examining the phonological coherence of certain sequences through the study of pauses, they aim to investigate the psycholinguistic validity of a corpus of FS previously identified through

learner-external criteria, either formal (Erman) or statistical (Dahlmann). They both interpreted an absence of pause or any other hesitation phenomena as evidence of absence of speech planning. However, in the framework of the present study, the learner-external approach adopted in both studies is not suitable as both Erman and Dahlmann regard the criterion of phonological coherence (examined through pauses) as secondary in the identification process and do not use it as the criterion to be applied first. As a result, if one followed the identification process used in these studies, one would miss out on some fluent sequences either not displaying idiomatic characteristics (Erman) or not frequent enough (Dahlmann), which would be problematic in this study as L2 learners might use holistic sequences which are not necessarily idiomatic or frequent.

Still, both studies are of paramount importance as they show the scope that the study of pauses' length and especially location can offer for a better understanding of language production. Dahlmann suggests that another possible way forward with regard to the psycholinguistic study of FS through phonological coherence is to approach it directly though fluency, i.e. take pause phenomena as a first step of the identification process, and then study what occurs between pause phenomena. This is precisely the methodological approach adopted in this study.

5.4.2.2 *Fluent pronunciation as necessary criterion*

Approaching FS identification directly through fluency, although rarely done, is not new since fluent pronunciation was the main criterion used by Raupach (1984) in his study of FS (which he calls formulae) in the oral productions of German learners of L2 French. Raupach unambiguously states that his approach is not linguistic but psycholinguistic and he associates formulae with planning units in language processing. He bases his method of identification on Goldman-Eisler's (1964: 116-117) distinction between newly-organised propositional speech and old automatic speech made of ready-made sequences and on her findings that pauses are more likely to occur in propositional than in automatic speech. As a first step, he proposes to list the strings uninterrupted by unfilled pauses and also to consider prosodic features such as intonation phenomena as possible unit markers. He then proposes to break these strings up into smaller segments by considering hesitation phenomena such as filled pauses, repeats, drawls and false starts in order to obtain 'possible candidates for processing units' (Raupach 1984: 117). He points out that other criteria could also be used for a

more detailed analysis such as changes in the articulation rate as well as frequency (defined as learner-internal frequency and not frequency counts in the target language).

There is one main problem with Raupach's method of identification through fluency. Since his identification method is based strictly on prosodic cues, there is a lack of clarity about the difference between fluent run displaying no characteristic of unity and formula unit. Raupach himself admits that 'not all segments produced within the boundaries of hesitation phenomena can be regarded as candidates for formula units' (1984: 119). For example, fluent runs such as *sont plus stricts* (are stricter) or *ne sont pas* (are not) cannot be considered as formula units because they do not have any semantic or functional unity except for the fact that they constitute a fluent run. Moreover, he also adds that in some cases, the fluent segments can easily be broken down into smaller units (ibid). For example, *je crois on peut dire* (I think we can say) can be further broken down into *je crois* and *on peut dire*. To break down speech segments into formula units, Raupach uses the fact that formula units can occur on their own elsewhere in the data so for example he finds several separate occurrences of *je crois* and *on peut dire* (although it could also be argued that *je crois on peut dire* has fused into a single psycholinguistic unit). However, Raupach remains silent on his way of discriminating between fluent runs which are not formulaic and formula units and when he mentions 'supplementary evidence' (1984: 117) needing to be supplied, he does not say which type. As a result, though criteria based on the phonetic and prosodic characteristics of the utterance are essential for the first stage of identification, they are insufficient and need to be complemented by additional criteria showing the holistic dimension of the unit. In this respect, it is interesting to notice that when looking at some of the sequences identified by Raupach, it is clear that he has resorted to additional semantic or functional criteria although he does not elaborate on how he proceeded.

However insufficient and imprecise Raupach's approach might be for our present purpose, his method of marking fluent runs is an effective first step in the process of identification of processing units when dealing with oral speech. Raupach's method raised an objection from Lin (2010) who suggested that the criterion of fluent pronunciation is not suitable for advanced L2 learners. According to her, the speech of advanced learners does not present enough disfluencies for the researcher to be able to isolate processing units within it. However, Lin's objection is undermined by the fact that the types of pauses Raupach recommends to use are very short. He himself used 0.3

second in his study but recommends using even shorter pauses of 0.2 second. Such short pauses cannot simply be equated with disfluencies and are likely to come up very frequently in the speech of advanced learners, as they would even in the case of native speakers (Riggenbach 1991). On the other hand, the absence of such short pauses can be regarded as indicating that a given sequence of words has been processed preferentially. As a result, using the criterion of fluent pronunciation when the pause threshold is as low as the one chosen by Raupach is an effective way of creating a subset of candidate FS and this was chosen to be the first and foremost identification criterion for the present study. This is an essential point to make since, although this criterion is insufficient since one cannot equate a fluent run with a processing unit: it has to be necessarily fulfilled for a sequence to be considered for formulaicity. This means that, although a sequence might fulfil all the other conditions that are about to be described, it will not be considered formulaic if it is not pronounced fluently as this would indicate that it has been put together on-line rather than processed as a unit. To sum up, though in need for additional identification criteria, fluent pronunciation is the necessary criterion for formulaicity as it is defined in the present study. The practical aspects of the identification of fluent runs and the way of dealing with pauses and other types of disfluencies will be detailed in the next chapter.

5.4.3 *Additional criteria*

As underlined previously, additional criteria must be applied on the subset of candidate FS obtained after the criterion of fluent pronunciation has been applied. Indeed, although fluent pronunciation shows ease of processing, all fluent sequences do not necessarily display unity. Consequently, the following question needs answering: amongst all the fluent multiword runs identified in the present corpus, which ones contain one or more processing units, which are not only processed easily but also possess a holistic quality, be it formal, semantic or functional? In order to answer this question, the characteristics that impart a holistic dimension to a given sequence must be identified. To come back to Hickey (1993)'s distinction between necessary and typical conditions, it can be said that on top of the necessary condition of fluent pronunciation, at least one typical condition showing a holistic dimension must necessarily be present for a given fluent sequence to be considered a processing unit.

It must be pointed out at this point that although the distinction between the speaker-external and the speaker-internal approach to formulaicity has been emphasised

previously (Chapter 1 section 1.3.1.) and is crucial to this study, criteria usually associated with a speaker-external approach such as irregularity, idiomaticity or frequency are also relevant to identify processing units within a learner-internal approach as they contribute to give a given sequence its holistic dimension. That is why, as long as they are applied on a subset of fluent sequences only, the following criteria can be used: semantic irregularity, semantic / functional unity as well as holistic mode of acquisition, as illustrated in the following sections.

5.4.3.1 *Irregular sequences with semantic unity*

Firstly, semantically irregular sequences have a holistic quality because their meaning only makes sense when the whole of the sequence is considered. This is the case because for such sequences, the meaning of the whole sequence does not correspond to the sum of the meanings of the parts. For example, as far as the metaphorical idiom *il pleut des cordes* ('it is raining ropes') is concerned, the meaning of the whole expression i.e. it is raining a lot, does not equal the sum of the meaning of its parts which would be that ropes are falling from the sky. In other words, semantically irregular expressions can only be considered in their entirety, which gives them a holistic quality. Moreover, due to their irregularity, these types of sequences are also unlikely to have been generated from scratch and are likely to be not only units showing ease of processing but also single lexical units. To this category can be added highly idiomatic constructions such as *ça m'est égal* (literally 'it is equal to me' = I don't mind) which also have a holistic mapping from form to meaning and which, although they are not strictly speaking irregular, are also unlikely to have been generated productively. In the same way, idiomatic phrasal verbs such as *avoir besoin de* (literally 'have need of' = need to*), avoir hâte de* (literally 'have haste of' = can't wait to), must be included in this category as they are grammatically irregular to a certain extent.

5.4.3.2 *Grammatically and semantically regular sequences with semantic or functional unity or holistic mapping of form to meaning/function*

Sequences do not, however, need to be irregular to display a holistic dimension. What is needed is just a clear semantic or functional unity. For example, this category can include a very wide range of sequences such as expressions to refer to common places (*à l'université* at university, *chez moi* at mine), time expressions (*l'année dernière* last year, *en ce moment* at the moment), expressions to introduce one's opinion (*à mon avis* in my opinion). There can also be multiword NPs referring to a single entity such as

lunettes de soleil (sunglasses). The criterion of semantic/functional unity can also include sequences finding their unity in their function as fillers (*je sais pas* I don't know, *quelque chose comme ça* something like that).

It can be noticed that the types of expressions given as examples also tend to display grammatical unity in the sense that they correspond to a full grammatical constituent such as a nominal phrase (NP) (*l'année dernière*) or a prepositional phrase (PP) (*à mon avis*). However this needs not be the case as what matters is the holistic form-function mapping, even if the form in question is not a grammatical unit as such. For example, a sequence such as *je pense que* (I think that) is made of a verb phrase (VP) and a subordinating conjunction. Nonetheless, it has a holistic quality which makes it a unit because the sequence in its entirety can clearly be mapped to one functional goal which can be described as 'introduce one's opinion'.

5.4.3.3 *Sequences getting their holistic status from the fact that they were learnt as wholes by the learners*

Although every learning experience has a unique quality, if one considers an homogenous group of learners having been exposed to the L2 in a comparable instructional setting, it is reasonable to suppose that some of the input they will have been exposed to will have some degree of similarity and that to some extent, they will all have been taught extremely commonplace sequences that can be described as 'necessary topics' (Nattinger and DeCarrico 1992) such as say your name *je m'appelle x*, ask what time it is *quelle heure est-il?*, say you like going to the cinema *j'aime aller au cinéma*, say the weather is nice *il fait beau*. Knowing the importance, in the British instructional context, of the rote-learning of common classroom routines which are highly formulaic (Mitchell and Martin 1997), many such sequences will have been taught holistically. Given the usage-based theoretical framework adopted in the present study, it can be claimed that, even when the learners have developed the ability to analyse them, these sequences retain their holistic nature.

5.4.4 *The complex status of frequency as an identification criterion*

Frequency of occurrence is one of the most common criteria to be used for identification in studies on formulaicity especially those that are corpus-based. As explained by Wray (2002: 25), the reasoning behind the use of frequency as an identification criterion is that the more often a string is needed, the more likely it is to be stored in prefabricated

form to save processing effort, and once it is so stored, the more likely it is to be the preferred choice when the particular message it expresses needs to be expressed. Since the preferential selection of the prefabricated form will actually suppress the frequency with which any other possible expression of the same message is selected, the contrast in frequency should be clear. When adopting a usage-based framework, as is the case in the present study, it appears that one cannot exclude a criterion based on frequency of occurrence.

However, although the importance of frequency cannot be ignored, its use as a key criterion for identification is not unproblematic. Firstly, one of the potential problems with using frequency as an identification criterion is that one has to make sure that the corpus one chooses to apply the criterion of frequency on is representative of the type of language users one is investigating. For example, if one focuses on formulaicity in L2 learners, then one can only work on a corpus of L2 learners' productions. Since the approach adopted in the present study is psycholinguistic and learner-internal, the only corpus that can be used is that of the learners' own productions since what is focused on is what is formulaic for these specific learners and therefore it would not make sense to check for frequency of occurrences in other corpora, even L2 corpora.

Within such a learner-internal approach, the construct of frequency can still be investigated in two ways: within the same learner (intra-learner frequency) or across all the learners under consideration in the study (extra-learner frequency). Intra-learner frequency, which was used as an identification criterion by Ejzenberg (2000), can be defined as the frequency of a given sequence within the same learner either in the same task or across tasks. In her diagnostic approach to FS identification, Wray (2008) adopts a similar speaker-internal perspective since one of the identification criteria she puts forward is akin to learner-internal frequency. Indeed she proposes to consider a sequence formulaic when "this 'precise formulation is the one most commonly used by the speaker when conveying this idea". As well as intra-learner frequency, it is also possible to consider inter-learner frequency i.e. the frequency of occurrences of a given sequence across learners especially in the context of a study like this one where the group of learners is relatively homogeneous in terms of proficiency and educational experience. This type of frequency is interesting in order to look at the sequences which all the learners are likely to have been exposed to through instruction and are thereby also likely to reproduce holistically. Inter-learner frequency can be related to Wray's identification criterion H in her diagnostic approach (2008: 120). Indeed Wray proposes

to consider a given sequence formulaic when "there is a greater than chance-level probability that the speaker will have encountered this precise formulation before in communication from other people". For example, in the UK context of the teaching of French, learners are all taught holistic sequences such as *il fait beau* (the weather is nice) or *j'aime écouter de la musique* (I like listening to music). For this type of sequence, inter-learner frequency can be applied to show that a given sequence is used by the majority of the learners under scrutiny, even though it is only used a small number of times by each of them. It emphasises the common store of automatised sequences the various subjects have as L2 learners having been exposed to a similar input through instruction.

Another problem with using frequency as an identification criterion is the fact that looking at the most frequent clusters in a given corpus does not allow for formal and semantic criteria to be taken into account. Consequently, one runs the risk of ending up with sequences without any holistic coherence if one proceeds by automatic extraction of frequent sequences in corpora. For example, amongst the most frequent sequences, one could obtain sequences such as *et je* (and I) which are not interesting to analyse as they do not have any formal, semantic or functional unity. This is why, in the present study, rather than looking for the most frequent clusters in the entirety of the data, the decision was made to apply frequency counts only on sequences previously identified through fluent pronunciation and the previously described additional criteria.

An additional problem with the use of frequency as an identification criterion is that of the arbitrariness of the frequency threshold: how many times should a given multiword sequence occur in order to be considered formulaic? It is a crucial issue since there are vast discrepancies across studies regarding the proportion of language that is viewed as formulaic as a consequence of the important differences in the frequency thresholds chosen as pointed out by Wray (2002: 28-29).For example, she refers to Altenberg (1998: 102), according to whom 80% of the London-Lund Corpus form part of recurrent word combinations and could therefore be viewed as formulaic. On the other end of the estimates, Butler (1997) evaluates that repeated phrases represent 12.5% of the spoken part of his corpus of Spanish. As explain by Wray, the reasons for such a lack of consensus lies in the different frequency thresholds applied by various researchers: from 2 occurrences for Altenberg to 10 occurrences for Butler.

The arbitrariness of thresholds is not the only problem with the use of frequency measures. Indeed, according to Wray (2002), raw frequency is simply not an adequate measure of formulaicity. According to her, in order to capture the extent to which a word string is the preferred way of expressing a given idea, we need to know not only how often that form can be found in the sample, but also how often it *could* have occurred. In other words, we would need a way of calculating the occurrences of a particular message form as a proportion of the total number of attempts to express that message. Calculating this kind of frequency ratio would be the only way to compensate for the fact that some messages are much more common than others, although this would obviously be very difficult to carry out practically. Finally some expressions might not be very frequent in spite of their formulaic nature, as is the case of many idioms such as *kick the bucket*. As a result, they might not be extracted by methods relying on a set frequency threshold.

Finally, raw frequency counts can be seen as overly simplistic if one considers, like in this study, that formulaicity can be situated at the more abstract level of the pattern rather than at the lower level of a particular word string. For example, for a certain number of sequences, composed of a fixed part and an open slot, formulaicity was considered to be at the higher level of the pattern rather than at the lower level of the exemplar. For example, in this study, as will be detailed later, the pattern '*c'est* + adjective' was considered formulaic and therefore all the exemplars of this pattern were considered formulaic, even when some exemplars of it such as *c'est gratuit* / *it's free* (2 occurences) or c'est *pratique / it's handy* (2 occurences) occurred much more rarely than other highly frequent exemplars such as *c'est important* (it's important) or *c'est difficile* (it's difficult). Similarly, in this study, the time expressions *le* + weekday or *en* + month are regarded as formulaic at the level of the pattern which means that not only frequent exemplars such as *le dimanche (9)* were considered formulaic but also less frequent ones such as *le lundi* (2). As a last example, verb phrases working as patterns such as *prendre* (take) + means of transport such as *prendre le train* (2) or *prendre l'avion* (1) must be dealt with in this way.

Because of all the limitations mentioned above, the decision was made, not to discard frequency as a criterion, but to use it as a reinforcing rather than necessary feature of processing units. In other words and to use Hickey (1993)'s terminology, frequency is considered as a graded criterion and not a necessary one: from the moment a sequence displaying a holistic quality (for one of the reasons given above) is used fluently, it is

reasonable to assume that it could have occurred frequently, given a suitable context. Nevertheless, since frequency is considered a graded criterion, the more frequent a unit is within the same learner or across learners, the more reliably its status as a processing unit will be considered.

5.5 Summary of the identification method used in the present study

As pointed out by Wray (2009: 40), as far as the difficult process of FS identification is concerned, 'one way forward is to take decisions, but remain vigilant and reflective about what they assume and entail.' Following Wray's advice, the identification method proposed in the present study can be summarised as follows:

1. Necessary criterion, applied first on the data in order to obtain a subset of candidate FS: Fluent pronunciation of the multiword sequence

More precisely, fluent pronunciation means that the multiword sequence is pronounced: without filled or unfilled pauses longer than 0.2 second, without any syllable lengthening and it does not contain any repetition or retracing. In the analysis of the data, fluent pronunciation can also go hand in hand with phonetic reductions or phenomena such as liaison and finally, some sequences might display an acceleration of the articulation rate. Some aspects complementing fluent pronunciation, such as acceleration of the articulation rate, phonetic reductions and liaisons will occasionally be taken into account and commented on but this will remain at the level of confirmatory evidence. These additional aspects however, will not be systematically investigated in the present study. More technical details about the concrete handling of the data will be given in the next chapter.

2. Necessary additional presence of one typical criterion showing the unity of the sequence: either 1) grammatical or semantic irregularity, 2) holistic form-meaning/function mapping or 3) likely presence of the sequence in the input received by the learners through instruction

As previously explained, because the identification method used in this study is hierarchical, this second criterion is only applied on the subset of fluent sequences obtained after the first step of the identification process.

3. Learner-internal frequency (frequency of occurrences of a given sequence within the same learner and / or learner-external frequency (frequency of occurrences of a given sequence across the learners). Frequency is not considered a

necessary criterion but is regarded as a graded criterion i.e. as strengthening the case for formulaicity in the identification process.

More details about the practical steps carried out as part of the identification process, as well as some problematic aspects, are outlined in the next chapter, which deals with the methodological aspects of this study.

Chapter 6. Methodology

The present chapter has three main objectives: to describe in detail how this study was carried out, to explain how the data collected was analysed and to account for all the methodological choices that were made. For clarity of purpose, it is important to summarise the main aims of this study:

1) Investigating FS use and its longitudinal development in English advanced learners of French
2) Analysing the relationship between FS use, fluency and lexical diversity.

For ease of reference, the various research questions outlined in chapter 2 and chapter 4 are gathered below.

1) FS use in advanced learners
    - How far do advanced learners of French use FS?
    - Does the type of task used have an effect on FS use?
    - What types of FS do they use?
    - Are there individual differences between subjects in terms of quantitative and qualitative FS use?
    - What are the implications of FS use in L2 learners for our understanding of language and the status of FS within it?
2) Longitudinal development of FS use, lexical diversity and correlation between the two
    - How does the use of FS by advanced learners of French develop during a seven months' stay in France?
    - How does their lexical diversity develop?
    - How does FS use develop in relation to lexical diversity? Is there a correlation between the two?
3) Longitudinal development of fluency and relationship between FS use and fluency
    - How does the fluency of advanced learners of French develop during a seven months' stay in France?
    - Does the development of FS use by L2 learners play a role in the development fluency during the year abroad? → Quantitative analysis
    - If it does, how does it take place? → Qualitative analysis
    - Which underlying mechanisms can explain this role?

4) Is there a relationship between the learners' engagement with the French language during their stay abroad and their development in FS use, lexical diversity and fluency?

Before explaining, in the second section of the present chapter, how the data collected for this study were analysed, the first section will present how the study was designed in order to address the research questions outlined above. Firstly, it will present the subjects who were investigated. Secondly, it will describe and account for the longitudinal learner-internal design chosen. Finally the tasks used in order to collect the data will be presented.

6.1 Study Design

6.1.1 *Participants*

Since this study investigates the use of FS by advanced English-speaking learners of French before and after a stay in France, I recruited second-year-university students of French who were about to go to France as part as their year abroad. Indeed, this type of student fitted the two requirements of being at an advanced level as well as being about to spend several months in France.

The participants were 5 native speakers of English who studied French as their degree subject at Newcastle University in the UK: Iris, Lily, Lola, Rose and Sally[4]. In the UK, a language degree lasts for 4 years, with a compulsory year abroad in the third year, which can be spent in one or two different countries depending on the degree programme. At the first time of data collection, the participants were in the last month of their second university year. At the second time of data collection, they were in the seventh month of their time in France. They were recruited on a voluntary basis during their second university year. There was a small financial incentive in order to compensate for their time, but on the whole, their volunteering was interpreted as a sign of motivation. Moreover, 2 participants studied linguistics as part of their degree and were enthusiastic about the idea of contributing to a research project in second language acquisition, though the precise aim of the study was not disclosed to the subjects.

The background in French of the 5 subjects was fairly homogeneous as they were all second-year students on the same degree programme and had gone through the same educational system. It is summarised in Table 2.

[4] The names chosen are pseudonyms in order to respect the participants' anonymity.

Table 2: Background of the five participants

	IRIS	LILY	LOLA	ROSE	SALLY
Number of years of Fr at primary school	NA	NA	4 years	NA	2 years
Number of years of Fr at secondary school	7 yrs	7 yrs	7 yrs	7 yrs	7 yrs
Length of stays in France prior to the year abroad in France	1 wk (Fr exchange), numerous family holidays but with little communication in Fr except with her mother (Fr teacher)	10 wks: 4 wks (2X2) (Fr exchange) + 6 wks (3X2): family holidays but with little communication in Fr	16 wks: 10 wks (school exchanges and stays with Fr families as an au pair) + 6 wks of family holidays but with little communication in Fr	6 wks (family holidays but with little communication in Fr)	1 wk with a Fr family Numerous family holidays but with little communication in Fr
Contact with native speakers of Fr during first 2 yrs at university	Minimal	Weekly conversations with Fr Erasmus Students	Weekly conversations with Fr Erasmus Students	None	Weekly conversations with Fr Erasmus Students
Input additional to university seminars	Minimal, occasional film or magazine	Quite frequent: films, magazines, French radio	Very little, except watching films	None	Some listening and reading but very limited
Other subjects studied at university	Economics	Linguistics	None	Law	Spanish and linguistics

As can be seen from Table 2, the participants were all girls who studied French for seven years in secondary school (age 11 to 18 years old). Three of them also received minor exposure to French at primary school. Though the participants all spent some weeks in France either for family holidays or as parts of school exchanges, there were some differences in the quantity of extra-curricular input they received. For example, Iris spent more time in France as part of her family holidays and sometimes spoke French with her mother who is a French teacher. In terms of French input in addition to their university classes, it can also be seen that there were some individual differences in the steps taken by the subjects to be exposed to the French language as Lily, Lola and Sally made regular contacts with French Erasmus students.

Despite the similar linguistic background of the participants and the fact they could all be broadly categorised as advanced learners, there were some differences in their proficiency level, as reflected by the different grades they obtained for the French language exam they undertook just before the time of data collection at the end of their second year at university: Sally and Lily got a first (78/100 and 71/100 respectively), Iris got a high 2.1 (67/100), and Lola and Rose got a 2.2 (58/100 and 56/100 respectively). These differences in proficiency confirm Tremblay (2011)'s claim that identical proficiency should not be assumed because of same level of study. These proficiency differences were not deemed problematic however. Indeed, they did not affect the methodological validity of the study as, the design being longitudinal, the same students were compared at time 1 and time 2. On the contrary, these differences ensured the ecological validity of the study as the role of the year abroad could be observed on a representative panel of second-year undergraduates in French.

6.1.2 *Longitudinal learner-internal study design*

6.1.2.1 *Rationale*

The present research project is an in-depth case study of 5 English-speaking advanced learners of French with a longitudinal pre-post study abroad design.

The design adopted for this study is longitudinal since it compares FS use, fluency and lexical diversity in the same advanced learners of French before and after a stay in France as part of the year abroad. This longitudinal design was chosen over a cross-sectional one for one main reason: as previously explained, at the core of this study is the idea that a FS is a learner-internal psycholinguistic construct. As a result, different

individuals are likely to have and use a different repertoire of FS, at least to a certain extent. Therefore, since FS are learner-specific, their potential development after a sustained period abroad can only be documented in the same individual learners, and a longitudinal design is the only way of ensuring such learner-internal coherence.

The study abroad context has been shown to be particularly prone to the development of fluency (Towell et al. 1996, Towell 2002, Freed et al. 2004; for more details on these studies, see chapter 4). Since one of the aims of this study is to investigate the link between FS use and fluency, the year abroad context was therefore chosen as the locus for this study and a pre and post study abroad research design was adopted. The study abroad context was also thought to be the context in which subjects were likely to receive the richest input from both a quantitative and qualitative perspective and would therefore be more likely to show development in terms of lexical diversity and FS use (Foster 2009, for more details, see chapter 4).

6.1.2.2 *Data-collection schedule*

Time 1

The first round of data collection took place at the end of the subjects' second year at university, in June 2009. This time was chosen so that it would be as close as possible to the end of the university year so that any development noticed at time 2 could reliably be traced back to their time spent in France and not to the final months of their second year at university.

Time 2

The second data collection session took place 10 months later i.e. at the beginning of April 2010. Since the subjects started their stay in France in September 2009, they had been in France for 7 months then. Data collection was carried out during a two-week field trip in order to visit the students where they were staying. Although the organisation of the field trip was complex, it was thought to be a much better solution than waiting for the subjects' return to university for their final year. Indeed, if time 2 of data collection had been postponed to the start of the students' final university year, the development potentially occurring during the year abroad might have been lost due to the length of time elapsed since the end of the stay in France, all the more so as some of the learners spent the rest of their year abroad in another foreign country such as Spain or Germany. In addition, if one supposes that immersion in the L2 environment is likely

to encourage the use of well routinised exchanges which are likely to be formulaic, one might also anticipate that their automatised nature might decrease quickly when not used in the language daily. As a result, interviewing the subjects at the end of their stay in France was considered a way of both maximising their length of exposure in the immersion context and testing them before they were removed from it. Finally, since one of the secondary aims of this study is to investigate the relationship between the subjects' linguistic development and their involvement with the French language, visiting the students in France was thought to be a good way of completing the information gathered in the language-contact questionnaire (see below) by getting a more qualitative insight into the way they immersed themselves in the French environment.

Time 3

Questionnaire (See 6.1.3.4. and Appendix A4)

At the end of their stay in France, the participants were asked to fill in a questionnaire about their use of French before and during their stay in France.

6.1.3 *Tasks and rationale*

Oral data were the only type of data judged suitable to answer the research questions of this research project. Firstly, spoken language, as opposed to written language, possesses some key characteristics that make it particularly suitable for the study of FS. As summarised by Miller and Weinert (1998: 22), spoken language is 'produced in real time, impromptu, and with no opportunity for editing'. As such, it is 'subject to the limitation of short-term memory in both speaker and hearer'. Because of these characteristics, spontaneous spoken language is the ideal locus to investigate the presence of FS in speakers. Indeed, due to the conjugation of the demands of on-line production and the processing limitations of the human brain, it is reasonable to assume that speakers are likely to resort to FS as much as they can when producing spontaneous speech in order to lighten the processing burden. Moreover, as previously explained, this study aims to establish how the use of FS contributes to the speeding up of speech production processes (Kormos 2006). However, the only way of better understanding the psycholinguistic mechanisms underlying speech production (Segalowitz 2010, see chapter 4 for further details) is through the detailed study of the external characteristics of speech, hence the necessity for oral data.

The study has a learner-internal focus in the sense that it aims to get an in-depth insight into the use and development of FS within individual learners. In order to get such a detailed insight, in addition to a longitudinal research design, it was necessary to collect a considerable amount of data per learner for several reasons. Firstly, the linguistic samples obtained had to be long enough to be representative of a given learner's speech. Moreover, although in this study, frequency of occurrences is not the main identification criterion of FS, frequency is still used as a graded criterion (see chapter 5). In other words, the more a given sequence occurs, the more it is likely to be formulaic. As a result, a greater amount of data makes the identification of FS more reliable. Additionally, one of the research aims is to use the construct of FS as a way of tapping into the issue of individual differences in language use and linguistic idiosyncrasies such as the preferred choice of a specific sequence in order to express a given semantic content. The investigation of such idiosyncratic lexical preferences is only possible if an important amount of data has been collected. Finally, with the quantitative analysis of the data in mind, the potential changes in the investigated variables between time 1 and time 2 are much more likely to be significant if a sufficient amount of data has been collected. Because of all these reasons, a design enabling an in-depth examination of a small number of learners was favoured over a more superficial examination of a higher number of learners as it was the only way of fully addressing the research agenda of investigating the use and development of FS within individual learners. This design makes this study stand out from other longitudinal studies with a pre/post study abroad design. For example, Freed et al. (2004) dealt with a higher number of learners of French (28) but analysed only 2 one-minute samples from each learner at both time 1 and time 2, in other words a total of 4 minutes per learner. By contrast, this study only focuses on 5 learners but analysed in detail an average of 50 minutes of each learner's oral productions at time 1 and time 2, i.e. a total of 100 minutes of speech per learner.

As previously mentioned, this study focuses on what learners usually say, rather than on what learners could say. As a result, and as will be noticeable from the presentation of the individual tasks below, all the tasks were designed in order to encourage spontaneous speech production, at least to the extent it can be spontaneous in an artificial data-collection setting. The topics selected were considered easy to talk about both from a conceptual and lexical perspective and it was therefore anticipated that the subjects would be able to speak at length about them as well as possess the necessary

vocabulary to do so. In other words, the objective was not to push the learners to produce complex utterances but to elicit common and FS-rich language from them.

Moreover, as will be explained in more detail in the description of each task, the tasks were chosen in order to ensure a compromise between, on the one hand, the need for a minimum of standardisation in order to allow comparisons across times and across learners and, on the other hand, the wish to elicit language as spontaneous and natural as possible in an artificial setting of data collection.

Each learner undertook five oral tasks: a general oral interview, a story-telling task and three discussion tasks about topics of general interest. Apart from the three discussion tasks which were carried out in pairs in the first round of collection, all the tasks were carried out on a one-to-one basis with the researcher. The same tasks were used at time 1 and time 2 in order to enable comparisons between the two times. The long time gap between the first and second time of data collection (10 months) was thought to be sufficient to avoid too much facilitative effect of having performed the task before, although such an effect cannot be avoided altogether. All the students had met the researcher a few times before the start of data collection so that they would all feel comfortable and relaxed when they were asked to perform the tasks for the first time.

All tasks were recorded digitally, and were of variable length. On average, the interview took 25 minutes, the story-telling task 4 minutes and each discussion task approximately 8 minutes each, making a total of a little less than one hour of recordings per subject at both times of data collection i.e. a total of nearly 1 hour forty minutes of spoken data per subject across both data collection times (including questions and comments by the researcher) i.e. a total of around 8.5 hours of oral data.

6.1.3.1 *General interview*[5]

This task was a one-to-one interview with the researcher and was chosen for several reasons. First, it aimed at generating speech about common themes such as the participants' personalities, their tastes and hobbies, their families, their past and present studies, as well as their plans for the future. The type of interview carried out was similar at time 1 and time 2 but there was some variation in the kinds of topics talked about with a focus on personal information at time 1 and on the subjects' experience in France at time 2. All the topics mentioned were assumed to be easy to talk about as well

[5] See appendix A.

as generate speech potentially rich in FS. Indeed as pointed out by Foster (2001: 79), the familiarity with the subject matter of the task is an important determinant in how many FS will be resorted to: 'the more familiar we are with the subject matter, the more likely it is that our memory will contain relevant ready-to-use language and the faster we are able to process it'.

Moreover, an interview was thought to be a satisfactory compromise between the need for controlling the task in order to enable comparisons across learners and across times and the will to elicit spontaneous language. As a result, although the learners were all asked a common set of questions at time 1 to ensure a degree of homogeneity, additional questions were added in response to some of the participants' answers in order to ensure a natural and dynamic flow of conversation. As mentioned above, the slight variation in the questions asked between time 1 and time 2 was unavoidable in order for the content of the interview to be relevant.

Finally, this general interview enabled the researcher to get some rich information about the participants' background, profile and relationship to the French language. Although the relationship between FS use and factors such as input and motivation is not part of the main research questions of the present study, it was thought that collecting background information on the participants could be useful in order to account for potential individual differences in FS use in addition to the questionnaire they were asked to fill in.

6.1.3.2 *Discussion tasks[6] on topics of current interest*

The 3 discussion tasks were structured along the same format. The learners were given a sheet showing (in French) a topic of current interest (alcohol, obesity and the reasons for learning a foreign language), together with four suggested measures to tackle the issue in question. They were asked each time to debate the pros and cons of the suggested measures, to decide which one (s) they thought were likely to be the most or the least effective and to propose further suggestions in relation to the issue under consideration. The learners were then free to make additional comments about the topics.

[6] See appendix A. These three tasks were taken (with permission) from the FLLOC (French Learner Language Oral Corpora) project, http://www.flloc.soton.ac.uk/.

In the same way as for the interview, the tasks were controlled to some extent while still allowing for the elicitation of relatively spontaneous speech. In more detail, the 3 topics selected were the following: Evaluate the best measures to reduce alcohol consumption by young people, evaluate the best measures to fight obesity amongst young people and evaluate the best reasons for learning a foreign language and French specifically.

Similarly to the general interview, the topics evoked were likely to be familiar to the participants, which might lead to the production of FS. Moreover, the participants were asked to evaluate different ways of responding to the various issues. For example, with regard to the issue of alcohol consumption amongst teenagers, the learners were asked to rate the following four measures: 1) Give a more sever fine to shops that sell alcohol to teenagers 2) Raise the price of alcoholic drinks 3) Inform young people about the long-term effects of alcohol 4) Lower the legal age for alcohol consumption to 16 years old. They were also told that they were free to add further suggestions and to discuss the topic more generally. Since it was thought that the subjects would be familiar with this argumentative style of exercise, it was anticipated that these tasks would generate FS to introduce one's opinion as well as expressions of agreement or disagreement.

6.1.3.3 *Story-telling task from a picture story book*[7]

The subjects were asked to tell a simple story based on a picture story book about a family spending a holiday at the Loch Ness. This task was chosen for several reasons. Firstly, it is more controlled than the other tasks used in this study since it is constrained by the story line. As a result, it makes comparisons of the different variables between time 1 and time 2, but also across learners, easier and more reliable. Moreover, because the task is both descriptive and narrative, it was used in order to elicit data of a different type to see whether it would lead to FS specific to narratives and descriptions with the purpose to examine whether FS use might vary across different genres. Although more constrained than the other tasks used in this study, this story-telling task was not challenging for the subjects. Indeed, contrary to some story-telling tasks which can be very challenging for the learners especially with regard to the type of vocabulary needed, this task was considered relatively easy from a lexical perspective as it was initially designed by the FLLOC project to be able to be done by learners of all levels, including beginners. It was anticipated that some of the story's pictures would elicit common FS

[7] See Appendix A. Like the discussion tasks, the story-telling task was the same as the one used for the FLLOC project.

learnt as school such as *partir en vacances* (to go on holiday), *faire du shopping* (to do some shopping to go shopping), *être en colère* (to be in anger to be angry), *avoir peur* (to have fear to be scared), etc.

6.1.3.4 *Questionnaire*[8]

The questionnaire, which was administered at the end of the learners' stay in France, was a shortened and simplified version of the Language Contact Profile used by Freed, Dewey, Segalowitz and Halter (2004). The Language Contact Profile was designed to develop a clearer sense of how students spend their time, with whom they interact and in what language, in various study contexts amongst which the study-abroad context.

The adapted version of the questionnaire used in this study contained three different sections. The first section was about background information on the participants, especially in relation to their use of French before going to university. The second section dealt with their use of French at university before their stay in France. The third section was concerned with their time spent in France, as well as questions about the quantity of input they received. It also contained more qualitative questions about their motivation and their qualitative evaluation of their time in France. The questionnaire was made simpler and shorter than the Language Contact Profile for two main reasons. Firstly, the students' engagement with the target language is not the main research question of this study. Consequently, rather than serving a detailed study of the input received by the subjects in France, the aim of the questionnaire was simply to gather general information about the subjects' quantitative and qualitative use of French in the eventuality that some differences in their exposure to French and engagement with their time abroad might contribute to explain differences in their linguistic development and in particular their use of FS. Secondly, the original Language Contact Profile, although extremely precise, has the disadvantage of being very long and repetitive. This is why it was feared that filling in such a long questionnaire would be tedious for the participants who might be tempted to rush and not take the time to answer the questions as accurately as necessary.

6.2 Methodology used for the analysis of the data

The following section has 2 main objectives. Firstly, it describes how the sound files were transcribed and annotated in order to answer this study's research questions.

[8] See appendix A.

Secondly, it presents the measures that were used to operationalise the variables investigated in this study: formulaicity, fluency and lexical diversity.

6.2.1 *Transcription and annotation*

6.2.1.1 *Identification of fluent runs*

As was described in the previous chapter, fluent pronunciation is considered a necessary identification criterion in the present study and is therefore the first to be applied. In other words, a given multiword sequence must imperatively be pronounced fluently to be considered a potential FS. But how can 'fluent pronunciation' be defined and operationalised? In order for a given sequence to be considered fluent, it has to be pronounced as an uninterrupted whole. In this study, the absence of interruption implies several requirements: the sequence should not be interrupted by a pause, a drawl (lengthening of a syllable), a word repetition or a rephrasing of the message.

Interruption by a pause

As described in chapter 4, when dealing with pauses, two main questions need answering: the types of pauses to be taken into account as well as the minimum length of a pause. In this study, silent and filled pauses are both considered signs that processing is not automatic and as a result, a given multiword sequence cannot be regarded as an FS if it is interrupted by a pause. A silent pause is an interruption of the normal speech flow, meaning that there is no acoustic signal. Although there are various ways of defining filled pauses (see chapter 4), in the present study, they will be limited to non-lexical fillers such as *euh*, *uhm* etc..., which are not recognised as words and contain no semantic information. Some researchers such as Riggenbach (1991) have listed lexical fillers such as *you know* or *I mean* as types of filled pauses because although they are words, they contribute hardly any semantic information. However, in the present study, these types of fillers will not be considered as pauses for several reasons. Firstly, although they do contribute little semantic information, they still have a semantic content and should not be treated in the same way as non-lexical fillers. Moreover, lexical fillers have been shown to be commonly resorted to by L2 learners as communication strategies when they are trying to produce speech and their role needs to be studied carefully within the study of fluency. Finally and very importantly, as will be shown in section, many lexical fillers can be regarded as FS so lexical fillers are part of the main research question being investigated in this study.

As previously explained, in this study, pauses are not just considered as evidence of disfluency but are also regarded as windows into the workings of on-line processing. In other words, a pause is considered interesting not just because it is a sign that a learner is struggling with expressing a given semantic content. It is also regarded as relevant for the discrimination between what seems to be processed automatically or not automatically. In other words, in this study, pauses play a central role in the identification of FS. For this second psycholinguistic purpose, one should choose a cut-off point as low as possible because even very short pauses can be considered as showing that a sequence is not automatic. However, as explained in chapter 4, if the cut-off point chosen is too low, normal pronunciation phenomena such as for example the stop phase of geminated plosives might be misidentified as pauses. This is why the minimum cut-off point for a pause chosen for this study is 0.2 second, as was recommended by Raupach (1984) and used by Dahlmann (2009). This cut-off point is lower than many thresholds previously used in the literature. For example, Freed et al. (2004) used 0.4 sec, Towell et al. (1996) and Towell (2002) used 0.28 sec and Wood 2010 used 0.3 sec.

To conclude, 0.2 was chosen as it was considered the best compromise between, on the one hand, identifying pauses as short as possible in order to notice any evidence of on-line processing and, on the other hand, avoiding the risk of misinterpreting normal pronunciation phenomena as pauses.

Interruption by lengthening of the syllable

As described in chapter 4, phenomena of syllable lengthening (also called drawls or sound stretches) are generally considered as pauses in the literature. As rightly pointed out by Dahlmann (2009), the problem with syllable lengthening is that it is quite difficult to measure, as a normal syllable/sound length initially needs to be defined in order for an extended sound to be identified. Despite the importance of sound stretches, these methodological difficulties have meant that procedures for their identification have received little attention to date. For example, in Towell et al. (1996) and Towell (2002), although syllable lengthening was marked on the transcripts, it was not considered as a pause or measured. Dahlmann (2009) suggests the following possible practical identification and measurement method. Referring to Zellner (1994: 47), she states that the duration of a typical syllable is around 0.2-0.3 second. Therefore, to be considered stretched the whole syllable has to be considerably longer, and she

accordingly proposes a possible cut-off point could be 0.4 s. The present study will use the same method of measuring sound stretches. It would not be feasible to measure all the syllables to check they are pronounced in less than 0.4 second. However, this is not a problem as a sound stretch is easy to notice by ear so that measurement can only be carried out on the syllables which appear to be remarkably long. Therefore the measurement of sound stretches in the present study was realised in two steps: first, a lengthening of the syllable was noticed and then it was measured. In terms of the marking of fluent runs, a sound stretch was considered as marking the end of a fluent run, which implies than any syllable pronounced after it was treated as the start of a new run. If the sound stretch was 0.4 sec then any speech after it was marked as a new run. If the sound stretch was longer than 0.4 sec, the remaining of the sound stretch time after 0.4 sec was marked as pause time until the start of the next speech run.

Interruption by the repetition of a word or by a repair

In the present study, the repetition of a word is considered as an interruption of a fluent run and seen as the start of a new run even when there is no pause between the repetition and the word repeated. For example, a sequence not interrupted by any pause but containing a repetition such as '*c'est c'est intéressant*' '(it's it's interesting') is regarded as two fluent runs '*c'est*' and '*c'est intéressant*' ('it's'+'it's interesting'). In the same way, any retracing of the message is also seen as the start of a new fluent run, even when there is no pause. For example, a run such as *c'est* // *il y a une famille* ('it's // there is a family') was regarded as two runs '*c'est*' (it's) and '*il y a une famille*'('there is a family'). This way of taking into account runs and retracing in the delimitation of runs is common practice and has been used in previous studies on fluency, though not all. For example, Towell et al. (1996) did not take into account repetitions or repairs in the marking of fluent runs. However, Freed et al. (2004) looked at repetition-free and grammatical-repair-free speech runs and in her micro-analysis of fluency, Riggenbach (1991) gives as much importance to repair phenomena (such as repetitions and false starts) as to hesitation phenomena such as silent and filled pauses.

6.2.1.2 *Choice of software*

One of the main objectives of this study is to investigate the longitudinal development of FS use in relation to that of fluency and general lexical development. As a result, the aim of the data analysis was to have full transcripts of the learners' speech but also to identify fluent runs (i.e. runs uninterrupted by pauses or other hesitation and repair

phenomena), to count the number of syllables per fluent run, to identify FS within the fluent runs and to count the number of formulaic syllables within each run.

Because of these multiple requirements, the software Praat was selected to transcribe the data (http://www.fon.hum.uva.nl/praat/). Praat (which means 'talk' in Dutch) is a free software program created by Paul Boersma and David Weenink of the Institute of Phonetics Sciences of the University of Amsterdam. It was designed for the phonetic analysis of speech and the reconstruction of acoustic speech signals. Praat was selected to deal with the data as it allows for the annotation of sound files in a way tailored to the aims of this study. Firstly, in the same way as Adobe audition, it is an excellent tool for the precise measurement of pauses thanks to the easily handled zooming device which enlarges the display of the sound wave, thus enabling very short pauses to be measured. However, it was more adapted to the present study than Adobe Audition as, thanks to it, a sound file can also be annotated by adding as many tiers as needed. This was highly convenient since different tiers could be used for different purposes such as marking pauses and fluent runs, transcribing utterances and identified FS orthographically and counting syllables. In this way, the entire annotation of the sound file (i.e. pause measurement but also orthographic transcription and annotation of FS) could conveniently be done on the same file.

6.2.1.3 *Annotation of sound files*

There were 5 sound files per learner per time of data collection, that is to say 10 files per learner and a total of 50 files to annotate. Each sound file was annotated following the same methodology. Figure 2 displays an example of a Praat screen which shows 15 seconds of an annotated sound file, taken from Iris French 2.

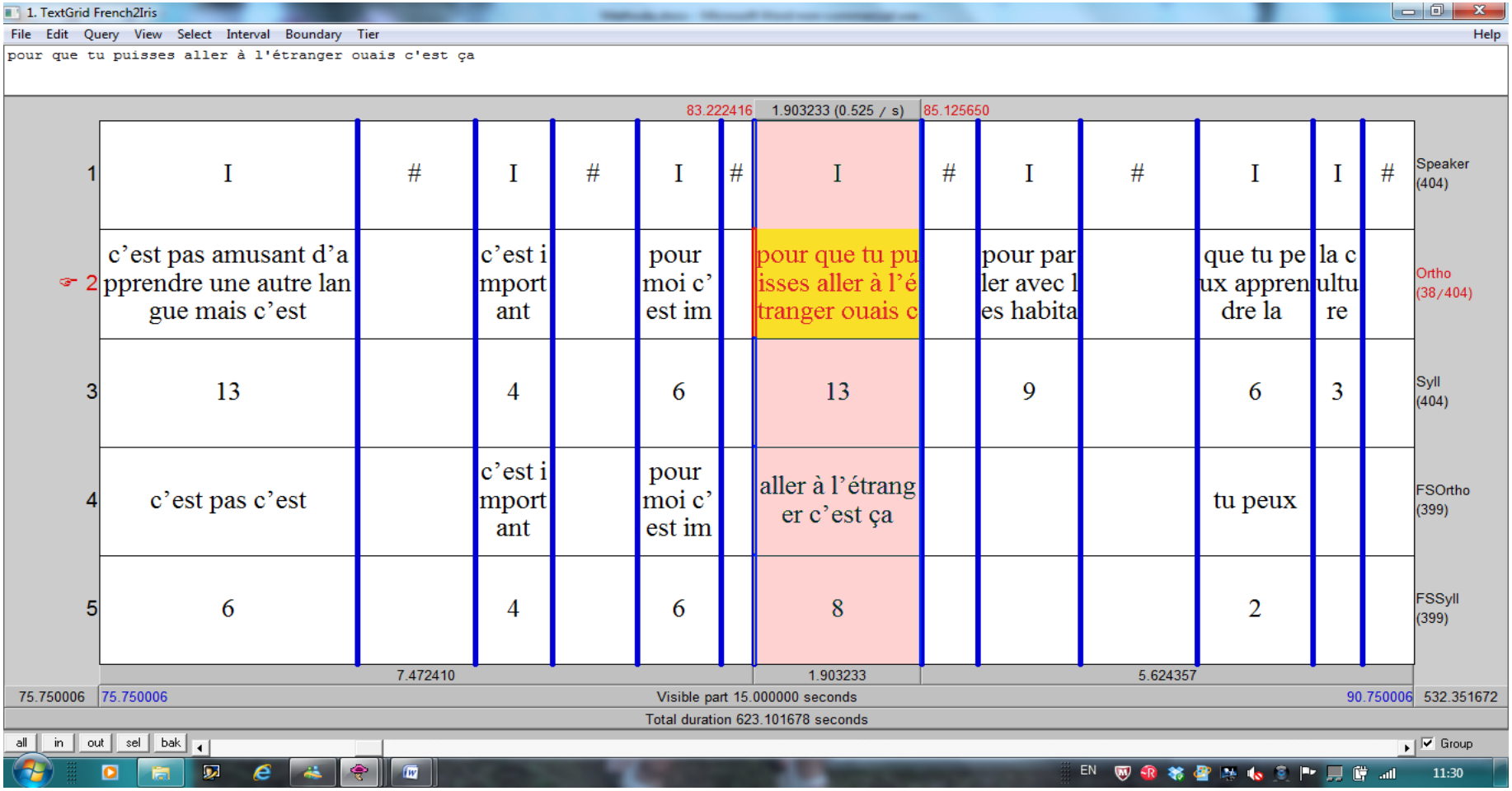


Figure 2: Example of an annotated Praat script (visible part= 15 seconds of the sound file)

Tier 1

The first tier was used to mark pauses, runs of fluent speech as well as irrelevant elements to be discarded from measurement (see below). As illustrated on Figure 2, the initial of the speaker was used to mark fluent speech (I for Iris on Figure 2), pauses were marked with the symbol # and irrelevant elements discarded from measurement were marked with the symbol *. As previously explained, the minimum cut-off point used was 0.2 seconds: all the pauses as long as or longer than 0.2 seconds were marked on each file. Silent and filled pauses were marked in the same way since, as pointed out earlier, what matters is the fact that a run is interrupted rather than the type of pause. Figure 3 illustrates the zoom-in function of Praat, which enables for a very short section of the sound file to be selected (for example 3.75 seconds as shown by Figure 3), thereby allowing for the very precise marking of pauses. Although Praat possesses a function allowing for the automatic annotation of pauses, this function could not be used for this study for two reasons. Firstly, although the sound files were recorded in very good auditory conditions, they were still not laboratory conditions and as a result, the sound files were not clean enough for the automatic function to be able to work precisely. Secondly, since filled pauses were considered as pauses, they needed to be annotated manually. Consequently, all the pauses, silent and filled, were marked manually. In the case of several pauses in a row e.g. one silent pause followed by a filled pause, only one pause was marked as again, what was considered important was the fact that fluent speech was broken.

The speech stretches between pauses were sometimes further cut down into shorter fluent runs since repetitions and retracing were considered as signals for the start of a new run, as were lengthened syllables. As a result, a fluent run was not necessarily always followed by a pause as there could be several consecutive runs prior to a pause.

Sound material considered irrelevant to the calculation of either speech or pause time was also marked. Therefore, questions or comments by the researcher, laughs, sentences uttered in English were all marked as irrelevant phenomena to be discarded from measurement.

Moreover, as prescribed by Riggenbach (1991), pauses over 3 seconds were marked as long pauses not to be taken into account in the calculation of pause time. When dealing with the data, Riggenbach's prescription really made sense as all the cases of long pauses corresponded to moments of communication breakdown or times where the subjects had nothing left to say about a given topic.

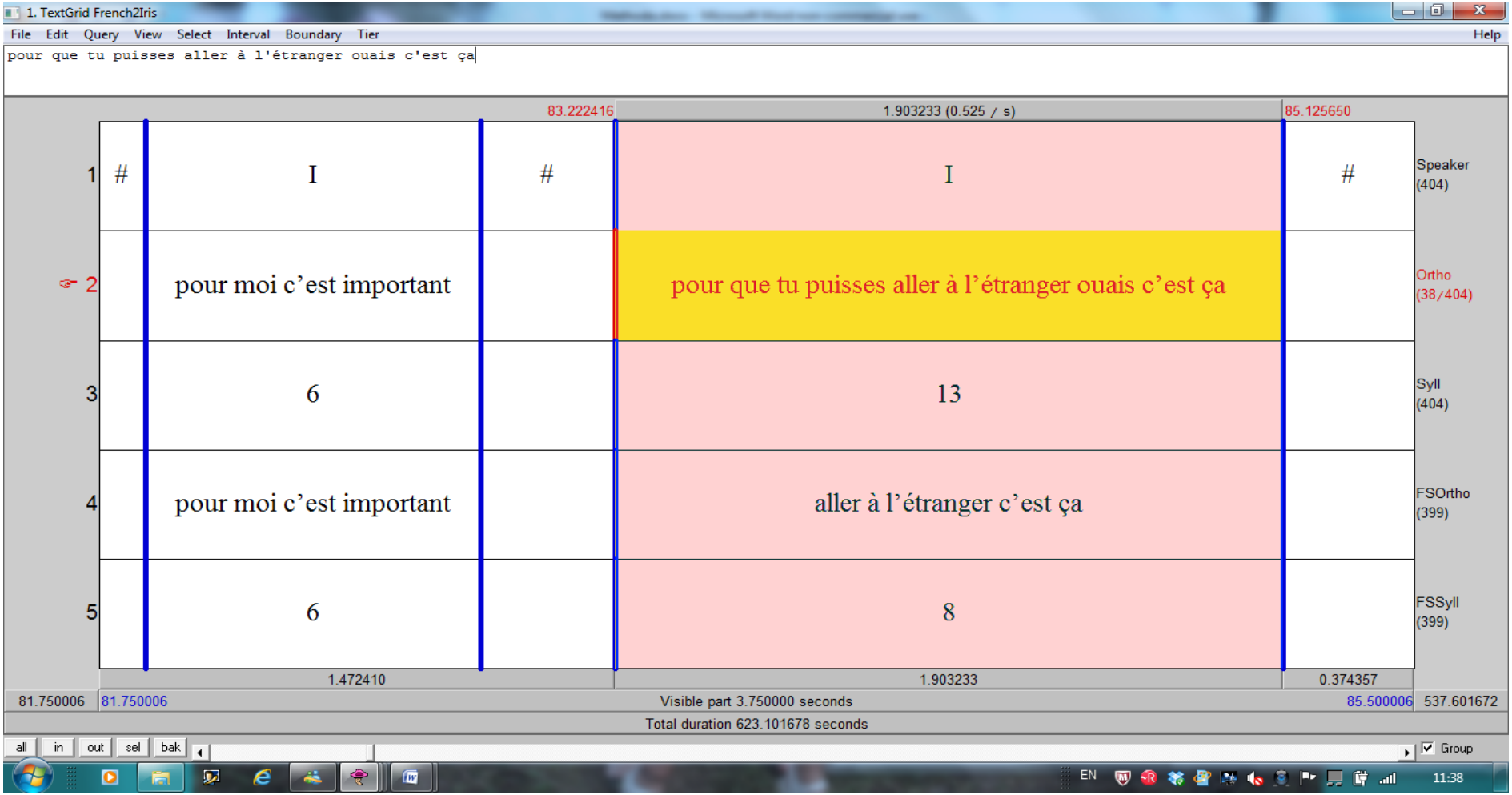


Figure 3: Same annotated Praat script (visible part 3.75 seconds)

In this respect, some pauses shorter than 3 seconds were also excluded from the calculation of pause time when it was clear that they did not correspond to any disfluency. For example, in the story-telling task, the pausing time spent turning the pages was excluded. In the interview as well as the discussion tasks, pause time between the end of the subject's utterance and the researcher's new question was also considered as irrelevant.

Tier 2

The second tier was used for orthographic transcription, which was carried out without any additional coding. Speech containing lexical or grammatical errors was transcribed in the same way as correct speech and errors were not coded as they are not the focus of this study. This orthographic transcription was then transferred onto a text file thanks to a Praat script written for that purpose.

Only the sound files collected at time 2 were orthographically transcribed on Praat as the first half of the files had already been transcribed in CHAT format, before the decision was made to use Praat for the treatment of the data. CHAT (http://childes.psy.cmu.edu/manuals/CHAT.pdf) is the transcription software of the CHILDES software (http://childes.psy.cmu.edu/). It is the standard transcription software in L1 acquisition and has also been used in L2 acquisition (Myles 2008). It follows specific guidelines and is compatible with the CLAN program which runs many useful analyses on corpora such as frequency counts and word clusters.

Examples of both kinds of transcripts (i.e of text files transferred from Praatfiles and in CHAT format) can be found in Appendix B.

Tier 3

The third tier was used to count the number of syllables in each fluent run. Phonetic reductions were taken into account in the counting of syllables. For example the elision of 'e' in *j' pense que* was taken into account and this sequence was counted as 2 syllables although its full form is constituted of 3 syllables. As was previously explained, repetitions were taken into account in the delimitation of fluent runs. They were also taken into account in the counting of syllables. For example, the utterance *c'est c'est intéressant*, was counted as a run of 1 syllable (*c'est*) followed by a run of 4 syllables (*c'est intéressant*), since the repetition of '*c'est*' was not taken into

account in the count of syllables of the second run. In the same way, syllables subsequently retraced were not counted. For example in the two runs *c'est// il y a une famille, c'est* was not taken into account in the counting of syllables as it is retraced.

Tier 4

The fourth tier was used for the written transcription of the formulaic sequences identified in some of the fluent runs thanks to the identification criteria described in the previous section: syntactic or semantic irregularity, regular sequences with semantic or functional unity and holistic nature of the sequence in the input. The identified FS were then transferred onto text files for further analysis thanks to a Praat script written for that purpose. Examples of lists of identified FS are given in Appendix C.

Tier 5

The fifth tier was used for counting the syllables of the FS identified in tier 4. In the same way as for the third tier, repetitions of a syllable belonging to a FS were not counted in the number of syllable. So for example, *c'est c'est une bonne idée* would be analysed in the following way: one run of 1 syllable (*c'est*) which is formulaic, and one run of 4 syllables (*c'est une bonne idée),* also formulaic, with the repetition of *c'est* not counted.

The decision was made to measure the quantity of FS in syllables rather than by counting the number of FS for 2 main reasons. Firstly, since the average quantity of speech was measured in syllables, it was thought to be more coherent to measure the quantity of formulaic speech in syllables as well. The second reason has to do with the difficulty in identifying the limits of a given FS in certain instances. For example, if one takes the multiword sequence *je pense que c'est intéressant* (I think that it's interesting), it could be argued that it is composed of the combination of 2 FS i.e. '*je pense que* and *c'est intéressant*' or that it has fused into one single sequence. Accounting for the quantity of formulaic speech in number of syllables is therefore a way of getting round this issue.

In relation to the precise counting of formulaic syllables, one issue emerged when analysing the data with regard to multiword verb phrases such as *avoir envie de* ('to have envy of' = to feel like). For example, if the learner uttered *elle a* (pause) *envie de* ('she has...envy of' = she feels

like), should the 2 syllables of *elle a* be included in the counting of the formulaic syllables since they were followed by a pause? The following decision was therefore made when considering such formulaic verb frames: what was taken into account for the counting of formulaic syllables were the inflected form of the verb frame, provided that it was pronounced fluently i.e. neither interrupted nor followed by a pause. For example, the fluently uttered sequence *elle a envie de* was counted as 5 formulaic syllables because it was considered that the whole exemplar with this particular inflected verb form was automatised. However, in the case of the 2 runs '*elle a* pause *envie de*', only the 3 syllables of *envie de* were considered and counted as formulaic as it was considered that only the lexical frame was formulaic as the inflected form was not automatised within this sequence. I will come back to the issue of the different levels of abstraction of FS in the discussion.

6.2.2 *Measures and analysis*

Although Praat is convenient to annotate sound files in a way tailored to the aims of this study, it cannot be used for numerical calculations such as sums and averages or for lexical analysis such as the measurement of lexical diversity. This is why the data annotated on the Praat files had to be transferred to text files and Excel files for further analysis. More precisely, the orthographic transcripts, as well as the list of FS identified, were transferred onto Text files and all the numerical data were transferred onto Excel files for further analysis. This was made possible thanks to two Praat scripts[9] which were specifically written for that purpose.

6.2.2.1 *Measure of formulaicity*

As explained earlier, the number of syllables as well as the number of syllables belonging to a FS was counted for each fluent run on all the Praat files. The data was then transferred onto Excel files and the following measures were carried out on each task for each learner:

- Total number of syllables
- Total number of formulaic syllables
- Total number of runs
- Total number of runs containing formulaic material

[9] I cannot be thankful enough to Dr Christophe Dos Santos for writing these two scripts for me.

Thanks to the above data, 4 measures of formulaicity were calculated in order to investigate the quantitative development of FS use in the subjects:

- Percentage of formulaic syllables out of total number of syllables (%FS)

Since the quantity of speech was counted in syllables for the fluency measures (see section 6.2.2.2), it was considered more precise and coherent to measure the quantity of formulaic speech in syllables too.

- Percentage of runs containing some formulaic material out of total number of runs (%RFS)

This measure is interesting because it gives an insight into the proportion of runs which contain formulaic speech. So although this measure is less precise than the previous one as it does not express the exact quantity of formulaic speech, it gives a more dynamic view of how FS are distributed in the speech of the subjects.

- Average number of formulaic syllables per run (ANR) out of the total number of runs

This average number is obtained by dividing the total number of syllables by the total number of runs. This measure is relatively similar to the percentage of formulaic syllables but as the previous measure (percentage of runs containing some formulaic material), it has the advantage of relating the construct of formulaicity (number of formulaic syllables) to that of fluency (number of runs), therefore enabling to inform the research questions more comprehensively.

- Average number of syllables belonging to a FS in fluent runs containing one or more FS (ANRFS)

This average number is obtained by dividing the total number of syllables out of the total number of runs which contain at least one FS. This measure is useful in order to track the development of the learners' ability to use a greater quantity of formulaic speech within a single run either by combining several FS or by using longer FS. With regard to the purpose of investigating the link between FS use and fluency, this measure is relevant to examine the effect of FS use on mean length of runs (see 6.2.2.2)

6.2.2.2 *Fluency measures*

From the data entered on Excel files, the following measures were carried out on each task for each learner:

- Sum of all the relevant pauses (i.e. excluding irrelevant pauses as previously explained), expressed in seconds
- Total speaking time, i.e. sum of fluent runs, expressed in seconds
- Total task time (including pauses but excluding irrelevant elements such as the researcher speaking, irrelevant pauses and phenomena such as the participant laughing, etc.)
- Total number of syllables uttered
- Total number of formulaic syllables, i.e. sum of the syllables belonging to a FS. The total number of formulaic syllables was considered as a more coherent measure of formulaicity than the number of FS as the quantity of speech was also calculated in syllables. Moreover, it was a way of avoiding the potential theoretical difficulty in deciding whether a given multiword sequence such as *je pense que c'est intéressant* (I think that it is interesting) is formed of several FS (in this case, *je pense que + c'est intéressant*).

Thanks to the measures outlined above, the following fluency measures were then calculated:

- Phonation/time ratio i.e. the percentage of time spent speaking out of total task time
- Mean length of runs expressed in syllables i.e. the average number of syllables per fluent runs
- Speaking rate expressed in number of syllables uttered per second (including pause time)
- Articulation rate expressed in number of syllables uttered per second (excluding pause time)
- ALP: average length of pauses in seconds
- PPM (pauses per minute): average number of pauses per minute

As previous explained in chapter 4, the fluency measures which are the most commonly used in studies are speech rate and mean length of runs. It might therefore seem unnecessary to resort to all the measures listed above in order to measure fluency. However, if one tries to link fluency measures to underlying cognitive mechanisms, as is the case in this study, these two measures

are not sufficient to understand which cognitive changes might be taking place when an increase in fluency is recorded. That is why the present study uses the same set of fluency measures as in Towell et al. (1996). As pointed out by Towell et al. (1996) (see chapter 4), speaking rate is a very general measure encompassing the working of all the phases of speech production i.e. conceptualising, formulating and articulating the message (see chapter 3 for a more precise description of phases of speech production in Levelt's model (1989, 1999)). That is why speaking rate needs to be complemented by additional measures in order to gain a deeper insight into which of the processing components the cognitive change might be taking place. Mean length of runs is a more precise measure than speaking rate as it suggests that proceduralisation has increased in the formulator as the speaker is able to 'formulate' longer runs of speech. However, according to Towell et al. (1996: 92-94), increase in mean length of runs alone is not enough to be interpreted as proceduralisation in the formulator since it could be argued that increased mean length or runs might be due to greater planning time. This is why precise measures of pauses, such as phonation/ time ratio, average length of pauses and average number of pauses per minute, are useful to get a more exact picture of what exactly is taking place when there is a general increase in speaking rate. Indeed, Towell et al. argue that the pattern which needs to be found to indicate greater proceduralised knowledge in the formulator is: increased mean length of runs with at the same time no increase in average length of pause and stability or increase in phonation / time ratio. If there is an increase in the mean length of runs with no parallel increase in pausing time, the improvements observed in the overall measure of the speaking rate can mostly be accounted for by changes in the mean length of runs. In terms of the Levelt model, this suggests that the changes are attributable to faster processing in the formulator, and therefore to the fact that linguistic knowledge has been proceduralised rather than to changes in the phases of conceptualisation or articulation.

6.2.2.3 *Measurement of lexical diversity*

Another research question of the present study concerns the development of lexical diversity in relation to that of formulaicity. As a result, an adequate way of calculating lexical diversity had to be found.

A measure very commonly used for the calculation of lexical diversity is the Type-Token Ratio (TTR), which is the ratio of the number of word types to the total number of word tokens.

However, as has been well reported (see for example Daller et al. 2007), the TTR measure is problematic as it does not take into account the fact that the longer someone speaks, the less varied their language is likely to be. As a result, the TTR measure is text- dependent and it tends to decrease with text length. Therefore, if one uses TTR on long texts, one is bound to misrepresent learners' lexical richness, since the learners who speak the most might wrongly appear to be less lexically-varied. Since the texts compared in the present study varied greatly between time 1 and time 2 and also across learners, TTR was judged unsuitable as a measure of lexical diversity. One of the alternatives to TTR, the Guiraud Index (1954), which corresponds to the ratio of types to the square root of tokens, was not selected either because it does not address the issue of the text-length dependency (Malvern et al. 2004).

Because of the limitations of the two above-mentioned measures, the measure D was chosen to investigate the development of lexical diversity in the data. McKee, Malvern and Richards (2000) provide a detailed description of the program but in brief, D was created to avoid text-length dependency and is based on an analysis of the probability of new vocabulary being introduced into longer samples of speech. D has been integrated within CLAN (Computerised Language Analysis), a program available through CHILDES and it is computable through the VOCD program. D has been validated as a reliable measure of lexical diversity across a wide range of language learners (Malvern et al. 2004). D is not, however, without shortcomings either. It has recently been claimed by McCarthy and Jarvis (2007) that D overcompensates for TTR's sensitivity to text length and according to them, principles of probability make it inevitable that D will increase as text length increases. However McCarthy and Jarvis show that in the range of 100-2000 tokens (in other words, a variation in length of one to twenty), length accounts for only 5% of the variance which means that although sample size does matter, it effectively makes very little difference over a small range. That is why they conclude that despite this problem of text-length dependency, 'D is undoubtedly a better performer than most alternative indices' of vocabulary diversity' (2007: 480). Since, in the present study, size differences between samples at time 1 and time 2 were not over a factor of 2 or 3, D was retained as a suitable measure of lexical diversity.

However, because of the issue of text length dependency, D was also calculated on the first 400 words of each task in order to control for the problems inherent to text length and make sure that

the results obtained for the entire tasks were reliable. Moreover, in order to measure lexical diversity for each learner for time 1 and for time 2, I used the average D value of all the tasks at time 1 and then the average D value of all the tasks at time 2 rather than creating a pool of all the texts at time 1 and time 2 and calculating D across this very long text, as this would have increased the chances for D to be less reliable.

As mentioned above, the VOCD program is only available through the CLAN software. Therefore, in order for the learners' productions to be measured for lexical diversity, the transcriptions in text files were converted into CHAT format thanks to the CLAN command TEXTIN.

Frequency of occurrences of specific FS was also counted thanks to the CLAN software and more specifically the command COMBO (see Appendix D for an example of search result)

6.3 Conclusion

As a conclusion to this methodology chapter, it is important to summarise how the learner-internal and psycholinguistic focus of the present study is reflected in the following key characteristics of the research design.

- It is a longitudinal design so that the development of FS can be compared within the same learners.
- It aims to collect a large amount of data per individual learner in order to get a representative insight into the language of these individual learners.
- All the tasks are oral so that the necessary identification criterion of phonological coherence can be applied.
- Because of the paramount importance of the definition of a fluent run in the identification of FS, particular attention was paid to the annotation of sound files and the very precise measurement of pauses thanks to Praat.

The results of this study are presented and discussed in the next four chapters, as follows. Chapter 7 presents and analyses the use of FS in advanced L2 learners and discusses its implications for our understanding of language and the status of FS within it. Chapter 8 presents

and discusses the longitudinal development of FS use and lexical diversity during the learners' stay in France as well as the correlation between the two. Chapter 9 presents and discusses the longitudinal development of fluency. It presents and analyses the correlations between FS use and the various fluency measures. It examines the effect of FS on fluency development by discussing their role at the level of underlying processing mechanisms. Chapter 10 deals with the relationship between the learners' development in FS use, lexical diversity and fluency, and their engagement with the French language while in France.

Chapter 7. FS use in advanced L2 learners

7.1 Introduction

This chapter focuses on FS use in advanced L2 learners and aims to answer and discuss the first part of the research questions of this study namely:

1) FS use in advanced learners

- How far do advanced learners of French use FS?
- Does the type of task used have an effect on FS use?
- What types of FS do they use?
- Are there individual differences between subjects in terms of quantitative and qualitative FS use?
- What are the implications of FS use in L2 learners for our understanding of language and the status of FS within it?

Before presenting these results, it is useful to summarise again the five tasks that were used for data collection. There were 3 discussion tasks (one about speaking French, one about alcohol consumption and one about obesity), one interview and one story-telling task (see chapter 6 for more details). In the present chapter as well as in the next chapter, these 5 tasks will be abbreviated as follows:

Fr	Discussion task about speaking French
Alc	Discussion task about alcohol consumption
Fd	Discussion task about food and obesity
Inter	Interview
LN	Story-telling task about the Loch Ness

Table 3 List of tasks and abbreviations

Each task will be followed by either 1 for before going to France or 2 for at the end of the stay in France.

- **Size of the corpus**

Before presenting the results, it is useful to get an idea of the overall size of the corpus of oral productions collected for this study. As shown by Table 4 below, the overall size of the corpus, i.e. across both times and all learners is 54505 words. Out of these 54505 words, 15543 words

were identified as belonging to a FS. The number of words as well as the number of words belonging to a FS per individual learner are also presented in Table 4.

	Iris	Lily	Lola	Rose	Sally	Group
Word number	13896	9753	10974	7165	12717	54505
FS words	4164	2388	3233	1661	4097	15543

Table 4: Number of words and words belonging to FS per learner

- **Statistical analysis**

The data that follow, as well as the data presented in chapters 8 and 9, were all analysed using repeated measures ANOVA; this was done by subjects with the independent variables being task (5 levels) and time (2 levels), and by task with the independent variables subject (5 levels) and time (2 levels). The results of these will be reported as F_s and F_t respectively. The statistical analysis aimed to measure: (1) the significance of the change between time 1 and time 2, (2) the significance of the differences between tasks and between subjects and (3) the significance of the differences between tasks and subjects with regard to the size of the change between time 1 and time 2. Where there were significant differences across subjects or tasks either for general performance or for the size of the change between time 1 and time 2, pairwise t tests were used to identify the sources of the differences. The individual results were obtained using a related t test for each learner over each of the 5 tasks.

7.2 Group and individual results

On average, i.e. after combining all the results across subjects, tasks and times, 27.77% of the subjects' speech was considered formulaic according to the definition of FS and the identification criteria used in this study. In other words, FS represent an important part (over a quarter) of the speech of the five advanced learners of French under scrutiny: Iris, Lily, Lola, Rose and Sally (thereafter abbreviated as I, Li, Lo, R and S).

Table 5 presents the group results for the 4 measures of formulaicity used in this study: percentage of formulaic syllables out of total number of syllables (%FS), percentage of runs containing some formulaic material out of total number of runs (%RFS), average number of formulaic syllables per run (ANR) and average number of syllables belonging to a FS in fluent runs containing at least one FS (ANRFS).

	%FS	%RFS	ANR	ANRFS
Group results	27.77	39	1.50	3.73

Table 5: Group results on the 4 measures of formulaicity

On average, across the 5 learners under scrutiny in this study, 39% of the speech runs contain at least one FS. On average, 1.5 syllables per run belong to an FS, bearing in mind that the average mean length of runs across all tasks and all learners is 5.15 syllables. If one only considers the 39% of runs which contain some formulaic material, then 3.73 syllables per run containing at least one FS are formulaic.

Table 6 presents, for each task and each learner, the number of syllables belonging to a FS and this as a percentage of the total number of syllables. The mean percentage per learner was obtained by calculating the mean of the percentages per task and not by using the raw numbers. This was done to give the same weight to each task. Similarly, the average percentage for the group corresponds to the average of individual average percentages in order to give the same weight to the 5 subjects.

	Iris		Lily		Lola		Rose		Sally	
	FSsyll	%	FSsyll	%	FSsyll	%	FSsyll	%	FSsyll	%
Fr1	202	30.3	58	28.8	136	32.4	99	24.5	154	26.5
Fr2	532	34.2	597	35	510	30.8	344	33	374	36.4
Alc1	294	30.8	139	24.6	222	30.5	96	21.2	457	36.8
Alc2	720	30.7	366	31.9	420	28.6	158	28	353	36.9
Fd1	62	23.9	103	29.8	73	22.6	53	15.5	216	32.7
Fd2	507	34.8	375	30.7	434	36.6	169	25.1	333	33.2
Inter1	667	34.5	552	23.6	636	29	381	18.3	1254	30.6
Inter2	1961	35.9	586	29.7	1143	33.8	584	26.2	1294	29.7
LN1	27	9	51	16	61	19.4	60	16	83	19.3
LN2	101	29.4	226	21	186	30.1	53	12.8	238	27.5
Average %		29.35		27.11		29.38		22.06		30.96
Average % Group	27.77%									

Table 6[10]: number of syllables belonging to a FS (FSsyll) per task, percentage of FS per task and averages of percentage of FS per learner across all tasks

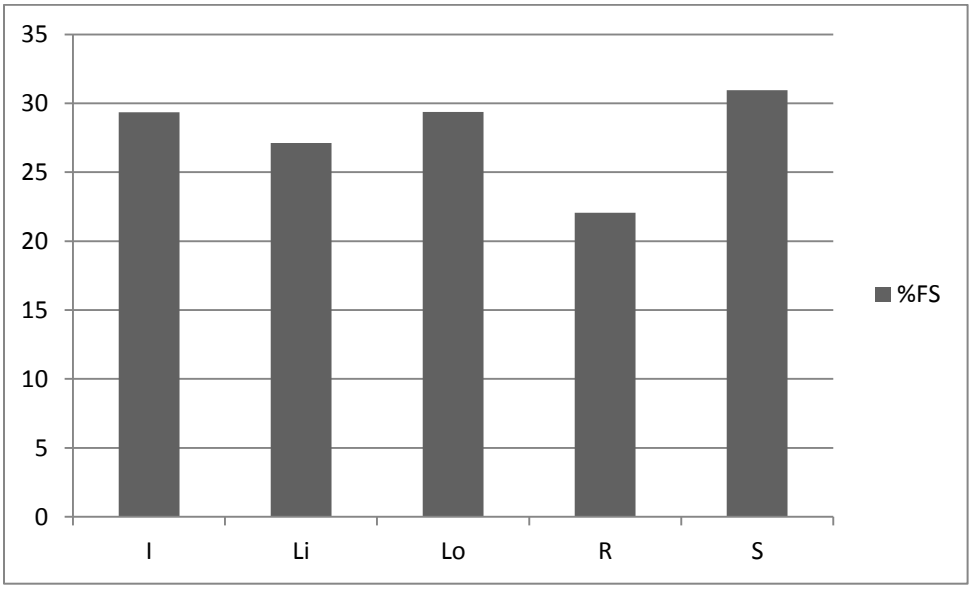
[10] For details about the acronyms for the task, see 7.1.

Table 7 summarises the results for individual learners for the 4 measures of FS use and presents the statistical significance of the differences between subjects in FS use.

	%FS[11]	RFS	ANR	ANRFS
Iris	29.35	42.19	1.72	4.04
Lily	27.11	37.54	1.43	3.73
Lola	29.38	42.53	1.63	3.77
Rose	22.06	25.65	0.83	3.15
Sally	30.96	47.07	1.90	3.95
Subjects df(4,16)	9.37 p< 0.001*[12]	31.19 p< 0.001*	35.53 p< 0.001*	7.36 p= 0.001*

Table 7: Individual and group results for the 4 measures of FS use and statistical significance of differences between subjects in FS use

As shown in Table 7, ANOVA showed that there were significant differences across subjects in the 4 measures of FS use ($p < 0.001$**)**, which means that at least 2 subjects performed significantly differently from each other. Let us take a closer look at the individual results for each measure to see what causes these differences.

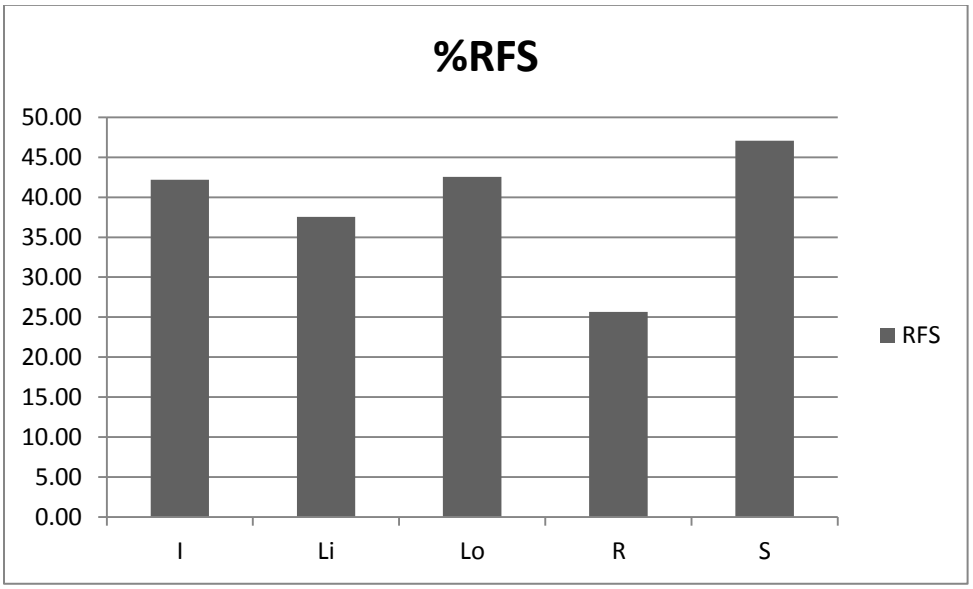


[11] The use of ANOVA with percentages has to be dealt with carefully. Indeed, it is recommended not to use ANOVA with percentages when they are based on binomial counts (in which the observations can only take two values). Moreover, a potential pitfall of percentages can be that they are less likely to satisfy the assumption of normality which is required to use ANOVA, since a large proportion of percentage data is often <20 or >80 (Wander Lowie (pc)). However, since the percentages in the present study are based on count data and their distribution is normal, it is safe to use ANOVA.

[12] For all the statistics, the asterisk indicates statistical significance i.e. that $p<0.05$

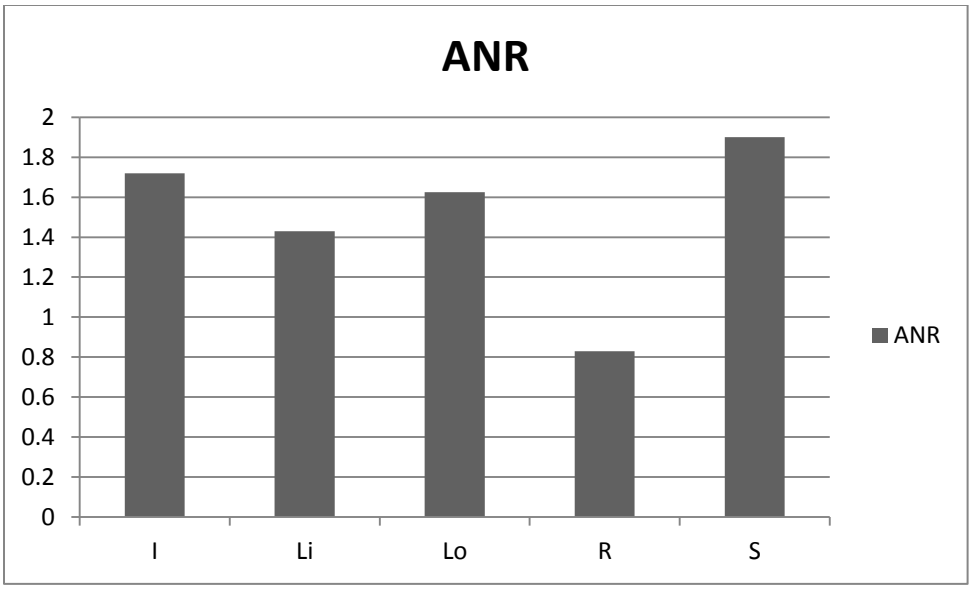
Graph 1: Percentage of FS out of total speech: Individual results combined across times

Graph 1 above illustrates the fact that the individual results for percentage of FS out of total speech are rather homogeneous as far as four of the learners are concerned. On average, Sally is the most formulaic subject (30.96%) but her rate of fomulaicity is hardly higher than that of Iris, Lily and Lola. The only learner who clearly stands out as the least formulaic is Rose with the lowest FS average (22.06%). Pairwise t tests show that overall Rose has a significantly lower percentage of FS than all the other participants ($p<0.015$ or less) who do not differ significantly ($p=0.06$ or more).



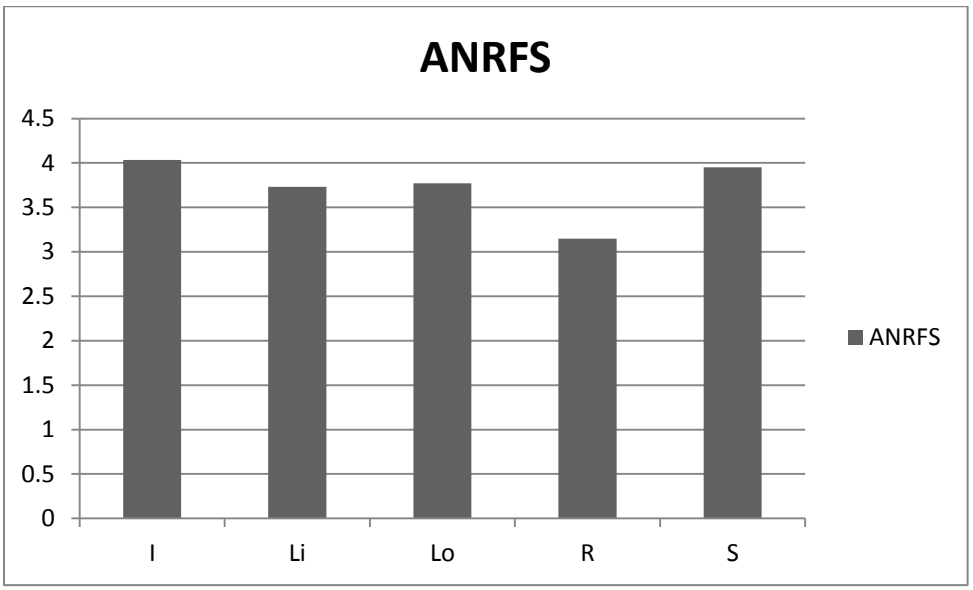
Graph 2: Percentage of runs containing at least one FS

As shown by Graph 2, for the percentage of runs containing at least one FS, the results are slightly less homogeneous. Again, Rose has a significantly lower percentage than all the other participants ($p<0.01$ or less) but Lily also differs significantly from all the participants except Iris ($p<0.03$ or less).



Graph 3: Average number of formulaic syllables per run (ANR)

For the average number of formulaic syllables per run (ANR) (Graph 3), Lily and Rose are significantly lower than the other 3 participants ($p<0.02$ or less) and are also significantly different from each other ($p=0.0008$).



Graph 4: Average number of formulaic syllables per run containing at least 1 FS

As shown by Graph 4, for the average number of formulaic syllables per run containing at least 1 FS (ANRFS), the results are relatively homogenous except for Rose. The pairwise t tests show

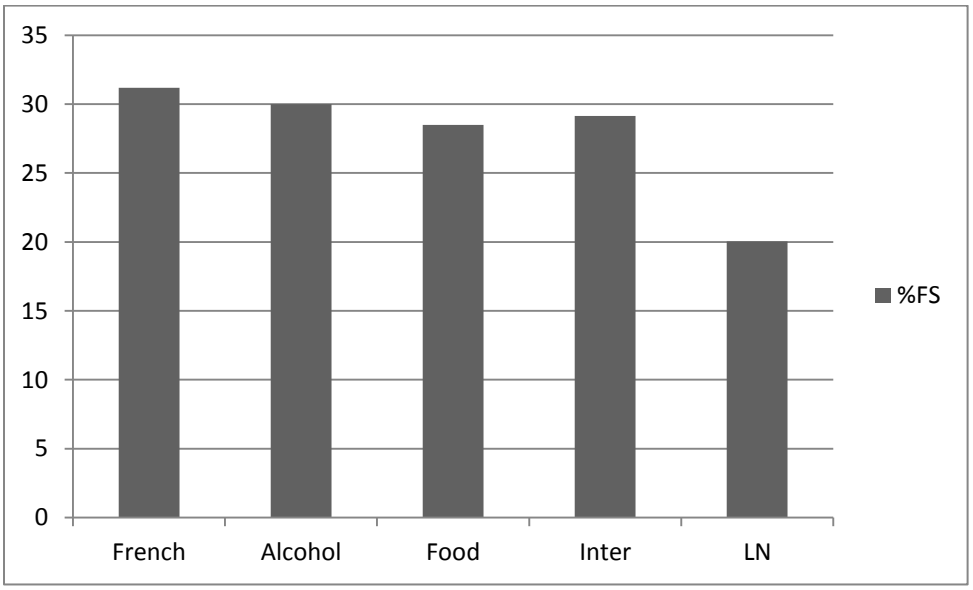
that overall she has a significantly lower average than all the other participants ($p<0.02$ or less) who do not differ significantly ($p>0.13$ or more).

7.3 Results by type of tasks

	French	Alcohol	Food	Interview	LN	Tasks df(4,16)
FSsyll	3006	3225	2325	9058	1086	NA
%FS	31.19	30	28.49	29.13	20.05	15.25 p< 0.001*
%RFS	46.1	41.54	39.01	42.835	25.495	29.56 p< 0.001*
ANR	1.82	1.55	1.53	1.70	0.91	25.65 p< 0.001*
ANRFS	3.87	3.62	3.82	3.84	3.49	23.29 p< 0.001*

Table 8: Number of formulaic syllables (FSsyll), FS use per task across all learners and statistical significance of differences between tasks

Table 8 presents the number of formulaic syllables per task, the results per task for the 4 measures of formulaicity used as well as the statistical significance of the differences between tasks for the 4 formulaicity measures.



Graph 5: Figure: Percentage of FS per task across all learners

As illustrated by Graph 5, which shows the percentage of FS per task, the results per task are also rather homogeneous. The only task which stands out as less formulaic is the story-telling task (LN). ANOVA (see Table 7) showed significant overall differences across the four tasks ($F_s(4,16)= 15.25$, $p<0.0001$), and pairwise comparisons show that LN is significantly different from the four other tasks ($p= 0.007$ or less) that do not differ among themselves ($p= 0.17$ or greater).

Pairwise comparisons carried out on the other 3 formulaicity measures confirmed the special status of the story-telling task (LN) compared to the 3 discussion tasks or the interview. Indeed, LN was also significantly lower than all the other tasks for the percentage of runs containing formulaic material (%RFS) and for average number of formulaic syllables per run (ANR). For ANRFS, it was significantly lower than the Interview and the discussion task about French.

To sum up the average formulaicity results across both times, about a quarter of the language of the advanced learners under scrutiny can be considered formulaic, according to the identification criteria used in this study. The 5 learners were found to be rather homogeneous except for one, Rose, who used significantly fewer FS than the rest of the group. Moreover, the figures show a task effect as the story-telling task (LN) was found to contain significantly fewer FS than the discussion tasks or the interview. The following section discusses these quantitative results as well as the task effects.

7.4 Discussion of the quantitative use of FS

7.4.1 *Percentage of FS*

According to this study, psycholinguistic FS represent about a quarter (27%) of the language of advanced L2 learners of French. It should be remembered that any figure with regard to quantitative FS use must be interpreted only in the context of the definition of FS as well as the identification method used. Indeed any change in the definition or the method of identification might affect the estimated number of FS.

As explained in chapter 2, the approach taken in this study is particularly important in the context of L2 acquisition, in which studies have mainly adopted a learner-external approach and dealt with the acquisition of idiomatic expressions in L2 learners, who have been shown to use few idiomatic expressions. The present results show that L2 learners' difficulty with mastering

idiomatic language should not be equated with the fact that they do not use psycholinguistic FS, as according to this study, these represent over a quarter of their language.

There are not many studies with which to compare these results as the quantity of psycholinguistic FS used by L2 learners has seldom been measured and when it has, other measures of formulaicity have been used. For example, Wood (2010) used a formula/run ratio to measure formulaicity. In his case study of a Japanese learner of English, Wood (2009) only found that she used around 12% of FS. The difference is probably due to the fact that the learner was less advanced than the learners of this study. Moreover, phonological coherence was only one criterion amongst others in his study. As a result, some psycholinguistic FS might have been overlooked. For example, the formulaic nature of some grammatically regular sequences such as common 'subject+ verb' units would probably not be spotted.

In brief, comparing the percentage of FS across studies is of limited usefulness. As we have seen previously, definitions (e.g. learner external or learner-internal) and identification criteria vary widely, as does the proficiency level of the learners studied. Another element which makes comparisons across studies difficult is task effects.

7.4.2 *Task effects*

Task effects have been found in studies on both fluency (Towell 2002) and lexical diversity (Skehan 2009). For example, Towell (2002) compared fluency results for 2 different types of tasks: a story-continuation and a personal adventure. For the personal adventure, the subjects were asked to recount something that had happened to them, with no restriction on the topic so that the subjects could choose incidents with which they were familiar. The story-continuation, however, was a much more controlled task which allowed less freedom as the subjects were given the beginning of a story and asked to continue. The results showed that all the fluency measures (SR, MLR and PTR) were between 13% and 27% higher for the personal adventure than for the story continuation. With regard to lexical diversity, Skehan found task effects on lexical diversity for both native and non-native speakers. He compared the results for D for a personal information exchange task, a narrative task and a decision-making task. For both native and non-native speakers, D was the highest for the decision-making task (90.6 and 52.9

respectively), followed by the narrative (75.2 and 46.9) and then the personal information exchange task, for which D was the lowest (45.6 and 36.1).

- **Impact of task type on the quantitative use of FS**

As described in 7.2, a task effect was found on the quantitative use of FS as the story-retelling task about Loch Ness contained significantly fewer FS than the other tasks (either discussion tasks or interview). For example, the average percentage of formulaic syllables across learners was 20.05% for the story-retelling task, whereas the average percentage of formulaic syllables for the 4 other tasks was 29.7%. As described in the quantitative results (7.2), pairwise comparisons showed that LN was significantly different from the four other tasks ($p = 0.007$ or less) for 3 measures of formulaicity used in this study: %FS, %RFS and ANR. For ANRFS[13], the story-retelling task was found to be significantly lower than the interview and the discussion task about French.

The lower quantity of FS in the story-retelling task is probably due to the fact that the content was less prone to FS use because it is a much more controlled task. It did not give the learners the opportunity to talk about themselves and did not require them to express their opinion, contexts in which they relied heavily on FS in the other tasks. However, the importance of the quantitative difference between the story-telling task and the other tasks was surprising to a certain extent: I expected the learners to resort to the use of fillers when they were finding it difficult to express a certain semantic content but they hardly did.

As shown by Table 8 and Graph 5 (for %FS), the formulaicity figures are comparable for the three discussion tasks. This is not surprising as these 3 tasks were structured along the same format. Moreover, although the interview was structured slightly differently from the discussion tasks, it also dealt with familiar conversation topics, which can account for the fact that the figures are similar to those for the discussion tasks.

- **Impact of task type on the qualitative use of FS**

Looking at the impact of task type on the qualitative use of FS, it appears that some types of FS occur specifically in certain tasks. For example, in the story-retelling task, one can find FS that

[13] See 7.2. for explanation of these acronyms

are specific to description and narration These sequences include time markers to organise the narrative sequence (*au début* (at the beginning), *à la fin* (at the end)), space markers (*au fond* (in the background)), sentence builders specific to description like *on voit* (one sees) or *on peut voir* (one can see). Moreover, the story-retelling is the only task in which the marker of the present continuous *en train de* can be found. On the opposite, opinion markers such as *je pense que* (I think that) are virtually absent from this task, since it did not give the learners the opportunity to offer opinions. Conversely, the discussion tasks are characterised by the high presence of FS used for argumentation and subjective judgements such as *à mon avis* (in my opinion). This is hardly surprising, since the learners were specifically asked what they thought about various issues. In brief, the difference in the types of FS used reflects the difference in the content of the discourse.

To conclude on the issue of task effect, the fact that learners use fewer FS in the story-retelling does not seem to be due to their lacking specific FS suitable for this type of task, since, as shown by the examples above, they are able to use FS specific to narration and description. Two key factors seem to explain the difference between the story-retelling task and the other tasks in terms of the quantitative use of FS. Firstly, the learners seem to use more FS when they are given more freedom in what to talk about than when the task is highly constrained. Secondly, tasks which give pride of place to subjectivity (either in terms of talking about oneself or give one's opinion), like the interview or the discussion tasks, encourage FS use more than prescriptive tasks, like the story-retelling, which do not require the learners to make subjective statements either in terms of opinions or tastes. It could be put forward that, the more the learners are left to talk about what they want, the more they rely on FS. Conversely, when they are required to describe specific things, they have less opportunity to integrate FS into their speech.

Finally, some methodological implications must be drawn from the task effects found in this study. Indeed, when making estimates of the quantity of FS used in a given corpus, one should bear in mind the nature of the task used for elicitation as it is likely to affect the learners' use of FS both quantitatively and qualitatively. In this respect, when investigating FS use, it seems problematic to only use one type of task, as was the case in Wood (2010) for example, as this might prevent the results from being generalised. Using different types of task, on the other hand, is a way of getting a more complete insight into the use of FS across different genres.

7.5 Typology of FS use in advanced L2 learners

This section aims to give an overview of the types of FS used by the learners in this study. There would be many possible ways of organising and classifying the corpus of FS identified in the oral productions of the five subjects under scrutiny in this study. One could organise it formally, for example according to the grammatical nature of the identified sequences. One could also organise it functionally, according to the communicative function fulfilled by the sequences. When tackling the task of classification, it quickly becomes apparent that it is not easy to be absolutely coherent for two main reasons. Firstly, even if one chooses a given organising principle e.g. functional, it is difficult no to resort to an additional one e.g. formal, at least as a way of structuring subcategories. Moreover, even when one managed to come up with a coherent way of classifying sequences, it is not always easy to decide whether a given sequence belongs to one category or another as in the case of certain sequences, their nature or function is not absolutely clear-cut and there is inevitably some overlap.

The organising principles for this study's typology are both functional and formal. Three main groups of FS are distinguished:

-referential FS: they are used to refer to various entities, be they objects ('sunglasses'), places ('at university'), times ('last year'), or even commonplace ideas ('it's a good idea').

-meta-discursive FS: they do not situate themselves at the level of the message itself but they are used to introduce one's opinion e.g. *à mon avis* (in my opinion). Fillers used to gain time when formulating a given message (e.g. *je sais pas* (I dunno)) also belong to this category, which also includes sequences used to structure narrations or descriptions.

-sentence builders FS[14]: they are the fixed part of sentence-building or phrase-building patterns composed of a fixed part and an open slot, and which provide a frame for the construction of a whole sentence such as *je pense que* (I think that) + clause or of a phrase such as *au niveau de*

[14] The term 'sentence-builder' is borrowed from Nattinger and DeCarrico (1992) who use it to refer to lexical phrases that provide the framework for whole sentences and which contain open slots. The term is used slightly differently here as the FS classified as sentence-builders here only provide a framework for parts of utterances. However, despite this slight difference, the term is borrowed all the same because of the same notions of ready-made framework with open slots.

(at the level of) + NP. It must be noted that, whereas the two previously outlined categories are functional, the category of sentence builders is a formal one.

As these three categories (i.e. referential, meta-discursive and sentence-builders) are very broad, they are further divided into subcategories. The typology that follows is not exhaustive, but provides a useful frame for understanding the role that FS play in oral production. A more exhaustive version of this typology can be found in Appendix E. In the typology that follows, only a few examples are given for each different type of FS. For each example of FS, the number in brackets indicates the number of occurrences of the FS across this study's corpus.

A. Referential FS

a. Whole sentence

Some examples of referential FS are whole sentences used to express an opinion or a commonplace idea. They could be described as 'useful ready-made sentences' in the sense that they are very common and can be used in numerous conversational contexts. They epitomise and exemplify the idea that part of the everyday language we use is made of common ideas expressed in common ways (Sinclair 1991).

At the formal level, these sentences vary in their degree of compositionality and analysability and therefore can be placed along a continuum: some sequences are entirely analysable and transparent in meaning such as: *je ne comprends pas* (I don't understand) or *c'est une bonne idée* (it's a good idea). Some are more idiomatic, especially the impersonal constructions starting with 'ça' such as *ça m'est égal* (this to me is equal=I don't mind).

i. Whole sentence expressing an opinion or a commonplace idea

1. Personal constructions

je suis (pas) d'accord ('I am (not) of agreement' I (don't) agree) 6

je ne comprends pas / *je comprends pas* / *j'comprends pas*[15] (I don't understand) 8

2. Impersonal constructions

[15] This FS has various forms depending on the level of its grammatical and/or phonetic reduction.

ça marche (it works) 16, *ça va marcher* ('it goes work'=it is going to work) 14, *c'est un problème* (it's a problem) 9

ii. Whole sentence expressing an opinion and following the structural pattern: *c'est/ c'était* + adjective

This was made as a subcategory of its own given the high frequency of the sequences following this pattern. In theory, these FS could have been classified in the sentence-builder category because they are composed of a fixed part (*c'est* or *c'était*) and a slot for the insertion of a variable item (the adjective). However, given the high frequency of many '*c'est* + specific adjective' exemplars such as *c'est difficile* (it's difficult) (32 occurrences) and the fact that they were always pronounced fluently, it was decided to classify these FS as whole sentences. Here are some examples of the '*c'est* + adjective' pattern with frequent adjectives:

c'est vrai (true) 17, *c'est important* 16, *c'est différent* 16, *c'est intéressant* (interesting) 14,c*'est super* (great) 13, *c'est normal* 12, *c'est facile* (easy)10, c*'est bon* (good) 10, *c'est mieux* (better) 10, *c'est (pas) pareil* (literally it's (not) identical= it's (not) the same) 9

iii. Ready-made whole sentences to express a 'necessary topic'

Learners also use full sentences to express a 'necessary topic' **(**Nattinger and DeCarrico 1992) for example related to the weather or a recurrent piece of biographical information:

il fait beau / du soleil, il faisait chaud / froid (literally 'it does' beautiful /some sun/ hot/cold=the weather is nice 1/ it's sunny 3/'it did' hot/ cold=it was hot 2/it was cold 1)

je joue au hockey (I play hockey) 3*, j'aime aller au cinéma* (I like going to the cinema) 3

b. Time, space and other referential FS complements

i. Place complements, usually prepositional phrases

Countries: *en Angleterre* (In England) 108, **Cities:** *à Paris* (in Paris) 29, **Various Places:** *à l'université* (at the university= at university) 35, **Spatial organisation:** *au fond* (at the back/ in the background) 4, **Chez + pronoun (strong form):** *chez moi* (at me=at/to mine) 23

ii. Time complements (nominal phrases and prepositional phrases)

NPs: *tout le/ l' temps* (all the time) 31*, l'année dernière* (last year) 11

PPs : *de temps en temps* (from time to time) 24**,** *en ce moment* (at the moment) 18

iii. Other types of complements (means etc….):

en vacances (on holiday) 13, *en modération* (in moderation) 5

c. Multiword NPs referring to a single entity

Noun + prep + noun : *transports en commun* (public transports) 2, *lunettes de soleil* (glasses of sun=sunglasses) 2,

Un/ une bon(ne) + noun : *une bonne expérience* (a good experience) 4

Noun+adjective : *année scolaire* (school year) 2

Definite article + noun : e.g. school subjects: *le français* (French) 43, or common groups of people: *les gens* (people) 69, *les jeunes* (young people) 20

d. Multiword VPs

parler anglais (speak English) 27, *faire du shopping* (to do some shopping) 7, *être en colère* ('to be in anger'=to be angry) 6, *prendre un verre* ('take a glass'= have a drink) 4

B. Meta-discursive FS

Discourse FS share the common point of facilitating discourse without bringing additional content to the message. In this sense, they can be said to be meta-discursive.

a. Fillers

The most two common fillers are: *je sais pas* (often reduced as *chais pas)* 90 and *je pense* (I think) 79

Other fillers include: *je crois* (I believe= I think) 46, *c'est ça* (it is that=yes/exactly) 38,*et tout* ('and everything', filler used at the end of utterances) 36, *et tout ça* ('and all this', filler used at

the end of utterances) 12, *des / les choses comme ça* (things like that) 34, *quelque chose comme ça* (something like that) 21

b. Prepositional phrases introducing one's opinion or structuring one's discourse

The following meta-discursive prepositional phrases are mainly used to:

- introduce one's opinion such as: *pour moi* (for me) 72, *a mon avis* (in my opinion) 12

-structure one's discourse e.g. *en fait* (in fact) 83, *par exemple* (for example) 57

c. Temporal expressions to structure narration

(et) après ça ((and) after that) 35, *et puis* (and then) 38, *à la fin* (at the end) 22

C. Sentence-builders

a. Express one's opinion + clause**:** *je pense que* (I think that) 215, *je crois que* (I believe that) 54, *je sais que* (I know that) 30, *je trouve que* (I find that) 14

It must be noted that most of the above sentence builders have been classified in this category because of their structure (fixed element + slot) but they could also have been categorised as meta-discursive FS as the function of many of them is to introduce the speaker's opinion and they do not directly contribute to the semantic content of the message.

b. Likes and dislikes + NP / infinitive: *j'aime bien* (I like well= I like) 38, *j'aime beaucoup* (I really like) 38

c. Existential expressions : *il y a* 262 (there is), *il y avait* 31 (there was)

d. Focus and presentative constructions e.g. *c'est* (it's) + NP+ qui/que (subject/object relative pronoun), *Il y a* + NP + *qui* (there is + NP + subject relative pronoun)

e. Necessary topics : *je m'appelle* (I me call= my name is) + name (10), *j'habite* (I live) + place (17)

f. Description : *on voit* (one sees= we can see) (17)**,** *on peut voir* (one can see=we can see) (8)

g. **Frequent 'subject+ verb units'** + complement**:** *je suis allée* (I am gone= I went) / *on est allé* (one is gone=we went) + place, *je vais* (I go=I'm going) / *on va* (one goes=we're going) / *ça va* (it goes=it's going) + infinitive verb to express the future
h. **Multiword prepositional phrases** + NP or infinitive **:** *au niveau de* +NP 10 : at the level of, *avant de* +infinitive 13 ('before of'=before), *en train de* +infinitive 18 ('in train of'=marker of an action in progress')
i. **Multiword quantifiers +** noun : *la plupart de / du / des* 19 (most +noun), *beaucoup de* (a lot of) 117, *plein de* (plenty of) 30
j. **Verb or phrasal verb + preposition+ infinitive verb /NP:** *commencer à* 9 (to start to), *avoir envie de* 23 ('to have envy of'=to feel like), *avoir besoin de* 20 ('to have need of'= to need)

The goal of the above typology was to illustrate briefly the main types of FS in this study's corpus. This typology is analysed in more detail in the next section.

7.6 Analysis of the typology

7.6.1 *Grammatically regular sequences*

Apart from a minority of idiomatic expressions, most sequences identified as FS in this study are grammatically regular. On the other hand, irregular or highly idiomatic sequences, though not absent from the typology, represent a minority of the sequences identified. It is clear that if the main identification criteria for FS had been based on idiomaticity, many sequences which have been included in this typology would not have qualified as formulaic.

As can be seen from the typology of identified FS, the most common 'subject + verb' sequences were regarded as formulaic, which might appear surprising. This decision was made because as the data was annotated and analysed, it clearly appeared that many common subject personal pronouns+ verb units both presented a processing advantage and worked as sentence builders like in the following examples: *j'ai* (I have) + age, *j'ai* (I have) + thing/person possessed, *je suis* (I am) + adjective or NP, *je suis allé* (I went) or *on est allé* (we went) + place, *on voit* (one/we can see) + thing which can be seen, *on peut* (one/we can) + infinitive verb.

It is not that L2 learners do not use FS, it is just that these FS are different from conventional idiomatic expressions. It appears from the results of this study that psycholinguistic studies investigating how L2 learners process idioms/ idiomatic expressions might be looking at the wrong sequences. Looking at further research, it would be a good idea for this type of studies to test the processing of sequences which have been previously identified in the productions of the same learners or at least in L2 learner corpora. Such a design (i.e. a design operationalising a learner-internal definition of FS) would guarantee that holistic processing is investigated for sequences which are known and used by L2 learners and not for sequences which are unlikely to be known by them.

7.6.2 *High frequency of sentence-builders and discourse FS and minority of idiomatic expressions*

This typology can be compared with the typology of Forsberg (2009), who compared the category distribution of FS in 3 different groups of French language-users: advanced learners (i.e. university students similar to the subjects of this study), very advanced learners (who had spent on average 4.5 years in France) and native speakers. Although Forsberg's approach to FS is learner-external and not learner-internal, it is interesting to compare her results with the results of this study as she also focused on advanced learners of French. Following Erman and Warren (2000), Forsberg classified the FS of her corpus into 3 categories: lexical FS (which correspond to referential FS in this study), grammatical FS (i.e. quantifiers and aspectual markers classified as phrase-builders in this study) and discourse FS (which correspond to both meta-discursive FS and sentence builders in this study). She found a significant difference between the group of advanced learners and the 2 other groups in the distribution of types of FS. The very advanced learners were found to use the same proportion of lexical FS as the native speakers, which according to Forsberg, was probably due to the role of input as they had been living in France for several years. On the opposite, the advanced learners used significantly fewer lexical FS than the 2 other groups and also overused discourse FS i.e. sequences such as opinion markers like *je crois que* (I believe that) or *en fait* (in fact).

The FS identified in this study would need to be coded according to their type in order to know how types of FS are precisely distributed across the corpus. In the meantime, however, even without having the exact numbers, it is clear that, compared to referential FS, there is a very high

frequency of meta-discursive FS and sentence builders (i.e. discourse FS in Forsberg's terms) in the present corpus. Indeed, it is striking that the average frequency of many referential FS is relatively low whereas the frequency of some discourse FS and sentence builders is extremely high. For example, the frequency of a sentence builder such as *je pense que* is above 200 whereas many referential NPs such as *lunettes de soleil* (sunglasses) or *année scolaire* (school year) only occur a couple of times in the corpus. The only referential FS that occur frequently are whole clauses which express commonplace ideas such as *c'est intéressant* (it's interesting) or *ça marche* (it works).

According to Forsberg's typology, the learners of this study can therefore be said to be advanced in the way types of FS are distributed across their oral productions. The only learner who stands out from the group in terms of distribution of types of FS is Sally, who uses many more idiomatic referential FS than the rest of the group, which is epitomised by the fact she has the highest score in both lexical diversity and lexical diversity within FS. There are indeed a lot of examples of idiomatic FS in her productions such as: *avoir les moyens* ('to have the means'=to afford), *ça m'a beaucoup plu* ('it me much pleased'(I liked it very much), *être à l'aise* ('to be at the ease'=to be comfortable), *pas grand-chose* ('not great thing'=not much), *ça me gêne* (it me embarrasses'=it makes me uncomfortable) , *faire pareil* (to do likewise=do the same), *prendre un verre* ('take a glass'= have a drink), *avoir l'air* ('to have the air'=to look/seem). In this respect, and according to Forsberg's distinction between 'advanced' and 'very advanced', Sally can be said to be starting to behave like a very advanced learner.

7.6.3 *Non-nativelike FS*

According to Forsberg (2009), non-nativelike FS are almost absent from advanced learners' productions and totally absent from very advanced learners' productions. That is why she considers that, for these learners, non-nativelike sequences are not a relevant issue to focus on. In this study, however, the learners, though advanced, still used various types of non-native FS.

- **Literal translation of the equivalent English expression**

In many cases, the non-correct nature of the sequences bears traces of L1 influence like in all the following examples:

Sur les nouvelles (Lo LN1), literal translation of 'on the news' instead of the idiomatic expression 'aux informations/ au journal télévisé'.

dans le soir (Lo LN1 and Inter 1): literal translation of 'in the evening' instead of the correct expression *le soir*

ce n'est pas le même (I Inter 1 and S Inter 1): literal translation of 'it's not the same' instead of the correct expression *ce n'est pas la même chose*

Lily uses several times *avoir une bonne soirée* (Alc2). This is an interesting example because it is only partly influenced by the English expression 'to have a good night'. Although Lily idiomatically translates 'night' into *soirée*, the verb is unidiomatically translated from the English 'have' instead of the idiomatic verb *passer*. In the same way, in *je l'aime bien* (I Inter 2), although '*aime bien'* is an idiomatic way of saying 'I like', the whole sequence is unidiomatically translated from the English expression 'I like it' whereas the correct idiomatic expression would be 'ça me plaît' (literally 'it me pleases').

- **Incorrect mix between 2 different expressions**

Some FS stem from the incorrect blend of two different sequences like in the following examples:

passer le temps avec ma famille (spend the time with my family): confusion between '*passer du temps avec quelqu'un*' (spend time with someone) and *passer le temps* (pass the time) (Li Inter 2)

en ce moment-là : confusion between '*en ce moment*' (at the moment) et '*à ce moment -là* (at that moment/ then) (I Alc 2) (I Inter 1)

elle n'a pas les besoins : mix between *avoir besoin* (have need=to need) and *avoir les* moyens (have the means to= to afford to) (I Inter 1)

le dimanche dernier (the last Sunday) : mix between *le dimanche* (the Sunday=on Sundays) et *dimanche dernier* (last Sunday) (I Inter 2)

- **Incorrect or unidiomatic sequences**

il y a quelques ans (there are a few years= a few years ago) instead of *quelques années* (S Inter 1)

une ou deux fois chaque année (one or twice every year) instead of *une ou deux fois par an*, *deux fois chaque semaine* instead of *par semaine* (S Inter 1)

à l'étrangère instead of *à l'étranger* (abroad) (S Inter 1)

pour la plupart du temps (for most of the time) instead of *la plupart du temps* (Li)

chaque lundi: grammatically correct but not as idiomatic as *tous les lundis* (literally 'all the Mondays) (Li Inter 2)

None of the learners know how to use the verb *pêcher* (to fish) and they all use incorrect expressions with *faire* (do)*: Fait du pêche* (R LN2), *faire la pêche* (I LN1, LN2): *fait de la pêche* (S LN1)

The expression *se passer* ('to go' or 'to happen') is also problematic as the learners tend to have automatised the form without the reflexive pronoun e.g. *ça passe bien* (intended meaning: it is going well) instead of *ça se passe bien* (Lo Inter 2), *ça ne passe pas ici* (intended meaning: it does not happen here instead of *ça ne se passe pas ici* (S Inter 1).

- **Non- pragmatic use of FS**

In some cases, what is not nativelike is not the FS itself but the way it is used by the learner. Indeed, some sequences are not always well used pragmatically. A very widespread example is the redundant use of *pour moi* (for me=in my opinion) as an opinion marker, which the learners tend to use non-pragmatically in conjunction with other opinion markers such as *je pense que*. Numerous examples of this non-pragmatic use of 'pour moi' are found in Iris' productions:

pour moi je pense que *tous ces euh ces aspects sont importants* (I Fr1) : for me I think that all these euh all these aspects are important

pour moi je pense que *les plats sains sont plus importants* (I Fd 1) : for me I think that healthy dishes are more important

Other examples are also found in the productions of the rest of the group:

pour moi personnellement j'ai voulu étudier la linguistique (for me personally I wanted to study linguistics) (Li Fr1)

pour moi je crois que *je préfère le sud* (for me I believe that I prefer the South) (Lo Fr2)

pour moi je crois que // je pense que *c'est pas mal* (S Alc1) (for me I believe that //I think that it is not bad): this last example is particularly interesting as it contains 3 opinion markers in a row as *je crois que* is retraced with *je pense que* .

- **Errors possibly caused by difficulties in breaking down the FS**

Some errors in the learners' productions might be due to their difficulties in breaking down the FS. For example, Lola seems to have difficulty in breaking the unit '*c'est* + adjective' as she has automatised the incorrect sequence '*tout c'est* (everything it is) + adjective' instead of the correct sequence *tout est* (everything is) as shown by the following examples: *tout c'est calme** (everything it is calm) instead of *tout est calme* (everything is calm) or *tout c'est très vert** (everything it is green) (Lo LN2) or *tout c'est fermé** (everything it is closed) instead of *tout est fermé* (Lo Inter 2).

Moreover, some grammatical errors might be encouraged by the automatic nature of some FS. For example, it is plausible that some article errors such as *faire un peu du* shopping* (S LN1) instead of *faire un peu de shopping (do some shopping)* or *j'ai un petit problème avec du* shopping* (R Inter 2) instead of *j'ai un petit problème avec le shopping* (I have a little problem with shopping[16]) are due to the fact that the sequence *faire du shopping* is so highly automatised in the learners that it makes them prone to use *du* in non-appropriate grammatical contexts.

Similarly, there could be a link between the use of FS and the incorrect use of a non-finite form or the lack of agreement in a form. In the following example, *la mère et un petit garçon aller en ville** (the mother and a little boy go (non-finite form) to town (Li LN1), *aller en ville* (to go to town) should have been conjugated and the non-finite form *aller* is incorrect. It could be argued that this error is due to the fact that Lily has automatised the non-finite form. Conversely, in the example *pour fais…fais du shopping** (to do (finite form) some shopping) (R LN1), the use of the finite form instead of the non-finite form might be due to the automatisation of the finite form in expressions such as *je fais du shopping* (I do some shopping).

[16] What Rose means here is that she is a shopaholic.

- **Over-reliance on some sequences**

In other cases, non-nativelikeness stems from the overuse of some sequences when learners over-rely on a limited set of sequences either across several tasks or in one single task. Such sequences can be said to act as 'islands of reliability' for the learners. The term was coined by Dechert (1983) who found that a German learner of English used certain FS as a platform for more fluent and accurate output. He described these recurrent FS that anchor the processes necessary for executing speech in real time as islands of reliability. The learners in this study all have a set of islands of reliability that they use recurrently. Table 9 below gives examples of some of their favourite islands of reliability.

This reliance on a limited set of sequences might be helpful with regard to their fluency, but it can also be regarded as non-nativelike because some sequences are overused. This is particularly the case when a sequence is overused within the same task. For example, in Inter 2, Sally relies a lot on *on est allé* ('one is gone'= we went) and in Inter 1, she repeats *je trouve que* (I find that) many times. Similarly, in Alcohol 2 and Food 2, Lily overuses *ça va marcher* (used 8 times in Alcohol 2 and 6 times in Food 2). Although *ça va marcher* (it's going to work) is idiomatic as such, its overuse does not sound nativelike. The reliance on particular sequences across all the tasks is less noticeable and therefore does not stand out as non-nativelike, except in the case of the two clearly overused sequences *il y a* (there is) or *je pense que* (I think that).

As illustrated by Table 9, although some islands of reliability are common to most of the learners, there are also individual differences in terms of which FS are the most relied on by individual learners for given communicative functions. For example, Iris relies a lot on *des choses comme ça* (things like that) as a filler (13 occurrences) whereas Lily and Lola prefer using *quelque chose comme ça* (something like that) (9 and 7 occurrences respectively).

Iris	Lily	Lola	Rose	Sally
Je pense que 56 *En fait 46* *Ça c'est 34* *Pour moi 30* *Il y a 21* *C'est juste 17* *Et tout 15* *Des choses comme ça 13* *Le problème c'est que 5*	*Il y a 59* *Je pense que 48* *Tu vois 17* *Ça va marcher 14* *Quelque chose comme ça 9* *En comparaison avec 7* *Il y a beaucoup de / les gens qui 7*	*Il y a 83* *Je crois que 45* *Je crois 29* *Ça c'est 28* *Je suis allée 28* *J'aime beaucoup 26* *J'ai beaucoup aimé 7(6 in inter 2)* *Je voudrais 18* *Je peux 14 (11* in Fr2*)* *Quelque chose comme ça 7* *On voit 8* *Il y a beaucoup de gens qui 5* *C'est / ce n'est pas un problème 7*	*Je pense que 53* *Il y a 45* *Je pense 33* *Je (ne) sais pas 12* *Avoir besoin de 7* *En fait 7* *C'est une bonne idée 6* *C'est important 7* *Je sais que 10* *On peut voir 5* *Nous avons +NP 6 (Inter 2)* *Je vais (+ infinitive) (mostly T2)* *C'est important de 6*	*Il y a 63* *Je pense que 49* *Ça c'est 30* *C'est* + adjective*:* *difficile 28, facile 18, sympa 20* *Pour moi 22* *Je trouve que 14* (most occurrences in Inter 1) *Et tout ça 10* *Les choses comme ça (6* in Inter 1*)* *nous sommes allés* (Inter1) *On était 5* *On est allé 7 (*Inter 2*)* *En train de (*9, all LN2*)*

Table 9: Islands of reliability

7.6.4 *Influence of input received at school on the learners' repertoire of FS*

It is obviously impossible to determine with certainty the link between input and intake (Schmidt 1995) but some sequences used by the learners are likely to be linked to the similar input they received as language learners in British schools. For such sequences, the distinction developed in chapter 2 between speaker-internal and speaker-external FS blends. Indeed, these sequences are likely to be formulaic both learner-externally and learner-internally: all the learners are likely to have automatised them because these sequences are highly resorted to in the linguistic environment they were all exposed to. For example, it is highly plausible that sequences corresponding to 'necessary topics' such as talking about the weather or talking about oneself were learnt at school where learners are encouraged to memorise ready-made routines such as:

Talking about the weather: *il fait beau* (the weather is nice) (I), *il fait du soleil* (it is sunny) (Lo and S) *il faisait chaud* (it was hot) (Lo), *il faisait froid* (it was cold) (S)

Talking about one's tastes/hobbies: *j'aime aller au cinéma* (I like going to the cinema) (Li), *je joue au hockey* (I play hockey) (S), *j'aime beaucoup faire du shopping* (I really like shopping) *(Lo)*

It is also very likely that many FS used to express one's opinion were automatised in the context of the school system, in which argumentative tasks requiring learners to express their opinions about a given topic are extremely common. Examples of such sequences include: *je suis (pas) d'accord* (I (don't) agree) or *je pense que* (I think that).

7.7 Individual differences in subjects' repertoire of FS

Many FS identified in this study are used by all the learners. For example, *à mon avis* (14) is a sequence which is used a few times by all of them. As mentioned in the previous section, the fact that the learners have a common repertoire of FS is not surprising given their similar learning profile. FS learnt at school are examples of such sequences, which can be described as being frequent across learners in the sense that they might not be highly frequent in the productions of each learner but they appear in the repertoires of all or at least a majority of them.

However, some multiword sequences are only frequent learner-internally as they are only used by one or two of the five learners. Of course, the absence of a given sequence in the speech of a given learner may be due either to chance or to differences across learners in the topics talked about. Therefore the absence of a given sequence in one

learner's productions should not be taken as evidence that it is altogether absent from their repertoire. As underlined by Wray (2002), frequency of occurrences depends directly on the contexts of subject matters brought up in a given speech sample. Therefore it cannot be excluded that a given sequence is absent from a learner's repertoire simply because there was no contextual opportunity for it to be used. Nevertheless, it is also likely that the absence or presence of some sequences in a given learner's productions point towards the fact that there are individual differences in the learners' repertoires of FS. In other words, some FS can be said to be idiosyncratic in the sense that they are only one or some of the learners' preferred way of delivering a given semantic content.

Firstly, some differences in the subjects' repertoires of FS are clearly linked to the subjects' different personal experiences. For example, for Rose who studies law, *la fac de droit* 'the law school' is a recurrent idiosyncratic unit. Sally, who spent her time in France in Paris, has automatised some sequences related to the vocabulary of Paris such as *le Quartier Latin* 'the Latin Quarter of Paris' In other words, subjects automatise sequences that are relevant to their daily life. In this case, lexical differences between subjects are a direct consequence of the fact that they are expressing different semantic contents. Such examples of idiosyncratic sequences due to differences in life experiences include: *jeune fille au pair* 'au pair' (Lo) , *manger à la cantine* 'eat at the canteen' (Lo, Li)*, faire du ski* 'to ski', *Irlande du Nord* (Northern Ireland) (R), *dans le cinquième* (in the fifth arrondissement of Paris), *dans le treizième* (in the thirteenth arrondissement of Paris), *la salle des profs* 'the staff room' (in a school) (Sally)

This link between the learners' idiosyncratic repertoires of sequences and their various life experiences is hardly surprising as it makes sense that the learners automatise sequences that are useful for them in their daily lives. What is more interesting is the fact that learners can also differ in the sequences they prefer to express the same semantic content or communicative function. For example, although most of the learners recurrently use the sentence builder *je pense que*[17] in order to introduce their opinion, Lola is the only subject who prefers to use the sequence *je crois que* in order to perform the same communicative function. Although *je pense que* is extremely frequent across the whole corpus (215 occurrences), Lola never uses this FS. Conversely, out of the 54 occurrences of '*je crois que',* 45 are from Lola's productions. Similarly, the 14

[17] For a case study of *je pense que*, see Appendix F.

occurrences of *je trouve que* are all uttered by Sally and the 4 other learners never use this FS.

The two tables below aim to illustrate the notion of lexical idiosyncrasies, with various examples of sequences used by only one learner (Table 10) or by 2 learners only (Table 11).

Lily	*La manière dont* (the way how) 4 *Tout de suite* (straight away) 2 **Pour la plupart* (most) 6 *Avoir hâte de* (to look forward to) 3
Lola	*C'est un vrai problème* (it's a real problem) 3 *Il y a du monde* (there is some world/crowd=a lot of people) 3 *Une bonne expérience* (a good experience) 3
Rose	*Bruit de fond* (noise of background=background noise) 2 *C'est une bonne idée* (it's a good idea) 7 **Etre honnête* (to be honest) 3 *Pas loin d'ici* (not far from here) 3 *Il semble que* (it seems that) 3 *Transports en commun* (transports in common=public transports) 2 *Année scolaire* (year scholarly=school year) 2 *Le droit* (law) 15 *Etre obligé(e) de* (be obliged to) 12 *Faire du ski* (do some ski=to ski) 3
Sally	*Je suppose que* (I suppose that) 3 *Ça m'a étonnée* (it surprised me) 2 *Ça me gêne* (it embarasses me) 2 *Ça craint* (it sucks) 2 *A peu près* (about/around)3 *Quoi d'autre* (what else) 4 *Par contre* 4 (by against=on the opposite) *Le lendemain* 2 (the day after) *N'importe quand* 3 (whenever) *Je trouve que* (I find that) 14 *Je dirais que* (I would say that) 3 *On était* (one was=we were) 5 *Un bon début* (a good start) 2 *Avoir l'air* (to have the air=to look) 4 *Prendre un verre* ('take a glass'=have a drink) 7 *Faire ses études* (do one's studies) 3 *Faire un master* (do a master) 3

Table 10: Idiosyncratic FS only used by one learner

En comparaison avec (in comparison with) 8	Li 7 I 1
C'est facile (It's easy) 10	S 6 Li 4
C'est pas grave (it's not serious=it's ok) 4	S 2 I 2
C'était super (it was great)10	S 4 I 6
Tu sais (you know) 5	S 4 I 1
Chaque semaine (every week) 5	Lo 3 S 2
En particulier (in particular) 8	I 5 R 3
En modération (in moderation) 5	Lo 3 I 2
J'ai beaucoup aimé ('I have much liked') 8	1 Li 7 Lo
On peut voir (one can see) 8	R 5 S 3
Au lieu de (instead of) 4	Li 2 S 2
Après avoir + past participle (after having) 3	Li 2 lo1
C'est juste (+NP) (it's just) 26	I 17 S9

Table 11: Sequences used by 2 learners only

Differences in the learners' lexical preferences can also be observed in the way some expressions are distributed across the learners. For example, out of the 72 occurrences of *pour moi* (used to introduce one's opinion), 30 are used by Iris, 22 by Sally, 10 by Louise whereas there are only 5 occurrences for Rose and Lily. Similarly, although *par exemple* (57) is used by all the learners, it is used much more by Lily (20) and Lola (21) whereas there are only between 4 and 7 occurrences for the other 3 learners. The filler *je crois* (45) is mainly used by Lola (29) and Sally (16) but is hardly used by the rest of the learners.

ça c'est ('this it is'= this is) is another interesting construction with regard to how it is distributed across the different learners. *ça c'est* can be described as 'double marking' as the subject, in this case the reduced form of the demonstrative pronoun *ce* is doubled by another demonstrative pronoun (*ça*). This construction is very characteristic of idiomatic French oral language. Although it appears very frequently in the present corpus (110 times), the occurrences are not distributed evenly across learners. It is used

frequently by 3 learners: Iris, Lola and Sally (34, 28 and 30 occurrences respectively). Sally is the only learner however, who already uses this construction frequently at time 1 (21 occurrences at time 1) whereas Lola uses it mostly at time 2 (1 occurrence at time 1 versus 27 at time 2). In the same way, *être en train de* (18) is only used by 3 learners (S 9, Lo 2 and I 7) and only occurs at time 2 (18 occurrences). Finally, out of 83 occurrences *of en fait* (in fact), it is used 46 times by Iris and hardly used by the other learners except Lily (14 occurrences) and Sally (13 occurrences). Moreover, all its occurrences only appear at time 2. *ça c'est, être en train de* and *en fait* are therefore good examples of differences across learners in how much they use a given sequence as well as lexical differences between time 1 and time 2 in terms of the frequency of a given FS.

To conclude on the issue of lexical differences between learners, it cannot be denied that the individual repertoires of FS display some common features across the 5 learners. This is not surprising considering the similar profile of the learners under scrutiny. However, the learners' repertoires also display some important differences either because some sequences are only used by one or 2 learners or because they are distributed very unevenly between them. Learners very clearly have their own sets of islands of reliability i.e. their preferred sequences on which they rely on to fulfil certain communicative functions or express certain semantic contents. It can therefore be said that, for certain FS, preferential processing works at 2 different levels. Psycholinguistically there is preferential processing because the sequence presents a processing advantage and is retrieved faster than other strings of words. Additionally, preferential processing also works at the lexical level i.e. given the contextual opportunity to express a given semantic content, one learner will tend to choose one particular FS over another.

The idiosyncratic nature of some of the FS identified in this study and, more generally, the individual differences in the learners' repertoires of FS are in line with the usage-based linguistic framework and its emphasis that language exists not only in the community of users but also in individuals. In this respect, due to the psycholinguistic nature of FS, the learners' different formulalects (Schmitt et al. 2004) can be interpreted as evidence of variation across individuals not only in lexical repertoires but also in internal cognitive representations. The longitudinal development of lexical diversity will be dealt with in detail in chapter 8 but a few remarks can be made here.

Moreover, the changes between time 1 and time 2 (either in terms of the appearance of new FS or the changes in the distribution of some FS) show that individual learners' representations are dynamic and in constant evolution, which can also be interpreted within a usage-based theory of language according to which our cognitive representations are constantly shaped by our experience of language.

It must be pointed out that lexical idiosyncrasies also need to be related to the learners' level of lexical development. Indeed, the more advanced learners, as they have a more lexically varied repertoire of FS, are more likely to use sequences that are not used by some of the less advanced learners. Therefore it makes sense that Sally, who is the most lexically varied of the group, as will be shown in the next chapter, possesses more idiosyncratic FS than the rest of the group.

The notion of lexical idiosyncrasies in FS use has methodological implications. Indeed, as described in chapter 2, psycholinguistic experiments with FS have tended to look at how speaker-external FS are processed by native speakers and language learners. However, considering the notion of individual differences in learners' repertoires of FS, it must be borne in mind that when a given pre-selected sequence does not seem to show any processing advantage for a given learner, it only means that it is not a FS for that given learner. However, it does not mean that the sequence tested is not formulaic for other speakers. Moreover, it does not mean either that that speaker does not use other FS as it could only be the case that this precise sequence is not formulaic for a given speaker who has other sequences in his/her repertoire.

7.8 Formulaicity at different levels of abstraction

When considering the typology of identified FS, it appears that some sequences have been automatised at a more or less abstract level. Indeed, while the learners use some sequences which are fixed and in a specific form, they also seem to have automatised grammatical frames and patterns at a more abstract level than fixed exemplars. For example, in the case of the pattern '*c'est* (it is) + adjective', it is the pattern which seems to have been automatised, and not only a specific sequence with a specific adjective. Indeed, given its high frequency throughout the corpus and the fact that it is always pronounced fluently whatever the adjective used, '*c'est* + adjective' can be considered formulaic at the level of the pattern. However, it is not clear at which level of abstraction each specific occurrence of '*c'est* + adjective' situates itself. For example, in the case of particularly frequent adjectives such as *important* or *intéressant* (interesting),

the sequence might be formulaic at a lower level, that is to say these specific sequences with a particular adjective are also stored holistically as exemplars by a given learner. As a result, within the same learner, some specific exemplars might be formulaic e.g. *c'est important* but others might only be formulaic at the level of the pattern e.g. *c'est bizarre*. In fact, it could even be the case that in the same learner, *c'est important* is sometimes produced as one unit and sometimes as the pattern '*c'est* + adjective' for example if the learner is still thinking about what they want to say and selecting the appropriate adjective. The level of abstraction of the same sequence might also vary across different learners. For example, the sequence *c'est intéressant* might be stored as a whole exemplar by a learner who frequently relies on it whereas it might only be a more abstract pattern for another learner.

The fact that some FS seem to be able to be automatised at different levels of abstraction lends itself well to a constructivist interpretation of language, as presented in Chapter 2. According to such a theoretical framework, as users of language experience tokens of language use, they categorise them at various degrees of abstraction (Bybee 2006, 2008). In our example of the pattern *c'est* + adjective, a learner will be exposed to many instances of it with various adjectives. Each exposure to a specific instance will reinforce both that specific exemplar and the pattern *c'est* + adjective. Depending on a learner's particular linguistic experience, particular exemplars of *c'est* + adjective will get a direct representation while others will not. A given sequence will therefore get a more or less abstract representation along a continuum of abstraction. The fact that the construction is represented at a more abstract level (*c'est* + adjective) does not stop particular exemplars being represented as well. In other words, speakers store both more or less abstract forms of the same construction simultaneously (Ambridge and Lieven 2011).

The typology presented in 7.5. also shows that more complex syntactic constructions / frames are recurrent and seem to have been automatised by the learners. This is for example the case for focus constructions i.e. constructions that are a means of 'highlighting particular constituents and the information they carry' (Miller and Weinert 1998: 190). Indeed numerous *it* cleft constructions such as *c'est* (it's) + NP qui/que (subject/ object pronoun) or presentative / existential structures can be found in the corpus. Towell et al. (1996: 111) also point out the recurrence in their corpus of these constructions, in which subordination is used not as a means of providing additional information but as a means of highlighting information.

Examples of formulaic syntactic frames used to highlight information include:

- **(il) y a/ avait (there is / was) + NP + qui (subject relative pronoun):**

y avait *plein d'mecs* ***qui.....*** *qui ont essayé de nous arrêter* (**there were** loads of guys **who**…who tried to stop us) (I Inter 2)

il y avait *une fille* ***qui*** *a trop bu* (**there was** a girl **who** has drunk too much) (Lo Alc 2)

- **il y a (there is) + NP + que (object relative pronoun):**

il y a *des choses* ***que*** *nous avons reconnues* (**there are** things **that** we recognised) (R Inter 2)

*je pense qu'****il y a*** *beaucoup de choses* ***que****....que les gens peuvent faire* (I think that **there are** a lot of things **that**…that people can do) (R Inter 1)

- **c'est (it's) + NP + que :**

c'est pas *le mot* ***que****.....je choisirais* (**it's** not the word **that**…I would choose) (I Fd2)

c'est *juste les chiffres* ***que*** *ça change* (**it's** only the figures **that** it changes) (S Alc 1)

Categorising the above constructions as formulaic is not consensual. For example, Towell et al. (1996), although they notice the importance of these focus constructions in their corpus, do not consider them formulaic and discuss them in the context of the development of their learners' syntactic competence. It is true that part of the following utterances is generated as they contain open slots which need to be filled. However, the fixed parts of the above constructions also follow a given pattern, which is why it is also possible to regard them as formulaic grammatical frames.

Moreover, another central argument can be put forward to argue for the fact that these structures have to be regarded as formulaic: in many examples of these structures, the open slot is filled with specific lexical items.

For example, in many occurrences of *il y a* +NP, the noun which fills the NP slot is preceded by the quantifier *beaucoup de* (a lot of). It can therefore be argued that for some learners, the more specific structure *il y a beaucoup de* + NP is formulaic, as illustrated by the following examples:

il y a beaucoup de *Français* ***qui*** *n'aiment pas l'Angleterre* (**there are a lot of** French people **who** don't like England) (Lo Fr2)

Il y a beaucoup d'*Anglais* ***qui*** *n'aiment pas la France* (**there are a lot of** English people **who** don't like France) (Lo Fr2)

Moreover, considering the more specific structure *il y a beaucoup de* + NP, in many of its occurrences, the NP slot is filled with *gens* (people) so again, it could be argued that the even more specific sequence *il y a beaucoup de gens qui* (there are a lot of people who) is represented directly in certain learners as illustrated by the following examples:

il y a beaucoup de gens qui *sortent* (there are a lot of people who go out) (Li Alc 2)

il y a beaucoup de gens qui*…qui boivent l'alcool* (there are a lot of people who…who drink alcohol (I Alc1)

The same interaction with specific lexical items can be noticed for the it-cleft constructions (*c'est* (it's) + NP + *qui*/*que (*subject/object relative pronoun). Indeed, many of its occurrences appear in conjunction with the indefinite pronoun *quelque chose* as in the following examples:

c'est quelque chose qui *m'a frappée* (**it's something that** stroke me) (Lo Inter 2)

c'est quelque chose qui *me montre que j'ai progressé* (**it's something that** shows me that I have improved) (I Fr 2)

c'est quelque chose que *j'aime pas trop* (**it's something that** I don't like very much) (Lo Inter 2)

c'est quelque chose que *je peux faire pour continuer à apprendre* (**it's something that** I can do to continue to learn) (Sa Fr2)

In total, *c'est quelque chose qui* appears 4 times in the corpus (used 3 times by Lola) and *c'est quelque chose que* appears 5 times (3 times used by Lola). It is therefore reasonable to argue that, at least in Lola's case, both constructions have a direct memory representation.

The interaction between a formulaic grammatical frame and specific lexical items is also well illustrated by the emphatic double marking construction NP *c'est que* (NP it is

that=NP is that). In most occurrences of this construction, the NP is *le problème* as illustrated by the following examples:

le problème c'est que *en fait y a // il a plein de maladies* (**the problem is that** in fact there is // he has loads of diseases) (I Inter 2)

le problème c'est que*....quand ils boivent*....(**the problem is that**…when they drink) (S Alc 2)

Therefore, the recurrence of the FS *le problème c'est que* shows that this specific sequence is probably represented directly, at least in some of the learners. For example, Iris uses this specific sequence 5 times in the corpus (across 3 different tasks), which shows that for her, it is more than likely to be formulaic. Again, the interaction between NP *c'est que* and the specific NP *le problème,* shows that this construction cannot be conceived as abstractedly as argued by the theoretical frameworks which argue for the independence of syntax.

Finally, most examples of the existential possessive construction *J'ai* + NP+ *qui* (I have + NP + who/which) occur with the NP *un(e) ami(e)* 'a friend' or its more familiar synonym *copain/copine* (mate) like in the following examples**:**

j'ai une amie qui *m'a dit que* (**I have a friend who** told me that) (S Fd 1),

j'ai une amie *ici* ***qui*** *fait* (**I have a friend who** does...) (Lo Alc 2)

j'ai une copine qui *habite à Lille* (**I have a friend who** lives in Lille) (Li Inter)

j'ai une très bonne copine qui *est allemande* (**I have a very good friend who** is German) (S Inter)

The other, less frequent occurrences are used to refer to either the learners' family members or their pupils.

j'ai *un cousin* ***qui*** *fume* (I have a cousin who smokes) (Li Inter)

j'ai *un petit frère* ***qui*** *ont* 2 ans* (I have a little brother who are* 2) (Li Inter)

j'ai *un élève* ***qui*** *est plus grand que moi* (I have a pupil who is taller than me) (Li Inter 2)

j'ai *des Premières* ***qui*** *parlent pas* (I have Sixth Form pupils who don't speak) (S Inter 2)

The interaction between J'ai + NP qui with *un ami* does not appear to be as strong as that between NP *c'est que* and *le problème* as there is more variety in the types of NP which fill the open slot. According to Bybee (2010) the distinct lexical items that occur in a slot in a construction constitute a category based primarily on semantic features. A category can be said to be more or less productive, depending on the number of items that are found to fill the open slot. Therefore, the NP in NP *c'est que* can be described as less productive than the NP in *j'ai* +NP *qui* as it seems it tends to be filled by *le problème* most of the time whereas the NP in *j'ai un ami qui* is filled with a higher number of types.

Moreover, Bybee also argues that some exemplars are central members of the category while others are more marginal. In this respect**,** *j'ai un ami* can be regarded as a prototype construction as *un ami* appears to be the central member of the category. Although it can be regarded as productive as it appears with a high number of NPs, in Bybee's terms, it is not very schematic as although the NP can filled with various NPs, it has a most frequent member (*un ami*) and the other NPs are also very semantically close to *un ami*, since they are either synonyms in a more familiar register (*copain/copine*), or lexical items related to the idea of relatives (*cousin, frère*) or more generally persons close to you in your daily life (for example one's pupils if one is a teacher). Bybee's hypothesis is that the more frequent member serves as the central member of the category and that new expressions tend to be formed by analogy with the more frequent member. Such hypothesis cannot be verified by the present data but what the present data exemplifies is the semantic closeness between all the NPs filling the slot in *j'ai + NP qui*.

All the above examples contribute to showing that many constructions situate themselves at an intermediate level between syntax and lexicon. Indeed, all the examples above follow a given grammatical pattern but tend to appear in conjunction with specific lexical items. This has important implications for our understanding of language. Indeed, all these examples give support to the notion that utterances are not necessarily produced by the most abstract rule possible (Ambridge and Lieven 2011). Instead, the important presence, in this corpus, of such constructions which display interaction between syntactic frames and specific lexical items, can be considered as evidence for great specificity in a language learner's cognitive representations of his or her L2.

To conclude, the study of the construct of FS is a way of observing the interaction of specific lexical items with specific grammatical configurations and getting an insight into the continuum between lexis and grammar.

Chapter 8. Longitudinal development of FS use and lexical diversity

This chapter focuses on the longitudinal development of FS use and lexical diversity and aims to answer and discuss the second part of this study's research questions namely:

- How does the use of FS by advanced learners of French develop during a seven months' stay in France?
- How does their lexical diversity develop?
- How does FS use develop in relation to lexical diversity? Is there a correlation between the two?

In order to answer the above research questions, the group and individual results for FS use are presented in the first section. The second section deals with the longitudinal development of lexical diversity. The third section discusses the development of lexical diversity within FS and the last section presents and discusses the correlation between FS use and lexical diversity.

8.1 Longitudinal development of FS use

8.1.1 *Group results*

Table 12 below summarises the main group results for FS use at time 1 and time 2 across the 5 tasks and 5 subjects: The results are given for the four measures of formulaicity used in this study: percentage of formulaic syllables out of total number of syllables (%FS), percentage of runs containing some formulaic material (%RFS), average number of formulaic syllables per run (ANR) and average number of formulaic syllables per run containing some formulaic material (ANRFS). Standard deviation, as well as minimum means (Min) and maximum means (Max) are given for tasks and for subjects. The name of the subject and task corresponding to the minimum and maximum values are given in brackets each time.

	Mean	SD (subject)	Min (subject)	Max (subject)	SD (task)	Min (task)	Max (task)
%FS T1	25.06	3.75	19.1 (R)	29.18 (S)	5.33	15.94 (LN)	28.78 (Alc)
%FS T2	30.48	3.32	25.02 (R)	33 (I)	3.71	24.16 (LN)	33.88 (Fr)
%RFS T1	31.86	7.04	22.88 (R)	41.40 (S)	8.21	18.35 (LN)	39.51 (Fr)
%RFS T2	46.1	10.41	28.42 (R)	54.32 (I)	7.79	32.64 (LN)	52.69 (Fr)
ANR T1	1.12	0.32	0.67 (R)	1.53 (S)	0.33	0.56 (LN)	1.41 (Fr)
ANR T2	1.88	0.53	0.99 (R)	2.31 (I)	0.37	1.26 (LN)	2.23 (Fr)
ANRFS T1	3.47	0.36	2.90 (R)	3.87 (I)	0.13	3.26 (LN)	3.58 (Fd)
ANRFS T2	3.99	0.35	3.40 (R)	4.28 (S)	0.22	3.72 (LN)	4.20 (Fr)

Table 12: Group results for the 4 measures of formulaicity

As shown on Table 13, ANOVA shows that the percentage of formulaic syllables (%FS) increased significantly after seven months in France ($F_s(1,4) = 79.94$, $p=0.001$; $F_t(1,4)=26.36$, $p=0.007$). There was no evidence, though, that the amount of change was significantly different across the tasks or across the subjects (interaction of task x time $F_s(4,16) = 0.72$, $p=0.59$; interaction of subject x time $F_t(4,16) = 0.24$, $p=0.91$).

	Analysis by subjects		Analysis by tasks	
Measure	Time df(1,4)	Time x tasks df(4,16)	Time df(1,4)	Time x Subjects df(4,16)
%FS	79.94 p= 0.001*	0.72 p= 0.588	26.36 p= 0.007*	0.24 p= 0.911
%RFS	21.12 p= 0.010*	0.18 p= 0.943	463.82 p< 0.001*	4.04 p= 0.019*
ANR	28.32 p= 0.006*	0.37 p= 0.824	428.08 p< 0.001*	5.63 p= 0.005*
ANRFS	8.38 p= 0.044*	0.47 p= 0.759	81.00 p= 0.001*	0.27 p= 0.895

Table 13: Statistical significance of the difference between T1 and T2 (Time) and of the size of the change between tasks (Time x tasks) and between subjects (Time x subjects)

The percentage of runs containing some formulaic material (%RFS) also increased significantly after seven months in France ($F_s(1,4) = 21.18$, $p= 0.0101$; $F_t(1,4)=463.82$,

p<0.0001). Although there was no evidence that the amount of change was significantly different across the tasks (interaction of task x time $F_s(4,16) = 0.18$, p=0.94); there were differences in the amount of change across participants (interaction of subject x time $F_t(4,16) = 4.07$, p=0.019). Pairwise comparisons showed that Iris' change was significantly larger than Sally's and Rose's (p (two tailed) = 0.002, and p= 0.044 respectively); none of the other pairwise differences were significant (p=0.054 or greater).

The average number of formulaic syllables per run (ANR) also increased significantly after seven months in France ($F_s(1,4) = 28.32$, p= 0.006; $F_t(1,4)=428.08$, p<0.0001). There was no evidence that the amount of change was significantly different across the tasks (interaction of task x time $F_s(4,16) = 0.37$, p=0.82); there were, though, differences in the amount of change across participants (interaction of subject x time $F_t(4,16) = 5.63$, p=0.005). Pairwise comparisons showed that Iris' change was significantly greater than Sally's, Lola's and Rose's (p=0.023 or greater) and Lily's change was greater than Rose's (p=0.005); none of the other pairwise differences were significant (p= 0.081 or greater).

ANOVA shows that the average number of formulaic syllables per run containing some formulaic material (ANRFS) increased significantly after seven months in France ($F_s(1,4) = 8.38$, p= 0.044; $F_t(1,4)= 81.00$, p<0.0001). However, there was no evidence that the amount of change was significantly different across the tasks or the subjects (interaction of task x time $F_s(4,16) = 0.47$, p=0.76; interaction of subject x time $F_t(4,16) = 0.26$, p=0.90).

8.1.2 *Longitudinal development of FS use: individual results*

8.1.2.1 *Percentage of formulaic syllables*

	Iris		Lily		Lola		Rose		Sally	
	FSsyll	% FS	FSsyll	% FS	FSsyll	% FS	FSsyll	% FS	FSsyll	% FS
Fr1	202	30.3	58	28.8	136	32.4	99	24.5	154	26.5
Alc1	294	30.8	139	24.6	222	30.5	96	21.2	457	36.8
Food1	62	23.9	103	29.8	73	22.6	53	15.5	216	32.7
Inter1	667	34.5	552	23.6	636	29	381	18.3	1254	30.6
LN1	27	9	51	16	61	19.4	60	16	83	19.3
Average % T1		25.7		24.56		26.78		19.1		29.18
Fr2	532	34.2	597	35	510	30.8	344	33	374	36.4
Alc2	720	30.7	366	31.9	420	28.6	158	28	353	36.9
Food2	507	34.8	375	30.7	434	36.6	169	25.1	333	33.2
Inter2	1961	35.9	586	29.7	1143	33.8	584	26.2	1294	29.7
LN2	101	29.4	226	21	186	30.1	53	12.8	238	27.5
Average % T2		33		29.66		31.98		25.02		32.74

Table 14: Number of formulaic syllables (FSsyll) and percentages of FS (%FS) per learner and per task at time 1 and time 2

Looking at the results per individual task and individual subject (Table 14), the percentage of formulaic speech increases between time 1 and time 2 for nearly all the

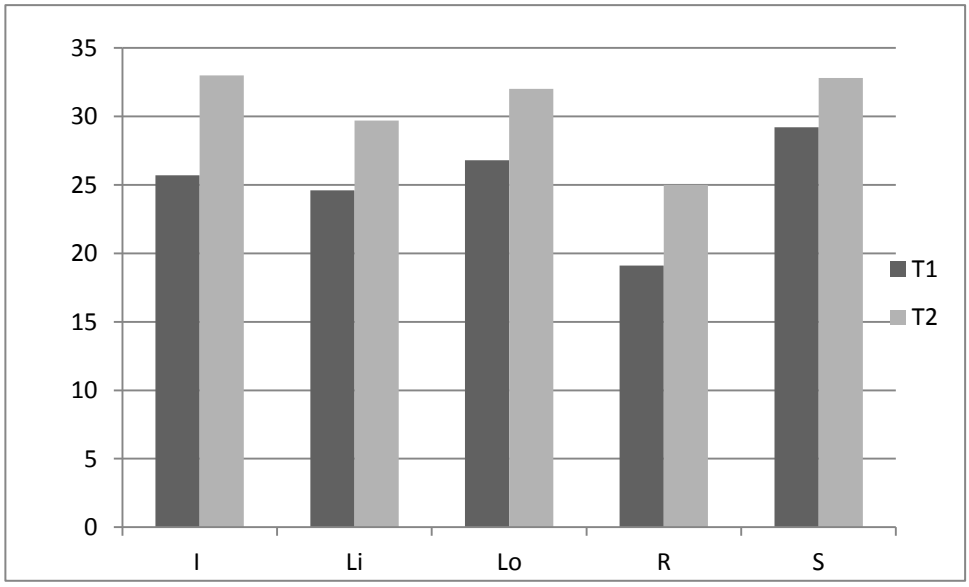
tasks as shown. Out of the 25 performances at time 1, there is an increase in the percentage of FS for 20 of them at time 2. In other words, 80% of the tasks see an increase in the percentage of formulaic speech between time 1 and time 2. The only exceptions are the alcohol task for Iris, the alcohol and French tasks for Lola, the story-retelling for Rose and the interview for Sally.

Table 15 below summarises the percentages of FS out of total speech for the individual subjects across the 5 tasks at time 1 and time 2 as well as the increase between time 1 and time 2 and the significance of this increase.

	T1	T2	+%	P value
Iris	25.7	33	+7.3	0.13
Lily	24.56	29.66	+5.1	0.01*
Lola	26.78	31.98	+5.2	0.18
Rose	19.1	25.02	+5.92	0.06
Sally	29.18	32.74	+3.56	0.19

Table 15: Increase in percentage of FS per learner between time 1 and time 2

This increase in percentage of FS per learner between time 1 and time 2 is illustrated by Graph 6 below.



Graph 6: Increase in percentage of FS per learner between time 1 and time 2

At time 1, Sally is the most formulaic since nearly 30% of her speech can be considered formulaic. Iris, Lily and Lola form a medium group with comparable rates around 25%. Rose is the least formulaic of the sample of subjects, with less than 20% of formulaic

speech. As shown by ANOVA; Rose is the only subject who differs significantly from the others.

Between time 1 and time 2, the percentage of formulaic speech increases for all the learners. There are some minor differences in how much increase takes place for each learner. However, ANOVA showed that there was no evidence that the amount of change was significantly different across the tasks or across the subjects (interaction of task x time $F_s(4,16) = 0.72$, $p=0.59$; interaction of subject x time $F_t(4,16) = 0.24$, $p=0.91$).

The increase in the percentage of FS used is only statistically significant in the case of Lily ($p=0.0101$) and very close to reaching statistical significance for Rose ($p=0.06$). However, the increase trends are strong for all the learners and can be considered significant given the high significance of the group results and the fact that ANOVA showed that the size of the changes between time 1 and time 2 was similar across the five subjects.

8.1.2.2 *Percentage of runs containing at least one FS*

Table 16 below shows the number of runs containing at least one FS, the percentage of runs containing at least one FS out of the total number of runs and the significance of the increase between time 1 and time 2.

	T1		T2		P
	RFS	%	RFS	%	
Iris	335	30	866	54.3	0.0017*
Lily	269	29.1	519	45.9	0.0026*
Lola	312	35.9	678	49.2	0.0037*
Rose	225	22.9	371	28.4	0.2344
Sally	593	41.4	607	52.7	0.8484

Table 16: Number of runs containing at least one FS (RFS), percentage out of tota number of runs (%) and significance

Graph 7: Percentage of runs containing at least one FS

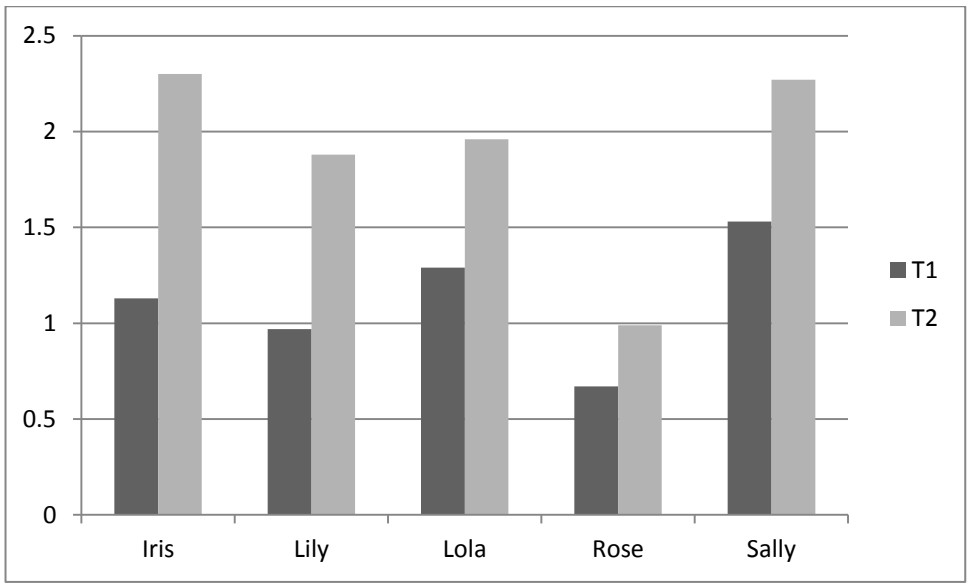
As illustrated by Graph 7, the percentage of runs containing at least one FS increases for all the subjects and as can be seen in Table 16, this increase is significant for 3 of the subjects: Iris, Lily and Lola. In terms of individual differences in the amount of change taking place, pairwise comparisons showed that Iris' change was significantly larger than Sally's and Rose's (p= 0.002, and p= 0.044 respectively); none of the other pairwise differences were significant (p=.054 or greater).

8.1.2.3 *Average number of formulaic syllables per run*

As illustrated by Graph 8 below, the average number of formulaic syllables per run increases between time 1 and time 2 for all the subjects and this increase is significant for all them except Rose as shown by Table 17.

	T1	T2	P value
Iris	1.13	2.3	0.0000*
Lily	0.97	1.88	0.0026*
Lola	1.29	1.96	0.0108*
Rose	0.67	0.99	0.0684
Sally	1.53	2.27	0.0060*

Table 17: Average number of formulaic syllables per run



Graph 8: Average number of formulaic syllables per run

Pairwise comparisons showed that Iris' change was significantly greater than Sally's, Lola's and Rose's ($p=0.023$ or greater) and Lily's change was greater than Rose's ($p=0.005$); none of the other pairwise differences were significant ($p= 0.081$ or greater).

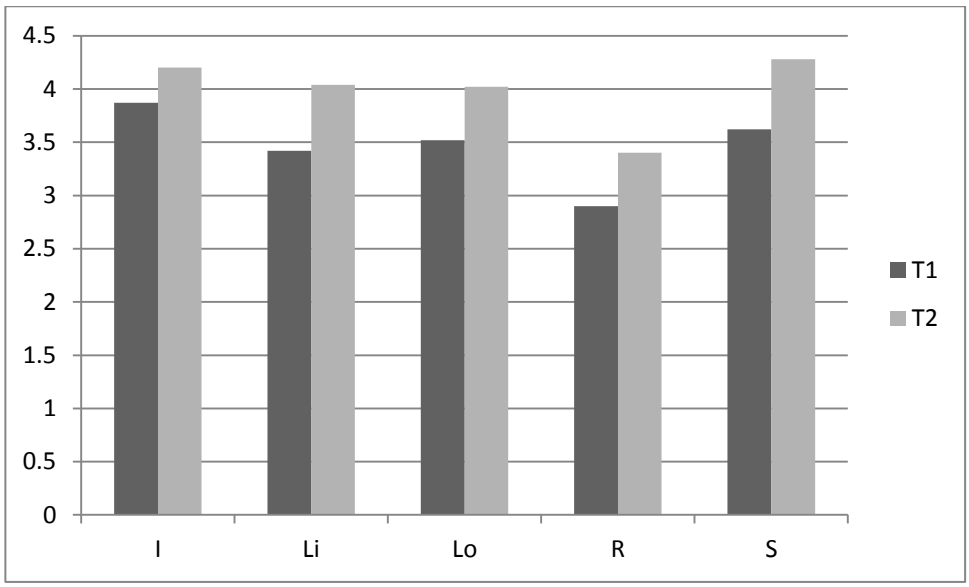
8.1.2.4 *Average number of formulaic syllables per run containing at least one FS*

As explained in the methodology section, the average number of formulaic syllables (ANFS) per runs containing at least one FS was calculated to see whether the subjects increased their ability to combine different FS within runs containing formulaic material.

	T1	T2	P value
Iris	3.87	4.20	0.2618
Lily	3.42	4.04	0.1394
Lola	3.52	4.02	0.1129
Rose	2.90	3.40	0.0300*
Sally	3.62	4.28	0.0009*

Table 18: Average number of formulaic syllables per run containing at least one FS, individual results

Looking at the results per individual learner, it can be seen that the average quantity of formulaic material within a run containing at least one FS increases for all the learners. This increase is statistically significant for Rose ($p<0.05$) and Sally ($p<0.01$). These results are illustrated by Graph 9 below.



Graph 9: Average quantity (in syllables) of formulaic speech within a run containing at least one FS, individual results across tasks

There was no evidence, though, that the amount of change in ANRFS was significantly different across the subjects (interaction of subject x time $F_t(4,16) = 0.26$, $p=0.90$).

8.1.3 *Summary of the results for FS Use*

	Iris	Lily	Lola	Rose	Sally	Group
%FS	√	√√	√	√	√	√√
%RFS	√√	√√	√√	√	√	√√
ANR	√√	√√	√√	√	√√	√√
ANRFS	√	√	√	√√	√√	√√

Table legend
√: increase but not significant
√√: statistically significant increase

Table 19: Summary of individual and group results for FS use

To sum up, there is an increase in the quantitative use of FS. Out of the 4 measures of formulaicity, this increase is significant for all the measures for the group results. For the individual results, they are significant on 3 of the measures for Lily, 2 of the measures for 3 of the learners (Iris, Lola and Sally) and only one of the measures for Rose (although %FS and ANR are also very close to statistical significance for Rose).

8.2 Longitudinal development of lexical diversity

8.2.1 *Types and tokens*

Table 20 below gives the number of types and tokens per learner and for the whole group at time 1 and time 2, to show the size of the corpus on which lexical diversity was measured. As explained in the methodology section, general lexical diversity as well as lexical diversity within the corpus of identified FS was calculated using the measure VocD.

	Iris		Lily		Lola		Rose		Sally		Group	
	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens
T1	197	1058	163	713	199	935	137	580	257	1881	494	19487
T2	368	3106	250	1675	300	2298	180	1081	399	2216	772	35019
Combined	441	4164	310	2388	359	3233	244	1661	498	4097	946	54506

Table 20: Number of types and tokens within the FS corpus per learner and for the group at time 1, time 2 and both times combined

8.2.2 *Group results*

As shown in Table 21, lexical diversity on the whole of the subjects' oral productions (D) increases between time 1 and time 2. As explained in the methodology section, the measure D, although it is assessed as the best available measure of lexical diversity is not exempt from text-length dependency. This is why, as a control measure, lexical diversity was also calculated on the first 400 words of each task. The results for D400 confirm the increase between time 1 and time 2. Finally, there is also a group increase for lexical diversity within the corpus of identified FS (DFS).

	Mean	SD (subjects)	Min (subjects)	Max (subjects)
D time 1	58.89	5.24	53.76 (I)	65.02 (S)
D time 2	71.46	11.54	54.10 (R)	86.20 (S)
D400 time 1	56.34	4.70	50.72 (I)	62.24 (Lo)
D400 time 2	64.90	10.44	50.78 (R)	78.66 (S)
DFS time 1	26.25	6.38	17.58 (R)	33.66 (Lo)
DFS time 2	32.96	9.65	20.35 (R)	44.64 (S)

Table 21: Group results for general lexical diversity (D), lexical diversity calculated on first 400 words of each task (D400) and lexical diversity within the corpus of identified FS (DFS)

As shown in Table 22, the increase of D between time 1 and time 2 is statistically significant ($F_s(1,4) = 9.91$, $p=0.035$; $F_t(1,4)=30.28$, $p=0.005$). The results of lexical diversity calculated on the first 400 words confirm the statistical significance of the increase in lexical diversity between time 1 and time 2 ($F_s(1,4) = 8.38$, $p=0.044$; $F_t(1,4)=44.05$, $p=0.003$). There was no evidence, though, that the amount of change in lexical diversity was significantly different across the tasks or across the subjects for D (interaction of task x time $F_s(4,16) = 0.93$, $p=0.474$; interaction of subject x time $F_s(4,16) = 2.83$, $p = 0.060$) or for D400 (interaction of task x time $F_s(4,16) =0.47$, $p=0.759$; interaction of subject x time $F_t(4,16) =2.45$, $p=0.088$).

	Analysis by subject		Analysis by task	
Measure	Time df(1,4)	Time x Tasks df(4,16)	Time df(1,4)	Time x Subjects df(4,16)
D	9.91 p = 0.035*	0.93 p = 0.474	30.28 p = 0.005*	2.83 p = 0.060
D400	8.38 p = 0.044*	0.47 p = 0.759	44.05 p = 0.003*	2.45 p = 0.088
	df(1,1)	df(4,4)	df(1,2)	df(4,8)
DFS	0.22 p = 0.721	2.49 p = 0.199	11.83 p = 0.075	8.55 p = 0.005*

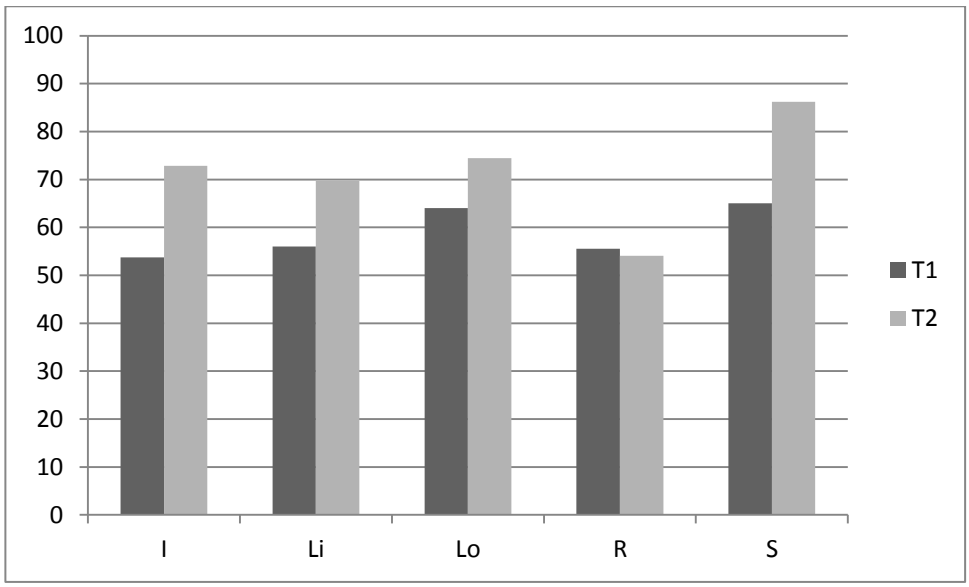
Table 22: Lexical diversity: difference between T1 and T2 and difference between tasks and subjects in size of the change between T1 and T2

The picture is different for lexical diversity within the corpus of identified FS. Although it goes up from 25.77 to 32.96, this increase is not significant across tasks or across subjects ($p>0.05$). Moreover, as shown on the table, ANOVA showed that there were statistically significant differences between subjects in the way they developed in relation to this variable ($p=0.005$).

8.2.3 *Individual results*

	Time 1	Time 2	P value
Iris	53.76	72.82	0.0193*
Lily	56.04	69.7	0.067
Lola	64.06	74.48	0.099
Rose	55.56	54.1	0.83
Sally	65.02	86.2	0.009*

Table 23: Lexical diversity (D) per learner at time 1 and time 2



Graph 10: Lexical diversity (D) per learner at time 1 and time 2

Taking a look at the results per learner (Table 23 and Graph 10), general lexical diversity increases for all the learners, except for Rose, for whom a slight non-significant decrease was recorded. Sally, who was the most lexically diverse at time 1, is also the subject for whom the rate of increase is the strongest as it increases by 32.6% between time 1 and time 2, which is highly statistically significant ($p<0.01$). Iris also significantly improves her lexical diversity score ($p<0.05$). Lily and Lola's rates of lexical diversity also increase, though to a lesser extent, with 24.4% and 16.3% increase respectively. This increase is significant for neither of them although Lily's increase is close to reaching significance ($p= 0.06$). These individual differences explain why the amount of change in lexical diversity across the subjects is so close to reaching statistical significance (interaction of subject x time $F_s(4,16) =2.83$, $p=0.060$), as illustrated by Table 22.

	Time 1	Time 2	P value
Iris	50.72	70.94	0.0135*
Lily	55.84	67.74	0.0930
Lola	62.24	72.12	0.1448
Rose	53.18	50.78	0.6225
Sally	59.74	78.66	0.0498*

Table 24: D400: lexical diversity (first 400 words of each task) per learner at T1 and T2

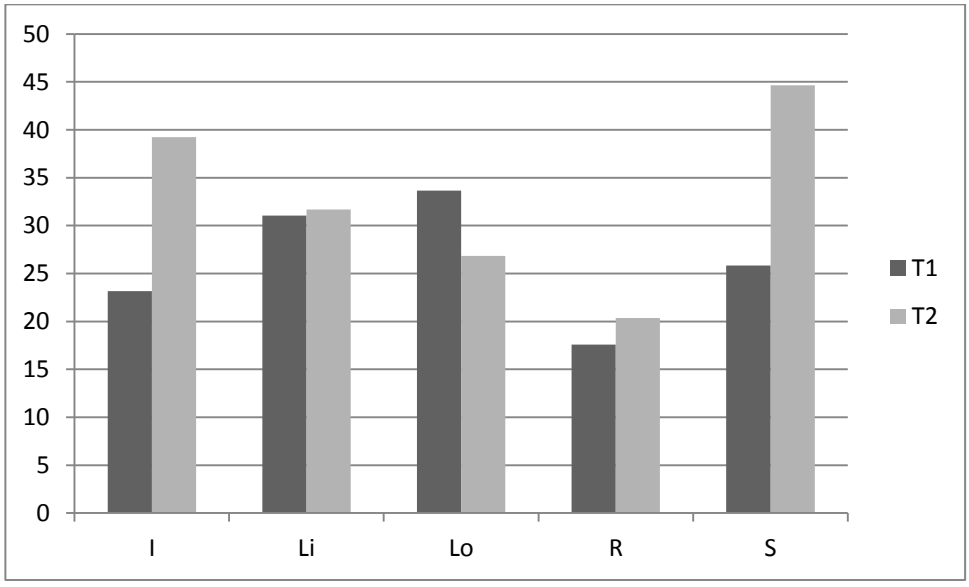
The control results for lexical diversity on the first 400 words (Table 24) confirm the results for lexical diversity calculated on the whole tasks as the same patterns of development can be observed for each individual learner.

As illustrated by Table 25 and Graph 11 , the individual results for lexical diversity within the corpus of identified FS confirm the more mixed picture of the group results which showed that the increase between time 1 and time 2 was not significant across tasks or across subjects ($p>0.05$) and that there were also statistically significant differences between subjects in the way they developed in relation to this variable ($p=0.005$).

	Time 1	Time 2	P value
Iris	23.15	39.22	0.037*
Lily	31.03	31.68	0.719
Lola	33.66	26.84	0.057
Rose	17.58	20.35	0.370
Sally	25.82	44.64	0.030*

Table 25: Lexical diversity within FS per learner at time 1 and time 2

Firstly, Graph 11 illustrates the fact that three different patterns can be observed. For two of the learners (Iris and Sally), there is an important increase in lexical diversity within FS, which is statistically significant ($p<0.05$) for both of them. On the other hand, for two of the other learners, Rose and Lily, the increase in lexical diversity within FS is much less important and is not significant. Rose's score only goes up by 15.8% and Lily's score virtually stays the same. Finally, in the case of Lola, there is a decrease in lexical diversity within FS since her score goes down by 20.3%, which is close to statistical significance ($p=0.057$).



Graph 11: Lexical diversity within FS per learner at time 1 and time 2

8.3 Discussion of the development of lexical diversity within FS

8.3.1 *Development of FS types and tokens*

As shown in section 8.1, there is a highly statistically significant increase in the proportion of FS used by the subjects after their stay in France. It is a general trend which can be observed for all the subjects. This raises the question of whether this increase is simply an increase in the number of tokens belonging to FS which were already present at time 1 or whether the learners have also acquired new lexical types of FS.

Table 26 below details the distribution of lexical types belonging to a FS:

- the total number of types belonging to a FS i.e. the number of types used across both times combined

-the number of types belonging to a FS used at both times i.e. types which are present at both time 1 and time 2

-the number of types belonging to a FS used only at time 1

-the number of types belonging to a FS used only at time 2

Table 26 also shows the percentage that each sort of type represents out of the total number of types.

	Iris		Lily		Lola		Rose		Sally		Group	
	Types	%	Types	%	Types	%	Types	%	Types	%	Types	%
Combined	441	100	310	100	359	100	244	100	498	100	946	100
Both times	124	28.1	103	33.2	140	39	73	29.9	158	31.7	320	33.8
T1 only	73	16.6	60	19.4	59	16.4	64	26.2	99	19.9	174	18.4
T2only	244	55.3	147	47.4	160	44.6	107	43.9	241	48.4	452	47.8
T1	197	44.7	163	52.6	199	55.4	137	56.1	257	51.6	494	52.2
T2	368	83.4	250	80.6	300	83.6	180	73.8	399	80.1	772	81.6

Table 26: Number and percentages of FS types for both times combined, number of types occurring at both times, at T1, T2, T1 only and T2 only

It can be seen that, for all the learners, the lexical types used only at time 2 represent the highest proportion out of the total number of types. Indeed, for the whole group, the types only used at time 2 represent nearly 48% of the total number of types. It cannot be

excluded that the large number of new types at time 2 is only an effect of the general increase in number of tokens at time 2, or that some of the new sequences at time 2 were already part of the subjects' FS repertoire at time 1. However, the fact remains that there is a high number of new FS types at time 2. In other words, the subjects not only use more of the same FS, they also use new lexical types of FS.

8.3.2 *Towards more idiomatic FS*

Although the difference between idiomatic speaker-external FS and psycholinguistic FS has been emphasised previously, there is definitely a shift towards a more idiomatic use of FS at time 2. Indeed, at time 2, one can see the occurrence of new FS which are idiomatic and which have clearly been picked up during the stay in France e.g. Sally: *on a des trous* 'we have holes' = we have gaps in our timetable. The acquisition of such colloquial sequences shows that the stay abroad is an effective way of being exposed to idiomatic input which would be more difficult to access in a home study context (Foster 2009).

One example of a new sequence at time 2 is *être en train de* ('to be in the process of' + infinitive = to mark the fact that an action is in progress). This sequence is entirely absent from the corpus at time 1 but is used recurrently by some of the learners (especially Sally and Iris) in the story-retelling at time 2.

New fillers also appear at time 2. Iris uses *et tout* (and everything) whereas Sally starts using *et tout ça* (and all that). At time 2, some of the learners (Iris, Lily and Sally) have also automatised common subject + verb units in the second person singular *tu*: such as *tu peux* (you can) (Li Fr 2, S Alc 2, I Alc 2), *tu vois* (you see) (Li Fr 2, Li Alc 2), *tu sais* (you know) (S Alc 2), *tu vas* ('you go' = you're going to) (I Alc 2, S Fd 2), *t'as* (you've got) (S Fd 2), *tu veux* (you want) (I Alc 2).

As can be seen from the above examples, the FS acquired at time 2 are not usually idioms or complicated idiomatic expressions. However, the acquisition of new fillers and the apparition of various verbal units with *tu* show that the learners' sociolinguistic competence has developed, in line with findings from studies investigating the benefits of residence abroad (Dewaele 2004; Regan, Howard and Lemée 2009).

8.3.3 *Development of general lexical diversity (D) and lexical diversity for FS (DFS)*

Several remarks can be made when comparing the scores for general lexical diversity and lexical diversity within the corpus of identified FS only.

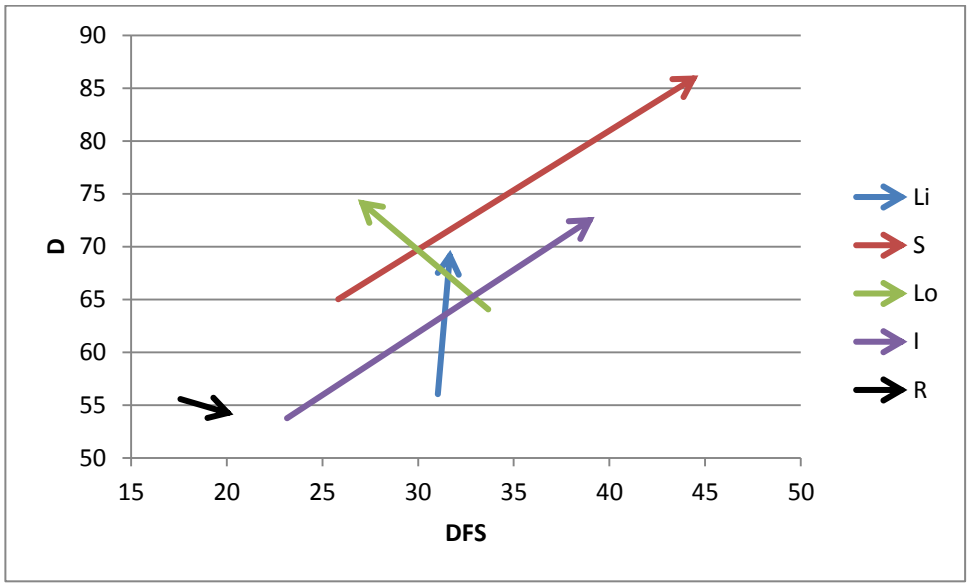
Firstly, overall, the mean score for DFS (29.5) is lower than for D (65.2). In other words, the repertoire of FS is less lexically varied than the subjects' general language: only a part of the subjects' lexicon has been proceduralised into automatically retrieved FS. This is not surprising as the range and variety of semantic content that needs expressing is far too wide for all language to be automatised into FS. Moreover, not all language needs to be automatised and it makes sense that automatic holistic form-function/meaning mapping develops only for a limited number of situations and contexts.

Secondly, when looking at the development between time 1 and time 2, there is a significant increase for D but the increase for DFS is not significant. Indeed, there is a general and statistically significant increase of overall lexical diversity D (from 58.89 to 71.46) and of the control measure for D, D400, which was calculated on the first 400 words of each task (from 56.34 to 64.9). On the other hand, although lexical diversity within the corpus of identified FS (DFS) goes up from 25.77 to 32.96, this increase is neither significant across tasks nor across subjects ($p>0.05$). In other words, lexical diversity increases less within the corpus of FS than on the whole of the subjects' output. This differential development between general lexical diversity and diversity within FS is interesting within the context of the hybrid nature of FS, which can be described as being composed of both declarative and procedural knowledge. In other words, a FS depends on lexical knowledge i.e. the items that constitute the FS need to be part of the learner's lexicon. It also depends, however, on the learner's ability to retrieve several individual lexical items as an automatic unit. The difference in rates of development between D and DFS seems to suggest that, in order for a given sequence to become automatised, the sequence must be part of the subjects' vocabulary in the first place. However, the automatic retrieval of the sequence belongs to the realm of performance and it takes some time before a given sequence becomes fully automatised as, apart from the cases when it has been acquired as a whole directly, it cannot be expected to get automatised straight away.

Moreover, with regard to the development of DFS, the group results mask important individual differences and ANOVA showed that there were statistically significant

differences between subjects in the way they developed (p=0.005). Iris' and Sally's increase is highly significant whereas Rose's and Lily's is not and there is a close to statistically significant (p=0.057) decrease for Lola.

Graph 12, which shows the relative development of D and DFS, illustrates these individual differences. For each learner the line joins 2 points which represent time 1 and time 2. The x axis represents DFS and the y axis represents D. For each learner / arrow, the x values of the first and second points represent the average values of DFS at time 1 and time 2. Similarly, the y values of the first and second points represent the average value of D at time 1 and time 2. The two points for each learner are joined by a line with an arrow indicating their endpoint at time 2, with the line starting at time 1. This way of representing conjointly the development of DFS and D aims to give a visual illustration of the learners' relative development in one variable in comparison to another. Graph 13, as well as several graphs in the next chapter on fluency, is designed according to the same principle.



Graph 12: general lexical diversity versus lexical diversity within FS

As illustrated by Graph 12, for Sally and Iris, the 2 variables develop hand in hand: in their case, the increase in general lexical diversity also applies to FS. In other words, lexical diversity within FS develops as quickly as general lexical diversity only for these 2 learners as they seem to be able to automatise a wider variety of sequences. As a result, for them, lexical diversity within FS still manages to go up, despite the importance of the quantitative development of FS use.

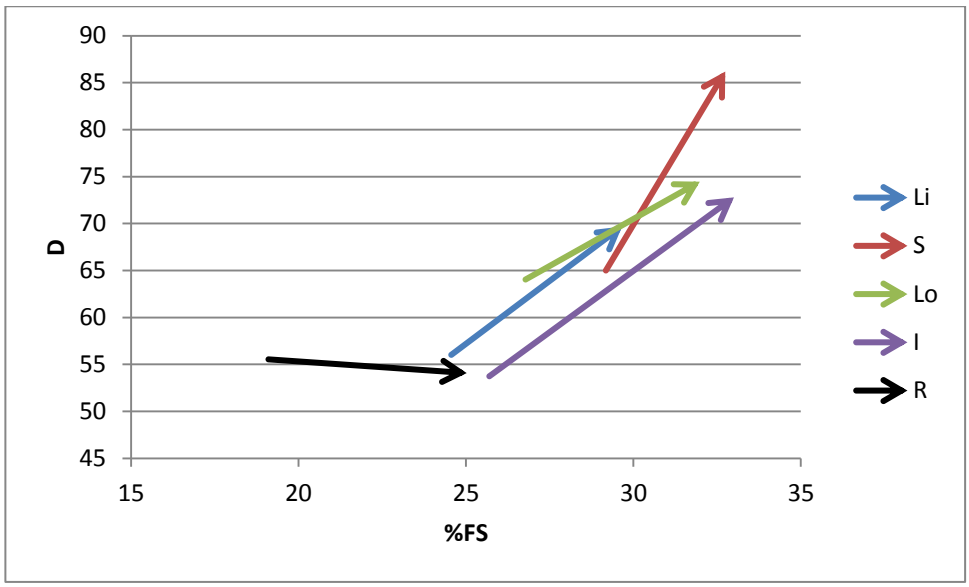
For Lily and Lola, the arrow goes up but not right: the general increase in lexical diversity does not translate into more lexically varied FS, DFS even decreases for Lola. This means that although, as shown in 8.3.1, their development in FS is also qualitative as they use a high number of new lexical types belonging to FS at time 2, it is probably not high enough in relation to the quantitative development of FS to prevent DFS from going down. As a result, the lexical diversity within their FS does not develop as quickly as their general lexical diversity.

For Rose, there is hardly any development on either of the variables, which shows that for her, there is hardly any lexical development whether it applies to automatised multiword sequences or not.

8.4 Correlation between FS use and lexical diversity

- **Statistical analysis of correlations**

This paragraph describes how the statistical analysis of correlations was carried out for the correlation between FS use and lexical diversity in this chapter but also for the correlations between FS use and fluency in chapter 9. For each of the five subjects, there are data available for five tasks. To use these data optimally, multiple regression was used over 25 data points with four dummy variables to code for the task with one of the variables in the correlation being used as the predictor with the second as the outcome variable. In this analysis, the t value on the coefficient for the predictor variable represents the 'importance' of that variable in predicting the outcome, while taking the relatedness of the texts into account and it has 19 degrees of freedom (*df*). In the presentation of the results, those t values are transformed into r values using the usual formula: $r = t/\sqrt{df + t^2}$. This is because r is the usual way of representing the strength of a relationship between two variables. The *p* values are two tailed *p*'s for the strength of the association.



Graph 13: Development of FS use in relation to development of lexical diversity

Graph 13 illustrates the comparative development of FS use (operationalised as percentage of FS) and lexical diversity. Rose shows a slight decrease in general lexical diversity at the same time as an increase in FS use. She is the only subject for whom the two variables pattern in this way. For the other subjects, FS use and lexical diversity develop in parallel.

r	D 1	D 2	D change
%FS 1	0.292 p=0.200	0.746 p=0.000*	0.619 p=0.003*
%FS 2	0.122 p=0.597	0.422 p=0.057	0.375 p=0.094
%FS diff	-0.181 p=0.433	-0.374 p=0.095	-0.292 p=0.200

Table 27: Correlations between %FS and lexical diversity (D)

As shown by Table 27, there is a significant correlation between FS use at time 1 and lexical diversity at time 2 ($r=0.746$ $p<0.001$): the more a learner is formulaic at time 1, the more they are likely to be lexically diverse at time 2. It must also be noticed that the correlation between %FS at time 2 and lexical diversity at time 2 is very close to reaching statistical significance ($r=0.422$ $p=0.057$). Moreover, there is also a significant correlation between %FS at time 1 and the size of the change in lexical diversity between time 1 and time 2 ($r=0.619$ $p=0.003$), which means that the more formulaic a learner is at time 1, the more likely they are to improve their lexical diversity between time 1 and time 2. This is illustrated in Graph 13: Sally, who has the highest score in %FS at time 1 is the learner for whom the increase in D is the sharpest.

The difference in pattern for Rose, as illustrated by Graph 13 can be accounted for in different ways. Firstly, her slower development could be due to individual differences such as motivation, aptitude, intelligence or working memory capacity. This could explain why both automatisation processes and the acquisition of new vocabulary are taking longer in Rose's case. Her individual circumstances might also be playing a part. Indeed, as will be described in chapter 10, from the results of the questionnaire, Rose is the learner who seems to have got the least input and output when she was in France, which is probably a factor in her slower development.

Moreover, the shape of the graph combined with the significant correlation between %FS at time 1 and D at time 2 as well as between %FS at time 1 and the change in D between time 1 and time 2, suggest a link between FS use and lexical diversity. It could be supposed that Rose's lower score in lexical diversity is partly due to the fact that she has the lowest percentage of FS use at time 1. Indeed, because she did not have many FS when she went to France, it is possible that she spent her time automatising more of her exiting vocabulary into FS and did not have the capacity to use her attentional resources to learn new vocabulary. Learning has been shown to involve automatising procedures in order to free attentional resources for new material (Towell and Hawkins 1994) but since human attention mechanisms and processing capacity are limited, these automatising procedures are likely to prevent cognitive resources to be spent elsewhere (Skehan 1998, Skehan and Foster 1999). In other words, one could put forward the argument that Rose's resources are all spent on the automatisation of her existing language and on the quantitative increase of her repertoire of FS and that, as a result, she has no resources left for increasing the size and diversity of her vocabulary. If such an interpretation is correct, Graph 13 acquires a more interesting dimension. Indeed, rather than just illustrating individual differences in the comparative development of FS use and lexical diversity, it also illustrates different stages of a same developmental path. In other words, although the fact that Rose is less advanced on the developmental path might be due to both individual differences and individual circumstances, it could also show that the automatisation process involved in FS is crucial for freeing attentional resources to move along the developmental continuum i.e. to use these freed resources to learn new material such as new vocabulary. It could therefore be argued that the graph and the analysis of correlations show that there is a link between the absolute value of %FS and the change in lexical diversity and that it is not until the learners have automatised a sufficient number of

sequences that they have enough free resources to acquire new vocabulary. In this respect, it is interesting to notice that Sally, who is the most formulaic subject at time 1 is also the one for whom development in lexical diversity is the sharpest and who is also the most lexically varied at time 2. It could therefore be put forward that, in her case, her high proportion of automatised sequences at time 1 encourages her lexical development, as FS use allows for her attentional resources to be dedicated to other cognitive processes such as the acquisition of new vocabulary. At time 2, Rose is only reaching the level of FS use of the others at time 1 so if one agrees that the graph illustrates a developmental path additionally to individual differences, it is plausible that, given time, that is to say once she has managed to increase her repertoire of FS, Rose will develop in a way similar to the other subjects.

8.5 Conclusion on the relationship between FS use and lexical diversity

To sum up, although the group results show that development of lexical diversity goes hand in hand with the increase in FS use, it is not the case with regard to the individual results. Indeed, Rose automatises more FS without any increase in her lexical diversity. For the other learners however, the results suggest that the development of FS is part and parcel of the development of their entire vocabulary. However, even for the 4 learners for whom general lexical diversity increases, there are differences in terms of the significance of this increase and the extent to which it also applies to the corpus of identified FS. For Lily and Lola, although there is an increase in lexical diversity, it is not significant and moreover, it does not extend to the corpus of identified FS. This pattern of results suggests that, in their case, the increase in FS use can be accounted for by a change at the performance level of automatic retrieval of units rather than at the level of lexical knowledge. In other words, the results suggest that they are using more FS at time 2 either because they are using more of the same FS already present at time 1 or because, between time 1 and time 2, they have become better at retrieving individual lexical items as multiword units. For Sally and Iris however, the increase in lexical diversity is statistically significant and also extends to the corpus of identified FS. Therefore, in their case, the pattern of results suggests that the increase in FS use can be accounted for not only by their using more FS already present at time 1 or by becoming better at retrieving pre-existing lexical items as units but also by acquiring FS composed of new lexical types. In other words, for these two learners, the joined development of FS use, lexical diversity and lexical diversity within FS suggests that they have

managed to either automatise new individual lexical units into multiword sequences very quickly or acquire some new vocabulary directly as multiword sequences.

In relation to the hybrid nature of FS, which belong both to the realm of competence (in terms of lexical knowledge) and performance (in terms of the learners' ability to retrieve lexical items as multiword units), the differences between learners in the results on lexical diversity allow to shed light on what underlies the increase in FS use. For some learners, the increase in FS use seems to be caused mainly by a change in the learners' performance i.e. their ability to retrieve vocabulary as multiword units whereas for other learners, the increase in FS use is also accounted for by lexical reasons additionally to increased automaticity.

The analysis of correlations also sheds some light on the relationship between lexical diversity and FS use and suggests that there is another way of interpreting the individual differences in the development of lexical diversity. Indeed, the significant link between FS use and development in lexical diversity and the shape of the graph illustrating the relative development of both variables suggest a single path of development, with learners varying in the stage they are at, at both time 1 and time 2.

It could be proposed that the developmental stages the learners go through at varying rates are the following:

1. Use attentional resources to learn new material
2. Proceduralise this new material, including automatise it into FS
3. Use the attentional resources freed by the automatisation of 'old' material to learn new material

The results therefore suggest that FS use in L2 learners could be regarded as a key variable to account for the acquisition of lexical knowledge since it is not until the learners have automatised a sufficient number of multiword sequences that they have enough free resources to acquire new vocabulary.

Chapter 9. Relationship between FS use and fluency development

This chapter deals with the relationship between FS use and fluency development and aims to answer and discuss the third part of this study's research questions:

- How does the fluency of advanced learners of French develop during a seven months' stay in France?
- Does the development of FS use by L2 learners play a role in the development fluency during the year abroad? → Quantitative analysis
- If it does, how does it take place? → Qualitative analysis
- Which underlying mechanisms can explain this role?

The first section presents the group and individual results for the development of fluency. The second section discusses these results. The third section presents the correlations between FS use and the various fluency measures. The fourth section is a qualitative analysis of the use of FS in order to account for its role in increasing fluency. The last section analyses how the impact of FS use on fluency can be conceptualised at the cognitive level of underlying processing mechanisms, by showing which stages of speech production it facilitates.

9.1 Fluency development

9.1.1 *Increase in quantity of speech*

The table below shows the increase in quantity of speech between time 1 and time 2.

	Iris	Lily	Lola	Rose	Sally
T1	3877	3316	3631	2974	5689
T2	10020	6437	7343	4191	7028
Total	13896	9753	10974	7165	12717

Table 28: Number of tokens per learner

If one compares the number of tokens at time 1 and at time 2, as shown in the above table, it is clear that the learners spoke much more at time 2. It must be pointed out that, for a restricted number of tasks, the difference in number of tokens between time 1 and time 2 can be qualified by the fact that some of the discussion tasks[18] were carried out in pairs in the first round of data collection, which means that the subjects were bound to utter fewer words. However, in spite of this methodological consideration, it is clear

[18] Discussion tasks on French, food and alcohol for Iris, Lily and Lola and discussion task on alcohol for Sally and Rose

that the subjects found it easier to express themselves at time 2. Although a familiarity effect cannot be excluded, this large increase in the number of tokens cannot be fully attributed to the fact that the same tasks were repeated given the fact there was a 10-month gap between the two rounds of data collection. Moreover, as far as the interview is concerned, the questions asked varied quite significantly between time 1 and time 2, as the focus was more on personal information at time 1 and more on the stay in France at time 2, therefore excluding the possibility of a familiarity effect.

On the whole, Iris and Sally are the subjects who talked the most, followed by Lola and Lily. Rose is the subject who talked the least. Iris is the subject for whom the change in quantity of speech is the most striking between the two times. Sally is the most even subject in terms of quantity of speech as the increase in number of tokens between time 1 and 2 is only of 24%. However, since she is the subject who spoke the most at time 1, it still represents a large increase.

9.1.2 *Fluency measures: group results*

In order to investigate the development of the global construct of fluency, the following fluency measures were calculated: Phonation-time ration (PTR) i.e. the percentage of time spent speaking out of total task time, mean length of run (MLR) i.e. the average number of syllables per fluent runs, speaking rate (SR) i.e. the average number of syllables uttered per second (including pause time) and articulation rate (AR) i.e. the average number of syllables uttered per second (excluding pause time). The way all these measures were calculated is described more precisely in the methodology section (see chapter 6).

	Mean	SD	Min	Max
PTR T1	62.80	2.34	59.97 (Li)	65.56 (Lo)
PTR T2	70.95	5.27	63.42 (R)	76.2 (Li)
MLR T1	4.30	0.62	3.50 (R)	5.06 (S)
MLR T2	6.00	1.28	3.80 (R)	6.95 (I)
SR T1	2.15	0.35	1.70 (R)	2.56 (S)
SR T2	2.93	0.65	1.83 (R)	3.48 (S)
AR T1	3.40	0.47	2.74 (R)	3.94 (S)
AR T2	4.11	0.81	2.90 (R)	5.08 (S)

Table 29: Fluency: group results

As can be seen from Table 29, there was a strong increase in the 4 fluency measures between time 1 and time 2.

Measure	Time (df1,4)	Time x Subjects (df4,16)
PTR	26.29, p= 0.007*	10.80, p< 0.001*
MLR	250.60, p< 0.001*	9.84, p< 0.001*
SR	384.00, p< 0.001*	13.70, p< 0.001*
AR	188.52, p< 0.001*	14.25, p< 0.001*

Table 30: Development in the 4 main fluency measures across all learners and all tasks

As can be seen from Table 30, the results of the repeated measures ANOVAs showed that the changes in the 4 fluency measures between time 1 and time 2 were highly significant ('Time' column). Moreover, as shown by the (Time x Subjects) column, which represents interaction between subjects and time, the amount of change differed between the participants. For PTR Iris and Lily show significantly greater change than the other three participants. For the other three measures (MLR, SR and AR) Rose shows significantly smaller changes than the other four subjects.

9.1.3 *Fluency measures: Individual results*

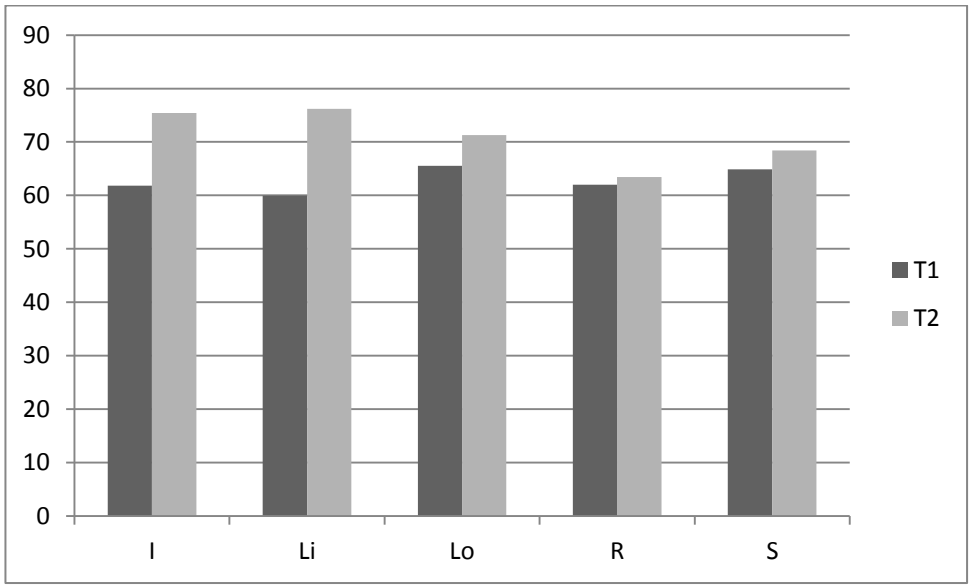
As shown by Table 31 and illustrated by Graph 14, Graph 15, Graph 16 and Graph 17 the individual results reflect the high significance of the group results.

		Iris		Lily		Lola		Rose		Sally	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
PTR	Mean	61.85	75.43	59.97	76.2	65.56	71.27	62.00	63.42	64.88	68.4
	Increase	+22.4%		+27%		+8.9%		+2.3%		+5.4%	
	P value	0.0073*		0.0018*		0.17		0.38		0.07	
MLR	Mean	4.24	6.95	3.92	6.27	4.74	6.15	3.50	3.80	5.06	6.85
	Increase	+63.9%		+61.5%		+29.8%		+8.6%		+36%	
	P value	0.0006*		0.0007*		0.0288*		0.08		0.0016*	
SR	Mean	2.27	3.29	1.88	2.89	2.30	3.14	1.70	1.83	2.56	3.48
	Increase	+44.9%		+53.7%		+34.8%		+7.6%		34.6%	
	P value	0.0001*		0.002*		0.0031*		0.69		0.0013*	
AR	Mean	3.68	4.35	3.14	3.82	3.50	4.4	2.74	2.90	3.94	5.08
	Increase	+18.2%		+21.6%		25.7%		+7.4%		+28%	
	P value	0.0017*		0.0035*		0.0004*		0.19		0.0006*	

Table 31: Fluency measures per learner across tasks at time 1 and time 2

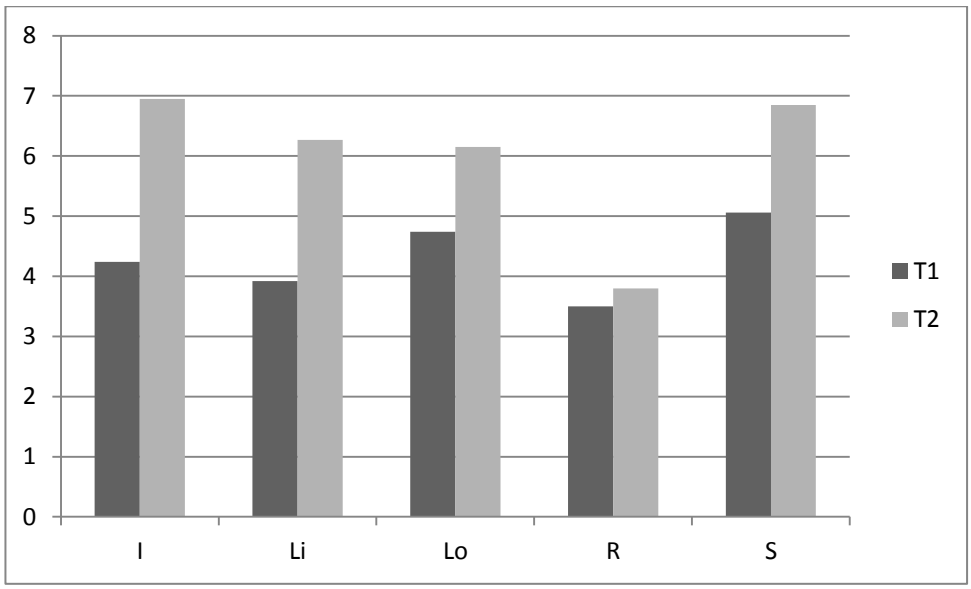
Between time 1 and time 2, as illustrated by Graph 14, Graph 15, Graph 16 and Graph 17, there is an increase for all the learners in all the fluency measures: phonation-time ratio, mean length of run, speaking rate and articulation rate. However, there are differences between the learners in the extent of this increase and the learners can be divided into three groups with regard to the extent and modalities of progression of their fluency. Iris and Lily are the 2 learners for whom the increase in fluency is the most striking. As shown in Table 31, for both of them, the increase on all the 4 different

measures reaches high statistical significance ($p<0.01$). ANOVA shows that for PTR, both Iris and Lily show significantly greater increase than the other three participants, as illustrated by Graph 14.

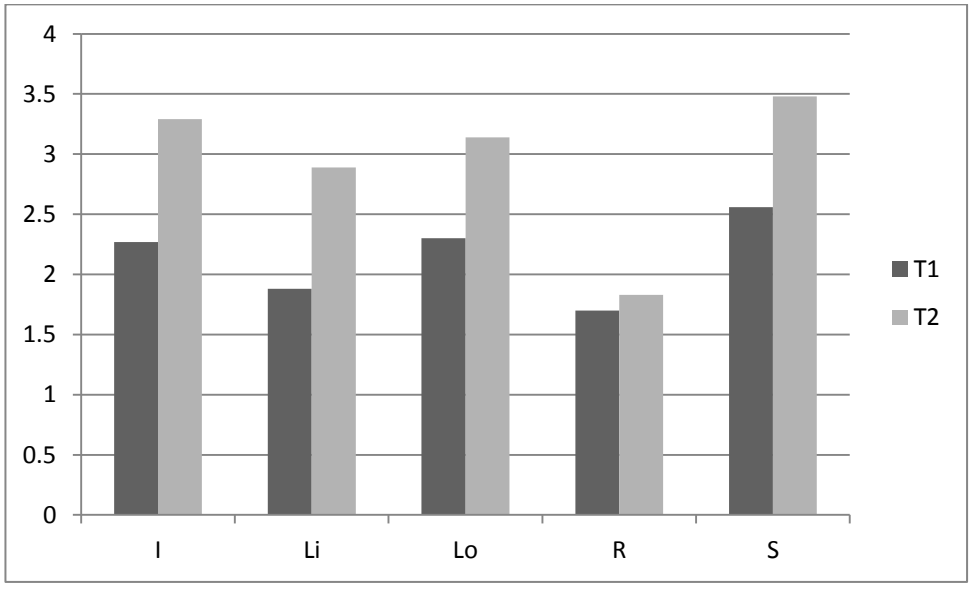


Graph 14: Increase in PTR per learner between time 1 and time 2

Lola and Sally also become more fluent but to a lesser extent, as for both of them, the increase on all fluency measures is highly significant except for PTR. The increase is MLR, SR and AR is highly significant ($p<0.01$ for these 3 measures except for MLR for Lola ($p<05$)). Although their increase in PTR is significantly smaller than Lily's and Iris' (see Graph 10), Sally's increase in AR is significantly greater than Lily's and Iris' and Lola's is significantly greater than Lily's, as illustrated by Graph 17.

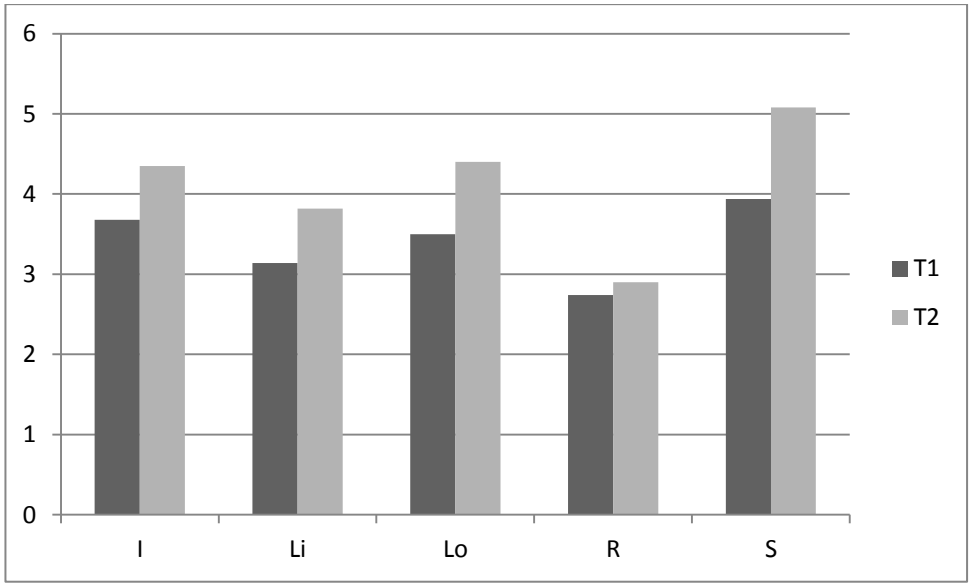


Graph 15: Increase in MLR per learner between time 1 and time 2



Graph 16: Increase in SR per learner between time 1 and time 2

As illustrated by the graphs, Rose is the learner for whom the least development happens and is the only subject for whom the increase in fluency is not significant on any of the 4 calculated measures, as shown in Table 31. In terms of the size of her change in the various fluency measures, she is the learner who differs the most from the others. Indeed, ANOVA shows that for MLR, SR and AR, the changes for Rose were significantly smaller than for the other four subjects, and for PTR, they were significantly smaller than Iris and Lily.



Graph 17: Increase in AR per learner between time 1 and time 2

9.1.4 *Pausing patterns*

Table 32 below gives the group results for average length of pause in seconds (ALP) and average number of pauses per minute (PPM). These results on average length and number of pauses are a way of better understanding how the increase in phonation-time ratio (percentage of time spent talking out of total task time) and speaking rate (number of syllables uttered per second including pause time) takes place as both these measures are directly dependent on the amount of pausing speakers do.

	Mean	SD (subject)	Min (subject)	Max (subject)
ALP Time 1	0.88	0.09	0.79 (S)	1 (R)
ALP Time 2	0.74	0.12	0.65 (I)	0.93 (R)
PPM Time 1	25.45	1.63	23.17 (R)	27.32 (I)
PPM Time 2	23.56	1.82	21.03 (Li)	25.32 (Lo)

Table 32: Group results for average length of pause in seconds (ALP) and average number of pauses per minute

As shown by Table 33 below, for ALP, the ANOVAs show that the reduction in average length of pause between time 1 and time 2 was significant ($F_t(1,4)= 7.76$, $p= .049$). The amount of change was also significantly different across the subjects, (interaction of subject x time $F_t(4,16) = 4.89$, $p=0.009$). PPM also showed a small but significant change ($F_t(1,4)= 8.04$, $p= .047$). The amount the participants changed also

varied significantly between subjects, as reflected by the high level of significance ($F_t(4,16) = 6.64$, p=0.002). For ALP, pairwise comparisons showed that Lily reduced her average length of pause significantly more than all the other participants apart from Iris, and Iris reduced her average length of pause significantly more than Sally. For PPM, pairwise comparisons showed that Lily reduced her number of pauses significantly more than all the other participants apart from Iris, and Iris reduced her number of pauses significantly more than Lola and Rose.

Measure	**Time** df(1,4)	Time x Subjects df(4,16)
ALP	7.78 p= 0.049*	4.89 p= 0.009*
PPM	8.04 p= 0.047*	6.64 p= 0.002*

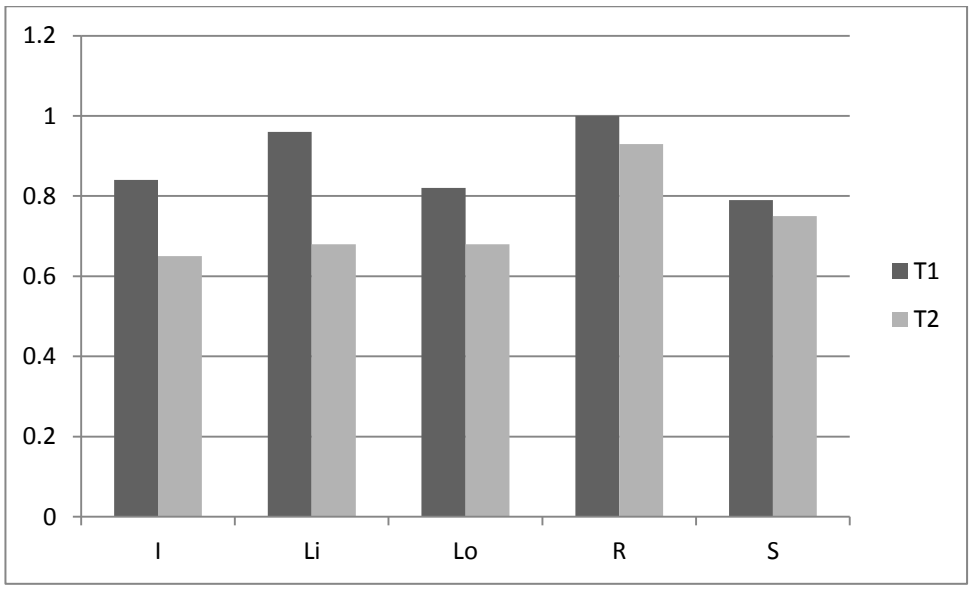
Table 33: Statistical significance of group results for pausing patterns

Let us take a look at the individual results to account for these differences between subjects.

	Iris		Lily		Lola		Rose		Sally	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
ALP	0.84	0.65	0.96	0.68	0.82	0.68	1	0.93	0.79	0.75
Change	-22.6%		-29.4%		-17%		-7%		-5%	
P value	0.0324*		0.0035*		0.12		0.53		0.4	
PPM	27.32	22.58	25.09	21.03	24.97	25.32	23.17	23.64	26.72	25.24
Change	-17.3%		-16%		+1.4%		+2%		-5.6%	
P value	0.0061*		0.0073*		0.79		0.78		0.0453*	

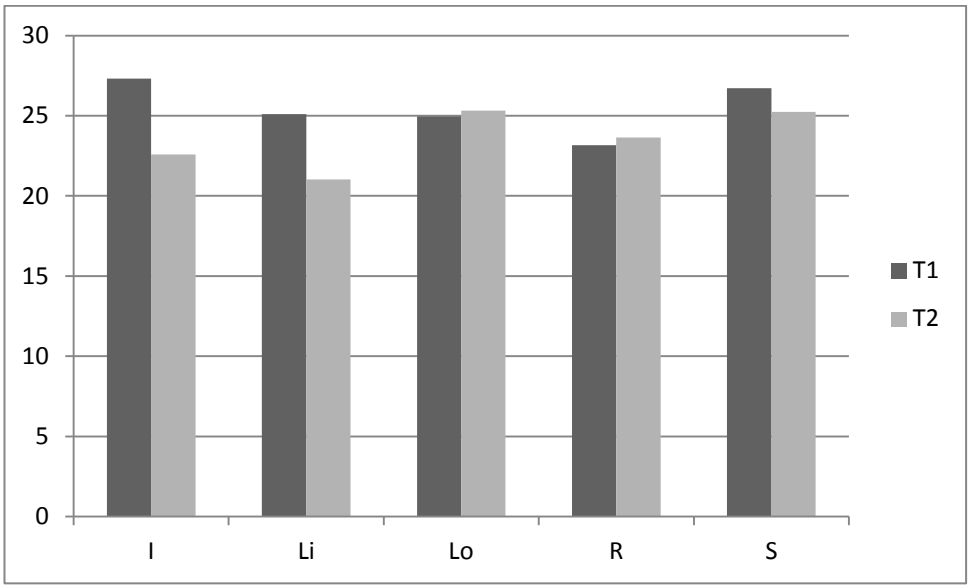
Table 34: Individual results for average length of pause (ALP) and average number of pauses per minute (PPM)

As illustrated by Graph 18, both Iris and Lily reduced their average length of pause more than the rest of the group. Table 34 shows that they were the only 2 learners for whom the reduction in ALP was statistically significant.



Graph 18: Average length of pause (ALP) per learner

Similarly, as illustrated by Graph 19, for the average number of pauses per minute, the decrease is the most striking for both Iris and Lily. This decrease is highly statistically significant for both of them ($p<0.01$), as shown by Table 34. The decrease in PPM is also significant for Sally ($p<0.05$).



Graph 19: Average number of pauses per minute per learner

These figures explain why the phonation-time ratio (PTR) of Iris and Lily increases much more than the other learners' and why this increase is significant only for these

two learners. For Rose, the lack of significance in the decrease of ALP, as well as the fact that there is actually a slight increase (though not significant) in the average number of pauses per minute also contribute to understanding the lack of significant increase in her PTR and SR. As far as Lola and Sally are concerned, neither of them reduces their pausing time as significantly as Iris and Lily. For Lola, the length of pauses does not decrease significantly and a slight increase (though not significant) in the average number of pauses per minute can be noticed. For Sally, the decrease in the average number of pauses per minute is significant but the decrease in ALP is not. In other words, although she pauses less often, the length of her pauses does not decrease significantly. The mixed pausing results for these two subjects contribute to understanding why for them, the decrease in PTR is not as important as for Iris and Lily and does not reach statistical significance. The pausing results also enlighten the results for SR i.e. the number of syllables per second (including pause time). Indeed, for Lola and Sally, the significant increase in SR is probably due less to the decrease in pause time and due more to the significant increase in AR. For Lily and Iris however, the decrease in pause time bears a larger influence on the significant increase in SR.

9.1.5 *Summary of fluency results*

To sum up the results for fluency, there is a general and highly significant increase in all the fluency measures used in this study. However, significant individual differences were found between subjects with regard to the size of the changes for certain measures. One learner, Rose, differed from the rest of the group as her increase in fluency was not significant on any of the measures. Iris and Lily's increase in PTR was higher than the rest of the group and Lola and Sally's increase in AR was higher than for the rest of the group.

	Iris	Lily	Lola	Rose	Sally	Group
PTR	√√	√√	√	√	√	√√
MLR	√√	√√	√√	√	√√	√√
SR	√√	√√	√√	√	√√	√√
AR	√√	√√	√√	√	√√	√√

Table 35: Summary of the group and individual fluency results

9.2 Analysis of the development of fluency

Except for one of the subjects (Rose) for whom the increase in fluency is not significant, the increase in fluency is highly statistically significant for all the other subjects for all the fluency measures used in this study: phonation-time ratio (PTR), mean length of

runs (MLR), speaking rate (SR) and articulation rate (AR). The general increase in all the fluency measures is in line with the findings in previous studies on the development of fluency (Freed et al. 2004).

It is interesting to compare the results obtained in this study with Towell et al. (1996) because they also focused on undergraduates before and after the year abroad and used the same fluency measures although the task they used was a story-retelling task whereas the majority of the tasks used in this study (the 3 discussion tasks and the interview) are of a less constrained nature. As shown by Table 36 the results of the present study are indeed very similar to Towell et al. (1996).

	Present study	Towell et al.	Present study	Towell et al.
	T1	T1	T2	T2
SR	2.15	2.28	2.93	2.61
AR	3.40	3.85	4.11	4.17
MLR	4.30	4.94	6.00	6.06
PTR	62.80	57.74	70.95	62.51

Table 36: Comparison of the fluency results with Towell et al. (1996)

The only measure for which there is a more important difference between the 2 studies is Phonation-time ratio, for which the figures are higher in this study compared to Towell et al.'s. This might be due to the difference in the nature of the tasks used since learners are more likely to pause more in the context of a story retell rather than in the context of an interview or a discussion on a familiar topic. Moreover, pauses longer than 3 seconds were not taken into account for measurement in the present study whereas they were in Towell et al's.

Compared to Towell et al., there is also more increase in the variables measured in this study. This might be due to the type of tasks used. Indeed the effect of the year abroad is more likely to be more obvious in less controlled tasks such as interviews and discussion tasks than in more constrained tasks such as story-retelling tasks. Tasks such as informal interviews are closer to everyday speech, in which the learners are likely to have improved during their stay in France. Moreover, the precise time of testing at time 2 is also a possible factor as Towell et al. did not test their learners while they were still in France and waited for the start of the next academic year after a summer vacation in England. It is therefore likely that their level of performance dropped through lack of practice by the time they were tested at time 2.

Interpretation of the increase in fluency in relation to Levelt's model and theories of automatisation

As explained in chapter 3, what is interesting in the measurement of the temporal variables of fluency is that they are a window into the cognitive processes at stake in speech production. According to Towell et al. (1996), increase in speaking rate can be regarded as evidence of a general automatisation of the processes of speech production. However, they argue that one needs to examine how the different fluency measures pattern in relation to each other in order to get a more detailed insight into which stage of speech production automatisation has taken place. According to them, increase in mean length of run without a simultaneous decrease in phonation-time ratio or an increase in average length of pause can be interpreted as evidence of proceduralisation in the formulator. Indeed, if there is such a configuration of fluency measures, increase in MLR cannot be accounted for by greater planning time. It must therefore be the case that processes in the formulator have become more effective. In this study, a similar pattern in fluency measures can be observed: there is a significant development in mean length of runs, accompanied by a significant increase in phonation-time ratio as well as a significant decrease in average length of pause. As a result, it can be argued that some automatisation has taken place in the formulator. In other words, mechanisms of language production have become more proceduralised.

Because in their results, the increase in MLR is comparatively more important than the increase in PTR and AR, Towell et al. (1996) argue that proceduralisation has happened mainly in the formulator rather than in the conceptualiser or the articulator. In other words, according to them, the changes have been mainly in the way linguistic knowledge is stored as procedural knowledge rather than changes in the way subjects have learnt to conceptualise the knowledge in the L2 or to articulate speech in the L2. However, in the present study, there is a statistically significant increase in all the fluency measures including AR and PTR and I will argue that, if one takes the argument of Towell et al. that automatisation has taken place in the formulation stage of speech production, one can also argue that automatisation has also increased for the conceptualisation and articulation stages. Indeed, the significant increase in AR can be interpreted as evidence of increased automatisation in the articulator. Moreover, the significant increase in PTR (i.e. the decrease in pausing time), although it obviously reflects increased automatisation in the formulator could also reflect greater automatisation in the conceptualiser, as it could be put forward that the learners have

become faster at conceptualising their message. I will argue that the use of FS can be related to faster conceptualisation

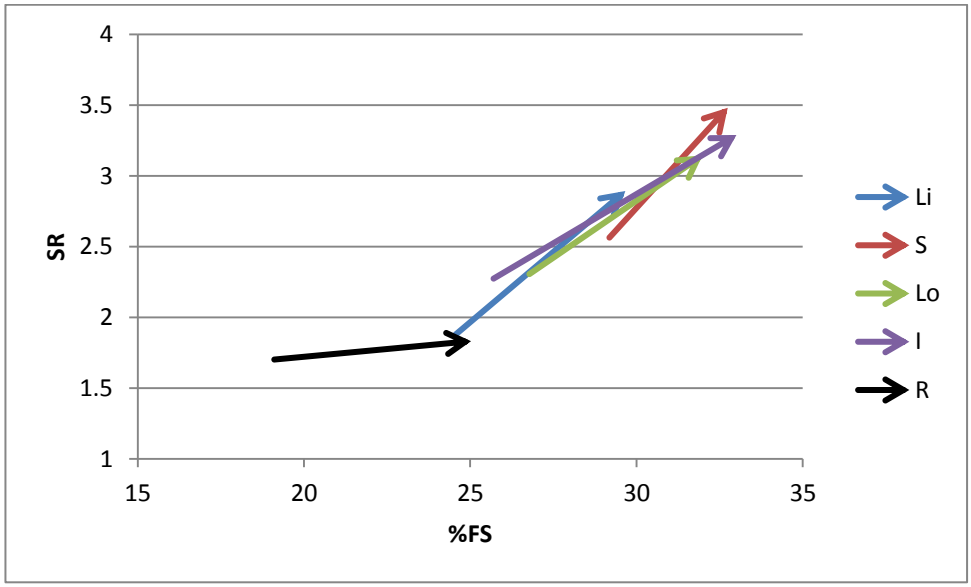
The next section investigates whether FS can be said to play a role in the increased efficiency of these different stages of speech production (formulation, conceptualisation and articulation) by focusing on the correlations between FS use and the various fluency measures.

9.3 Correlations between FS use and fluency variables

In order for the comparative development of FS use and fluency to be investigated, several graphs designed according to the same principle as those used for the correlations between FS use and lexical diversity (chapter 8) will be presented.

9.3.1 *Comparative development of FS use and speaking rate*

Graph 20 represents the development of the percentage of FS (%FS) relative to speaking rate (SR).



Graph 20: Development of percentage of FS compared to the development of speaking rate (SR)

As shown by Graph 20, the percentage of FS and the learners' speaking rate develop conjointly. All five learners increase their FS% (on the graph they move to the right) by similar amounts. All five learners also increase their SR: that is they move upwards on the graph, but they do so to a different extent: as shown earlier, Rose's SR changes significantly less than the other learners'.

The graph speaks for itself in the sense that 3 out of the 5 developmental lines are virtually superimposed, showing that the relative development of these 2 variables is similar for Iris, Lily and Lola. Sally's developmental path is very close to that of the 3 above-mentioned learners but the increase in SR in relation to %FS is comparatively larger. On the other hand, Rose's line stands out because her relative increase in SR is smaller and not significant, in addition to the fact that for her, the values for the variables are altogether lower. The other striking feature of this graph is that all of the observations appear to fall on a single curvilinear trajectory, with greater changes in SR with higher %FS. This is really interesting as it could have been otherwise since, statistically, there is no reason why there should be a single trajectory unless there is a single developmental path. In other words, it can be argued that, although Rose is at an earlier stage than the other learners, given time, as she automatises more FS, her speaking rate will increase since, as shown by Table 37, %FS and SR correlate.

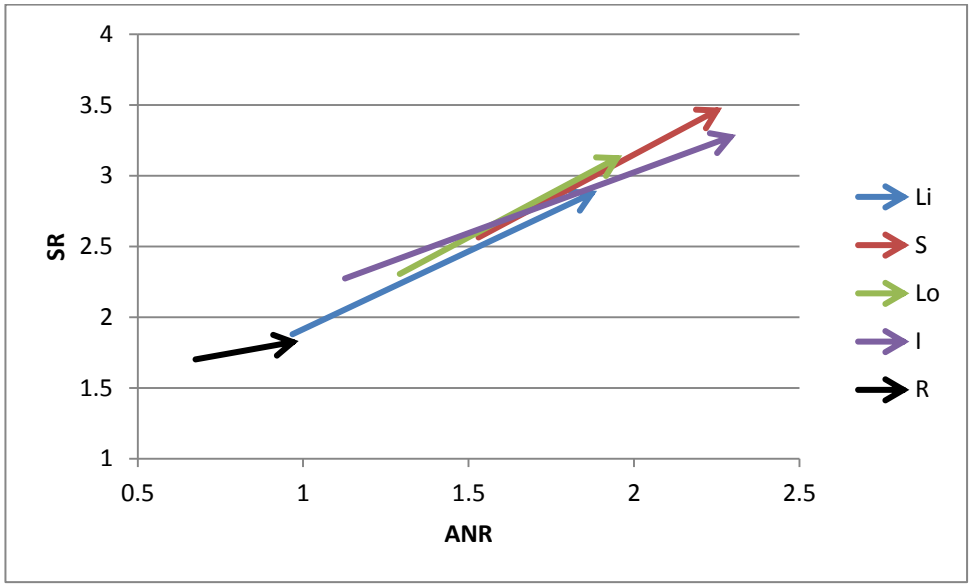
r	SR 1	SR 2	SR change
%FS 1	0.646 p=0.002*	0.661 p=0.001*	0.436 p=0.048*
%FS 2	0.542 p=0.011*	0.635 p=0.002*	0.492 p=0.023*
%FS diff	-0.184 p=0.425	-0.123 p=0.596	-0.022 p=0.923

Table 37: Correlations between speaking rate (SR) and percentage of FS (%FS)

Indeed, statistical correlations investigating the relationship between SR and %FS (Table 37) show that there is a significant correlation between the values of %FS at both time 1 and time 2 and the values of the SR at both time 1 and time 2. In other words the higher the percentage of FS, the higher the speaking rate is likely to be.

There is no significant correlation between the size of the change in %FS between time 1 and time 2 and the size of the change in SR. However, the importance of the change in SR correlates significantly with %FS at time 1 and time 2 ($r=0.436$, $p=0.05$ and $r=0.492$ $p=0.02$ respectively). In other words, learners who are, in absolute terms, more formulaic at time 1 and at time 2 are significantly more likely to undergo a greater change in SR between time 1 and time 2 than those for whom the level of FS use is smaller.

The above results are reinforced by the significant correlations between ANR (average number of formulaic syllables per run) and SR as shown by Table 38 and illustrated by Graph 21.



Graph 21: Comparative development of ANR and SR

Again, as shown by Graph 21, the superimposition of the lines representing each learner illustrates their similar development, except for Rose, for whom values for both variables are lower.

r	SR 1	SR 2	SR diff
ANR 1	0.859 p<0.001*	0.721 p<0.001*	0.335 p=0.137
ANR 2	0.732 p<0.001*	0.946 p<0.001*	0.802 p<0.001*
ANR diff	0.134 p=0.564	0.546 p<0.001*	0.728 p<0.001*

Table 38: Correlations between average number of formulaic syllables per run (ANR) and speaking rate (SR)

Table 38 shows that the values of ANR at both time 1 and time 2 correlate significantly with the values of SR at both time 1 and time 2. In other words, the higher the ANR at time 1 and time 2, the higher the SR at time 1 and time 2: learners who are more formulaic at both times, are more likely to have a higher speaking rate at both times.

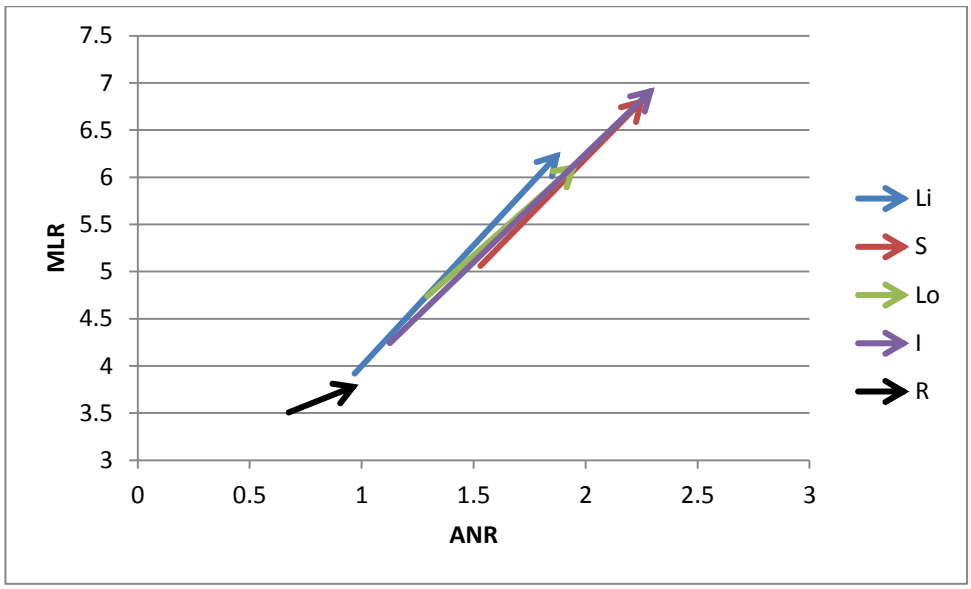
Moreover, not only a very significant correlation (r=0.80, p=0.00) was found between the value of ANR at time 2 and the importance of the change in SR but also between the size of the change in ANR and the size of the change in SR (r=0.73 p=0.00). In other

words, the higher the value of ANR at time 2, the greater the change in SR; but also, the greater the change in ANR, the greater the change in SR. This means that the learners who are the most formulaic at time 2 and who change the most in level of FS use between time 1 and time 2 are those who are the most likely to increase their speaking rate.

As was previously argued, increase in SR can be interpreted as evidence that the processes of speech production are carried out with more ease and efficiency. The highly significant correlations between FS use and speaking rate suggest a role for FS in the general development of fluency. However, as noted by Towell et al. (1996), speaking rate is a very broad measure of fluency and does not allow to pin down where proceduralisation has taken place in the speech production process (See chapter 3 for more details on Towell et al.'s line of argument). This is why the correlations between FS use and the other fluency measures need to be examined in order to pin down more precisely the role of FS on the development of fluency.

9.3.2 *Comparative development of FS use and mean length of runs*

Graph 22 illustrates the comparative development of ANR and MLR. In the same way as for the previous graphs, the superimposition of the developmental lines is striking and shows that all the learners except Rose develop in a very similar way. Again, as can be seen from the inclination of the arrow, the relative increase in MLR in relation to the increase in FS use is lower for Rose than for the other subjects and the values of both variables are also altogether lower than for the other learners.



Graph 22: Comparative development ANR and mean length of runs (MLR)

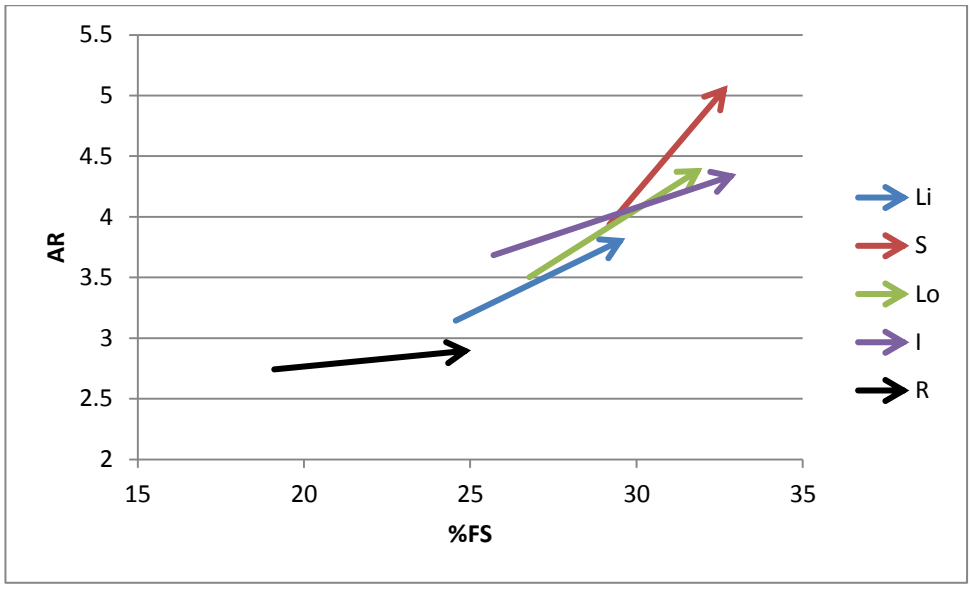
As shown by Table 39, the values for ANR at both time 1 and time 2 correlate significantly with the values for MLR at both time 1 and time 2. In other words, the higher the ANR, the higher the MLR. Moreover, the value for ANR at time 2 correlates significantly with the amount of change in MLR between time 1 and time 2: the higher the value for ANR at time 2, the greater the change in MLR between time 1 and time 2. Finally, the amount of change in ANR also correlates with the amount of change in MLR, which means that the more a learner increases their ANR between time 1 and time 2, the more they are likely to increase their MLR. All these results show that the development of FS use significantly contributes to the production of longer runs of speech.

r	MLR1	MLR2	MLR diff
ANR 1	0.888 p=0.000*	0.635 p=0.002*	0.152 p=0.510
ANR 2	0.581 p=0.006*	0.942 p=0.000*	0.731 p=0.000*
ANR diff	-0.091 p=0.694	0.624 p=0.003*	0.810 p=0.000*

Table 39: Correlations between ANR and MLR

9.3.3 *Comparative development of FS use and articulation rate*

As illustrated by Graph 23, there are individual differences in the learners' comparative development of percentage of FS and articulation rate. Indeed, as can be seen from the inclination of the arrows, Lola's and especially Sally's articulation rates develop more than the other learners' compared to the development of their % FS. Again Rose stands out as the values for her %FS and AR are both lower than for the other learners and, as shown by the inclination of the arrow, the comparative development of her AR in relation to her change in %FS is the lowest of the group.



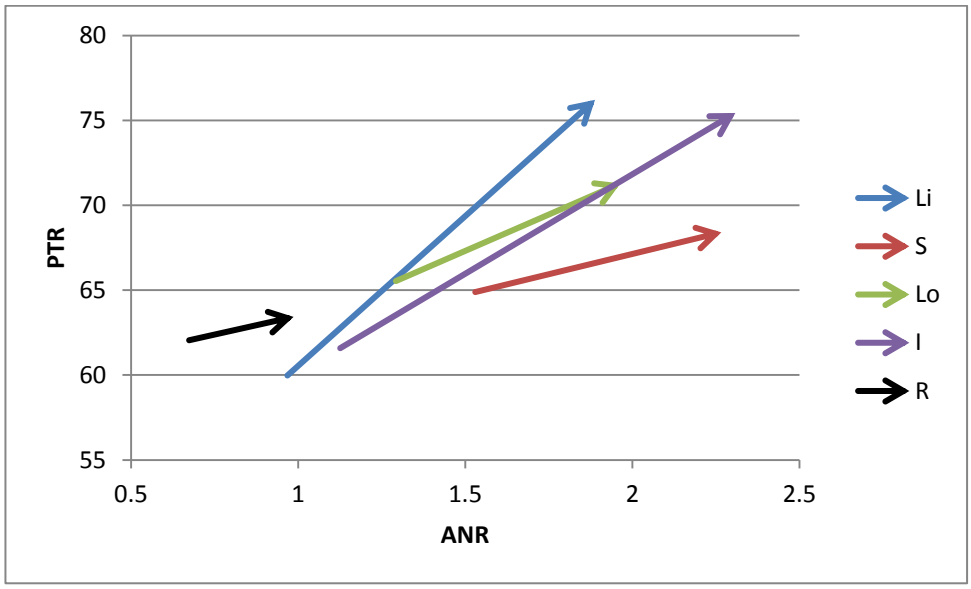
Graph 23: Development of FS use compared to the development of articulation rate (AR)

As shown by Table 40, significant correlations were found between the values for percentage of FS at both time 1 and time 2 and the values for AR at both time 1 and time 2 as well as between the values for percentage of FS at time 1 and time 2 and the importance of the change in AR. However, the correlation between amount of change in percentage of FS and amount of change in AR was not significant. In other words, it is the absolute values of percentage of FS (at both times) and not the amount of change in percentage of FS that are significant predictors of the absolute values of AR as well as the importance of the change in AR. The more learners use FS, the higher their AR and the more likely they are to increase their AR between time 1 and time 2.

r	AR 1	AR 2	AR diff
%FS 1	0.587 p=0.005*	0.664 p=0.001*	0.612 p=0.003*
%FS 2	0.686 p=0.001*	0.636 p=0.002*	0.436 p=0.048*
%FS diff	-0.011 p=0.962	-0.125 p=0.589	-0.235 p=0.305

Table 40: Correlations between %FS and AR

9.3.4 *Comparative development of FS use and phonation-time ratio*



Graph 24: Relative development of ANR and PTR

As can be seen from Graph 24, although all the learners develop in the same direction, since there is a significant increase in both variables, individual differences can be noticed in the comparative development of ANR and PTR. Indeed, the difference in inclination of the lines shows that for some subjects, in this case Iris and Lily, the increase in PTR in relation to the increase in ANR is stronger than for the other learners. For Lola, Rose and Sally, the developmental lines are nearly exactly parallel which shows that, although the values are lower for Lola than for Sally and are altogether much lower for Rose, the relative development of the 2 variables is similar for these 3 subjects and that in their case, the increase in PTR in relation to the percentage of FS is lower than for Iris and Lily, especially in the case of Rose.

r	PTR 1	PTR 2	PTR diff
ANR 1	0.587 p=0.005*	0.186 p=0.420	-0.205 p=0.373
ANR 2	0.109 p=0.638	0.638 p=0.002*	0.401 p=0.072
ANR diff	-0.421 p=0.058	0.656 p=0.001*	0.721 p=0.000*

Table 41: Correlations between ANR and PTR

As shown by Table 41, the values of ANR at time 1 and time 2 significantly correlate with the values of PTR at time 1 and time 2 respectively ($r=0.587$, $p<0.01$ for time 1 and $r=0.638$, $p<0.01$ for time 2). At both times, the higher the value of ANR, the higher the value for PTR. Moreover, the size of the change in ANR between time 1 and time 2 significantly correlates with the value of PTR at time 2 ($r=0.656$, $p=0.001$) and also with the size of the change in PTR ($r=0.721$, $p=0.000$). This means that the greater the change in ANR, the greater the change in PTR and the greater the value of PTR at time 2.

To sum up the correlations between FS use and fluency, FS use (operationalised here as percentage of FS (%FS) and average number of formulaic syllables per run (ANR), correlates strongly with all the fluency measures used in this study. For all the measures, the absolute values corresponding to FS use significantly correlate with the absolute values of the fluency measures. The higher the level of formulaicity, the higher the level of fluency. Moreover, the absolute values corresponding to FS use significantly correlate with the size of the change in all the fluency measures except for PTR. In other words, the higher the level of formulaicity, the greater the change in fluency. Finally, the importance of the change in level of formulaicity (operationalised as ANR) significantly correlates with the importance of the change in all the fluency variables except AR, which means that, on a general basis (except for AR), the greater the increase in formulaicity, the greater the increase in fluency.

It could be said that the correlation between formulaicity and fluency is not surprising given the fact that the main identification criterion of FS used in this study is phonological coherence (operationalised as fluent pronunciation). However, such a correlation is not as straightforward as it might seem. Indeed more fluently pronounced FS could also have meant more, or longer, pauses before or after them, which was not the case.

Finally, one has to comment on the remarkable shape of some of the graphs representing the learners' comparative development of FS use and fluency. Indeed, on several of them, all of the observations appear to fall on a single curvilinear trajectory, with greater changes in fluency measures with either higher %FS or higher ANR. These graphs, similarly to the graphs representing the relative development of FS use and lexical diversity (see 7), suggest that the learners are following a single developmental path although they follow it at different rates, as exemplified by Rose's slower development. This interpretation of a single developmental trajectory would of course need to be confirmed as it is only based on 5 learners. However, although this interpretation can only be tentative, it is based on a principled identification of FS in systematic large speech samples recorded before and after the learners' time in France.

9.4 Qualitative analysis to account for the link between FS use and fluency

All the results presented in 9.3 show strong correlations between FS use (in terms of percentage of FS and/or average number of formulaic syllables per run) and fluency (operationalised through SR, MLR, AR and PTR). The aim of this section is to present a qualitative analysis of the data in order to illustrate, with representative concrete examples taken from the corpus of this study, how FS use encourages greater fluency.

9.4.1 *Link between FS and reduction of pausing time*

The link between FS and the reduction of pausing time needs to be explored in order to understand the correlations between FS use and both SR and PTR i.e. two fluency measures which are heavily determined by the amount of pausing time.

9.4.1.1 *Reduction of pausing through fillers*

FS, when they are used as fillers, contribute to the reduction of pausing time, as fillers are used instead of pausing. The following examples epitomise how a filler, inserted within a given utterance, can contribute to the reduction of pausing time:

*de temps en temps il y a……**je sais pas** que des frites avec…*(from time to time there are…**I don't know**…only chips with.....) (Lo Fd2)

*je vais**…..je vais**….**je sais pas**….reconnecter avec l'Angleterre* (I'm going…**I'm going**….**I dunno**….to reconnect with England) (Lo Inter 2). In this example, the repetition of the sentence builder *je vais* is also used as a filling strategy.

The filler can sometimes be inserted between the repetition of the word after which the communication breakdown happens, as in the following examples:

*et le 2ème stage c'est…..**je pense que** c'est….à Annecy* (and the second placement it's…**I think that** it's…..in Annecy) (I Inter 1)

*j'avais trop peur parce que j'pensais que….**je sais pas** que….que j'pourrais pas faire mon travail* : I was really scared before I thought that…**I don't know** that… that I wouldn't be able to do my job (I Inter 2)

In the 2 examples above, a filler is inserted when the speaker encounters a communication difficulty. In both examples, the filler is followed by the repetition of the word used before it (*c'est* in the first example and *que* in the second example). So in both cases, the filler acts as a time-gaining strategy which enables the learners to gain more time for planning the next bit of speech.

Moreover, as illustrated by the 2 examples below, rather than producing a long pause, a filler can also be a way of closing an utterance when the speaker is left with nothing to say or does not know how to say what they want to say.

*mille livres en ce moment **je pense ou quelque chose à peu près comme ça*** (1000 pounds at the moment **I think or something kind of like that**) (S Fd 2)

je sais pas** la littérature **les choses comme ça (**I dunno** literature **things like that**): 2 fillers in one run (S Inter 1)

The 2 above examples, on top of containing a filler placed at the end of the utterance, also illustrate the use of several fillers in one single utterance.

All the above examples illustrate how the use of one or more fillers, though it does not eliminate pausing time completely, can contribute to its reduction. They obviously contribute to the increase in MLR as they make the runs longer. Moreover, it must be pointed out that in all these examples, the fillers are not completely empty of meaning but are used to edge statements and express uncertainty.

9.4.1.2 *Reduction of pausing time through repetitions*

Additionally to the use of fillers, pausing time is reduced through the repetition of whole FS or parts of FS.

- **Repetition of the whole FS**

Repeating a FS is a way of gaining time in order to finish formulating one's message, but without pausing for too long, like in the example below.

en ce moment en ce moment… j'attends (at the moment at the moment…I'm waiting) (R Inter 2)

In many cases, the FS which is repeated is then followed by one or more words. In this respect, this time-gaining strategy through repetition can be integrated within a more general strategy of incremental sentence building.

- **Repetition of the whole FS with an addition**

je sais que*…****je sais que*** *je….je suis heureuse* (**I know that**…**I know that** I…I'm happy) (R Fr 2)

les deux autres sont ***en train de****….****en train de*** *peindre* (the other two are **in the process of…in the process of** painting) (I LN2)

c'était difficile…c'était difficile à rencontrer des Français (it was difficult….it was difficult to meet French people) (S Inter 2)

ce n'est pas un problème….. ce n'est pas un problème majeur en France (it's not a problem…it's not a major problem in France) (Lo Inter 1)

un peu parce que ***je ne sais pas….je ne sais pas*** *où…… où j'irai* (a little because **I don't know…I don't know** where...where I will go) (Lo Inter 1)

- **Repetition of the last word of the FS:**

There are also many examples of repetitions of the final preposition of the FS when the FS is a sentence-builder

j'aime beaucoup l'idée de…..de le faire… (I really like the idea of….of doing it) (S Fd 1)

au niveau de*…****de*** *ce qu'on pense* (**at the level of…of** what we think) (Li Fd 2)

fait attention à*…****à*** *faire les choses* (**pays attention to…to** doing things) (R Inter 1)

ils ont hâte de *….****de*** *voir* (literally **they have haste to…to** see= they look forward to seeing) (Li LN 2)

In many examples, the repetition of the preposition is necessary for the elision of the preposition to take place:

j'ai besoin de*….****d****'avoir une formation* (**I need to…to** 'have a training'=to be trained) (R Inter 2)

au niveau de / du *remboursement* (**at the level of/of** the refund) (I Alc 2)

All the above examples show that whole or partial repetitions of FS can be regarded as part of a general strategy of incremental sentence building on the part of the learners. FS therefore play a central role in the sentence-building process when the learners proceed by repeating the FS and making progressive additions to them in order to build a sentence.

9.4.1.3 *FS as an easy way of starting an utterance*

FS can be used as an easy and time-efficient way of starting an utterance even when the end of the message has not been fully planned yet. Many FS classified as sentence builders fit this purpose especially well. They are a way of not only starting a message but also of gaining time while the rest of the message is constructed. In this respect they can be regarded as time-gaining strategies which also contribute to the reduction of pausing time since they enable some utterances to be started faster than if they were preceded by long pauses.

c'est bizarre que*…que le personnel…..essaye de ……prévenir* (I Alc 1)

c'est vraiment difficile de*…..s'impliquer dans la vie culturelle* (S Fr 1)

Fillers are also used as ways of starting an utterance:

j'sais pas *c'était…..j'crois qu' c'était……excitant* (L Inter 2)

je sais pas *donc c'est pour ça* (S Inter 1)

It must be pointed out that sentence builders as entries into utterances are used in a particularly idiosyncratic way by Sally. Indeed, on several occasions, she noticeably slows down her articulation rate when she uses an FS at the start of an utterance like in the following examples: *Je pense que* (Fd1), *je dirais* (Alc2), *il y avait* (Inter 1), *pour moi* (Inter 1). In all these examples, Sally clearly uses FS as a way of starting an utterance while gaining time to think about the rest of the message.

Although they might be efficient in Conversation Analysis terms, some FS used as a way of starting an utterance are not always that efficient in the sense that they are false starts that need to be rephrased like in the following examples:

je crois / je voudrais…..j'ai pas….je sais pas si j'ai progressé beaucoup (Lo Fr 2)

In the above example, 3 FS are in fact false starts before the message is finally expressed

*c'est vrai que…c'est vrai que….**il y a**….si on est touriste dans les lieux touristiques* (Lo Fr 2)

In the above example, *il y a,* instead of working as a sentence builder, is a false start, which is followed by retracing.

The most frequent FS used as a false start is *c'est* as illustrated by the following examples:

***c'est**/je voudrais être dans une ville* (Li Inter1)

***c'est** / je pense qu'il y a* (I Alc 1)

***c'est** / j'aime bien la variété* (I Inter 1)

*et **c'est**/je serai réceptionniste* (I Inter 1)

*je pense que **c'est / c'est** / il y a plus de pres / de pression* (I Alc 1)

It can be pointed out that many common subject+verb units that have been identified as FS in this study work as false starts. In the following examples, it appears that some verbal forms are more automatised than others like in the following examples:

***elle est** / elle parle* (I Fr 2)

*à l'école **on a**…..on avait* (I Inter1)

***il a** / il avait continué à payer* (I Alc2)

***on est**…on était obligé de….d'apprendre le français* (R Fr 2)

In the above examples, FS are more automatised because they contain either a more common verb (*est* versus *parle*), or a more common tense (present versus imperfect)

9.4.2 *Negative effect of lack of automatisation on fluency*

After illustrating the positive effect of FS use on fluency, it is interesting to take a look at sequences which would have been considered as formulaic if they had been delivered in an automatic way. However, these sequences were not considered formulaic in the productions of certain learners because they were uttered in a hesitating way or there was variation in their form, which, in both cases, was interpreted as a lack of automatisation. It is essential to point out that both occurrences of variation and hesitation go hand in hand with pauses. For example, in the case of Rose, there are many instances when she has not automatised a sequence e.g. prep + Noun for countries. For example, at time 1 she keeps hesitating between *en France* and *à France* or *en Newcastle* and *à Newcastle*. In the same way, she has not automatised units such as *à pied* as she produces both *à pied* and *au pied*. Other examples of variation can be found in Lily's productions *à / en Londres* (Inter 1); *jouer le/du tennis* (Inter 1 and Fd 2). In these cases, it is possible to argue that the problem is situated at the level of metalinguistic knowledge (e.g. the learner does not know that the preposition preceding a town should be *à* and not *en*) rather than at the level of automaticity of retrieval. It must be pointed out that the 2 are linked since if a learner is hesitating between 2 conflicting rules, it means that they have not stabilised on one yet, and therefore cannot have automatised it. However, whatever the root of the variation/hesitation/variability may be, what is important is that these sequences are not automatic FS, as they are in the productions of some of the other learners and that, as a result, this lack of automaticity goes hand in hand with hesitations and occurrence of pauses, thereby having a detrimental effect on the learners' overall fluency like in the following example: *j'habite dans la….à la campagne ?* (I Inter 2)

It must be pointed out that such example of dysfluent variations and hesitations can still be found at time 2 in the case of some (or all?) of the learners e.g. *jouer le / du tennis* (Li Fd 2), *à/au début* (Li Inter 2), *au / en bus* (Li Inter 2)

Moreover, even in the case of correct sequences such as *tous les jours*, *tout le monde*, *beaucoup de gens*, *je n'ai pas le temps*, there can be a lack of automatisation which manifests itself through micro-pauses within the sequences. These sequences could be regarded as weak psycholinguistic FS in the sense that they are sequences in the process of becoming automatised but the strength of the association between the various members of the sequence is still very weak Hay (2001, 2002).

To conclude, the use of FS contributes to the reduction of pausing time in varied ways. FS also encourage fluency by contributing to the increase in MLR, as demonstrated in the next section.

9.4.3 *Increase in MLR thanks to FS use*

As was previously explained, increased MLR without increase in pausing time shows that the subjects are able to process more language within a single speech run without any additional planning time. As a result of this particular configuration of fluency measures, it has been argued that proceduralisation has taken place in the formulator. As Towell (2002: 121) puts it, 'if the subject manages to produce longer runs without at the same time pausing more, it can plausibly be argued that there must be some change in the way language is stored in order for the speaker to produce more of it in a single run'. Towell contemplates 2 (possibly complementary) possibilities to account for this linguistic change. Firstly, the fact that the learner has created ' productions' in the formulator (i.e. condition/action pairs which are very fast to access, see section 3.5.2.1.) and secondly, the fact that 'the learner has created formulaic language, perhaps stored in the lexicon, which allows quick access to standard phrases'. The results of this study can be interpreted as confirming Towell's hypotheses that FS play a role in the increased efficiency of the formulator. Indeed, significant correlations were found between FS use and MLR, suggesting that FS use play a role in the increase of MLR.

To a certain extent, such an influence of FS use on the development of MLR is to be expected considering that one of the main characteristics of FS (as they are defined in this study) is that they are sequences pronounced fluently and uninterrupted by pauses. However, this influence is not necessarily straightforward as the fact that a subject uses more FS does not necessarily imply that they are uttering longer runs. Indeed a subject could use a lot of short FS, preceded or followed by a pause, which would not necessarily result in them producing longer fluent runs of speech. Nonetheless numerous examples can be found in the subjects' oral productions which show that FS use promote longer fluent runs uninterrupted by pauses.

Firstly, the detailed examination of the longest fluent runs produced by the subjects shows that they often correspond to several FS combined together like in the following examples:

c'est difficile de temps en temps à Newcastle il y a beaucoup de gens qui (it's difficult from time to time in Newcastle there are a lot of people who) (Li Fr 2): 19 syllables:

j'aime bien l'économie *mais* ***c'est très difficile***: (I like well the economy (I like economy) but it's very difficult) (I Inter 1): 11 formulaic syllables

o***n a pas beaucoup de choses en commun on a pas envie de*** *passer beaucoup de temps ensemble*: we don't have much in common, we don't feel like spending a lot of time together (I Inter 2): 16 formulaic syllables

je crois qu'en France en ce moment: I think that in France at the moment (Lo Alc 2) : this sequence made of 3 FS combined together allows the speaker to utter already 8 syllables even before expressing the main part of the message.

j'ai beaucoup aimé la ville *parce que* ***il faisait chaud*** (I really liked the city because it was hot) (Lo Inter 2): run of 13 syllables, 11 of which were identified as formulaic.

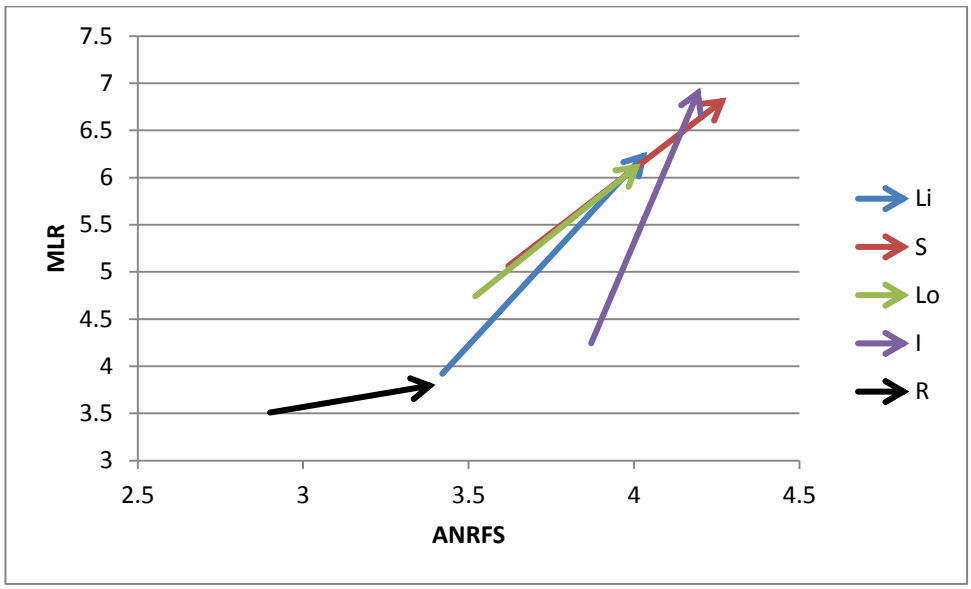
oui ***je crois en Angleterre on peut dire*** *ah* ***je vais faire du shopping c'est*** *mon hobby*: (yes I think in England one can say ah I go do some shopping (I'm going shopping) it's my hobby) (Lo Inter 2): run of 21 syllables, 16 of which were identified as formulaic.

je vais à la fac en Angleterre je n'ai pas besoin d*'aller ici* (I go to university in England I don't need to go here) (Lo Inter 2): 15 formulaic syllables

As shown in 9.4.1.1, another modality of the production of longer runs thanks to the use of FS is through the insertion of fillers in fluent runs. For example, Sally, the most fluent learner of the group recurrently integrates fillers within fluent runs, which add syllables to the syllable count like in the following examples. In Inter 1, she uses *je sais pas* several times at the end of fluent run, which adds to the syllable count.

The above examples show that the combination of several FS within a single run of fluent speech and the insertion of fillers encourage longer runs of speech. This qualitative analysis is confirmed by Graph 25 which represents the comparative development of ANRFS (average number of formulaic syllables per run containing at least one FS) and MLR. As previously explained, ANRFS aims to examine the development of the subjects' ability to use more formulaic material within a single run either by using longer FS or by combining several FS. The graph shows that ANRFS and MLR develop hand in hand. Again, individual differences can be noticed in the inclination of the arrows. Iris's development in MLR relatively to ANRFS is greater

than for the other learners. Again, Rose's development in MLR relatively to ANRFS is smaller than for the rest of the group, and the values for her on both variables are lower.



Graph 25: Comparative development of ANRFS and MLR

r	MLR 1	MLR2	MLR change
ANRFS 1	0.521 p=0.015*	0.513 p=0.017*	0.258 p=0.259
ANRFS 2	0.473 p=0.030*	0.721 p=0.000*	0.540 p=0.012*
ANRFS diff	-0.145 p=0.530	0.078 p=0.736	0.194 p=0.401

Table 42: Correlations between ANRFS and MLR

As shown by Table 42, significant correlations were found between the values for ANRFS at both time 1 and time 2 and the values for MLR at both time 1 and time 2 as well as between the values for ANRFS at time 2 and the importance of the change in MLR. However, the correlation between amount of change in ANRFS and amount of change in MLR was not significant. In other words, it is the absolute values of ANRFS (at both times) and not the amount of change in ANRFS that are significant predictors of the absolute values of MLR at both times. Moreover, the higher the value of ANRFS at time 2, the greater the change in MLR is likely to be. These significant correlations confirm the link between FS use and MLR put forward in the qualitative analysis.

9.4.4 *Link between FS use and increased articulation rate*

As presented earlier in the fluency results, there is a significant group increase in the articulation rate (AR). This is why, although, according to Towell et al. (1996), proceduralisation mainly happens in the formulator, according to this study, it also happens in the articulator. Moreover, the results of the correlations above have shown that there is a significant correlation between FS use and AR: the higher the percentage of FS, the higher the articulation rate. Without doing further quantitative analyses, it is difficult to describe in detail how the use of FS affects the AR. For example, does the increase in AR only apply to the FS itself or does it have a global positive impact on the speed of articulation of the syllables around the FS? In order to answer these questions reliably, one would need to compare the articulation rates in different types of runs e.g. runs entirely formulaic, entirely non-formulaic, composed of both formulaic and non-formulaic material. Such detailed analyses could not be performed for the present study although they would be of interest for further research on the link between FS use and AR.

At the level of the qualitative analysis however, there are some striking examples in the data (in the sense that they are noticeable by ear) of some FS which coincide with an increase of the articulation rate. This is particularly obvious in the case of Rose, who clearly accelerates her articulation rate when pronouncing certain sequences. It is possible that it is particularly noticeable in her case as she also happens to be the least fluent of the group. Examples of the FS she uses with an increased AR include: *je pense que* (I think that), *il semble que* (it seems that), *par exemple* (for example), *je ne sais pas* (I don't know), *joue un rôle primordial* (plays a major role), *faire du ski* (to ski). A few examples can also be noticed in the productions of Lola e.g. *par exemple* (for example), *de temps en temps* (from time to time), *c'est amusant* (it's fun).Such striking examples could only be found in the case of these 2 learners.

- **FS use and speeded up articulation through phonetic reductions**

Additionally to the acceleration of the articulation rate, another modality of the speeding up of the phase of articulation through FS use has to do with phonetic reductions. The most widespread reduction, which can be noticed in all the learners' productions, concerns the filler *je ne sais pas*. Several reduced forms of this can be observed: the elision of the [ə] of the negative element *ne* like in *je n'sais pas* (I Inter 1, Lo Fd 2 and Lo Inter 2), or the highly reduced form *chais pas* ☯♦♏□♋☸in which the negative

particle is dropped and the personal pronoun *je* becomes [⧫] as it undergoes schwa deletion and devoicing (I Fd 2 + S Inter 2, Lo Inter 2). It must be noticed that although one example given above is taken from time 1, most of the examples of phonetic reduction are at time 2. For example, the 9 occurrences of *chais pas* are all found at time 2.

Reductions in FS can take the form of the dropping of one of the words in the sequence especially in impersonal expressions such as *il faut* or *il y a* as in the following examples: *faut que* (Li Alc 2), *faut faire un peu d'efforts* (I Inter 2), *y a plein d'* (S Inter 2) and *y avait* (I Alc 2). Phonetic reductions also occur in the personal pronouns of very common 'subject+ verb' units. For example Iris and Sally sometimes do not pronounce the 'l' in *il/ils* in common subject+ verb' units such as *ils ont* (I Alc 1) pronounced [izɔ̃],: *il faut* (I Alc 2 and Inter 2) *ils vont* (I Alc 2) pronounced ☯ivɔ̃], *ils sont* (I LN2 and S Inter 2) pronounced [isɔ̃]. Moreover, *je* is very often reduced in common sentence builders such as *j'pense que* (I Fd 2), *j'trouve que* (S Alc 1), *j'voudrais* (Lo Inter 2). There are also examples of reduction of articles of prepositions within the FS: e.g. *tout l'temps* (Li Fr2), *à la fin d'l'année* (S Inter 2). Finally there are also examples of reductions of a vowel inside the word of an FS such as *ça s'rait* (it would be) (Li Fr2) or, *la s'maine prochaine* (Lo Inter 2).

All the above examples are taken from time 2. They show that, at the end of the learners' time in France, articulation has become more proceduralised for these given sequences a. As explained by Bybee (2002: 216), phonological production is a neuromotor behaviour and phonetic reductions show that this neuromotor behaviour becomes more fluent with practice. Indeed, with practice, some changes take place in the sequences of articulatory gestures, which results in phonological alternations.

In the context of a usage-based framework, however, phonological reductions are not just simple illustrations that articulation is performed faster. According to Bybee (2002), they also have to be regarded as 'evidence for exemplar storage of multiword sequences'. This is due to the fact that phonetic reduction interacts with chunking (Bybee 2010). As proposed by Hay (2001, 2002), each instance of direct access of a multiword unit strengthens the path of access to that whole unit and weakens the access through the component parts. The stronger the sequence will become as a unit, the less associated it will be to its component parts, which results in a gradual loss of analysability. The process of phonological reduction, during which whole (e.g. *ne)* or

parts (e.g. schwa deletion) of component members disappear, epitomises this loss of analysability. Indeed, the fact that some of the members of the units disappear either partly or completely shows that the sequence is now accessed as a holistic unit and not through its individual component parts.

It must be pointed out that examples of phonetic reductions are only found in the case of a limited number of FS. It is beyond the scope of this study to try and account for why this is the case. Very briefly though, reductive sound changes have been shown to affect high-frequency words and phrases earlier and to a greater extent than low-frequency ones. For example, *don't* reduces the most in phrases like *I don't know*, in which it occurs the most often (Bybee and Scheibman 1999). It has also been shown that discourse markers are prime candidates for phonetic reductions (Bybee and McClelland 2005). This would explain why, in this corpus sequences such as *je ne sais pas, il y a* or *il faut* are particularly subjected to phonetic reductions.

9.5 Integrating FS in models of speech production

9.5.1 *Speeding up of encoding*

The aim of this section is to illustrate, with examples from the corpus of identified FS, how the use of FS facilitates the formulation of speech. It was hypothesised in chapter 3 that FS use might speed up the formulation stage in two different ways according to the type of FS involved: by speeding-up of grammatical encoding alone or by bypassing grammatical and morpho-phonological encoding entirely. These two modalities of speech facilitation can be illustrated by many examples from the corpus of FS identified in this study.

Firstly, as detailed in the typology of identified FS, the learners use a number of formulaic frames which are partly-fixed i.e. composed of some fixed parts and some slots that have to be filled. Examples of such formulaic frames include all the multiword VPs such as *faire du shopping* (to do some shopping), *être en colère* ('to be in anger'=to be angry), or *prendre un verre* ('take a glass'=have a drink) as well as sentence builder VPs such as *avoir besoin de* (to have need of=to need). Chapter 3 explained how in Levelt's model, encoding operations are largely controlled by the grammatical properties of the lemmas. Lemma retrieval is thought to activate dedicated syntactic procedures which set up the appropriate grammatical frame. If there is already a syntactic frame, however, the phrase is already partially built and some of the syntactic procedures do not need to take place. If one takes *je fais du shopping* as an example, it

is only the verb which needs to be inflected for the correct person and tense. The rest of the phrase building procedure is already completed since the correct lexeme for the masculine partitive article is already selected as well as the lexeme *shopping*.

It is easy to understand how the fixed part of such a frame prevents many potential problems for L2 learners. Indeed, selecting the appropriate determiner and grammatical gender is a well-known problem for L2 learners, which is at the source of many errors and speech hesitations. In a formulaic frame such as *faire du shopping*, however, the learner does not need to select the correct determiner at the lemma or the lexeme level. At the lemma level, all the lexical pointers are already indexed for diacritic features so the learner does not need to select the correct features (masculine form of the partitive article). As a result, the learner does not have to wonder which type of article they should use or whether the masculine or feminine form of the determiner should be retrieved: in Levelt's terms, the determiner is already indexed with the correct diacritic features. Moreover, encoding is also facilitated at the lexeme level. Indeed there are cases where, although learners have not encountered problems at the level of the surface structure (i.e. they are able to index the lexical pointers of the lemma with the right features (for example here the masculine form of the partitive article), they are unable to retrieve the appropriate form for it at the level of morpho-phonological encoding i.e. the cannot retrieve the form *du*. In the case of the fixed part of a formulaic frame, however, the retrieval of the correct morpho-phonological form is automatic, which means that learners are not faced with problems with retrieving the correct lexical item at the lexeme level.

As explained in chapter 3, Levelt compares grammatical encoding to a process of solving simultaneous equations because the emerging syntactic structure should simultaneously satisfy all the constraints required by the various lemmas. In the context of the use of prefabricated frames, it can then be said that speech production is speeded up because some of the 'equations' involved in creating speech are already solved.

Additionally to the use of syntactic frames, the typology of FS used show that the learners use a wide range of fixed phrases or whole sentences. Examples of such FS include: *c'est une bonne idée* (it's a good idea), *ça va marcher* (it's going to work), *à mon avis* (in my opinion). If one argues that these sequences have a direct memory representation as their ease of retrieval suggests they have, they can be thought of as allowing for the speeding up of both grammatical encoding and morpho-phonological

encoding. In other words, they enable a global speeding-up of the whole of the encoding phase.

9.5.2 *Formulaicity at the conceptual level*

Moreover, it was argued in chapter 3 that, in the case of formulaic fixed phrases, the chunking process has also happened at the conceptual level in the sense that these FS express a meaning/function which can be conceptualised as one unit (e.g. give one's opinion, say one agrees). Such an automatic link between a simple conceptual unit and a ready-made phrase or sentence is likely to happen in the case of frequent and common messages. Therefore, these formulaic phrases which allow for the speeding-up of the whole of the encoding phase are likely to be found when the learners are talking about relatively common topics and expressing widespread ideas. It can therefore be argued that the conceptual phase of speech production is also speeded up as the content of the message is so common and predictable that it has become a kind of conceptual chunk. Formulaicity situates itself both at the conceptual and linguistic level as the multiword linguistic chunk corresponds to a conceptual chunk. This is interesting in the context of the notion of formulaicity conceived of as 'lack of originality' (chapter 2): many automatised sequences correspond to a formulaic idea. In other words, increased fluency does not seem to go hand in hand with increased complexity at the conceptual level.

Chapter 10. Potential effect of the year abroad on development

The aim of this chapter is to relate the learners' development in FS use, lexical diversity and fluency, to a potential role of the year abroad and more particularly to the modalities of the learners' stay in France. The first section presents a summary of the learners' individual patterns of development in the three variables investigated. The second section asks whether this development can be interpreted as being due to their stay in France. The last section aims to relate the learners' development in the three variables to the specific characteristics of their experience in France in terms of their motivation, the quantity and type of input they were exposed to, and the general quality of their experience in France.

10.1 Summary of patterns of development for individual learners

Table 43 below summarises the significance of the individual and group changes in the different variables.

	Iris	**Lily**	**Lola**	**Rose**	**Sally**	**Group**
Quantitative FS use	√√	√√√	√√	√	√√	√√√
Fluency	√√√	√√	√√	√	√√	√√√
Lexical diversity within FS	√√	X	X	X	√√	X
General lexical diversity	√√	√	√	X	√√	√√

Table legend
X No increase or non-significant decrease
√ Slight increase but not significant or significant increase on only one measure (for FS use)
√√ Significant increase or significant increase for 50% of the measures (for fluency and FS use)
√√√ Significant increase on 3 of the 4 measures (for FS use) and on all of the 4 measures (for fluency)

Table 43: Summary of individual and group results

To sum up, putting all the variables together, 3 different profiles of development can be distinguished in the 5 subjects under scrutiny in this study:

- **Iris and Sally**

For these two subjects, there is an increase in all the variables investigated: they both become more fluent with statistical significance on 3 of the 4 fluency measures, their quantitative use of FS increases with statistical significance on 2 of the 4 formulaicity

measures, their use of FS becomes more varied lexically, and their general lexical diversity also increases significantly.

- **Lily and Lola**

For these two subjects, there is a significant increase for all the fluency measures. There is an increase in quantitative use of FS with statistical significance on 2 of the 4 formulaicity measures. General lexical diversity increases but not significantly. However, lexical diversity for FS does not increase and even decreases in Lola's case.

- **Rose**

For this subject, there is no significant increase on any of the investigated variables. There is a slight increase in fluency but it is not significant. There is an increase in the quantitative use of FS but it is not significant except on only one of the 4 formulaicity variables. Finally there is no increase in general lexical diversity or lexical diversity for FS.

10.2 Increase in the different variables and potential role of the year abroad

The study only focused on subjects spending several months in France and did not compare their linguistic development with subjects pursuing their studies at home. Therefore, strictly speaking, it cannot be claimed that the changes noticed between time 1 and time 2 in the subjects' formulaicity, fluency and lexical diversity are due to the effect of the year abroad per se. Indeed, it could be claimed that the recorded increases are simply due to their linguistic development over time and that these increases would have been the same had the subjects remained in their usual university instructional setting.

Being aware of this, it is still reasonable to suggest that although other factors might have been at play, the year abroad is a context which is particularly favourable to the development of the variables studied in this study. The role of the year abroad in the development of fluency has been well documented and the results in the present study are in line with previous findings (Towell 2002, Freed 2004). For example, in Towell (2002), the longitudinal study of the L2 advanced learners of French spans over 4 years. However, Towell underlines that the changes between year 2 and year 3 (i.e. between before and after the year abroad) are much more important and significant than the changes between year 1 and year 2. The year abroad context has also been shown to play a part in the development of lexical diversity and again, the results in this study

confirm previous findings (Foster 2009). As described in chapter 7, some lexical expressions present at time 2 clearly are clearly linked to the subjects' time in France and the input they received there. More precisely, some lexical items or phrases are related to the subjects' experiences for example where they lived or where they worked.

Since the construct of FS (defined psycholinguistically) is both linked to automatisation and lexical development, it seems reasonable to assume that if the year abroad is a favourable context to the development of both fluency and lexical diversity, it is likely to be also favourable to the development of FS use. The strong quantitative increase in FS use during the year abroad as well as the presence of new FS at time 2 (some of them clearly linked to the learners' experience in France) show that it is very likely to be the case.

10.3 Potential link between the modalities of the participants' stay in France and their development in the different variables under scrutiny

Table 44 below presents a summary of the answers given by the participants in the questionnaire about their time in France. It gives details about their degree of immersion in the French environment and gives an estimation of the number of hours they spent speaking French as well as the additional French input they received.

10.3.1 *Summary of the modalities of the participants' stay in France*

	Iris	**Lily**	**Lola**	**Rose**	**Sally**
Position in France	Worked in a youth hostel for 4 months and then in a scientific foundation	Language assistant in a primary school	Language assistant in a 'challenging' secondary school	Erasmus student in law	Language assistant in a secondary school
Living arrangement	Lived in a youth hostel and then on site at the foundation	Rented a room in the house of a French family along with 2 other French students	Rented a room in a 'foyer' for young workers	Shared a flat with a friend from Newcastle University	Shared a flat with a French couple (first month on her own in a studio)

Language spoken in the place of habitation	English at the youth hostel, French at the foundation	French	French / English (but very few interactions altogether)	English	French
Estimated weekly hours of speaking French	39	38	14	7	34
People spoken to in French and estimated weekly hours of speaking French to them	French friends and flatmates at the Foundation (17), colleagues and customers (21), people in shops (1)	People with whom she lived (7), French friends (14), friends non-native speaker of French (14), teachers at school (2), people in shops (1)	French friends (6), friends non-native speakers of French (3), people with whom she lived (1), teachers at school (3), people in shops (1)	French friends (2), University staff and students (2), people in shops (3)	Flatmates (7) French friends (6), friends non-native speakers of French (14), teachers at school (4), people in shops (3)
Days away from France during the stay	7	25	14	35	21
Additional input or activities in French	Watched films or TV every day, read newspapers or magazines and listened to the radio once a week, wrote in French every day (emails to clients)	Watched films or TV every day, listened to the radio every day, read novels every day and magazines and newspapers twice a month	Watched film / TV once a week, read newspapers/ magazines once a week and listened to the radio every day	Watched film / TV twice a month, read newspapers/ magazines twice a month, listened to the radio every day and wrote in French every day for university	Watched film / TV once a week, read newspapers/ magazines twice a week, read novels every day, listened to the radio twice a week

Goal before the year abroad	Improve her French	Become more confident orally, meet new people	Improve her French	Survive!	Improve her French
Motivation (according to the subjects)	Very motivated but only became really proactive in the 2nd part of her stay	More and more motivated to immerse herself as the year went on	Not very motivated to immerse herself, spent a lot of time on her own	First goal was to study law, getting immersed in French context was secondary	Very motivated
Problems encountered	Found it difficult to make French friends in the first part of her stay	Difficulty in expressing herself	Difficulty with getting what she wanted from shops	Difficult to understand the lectures	Unsuccessfully tried to join a sports team
Strategies deployed to deal with these problems	Joined a conversation exchange network	Dealt with language issues by looking for the vocabulary	No particular strategy	Got some help from French students who lent her their notes	Decided to leave sport aside and focus on meeting new people
Regrets	Not to make French friends quicker	Wishes she had met more people earlier	Wished she had gone to a different area of France	Wishes she had lived with French people	Wishes she had focused on French and not sport from the start
Perceived areas of improvement	Speaking and listening	Speaking	Comprehension	Speaking	Speaking, listening and colloquial vocabulary
Perceived area(s) of non-improvement	Writing	Writing	Writing, vocabulary	Writing and grammar	Formal writing

Table 44: Summary of the modalities of the participants' stay in France

Iris	Very positive year mainly due to her second work placement in an ideal setting where she socialised with a lot of native speakers of French. Had such a positive experience in her second work placement that at the second time of data collection, she was planning to return there the following summer
Lily	General feeling of growing strong satisfaction as the year went on and she met more and more people
Lola	Challenging year due to a very difficult work placement and a living arrangement not prone to social contacts. Nearly went home before Christmas. General feeling of not having improved as much as she should have due to all these difficulties
Rose	Despite her awareness of not having immersed herself very much in the French context, general feeling of satisfaction with her year after being highly anxious about it. Proud of having managed to deal with following lectures in French etc...
Sally	Feels that she definitely made the most of her stay

Table 45: General feeling about the year abroad

10.3.2 *Learner profiles in terms of motivation and interaction with French*

- **Iris, Lily and Sally**

As can be seen from Table 44 and Table 45, Iris, Lily and Sally are the three subjects who made the most of their time in France in terms of how much they immersed themselves in the French environment. Despite all needing some initial time to start and meet people, they managed to create a very rich social and linguistic environment for themselves. When visiting them in France, I was impressed by their happiness and enthusiasm about their daily life in France. Iris had very rich social interactions through her second work placement where she was in contact with native French speakers all day. Lily and Sally had some interactions with the teachers in their schools but mainly focused on their social life in order to speak as much French as possible. Lily and Sally's high level of motivation and desire to interact in French can be noticed in the fact that, when they could not socialise with French native speakers, rather than socialising with native speakers of English, they socialised with assistants from other countries and spoke French with them.

- **Lola**

Despite an initially high motivation to improve her French, Lola was faced with a number of issues that made her immersion in the French context much more challenging and frustrating than she had anticipated and more difficult than the three above subjects. She was disappointed with the geographic area where she was appointed as an assistant

as she felt it was not welcoming to foreigners, contrary to the areas where she had previously stayed in France. She did not feel welcome in the high school where she worked, neither by the students nor the teachers and her type of accommodation (a foyer for young workers) was not prone to rich social interactions as people there tended to keep to themselves, as did Lola. When visiting her in France, I noticed her general sense of frustration and feeling of melancholy.

- **Rose**

Rose differs from the other participants in many respects. Firstly, she was the only participant for whom immersing herself in the French context was secondary compared to her primary motivation which was to study law. Moreover, she was highly anxious about going to live in France and as a result chose to share a flat with a fellow native speaker of English. Finally, she did not actively seek contact with native speakers of French. She only had contact with French people when this contact was necessary for her studies or her daily life. For example, she made contact with French fellow students to get help from them or phoned French companies when she needed to sort out various bills. The number of hours she reported speaking French weekly stands out from the other learners' as it is much lower.

10.3.3 *Link between the learners' linguistic development and the modalities of their stay in France*

Relating linguistic development to motivation on the one hand, and to quantity and quality of input on the other hand, is extremely complex. Moreover, it is obviously impossible to draw definite conclusions based on five learners only. Therefore the interpretations that follow are tentative and are by no means intended to account entirely for the learners' development in the different variables investigated. However, while considering, on the one hand, the modalities of the learners' stay in France (Table 44 and Table 45) and, on the other, their development in FS use, lexical diversity and fluency (Table 43), one cannot help noticing certain trends. Firstly, the two learners who improved the most with a statistically significant increase in all the variables are both characterised by their high level of motivation as well as the extremely positive character of their time in France. They clearly made the most of their stay in France and totally immersed themselves in the French context. On the other hand, Rose, i.e. the learner who improved the least in all the variables, is the only learner: (1) whose primary motivation was not to improve her French but to study law; (2) who lived with

a fellow native speaker of English; (3) who was characterised by a high level of anxiety in relation to her stay in France as suggested by the fact that her goal was 'to survive'; and (4) who interacted the least with native speakers of French and was exposed to the fewest hours of input.

Trends are less striking for Lily and Lola whose development is relatively similar but who differ in terms of the quality of their experience in France since on the whole, Lily had a much more positive experience than Lola and was reportedly exposed to much more input.

To conclude, interpreting the link between the learners' development in all the variables and the different contexts of their stay in France can only be tentative given the small number of participants and the high number of confounding factors. However, one might still say that some of the trends noticed between the learners' degree of immersion in the French context and their development in the various variables are in line with Freed et al. (2004: 298) who state that 'it is not the context per se that promotes various types of learning but rather the nature of the interactions, the quality of the experiences, and the efforts made to use the L2 that render one context superior to another with respect to language gain'. Indeed, although the learners under scrutiny in this study all spent several months in France, in effect, they had quite different experiences, which undoubtedly played some part in their development in the variables investigated and might contribute to explain, though only to a certain extent, why some of them developed more or less than others.

Chapter 11. Conclusion

11.1 Rationale, goals and methodological contribution

In the present longitudinal study, I set out to examine the presence, nature and role of formulaic sequences, defined psycholinguistically, in advanced English learners of French, before and after a year spent in France.

This study is based on an essential distinction between speaker-external FS and speaker-internal FS, i.e. between sequences which are formulaic in the language (because of formal criteria or because they occur frequently) and sequences that present a processing advantage for a given speaker. The rationale for this study stemmed from several observations:

1) A majority of previous studies on formulaicity have tended to take for granted that FS defined speaker-externally are also psycholinguistic realities.
2) Most of the research into formulaicity in advanced L2 learners has been characterised by its speaker-external focus, investigating L2 learners' knowledge of speaker-externally defined, idiomatic FS.
3) The role of FS on the development of L2 fluency has hardly ever been investigated, as studies have tended to explain increases in fluency by proceduralisation of grammatical knowledge.

Consequently, this study had several goals.

1) To adopt a clearly speaker-internal approach, emphasising its distinctiveness from speaker-external approaches

2) To investigate the presence and nature of speaker-internal FS from a learner-language perspective without projecting native-speaker assumptions on L2 learners. In other words, rather than testing the processing of speaker-external FS unlikely to be known by L2 learners, this study focused on identifying speaker-internal FS in the oral productions of L2 learners

3) To investigate the longitudinal development of FS during the year abroad and its relationship with the development of lexical diversity and fluency

At the methodological level, the main contribution of this study situates itself at two levels.

1) The study design allowed for the investigation of FS from a learner-internal and psycholinguistic perspective. Firstly, the longitudinal design ensured that the development of FS could be analysed within the same learners. Moreover, the case-study design which focused on a limited number of learners allowed for a large amount of data per individual learner to be collected in order to get a representative insight into the presence, nature and role of FS in the language of these individual learners.

2) This study adopted a hierarchical method of identification with a necessary criterion based on phonological coherence so that the focus on psycholinguistic advantage would be respected. That method of identification was rigorously applied thanks to a precise annotation of sound files using the software Praat.

11.2 Status of the L2 sequences identified in this study

This section summarises the claims that can be made about the sequences identified as formulaic in the present study. It also describes which types of multiword sequences were not classified as formulaic.

Firstly, the sequences identified as formulaic are easily processed as shown by the fact that they are always pronounced as fluent wholes (i.e. uninterrupted by pauses or repetitions or retracing). What cannot be claimed however is that they are lexical units stored whole in the lexicon. It is possible that some of them might be but it cannot be empirically tested.

The second claim that can be made about the sequences identified as formulaic is that they have a holistic quality i.e. they possess at least one characteristic that gives them the status of units. This holistic quality can be accounted for by one of the following reasons: 1) they are either grammatically or semantically irregular and as such can only be considered in their entirety 2) there is a clear mapping between their form and their function / meaning 3) they have been learnt as wholes by the learners.

The third claim that can be made is that each identified sequence is only formulaic for the learner who produced it as formulaicity in the present study is defined speaker-internally. However, many sequences are formulaic for several or even all the learners, which reflects their common profile of instructed advanced learners of French.

Contrary to the sequences described above, the following types of sequences were not identified as formulaic:

- Sequences presenting a holistic quality but which also present signs of processing effort such as pauses, repetitions or rephrasing. For example, a sequence such as *à mon avis* (in my opinion), which possesses a holistic quality for some of the reasons described above, was not counted as formulaic in the present study when it presented signs of processing effort in the way it was uttered.

- Fluent stretches of speech which do not possess a holistic quality i.e. *et je* (and I) because, as explained in chapter 5, a formulaic sequence in this study is a more precise construct than a simple fluent run of speech as it also needs to possess a holistic dimension.

11.3 Main results and conclusions

11.3.1 *FS use in advanced L2 learners and its implications for our understanding of language*

27 % of the language of advanced L2 learners consists of formulaic units presenting a processing advantage. Most of these sequences are grammatically-regular sequences and only a minority are idiomatic in the traditional sense of the term. Learners use different types of FS: referential, meta-discursive and sentence-builders. They use them to refer to various entities but mainly to express their opinions and talk about common topics such as what they like or where they live. A clear link was therefore found between FS and formulaicity at the conceptual level: FS are often resorted to when the learners express very commonplace conceptual content. This explains why a clear task effect was found for the story-retelling task which contained significantly fewer FS than the discussion tasks or the interview: learners use fewer FS when they are constrained in what they need to say than when they are given more freedom about the topic of conversation. Despite the advanced level of the learners under scrutiny, some FS presenting non-nativelike characteristics were identified and the learners were also found to over-rely on some sequences that native speakers would not use much.

The results of the analysis of the corpus of identified FS had several implications for our understanding of language. Firstly, the study of FS in advanced L2 learners allowed for an insight into the idiosyncratic nature of the learners' repertoires of FS. Indeed, although some FS are common to all learners, probably due to their similar linguistic background as well as the fact they performed the same tasks, there are important individual differences in the learners' repertoires of FS: learners have lexical

preferences for the expression of certain semantic content. Indeed, some sequences are used only by one or two learners, and even within the FS used by all or most of the learners, there were clear differences in their distribution. FS therefore epitomise the link between our linguistic representations and our individual experience with language.

Moreover, the changes in the corpus of FS between time 1 and time 2 exemplify the dynamism of linguistic representations: new FS appeared, amongst which some were more idiomatic than at time 1. There were also some changes in the distribution of FS already present at time 1, as some heavily relied-on sequences at time 1 disappeared at time 2 and, conversely, less frequent FS at time 1 became more frequent at time 2. In other words, FS do not disappear when the learners' lexical diversity increases, they become more varied and sophisticated.

The corpus of identified FS showed that a processing advantage could also be found for more abstract FS — that is, formulaic frames with open-slot(s) — although it is difficult to determine the level of abstraction of such sequences, i.e. whether they are separate lexical FS rather than a formulaic frame with slots. What the analysis of the corpus of FS revealed is the very close level of interaction between syntactic structures and the lexicon, as some FS with open slots clearly tend to appear in conjunction with specific lexical items. The study of such FS is therefore a way of getting an insight into the continuum between lexis and grammar.

These implications all point towards the idiosyncratic nature and dynamism of linguistic representations, as well as the non-dichotomous nature of lexis and grammar, which seem to be located on a continuum, in support of a usage-based conception of language.

11.3.2 *Longitudinal development of FS use and its relationship with the development of lexical diversity and fluency*

The results of the longitudinal study showed a statistically significant group increase in the three variables investigated: FS use, lexical diversity and fluency. One learner stood out from the group as although her level of formulaicity increased, there was hardly any increase in fluency and lexical diversity.

There was a statistically significant correlation between FS use and lexical diversity: the more learners used FS, the more they were lexically diverse. There was also a significant correlation between the quantity of FS at time 1 and the amount of change in lexical diversity. In other words, the more formulaic learners at time 1, such as Sally,

are those who were the most likely to increase the most in lexical diversity. This explains why the learner with the lowest percentage of FS at time 1 is the one whose lexical diversity develops the least. To account for that learner's pattern, the interpretation that a certain level of automatisation is necessary in order to have enough cognitive resources to acquire more vocabulary was put forward. It was suggested, however, that given time, that learner would develop in the same way as the others, as the graph representing the relative development of FS use and lexical diversity suggests a common developmental trajectory with learners varying in their rate of development.

Statistically significant correlations were also found between FS use and fluency. The absolute values corresponding to FS use significantly correlated with the absolute values of the fluency measures: the higher the learners' level of formulaicity, the higher their fluency. Moreover, the absolute values corresponding to FS use significantly correlated with the size of the change in all the fluency measures except for PTR. In other words, the higher their level of formulaicity, the greater their increase in fluency. Finally, the importance of the change in FS use significantly correlated with the importance of the change in all the fluency variables except AR, which means that, on a general basis, the greater the increase in formulaicity, the greater the increase in fluency.

The qualitative analysis showed that FS use plays a role in increasing fluency by contributing to the reduction of pausing time, enabling longer speech runs, as well as the speeding up of the articulation rate. These quantitative and qualitative results suggest that, at the internal level of processing mechanisms, FS play a facilitating role at various stages of speech production. Indeed, they speed up or bypass formulation at both the levels of grammatical encoding and morpho-phonological encoding. They also speed up the articulation stage. Finally, because many FS tend to express very formulaic, unoriginal content, FS can be thought of as speeding up the stage of conceptualisation too.

The last chapter presented a summary of the different learners' development in FS use, lexical diversity and fluency and attempted to relate it to the different contexts of the learners' stay in France in terms of how much they engaged with the French language. Although the analysis is only based on 5 learners, which limits its generalisability, it showed that the extent to which the learners immersed themselves in the French context varied hugely and that there was a link between the degree of learners' interaction with the French language and their development. This study reinforces previous conclusions

that the study-abroad context is not sufficient in itself to guarantee rich linguistic interactions, and factors such as personality and motivation but also chance and luck (in terms of living arrangements or number of people met) play an important role in the potential linguistic benefits of a stay abroad.

11.4 Limitations

Some limitations of this study are directly related to the nature of the construct under scrutiny. Firstly, FS are inherently difficult to investigate empirically. Indeed, since they are a psycholinguistic reality, any empirical evidence in relation to them will necessarily be indirect. It was thought that the best way of investigating learner-internal, psycholinguistic FS was to examine a large amount of data per learner and establish 'fluent pronunciation' as a necessary identification criterion. The downside of this learner-internal design, however, is that, because of the large quantity of data analysed per learner, as well as the very time-consuming identification method based on the precise annotation of pauses, only 5 learners could be investigated. This makes the conclusions of this study tentative, and generalisability to other advanced learners of French remains to be confirmed.

Finally, in a study placing formulaicity within a usage-based theory of language which conceptualises FS along a continuum from what is constructed on line to what is formulaic, it might seem paradoxical to set out to identify the proportion of FS in a given corpus as it implies making a distinction between what is formulaic and what is not formulaic. Drawing such a border between what is formulaic and what is not formulaic was thought to be a necessary methodological device in order to get an insight into the phenomenon of formulaicity. It does not imply, however, that formulaicity is not thought of as a graded notion or that speakers' repertoires of FS will not evolve with time.

11.5 Methodological implications and directions for future research

The results of this study show that speaker-internal FS represent a considerable part of learner-language but that more research is needed to characterise them better. For example, the present corpus could be coded for FS types in order to get a more quantitative insight into the distribution of FS types within and across learners and how it evolves during an extended period. Phonetic reductions would also need to be coded more systematically in order to better understand how they interact with FS.

The majority of speaker-internal FS identified in this study are very different from the idioms and idiomatic expressions that have been used as stimuli in various psycholinguistic experiments investigating the processing of FS in L2 learners. Within a speaker-internal and second-language acquisition approach to FS, future researchers therefore need to ensure that the stimuli investigated in processing studies are relevant in the L2 context. In this respect, a strict learner-internal design could be envisaged for psycholinguistic experiments investigating the processing of FS by L2 learners. In other words, learners' own productions could be used as the corpus from which to extract stimuli for further psycholinguistic experiments on the receptive processing of such sequences. This would guarantee both a learner-internal and learner-language perspective on the study of FS. Although such learner internal designs are complex and time-consuming, it is worth developing them, given the central importance of FS for our understanding of language and language acquisition.

Appendix A. Tasks

a. General interview (Inter)

English translation of the selection of questions asked at time 1

Describe yourself, especially your personality.
Where are you from? Tell me about your home town and your old school. What did you like/ dislike about them?
Describe your student life now and its advantages and drawbacks.
Would you say it is better to be a student in France or in the UK? Why?
What would you like to do after you have graduated?
Tell me about your best / worst holidays.
What are your hobbies, your passions?
Tell me about your favourite film or a film you have seen recently.
Tell me about someone (famous or not) that you really admire / dislike.
In your opinion, what are todays society's major problems?

English translation of the selection of questions asked at time 2

Can you introduce yourself again? Tell me about where you're from, what you study, your hobbies.
Can you tell me about how this year in France is going. What do you do? Where do you work? How is your week organised?
Have you met a lot of people?
What were your first impressions of France when you arrived? Were there cultural differences that you noticed?
Would you like to live in France one day? Why?
What are your plans for the next months before going back to university?
Did you get visitors?
Did you get the opportunity to visit other places in France?
How would you rate your experience here? Would you do it again? Are there things you would now do differently?

b. Discussion tasks (French, Alcohol and Food)

For each of the 3 discussion tasks, the learners were given the following instruction: Please rank the suggested reasons/ measures from what you think is the best/ most helpful to the worst/ least acceptable according to you. Feel free to add further suggestions and conclude by discussing the topic more generally.
The learners were given the topic and suggested reasons/ measures in French as follows:

i. French

Pourquoi apprendre à parler le Français quand tout le monde parler anglais ?

1. Pour pouvoir aller à l'étranger et parler avec les habitants du pays
2. Pour pouvoir avoir un meilleur emploi
3. Pour apprendre à connaître une autre culture
4. Parce que c'est facile et amusant d'apprendre une langue

ii. Alcohol

Comment réduire la consommation d'alcool chez les mineurs ?

1. Donner une amende plus sévère aux magasins et aux bars qui vendent de l'alcool aux adolescents
2. Augmenter le prix des boissons alcoolisées
3. Alerter les jeunes sur les effets à long terme de l'alcool
4. Abaisser l'âge légal de la consommation d'alcool à 16 ans.

iii. Food

Que peuvent faire les écoles pour combattre l'obésité ?

1. Interdire les distributeurs automatiques de chips, boissons gazeuses et sucreries
2. Introduire plus de cours de sport
3. Augmenter les prix des aliments trop sucrés, gras ou salés.
4. Avoir un plus grand choix de plats qui sont bons pour la santé à la cantine de l'école

English translation

i. French

Why learn French when everyone speaks English?

1. To be able to go abroad and talk to the locals
2. To get a better job
3. To get to know another culture
4. Because it's easy and fun to learn a foreign language

ii. Alcohol

How can under age alcohol consumption be curbed?

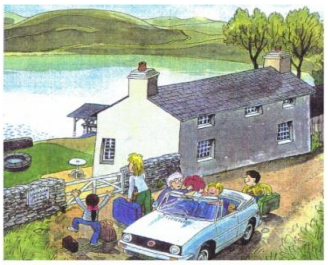
1. Give a higher fine to shops and bars which sell alcohol to teenagers
2. Raise the price of alcoholic drinks
3. Warn young people about the long-term effects of alcohol
4. Lower the minimum legal age for alcohol consumption to 16

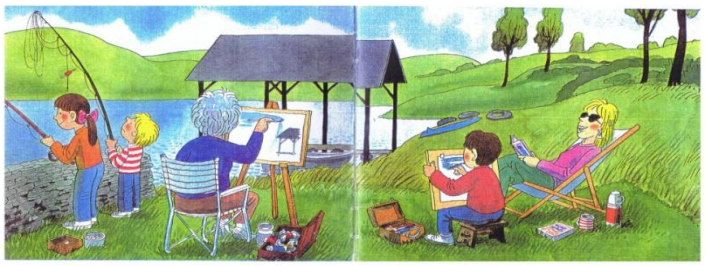
iii. Food

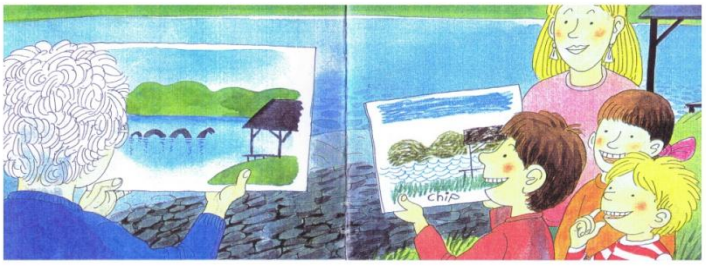
What can schools do to fight obesity?

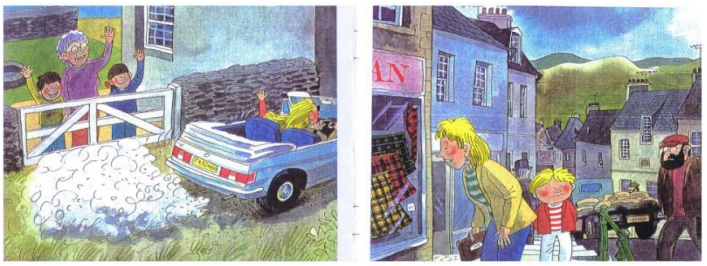
1. Ban vending machines of crisps, sodas and sweets
2. Have more sport classes
3. Increase the price of food which is too sweet, fatty or salty
4. Have a greater selection of healthy dishes at the school canteen

c. Story-retelling (Loch Ness)[19]

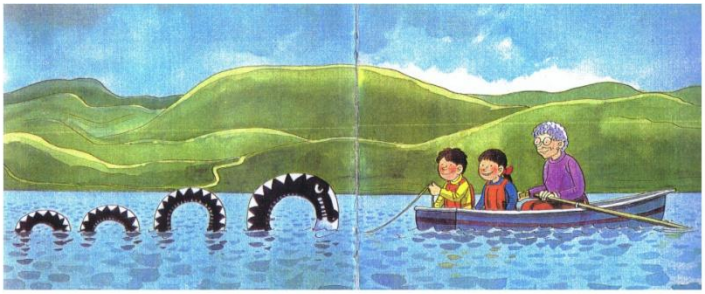


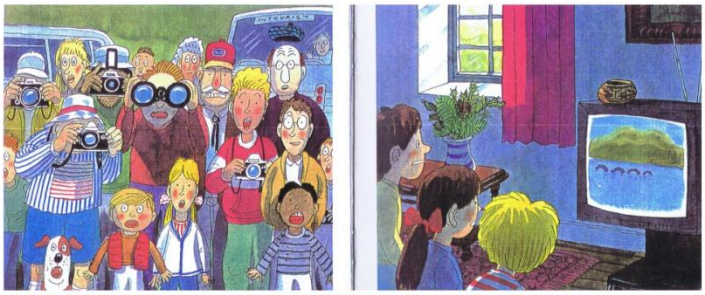


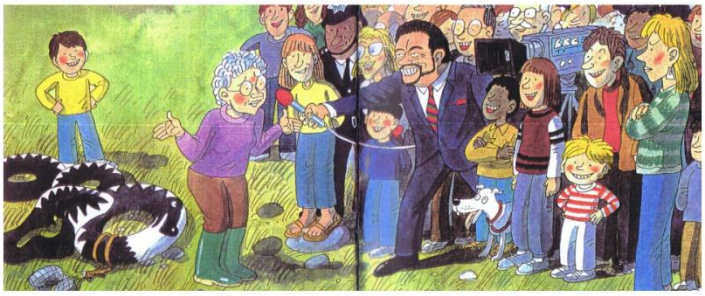


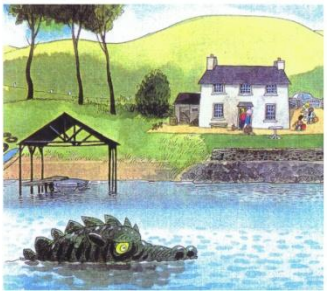


[19] Illustrations by Alex Brychta for *A Monster Mistake* by Roderick Hunt (Oxford Reading Tree, 2003) used by permission of Oxford University Press.









d. Questionnaire, adapted from the Language Contact Profile (Freed, Dewey, Segalowitz and Halter 2004)

PART 1: BACKGROUD INFORMATION

1. Before going to University, did you study French at primary school?

If yes, for how many years and how many hours a week?

2. Before going to University, did you study French at secondary school?

If yes, for how many years and how many hours a week?

3. Please state all your stays in France prior to the year abroad:

DATE	LENGTH	CONTEXT

PART 2: USE OF FRENCH AT UNIVERSITY BEFORE THE STAY ABROAD

1. In your first and second year at University, apart from speaking French during class or when you went to France, did you speak French on other occasions for example with Erasmus students? If yes, how frequently?

2. For each of the items below, choose the amount of time you spent on average doing each activity in French **BEFORE THE YEAR ABROAD.** Please answer

the question by using one of the followings: Never / A few times a year / Once a month / Twice a month / Once a week / Twice a week / Every day

Please give details if you can

Activity	**Situation**
Watching French films, French web videos or French language television	
Reading French language newspapers, magazines or Internet pages	
Reading novels in French	
Listening to French radio, podcasts or songs in French	

PART 3: YEAR ABROAD: TIME SPENT IN FRANCE

1. Date of arrival in France

2. Date of departure from France

3. Periods when you were away from France for holiday elsewhere

4. Describe your living arrangement

5. If you lived with other people, which language did you speak with them? How often did you speak with them?

6. If you spoke French with them, how many hours a week approximately?

7. Please specify: How many days per week and on average how many hours per day you used French in the situation indicated. Just write N.A if a situation does not apply to you.

Situation	How many days per week	How many hours per day (approximately)
I spoke French to friends who were native speakers of French		
I spoke French to friends who were not native speakers of French but were not native speakers o English (e.g. German language assistants)		
I spoke French to the native speakers with whom I lived		
I spoke French to the teachers and other staff in my school		
I spoke French to the staff and students in the university where I studied		
I spoke French to my work colleagues		
I spoke French to French people in shops, cafes etc…		
Other situations in which I spoke French (Please specify)		

8. How much time did you spend approximately doing the following:

Situation	How many days per week	How many hours per day (approximately)
Speak English to native speakers of English who were in France at the same time		
Speak English (on the phone or the Internet) to my friends and family back home		
Speak English to native speakers of French		
Speak English to native speakers of other languages than French		

9. How often did you approximately do the following activities? Please answer the question by using one of the followings:

Never / A few times a year / Once a month / Twice a month / Once a week / Twice a week / Every day

Situation	Frequency
Watching French films, French web videos or French language television	
Reading French language newspapers, magazines or Internet pages	
Reading novels in French	
Listening to French radio, podcasts or songs in French	
Writing in French	

10. How often did you approximately do the following activities? Please answer the question by using one of the followings:

Never / A few times a year / Once a month / Twice a month / Once a week / Twice a week / Every day

Situation	Frequency
Watching films, web videos or television in English	
Reading newspapers, magazines or Internet pages in English	
Reading novels in English	
Listening to English radio, podcasts or songs in English	
Writing in English	

11. What did you want to achieve during your stay abroad? Do you think you succeeded?

12. How motivated to immerse yourself in the French context would you say you were?

13. Did you spend a lot of time on your own?

14. Did some things frustrate you while you were there? Did you develop any strategies to deal with these problems?

15. Do you feel that you made the most of your stay? What would you do differently if you could do it over again?

16. On what aspect(s) of your language would you say you improved the most?

17. On what aspect(s) of your language would you say you did not really improve?

Appendix B. Transcripts

This appendix regroups several transcripts of the learners' productions. Given the size of the corpus (55000 words), only one learner's transcript per task can be provided for reasons of space. Times 1 and time 2 are given for the same learner.

As described in chapter 6, some of the discussion tasks at time 1 were carried out in pairs. Therefore, the example of the alcohol task was performed by Lily and Lola, whose separate performances at time 2 are also provided.

As explained in chapter 6, the transcription format used at time 1 is CHAT. CHAT transcription conventions can be found in Appendix G. The format at time 2 is a normal text format as the transcriptions were transferred from Praat onto text files. The visual advantage of this second format is that each new line corresponds to a new fluent run.

The FS identified for the sample of transcripts given below are all gathered in Appendix C.

French discussion task, time 1, Sally, CHAT format

@Begin
@Languages: fr
@Participants: Sal Subject, Car Investigator
@ID: fr|ncl|Sal|20;00.00|female|Y2||Subject||
@ID: fr|ncl|Car||female|||Investigator||
@Date: 15-JUN-2009
@Location: NCL
@Situation: French Discussion Task
*Sal: euh pour moi je suppose que je [//] j' apprends le français parce que: j' aime beaucoup la langue .
*Sal: euh mais aussi j' aime beaucoup la culture la: [//] l' abilité de [/-] c' est pas ça la compétence ?
*Car: la capacité .
*Sal: la capacité c' est ça .
*Sal: pour quand je vais à l' étranger (.) de pouvoir parler avec les habitants du pays pour [/-] parce que si on va à un pays où on ne parle pas la langue c' est vraiment difficile de s' impliquer dans la vie culturelle la vie sociale même juste pour les vacances.
*Sal: je pense que les raisons pour lesquelles on apprend le français dépendent de la personne .
*Sal: euh évidemment ici toutes les raisons sont les bons raisons .
*Sal: euh pour pouvoir avoir un meilleur emploi .
*Sal: je pense que ça pourrait être intéressant parce que: si on apprend la langue juste pour améliorer (.) euh la vie de travail c' est pas nécessairement le meilleur [//] (.) la meilleure raison de le faire .
*Sal: parce que c' est pas pour l' amour pour [//] de la langue .
*Sal: c' est juste (.) je sais pas comment on peut expliquer .
*Car: par opportunisme .
*Sal: oui c' est ça .
*Sal: donc ça va si (.) on est quelqu' un qui peut travailler très dur sur quelque chose qu' on ne pas [//] qu' on n' aime pas (.) forcément .
*Sal: mais pour moi je ne pourrais pas le faire euh parce que si je vais

faire quelque chose je veux (.) m' amuser bien quand je le fais .
*Sal: ça aussi je pense que ça c' est une très bonne raison .
*Sal: parce que c' est vraiment difficile de comprendre une culture sans le cinéma (.) la musique (.) les livres .
*Sal: et toutes ces formes de la culture normalement sont dans la langue du pays .
*Sal: donc encore c' est difficile de (.) vraiment connaître un pays ou une culture sans la langue .
*Car: on peut pas vraiment séparer les deux .
*Sal: c' est presque le même .
*Sal: amusant je suis d' accord facile je ne sais pas .
*Sal: ça dépend euh quand on aime faire quelque chose (.) ça n' importe pas si c' est difficile ou facile .
*Sal: mais je pense que (.) apprendre une langue parce que c' est facile ça c' est pas une bonne raison c' est comme (.) je sais pas je peux pas penser d' une synonyme d' une (.) équivalent .
*Sal: mais c' est utile si c' est facile mais c' est pas la raison de faire quelque chose .
*Sal: je déteste la phrase tout le monde parle anglais [/] parler anglais parce que c' est pas vrai .
*Car: tu verras que beaucoup gens ne parlent pas .
*Sal: et aussi c' est que c' est impoli d' aller à un autre pays et de (.) penser que les habitants là bas vont parler ton langue [//] ta langue .
*Sal: c' est impoli c' est pas comme (.) s' il y avait quelqu' un qui est venu ici euh mais qui n' essayait pas de parler l' anglais [//] parler anglais tout le monde dirait mais tu es ici pourquoi tu ne le fais pas ?
*Sal: et je pense que (.) c' est une attitude vraiment (.) égoïste mais on peut le voir très souvent .
@End

French discussion task, time 2, Sally, text format

pour moi
c'est
parce que j'aime bien les langues
j'aime bien découvrir les autres cultures et rencontrer les autres gens
et
c'est très intéressant de faire ça
dans une autre pays
et c'est
tu peux pas vraiment le faire si tu parles pas la langue
il y a des choses culturelles
qui sont exprimées par la langue
donc si t'essayes de
de
de parler de ça dans une autre langue ça marche pas
ouais bien sur j'aimerais b
j'aime bien aller à l'étranger et parler avec les habitants du pays
et aussi apprendre à connaitre une autre culture
pour moi

d'avoir un meilleur emploi
je sais que le fait que je parle français va m'aider
mais c'est pas pour ça que je l'ai appris
ça c'est une un bonus et c'est cool mais c'est pas l'objectif pour moi
un bonus c' est
c'est cool mais c'est pas l'objectif pour moi
j'aimerais bien être interprète
je sais que c'est dur
je sais même pas si j'ai
je vais avoir un assez bon niveau
mais je vais essayer
et aussi
si
parce que j'ai une année de plus à Newcastle
et après
il faut décider ce que je veux faire
et si
je veux être interprète il faut faire un master
si je fais un master en Angleterre c'est un an
mais je vais être en Angleterre
et je vais parler anglais tout le temps sauf en cours
anglais tout l' temps sauf en cours
par contre si je viens à Paris en France en Espagne je sais pas
pour faire un master
j'sais pas si c'est deux ans en Espagne mais c'est deux ans ici
mais c'est deux ans
en France
donc même si j' réussis pas même si j'décide c'est
c'est pas pour moi
j'ai pas perdu le temps après que
j'aurai amelioré mon français
j'aurai rencontré du monde
et caetera
ah ouais ouais
ouais c'est sur
c'est juste une question de quand ou pourquoi
pourquoi
ça m'a beaucoup plu oui
ça dépend
vraiment c'est amusant
j'aime bien
mais je dirais pas qu' c' est facile
mais c'est pas que c'est difficile non plus
il faut juste s'appliquer il faut juste
j'ai appris plein d'vocabulaire des trucs
que j'aurais pas appris chez moi
une poêle une casserole
j'sais pas une serre tête les choses que j'aurais pas appris chez moi
mais aussi
c'est
j'ai
j'avais pas peur de faire des erreurs d'abord mais

quand même j'ai encore moins main(te)nant
je suis beaucoup plus à l'aise
je sais que de temps en temps je vais avoir du mal à m'exprimer
mais
surtout je crois que c'est grâce à mes amis que j'ai fait ici
il y a une fille qui a passé quelques mois aux Etats-Unis
plusieurs qui vient une qui vient des Antilles un qui vient du Congo
donc ils comprennent
que c'est pas toujours facile de s'adapter à
une vie nouvelle
et comme ça ils
peut-être ils sont
ils acceptent plus facilement
que de temps en temps je fais des erreurs je
je peux pas m'exprimer je
que je cherche des mots et tout
ce sont des gens ouverts
c'est plus facile peut-être
j'ai lu
les livres de
les quatre premiers
quatre
non les quatre premiers livres de Harry Potter en
non en français
et j'ai souligné
pas tous les mots que je connaissais pas
mais tous les mots qui
m'empêchaient de comprendre le
l'essentiel
alors c'est un livre
c'est le livre que je connaissais bien
déjà donc
je comprenais le sens quand-même
mais les
peut-être les mots qui
que j'ai
j'avais vus plusieurs fois j'dis non ça j' comprends pas
et j'ai commencé
en fait je lis beaucoup plus vite parce que je lis dans le train
quand j'attends quelqu'un et tout ça
donc je lis beaucoup plus vite que j'ai le temps de chercher dans le dictionnaire
j'vais pas chercher tous les mots que j'ai soulignés mais quand-même je sais qu'ils sont
là donc
quand j'suis chez moi l'année prochaine
c'est quelque chose que je peux faire pour
continuer à apprendre
deux
je suis allée que deux fois j'vais pas trop cinéma en fait
j'ai vu
la deuxième
le deuxième film de Twilight
parce que je voulais le voir mais j'ai vu

je l'ai vu en VO
j'ai vu un film
un film qui s'appelait Mère et Fille
qui était intéressant
mais
y avait pas beaucoup d'histoire
c'était intéressant mais
pas grand chose
pas
pas vraiment parce que je préfère sortir mais
j'aime bien un diner presque parfait
parce que ça c'est à l'heure ou je rentre chez moi
alors je peux me détendre et je le regarde pour une heure
et aussi les
les matins avant d'aller à l'école
de temps en temps je regarde il y a des
séries américaines
en français
et c'est juste
pas je
je l'allume et je
je me prépare
je
je prends ma douche et tout ça mais c'est là et je l'entends

Alcohol discussion task, Lily and Lola, time 1, CHAT format

@Begin
@Languages: fr
@Participants: Lol Subject, Lil Subject, Car Investigator
@ID: fr|ncl|Lol|20;00.00|female|Y2||Subject||
@ID: fr|ncl|Lil|20;00.00|female|Y2||Subject||
@ID: fr|ncl|Car||female|||Investigator||
@Date: 25-JUN-2009
@Location: NU
@Situation: alcohol
*Lil: comme quelqu' un qui j' aime boire d' alcool euh je ne pense pas que en augmentant [/] en augmentant [/] en augmentant le prix des boissons alcolisées je ne pense pas ce ce mesure va aider parce que euh si les [/] si les mineurs veulent boire l' alcool ils boivent buvent +/.
*Lol: boivent .
*Lil: +, boivent d' alcool si le prix est plus élevé ou non .
*Lil: et aussi ces mesures euh .
*Lol: ce serait énervant pour les personnes de plus de dix huit ans .
*Lil: oui oui euh et aussi à la fac euh je vois certains gens qui boivent plus d' alcool et par conséquent ils n' arrivent pas au cours euh le [/] le matin prochain .
*Lil: mais pour ces gens c' est un problème mais pour les autres gens ce n' est pas un problème si [/] euh si on sait l' effet de l' alcool .
*Lil: je pense que le gouvernement va alerter ces gens aux conséquences de [/] de boire euh trop d' alcool .
*Lil: mais c' est un choix personnel .
*Lil: et le gouvernement ne peut [//] ne peut pas demander aux gens de ne boire pas d' alcool donc c' est difficile .
*Lol: je ne pense pas qu' il y a aille +/.
*Car: qu' il y ait .
*Lol: +, ait une solution facile ou simple .
*Lol: c' est dans notre culture .
*Lil: c' est il existe aussi en Europe mais beaucoup plus en en Grande_Bretagne .
*Lol: et peut_être s' il y aurait euh des amendes pour des parents des gens ou pour des gens aussi .
*Lol: mais je ne sais pas s' il y a une solution .
*Lol: euh je sais qu' il y a des gens de moins de seize ans qui qui trouvent quelqu' un sans abri le donner leur [//] lui donner argent pour acheter euh des boissons de bière euh pour eux .
*Lol: donc euh si euh on change euh l'âge de consommation d' alcool à seize ans ce [/] ce n' aurait pas un effet peut_être ça pourrait résulter dans des gens de dix ans et onze ans de boire d' alcool .
*Car: quelles sont les autres choses ?
*Lol: alerter les jeunes sur les effets à long terme de l' alcool .
*Lil: c 'est important mais euh dans les paquets de cigarettes il est vrai qu' il y est les euh [/] les alertes comme ça mais les fumeurs continuent à fumer .
*Lil: peut_être si: euh si les fumeurs les boisseurs ?
*Car: les buveurs .

*Lil: les buveurs peuvent voir quelqu' un qui est tombé à mort à cause de ces produits face à face ça serait plus effectif .
*Lil: mais (.) c' est difficile c' est c' est vraiment un choix personnel .
*Lil: pour le gouvernement c' est efficace de euh réduire les effets négatifs au [//] de la société de l' alcool par exemple les euh débacles euh dans les pubs fights ?
*Lol: je crois aussi que nous avons besoin de plus de facilités dans les villes pour les jeunes parce que euh dans ma ville à [/] à Bath il y a un cinéma mais il est six livres pour euh une personne .
*Lol: donc ce n' est pas très euh très facile pour une personne de moins [/] de moins de seize ans euh euh de aller [//] d' y aller .
*Lol: donc peut_être si on aurait plus de euh de patinage de ice skating ring de bowling euh pour des jeunes à utiliser.
*Lol: mais aussi je crois que c' est euh [/] c' est la cause est dans la [//] le caractère des personnes euh de Grande_Bretagne parce que nous sommes euh euh une race assez timide donc nous utilisons euh l' alcool peut_être pour avoir plus <de confiance> [//] de la confiance .
*Lol: je ne crois pas qu' il y a le même problème en Europe .
*Lol: mais ça ces points de vue changent je crois que ça change un peu en ce moment parce que quand je suis allée à Toulouse l' année dernière et je suis allée dans les bars dans le centre ville avec euh mon amie j' ai vu des euh des Français qui ont trop bu des mais c' était plus des hommes .
*Lol: donc ici maintenant c' est plus de femmes qui boivent trop .
*Lol: mais peut_être euh euh ça va changer en Europe .
*Lol: mais euh j' espère que non .
*Car: est ce que vous pensez que la vie étudiante peut mener à l' alcoolisme ?
*Lol: oui bien sûr .
*Lil: oui parce que toutes les activités du soir euh inclure ?
*Car: incluent .
*Lil: incluent l' alcool pour la plupart du temps par exemple si euh si c' est mon anniversaire euh c' est probable que moi et mes amis euh aller au restaurant où on achète du vin ou des bières euh sinon euh on va en ville au pub et au boire [//] au bar .
*Lil: et donc c' est évident dans la plupart des nuits les étudiants boivent d' alcool .
*Lol: par exemple si Freshers' week au début de l' année chaque nuit il y a des euh des évènements qui sont euh avec euh qui a l' alcool au centre des activités .
*Lol: et par exemple moi je n' aime pas beaucoup boire parce que je n' aime pas le goût .
*Lol: et j' ai des problèmes avec des amis qui pensent pourquoi est ce que [//] comment est ce que tu peux euh euh t' amuser ?
*Lol: et donc euh ce n' est pas mon problème c' est leur problème .
*Lol: mais oui euh oui à Newcastle par exemple il y a beaucoup beaucoup de de bars quand on peut acheter l' alcool pour un prix très baissé ?
*Lol: et peut_être c' est ça dans trente quarante ans euh on aura plus de gens avec euh des problèmes de [/] de la foie .
*Lol: j 'ai [/] j' ai euh [/] j' ai lu dans les journals et aux informations qu' il y a plus de personnes de vingt ans qui ont euh

le cancer de foie et les choses comme ça .
*Lil: j' ai lu article qui a dit que il y a un homme à l'université euh peut_être dans le sud de l' Angleterre quelque chose comme ça .
*Lil: et il était partie d' un jeu et chaque euh quarante cinq secondes il a dû euh bu euh [//] boire une shot d' alcool de vodka et après quarante minutes il était mort .
*Lil: et c' est choquant .
*Lil: c' est difficile parce que (.) la plupart des gens j' espère ne veulent pas le faire mais il y a certains gens qui pensent que ce jeu c' est normal .
*Lol: en France euh on boit pour [/] pour si on aime le gôut pour savour ?
*Lol: par exemple le vin .
*Lol: moi j' aime les cocktails pour les goûts et pour les couleurs et caetera .
*Lol: mais je ne veux pas être ivre et je ne veux pas tomber dans la rue .
*Car: est ce que vos trouvez qu' il y a une pression sur les gens ?
*Lol: oui bien_sûr .
*Lil: parce que je n' ai pas beaucoup d' argent et donc si euh je vais en ville euh de temps en temps je ne bus pas je prends de l' eau .
*Lil: et les gens me moquent ils dis oh pourquoi est ce que tu prends de l' eau ?
*Lil: c 'est facile euh de dire euh que tu (.) [/] tu dois rouler euh la voiture parce que c' est une raison importante mais à part de ça.
*Lol: et pour eux c' est normal par exemple euh nous avons fait euh nos examens et euh chaque [/] chaque jour il y aura beaucoup d' étudiants dans les bars <qui va> [//] qui vont célébrer la fin des examens .
*Lol: mais c' est [//] ça c' est bien bon bien .
*Lol: mais je crois que c' est un problème quand on boit chaque journée beaucoup d' alcool .
*Lol: par exemple j' ai un ami qui [/] qui doit boire chaque journée .
*Lol: et à mon avis ce n' est pas une très bonne chose pour lui .
*Lol: mais et aussi c' est assez cher de boire .
@End

Alcohol discussion task, Lily, time 2, text format

donc pour réduire la consommation d'alcool chez les mineurs
oui je pense que si
il y avait
une amende plus sévère aux magasins
et aux bars qui vendent de l'alcool aux adolescents je pense que ça va marcher parce que les
les bars et les magasins peuvent oublier le
le droit de
de vendre de l'alcool je pense que ça va
ça va vraiment marcher quand
quand j'avais 16 ans
j'ai pu aller au
au bar facilement mais
parce que les
les lois ont

a changé quand j'avais
17 ans et
tous les bars ont dit non il faut qu'on
on vérifier
et donc
oui
ça va marcher parce que
si on peut pas acheter d'alcool on peut pas le boire
et
je pense que pareillement
également
au niveau des
de la nourriture
si les prix
des boissons alcoolisées étaient augmentés je pense que c'est pas juste pour les autres
peut-être ça va marcher un peu mais si les gens veulent boire
ils vont boire
boire
ça va
peut-être ça va marcher un peu parce que tu peux pas boire
boire
beaucoup
si t'as pas l'argent
et aussi si
il y a les
les peut-être les alcooliques
qui peut
vont dans les écoles pour parler aux jeunes peut-être ça va marcher pour
parce que
les tout le temps les
les professeurs qui disent ah faut que tu
tu boives pas beaucoup parce que c'est mal et c'est
et s'il y a vraiment quelqu'un qui est pre
presque mort à cause d'alcool je pense que ça va marcher si tu as quelqu'un
à cô
face à toi qui dit ah non tu vas
tu vas mourir si tu fais ça
ça va marcher
la plupart
non je pense
peut-être si
l'âge légal était abass
abaissé jusqu'à 16 ans
peut-être ça va
je pense que les
les
les jeunes qui ont 18 ans qui commencent à
aller au
au bar ils
ils boivent trop
ils
vomit ils fait n'importe quoi

peut-être si ils le font quand ils avaient 16 ans
peut-être ça va passer trop ra
plus rapidement
mais ça va
si c'est pas si
ça va marcher mais par exemple quand
quand
j'ai commencé
à boire peut-être je faisais oh c'est cool
mais
après c'est
c'est normal donc
tu bois pas trop
mais
oui je pense si c'est vraiment quelque chose au niveau de la famille
quand j'étais
jeune quand j'avais 14 ans je
j'ai
j'ai pu
boire un verre de vin avec ma famille et donc
j'ai
c'était pas quelque
quelque chose
je peux pas faire donc c'était pas quelque chose
très
excitant très
fantastique donc c'était oui ça va
mais s'il y avait
s'il y a les
les parents qui dit ah non faut que tu boives pas
donc tu veux boire
donc oui c'est pour les familles je pense
je pense qu'au fond s'il y a moins
moins de jeunes qui
boit
trop
beaucoup les gens qui fait de binge drinking
d'un côté il y a beaucoup de gens qui sortent
pour le soir
ils boivent beaucoup chaque
chaque soir c'est le but
du
de sortir c'est d'être bourré
et j'aime pas ça du tout j'aime pas
je
j'aime
boire si je veux boire mais j'aime pas
boire d'être bourré c'est
c'est
t'as pas besoin de te bourrer de
d'avoir une bonne soirée
et donc je pense qu'au fond c'est

c'est plutôt comme ça tu bois
assez souvent peut-être tu
tu as un verre de vin avec le déjeuner et
mais c'est
c'est pas que
les jeunes aussi les jeunes peut-être boit avec le
leur famille et je pense que c'est
ça marche parce que
il y a les gens qui
les Français qui
boit trop mais pas aussi qu'en Angleterre je pense
donc je pense que si
quand on sort en France on va dans les bars
on boit mais le but c'est pas d'être bourré le but c'est de
de
c'est d'a
d'avoir une bonne soirée
soirée intéressante et peut-être on va boire peut-être pas donc
mais alors comme ça c'est
l'alcool c'est
c'est plus
plus important
comme les émissions télévisées les choses comme ça
peut-être oui c'est
oui c'est vrai que
les Britanniques sont trouvés d'être timides donc il faut que l'on
on doit boire pour avoir une bonne soirée pour
pour
s'amuser
mais
et où ça marche
je sais pas si c'est exactement vrai je pense que c'est l'image de Grande-Bretagne de
boire beaucoup et
parce que c'est l'image et parce que c'est normal
tout le monde le fait
pour l'alcool ils
avec moi ils boivent pas sauf si c'est une occasion spéciale
c'est-à dire
quand on a mangé le repas de Noel mais aussi
ils boivent pas le soir et dimanche déjeuner ils boit
l'alcool donc ça c'est intéressant
parce que ma famille en Angleterre
mes parents boit le soir mais pas le déjeuner sauf si c'est Noel ou
le
dîner c'est plus important
rapport de la nourriture
ils mangent comme une
comme je pense que toutes les fran
les Français
mangent
les
les plats va

variés et beaucoup
de plats pour chaque repas
poisson j'aime
je l'aime beaucoup ici parce
je les aime beaucoup ici parce que
il y a
plus de variation au
en Angleterre
il y a
pour la plupart
beaucoup des restos italiens
et aussi il y a
c'est bon parce que les restos
indiens et
comme ça mais en France j'aime
j'aime bien la nourriture française
c'est bon
c'est assez cher mais c'est
aussi les bars sont sympas parce
parce que
ils sont plus beaux à l'intérieur
il y a un bar
avec le tableau noir et on peut
prendre les craies et on peut écrire les choses et c'est sympa
mais on peut
porter les dés pour jouer et
on n'a pas les choses pareilles en Angleterre
les bars sont un endroit pour boire c'est pas pour s'amuser c'est pas pour être joli

Alcohol discussion task, Lola, time 2, text format

je crois que on devrait alerter les jeunes sur les effets à long terme de l'alcool
par exemple
de parler avec quelqu'un qui
qui était blessé et qui sont tombés ou quelque chose comme ça ou qui
qui avait des problèmes
des grands problèmes avec l'alcool
mais même ça je ne sais pas si ça pourrait marcher je crois que
c'est dans notre culture
c'est quelque chose que les gens utilisent comme
ils voudraient avoir plus
plus la confiance ou ils voudraient
s'amuser
par exemple j'ai vu quelque chose l'année
la semaine dernière il y avait
plus de mille
mille Britanniques qui est allé en Espagne
pour quelque chose du sport mais il y avait beaucoup de photos d'eux dans les rues

ils étaient complètement
bourrés
et

oui c'est
apparemment ça c'est
ça c'est les Britanniques
ça c'est nous
donc je ne sais pas s'il y a quelque chose qu'on peut faire parce que c'est
c'est accepté dans notre culture
si on boit beaucoup
je crois que
parce que moi je ne bois pas beaucoup et donc c'est moi qui a des problèmes avec ceux qui pensent
qui dit pourquoi
pourquoi est-ce que tu ne bois pas ou pourquoi
tu ne veux pas être
je n'sais pas pourquoi tu ne veux pas
boire six cocktails ce soir
parce que je ne veux pas tomber je ne veux pas être idiot
mais je sais que
mon frère il a 17 ans
donc c'est pas légal pour lui
ses amis
quand ils
parce que maintenant il a des amis qui ont
18 ans
mais quand il était plus jeune
il y avait des amis qui ont payé
ceux qui sont sans abri pour acheter de l'alcool pour eux
que ça
ça me choquait mais
ils ont fait ça donc
je n'sais pas je crois que
je n'sais pas s'il y a
une
une solution
augmenter les prix oui peut-être parce que
dans les supermarchés en Angleterre on peut acheter
d'la bière
pour je n'sais pas des prix stupides
moins que
de l'eau
par exemple
et oui peut-être si ce serait
plus
plus cher mais je ne sais pas si les supermarchés pourraient faire ça
parce que
s'ils font ça ils vont perdre des
des conso
des clients
peut-être et aussi ils ont
ils ont je sais pas des liens avec
les organisations
alcool
d'alcool

oui donc
et aussi le gouvernement il
il voudrait le
il veut la taxe
pour
sur l'alcool
donc ils n'ont pas vraiment
ils n'ont pas
il n'y a pas quelque chose pour
les persuader d'augmenter les prix
et
je
je ne peux
je ne sais pas s'ils vont baisser l'âge légal je crois
non parce que comme ça on aurait
de plus en plus de jeunes personnes qui
qui
qui boivent de l'alcool
je crois que 18 ça va
oui c'est l'âge quand on peut faire tous les autres choses et on va à la fac donc
aux Etats-Unis comme c'est 21 ans on est
on est déjà à la fac donc on boit bien sûr avec des amis
donc ça ne marche pas
oui je
je n'sais pas parce qu'avec les drogues
c'est
c'est plus simple peut-être parce que on peut parler avec
ceux qui avaient des vrais problèmes mais
qu'est-ce que c'est un vrai problème avec l'alcool
est-ce que
parce que il y a ceux qui fait du binge drinking
donc
est-ce que ça c'est un vrai problème pour eux
je n'sais pas je crois qu'on verra les faits dans
les 5
dans
des
20 ans 30 ans
on verra
s'il y a des grands problèmes avec les gens
oui je crois qu'en France en ce moment
le
le binge drinking
n'est pas un phenomenon ici mais ça change
peu à peu ça change
non pas vraiment mais normalement il dit
oh je suis allé à
à une fête
j'ai bu beaucoup quelque chose comme ça
donc je crois que ça change ici ils voient
parce qu'il y a l'influence américain
en France aussi et

ça change le
tout la
la société ici
parce que avant c'était les Français
boivent un peu
un peu de vin avec leur repas c'était tout mais maintenant je crois que ça change
c'est pas
c'est pas la même niveau qu'en Angleterre
bien sur que non c'est pas la même niveau mais
mais ça change aussi en France
avec les Fran
ça dépend parce que les Français à Lille ils boivent
ils ont0
ils avaient de leur voiture avec eux donc ils ont bu
une bière
et puis
ils ont retourné chez eux
pour travailler le prochain jour mais quand je suis allée à Reims pour voir une amie là bas
elle boit beaucoup
et ses amis ont
ont bu beaucoup beaucoup et
donc je crois ça dépend des
des gens
en fait
oui ici je suis allée à une fête il y avait une fille qui a trop bu
mais
et les autres a pensé qu'elle était une idiote
pour faire ça mais avec d'autres amis
ça c'est normal ça c'est
leur habitude donc ça dépend des gens je crois ici
mais je sais que bien sûr les Anglais boivent
plus que les Français
ah c'est différent ici
c'est différent oui
chez nous
parce que
ici c'est
oui je suis allée à la fac ici
dans quelques cours aussi
et
c'est
c'est ouais
c'est très différent
pas très différent mais différent parce que
je sais pas c'est pas parce que il y a beaucoup qui
qui habitent près de Lille
par exemple donc
ils retournent chaque week-end chez eux
mais moi je retourne tous les deux mois
chez moi parce que j'habite trop loin
et normalement on habite dans des

des grandes collocations ici c'est des petits appartements ou tout seul
aussi
en Angleterre c'est plus
le travail est plus indé
indépendent
que
ici on prend des notes et puis
on les apprend pour l'examen
mais chez nous
le prof dit ah j'ai
j'ai un livre
va à la bibliothèque
pour prendre
pour prenez
pour prendre des notes
oui donc c'est
et
ici on fait
30 35 heures
à la fac chaque semaine
chez nous c'est entre 10 et 15
mais on a besoin de travailler toute seul
donc
oui c'est différent c'est différent système
je
je préfère notre système parce que je n'aime pas des cours de 3 heures
comme mes amis ici
mais je crois que c'est bien si on
si on a besoin de rechercher quelque chose tout seul
aussi on aime que notre système
avec notre système on peut faire
ce qu'on veut et puis
on a tous les choix donc ici
le commerce maths sciences ce sont
plus
importants
que
je n'sais pas littérature
histoire
j'ai un ami ici qui fait
le commerce
parce que il voudrait travailler
dans le secteur financier même que il adore histoire
mon frère il va faire histoire à la fac
et il voudrait travailler dans le secteur financier donc il va faire
histoire pour 3 ans puis je n'sais pas un an 2 ans
de
du commerce
à un collège
ou quelque chose comme ça
donc j'aime que nous avons le choix
de faire tous les choses différents donc moi

je ne veux pas être prof de français
peut-être mais c'est pas
c'est pas mon but donc
j'ai le choix de faire
quelque chose dans un autre secteur

Food discussion task, Rose, time 1 CHAT format

@Begin
@Languages: fr
@Participants: Ros Subject, car Investigator
@ID: fr|ncl|Ros|20;00.00|female|Y2||Subject||
@ID: fr|ncl|car||female|||Investigator||
@Date: 15-JUN-2009
@Location: NCL
@Situation: Food discussion task
*Ros: je pense que c' est important pour l' école de euh prendre un rôle euh (..) important dans la lutte de [//] contre l' obésité .
*Ros: mais euh je crois que c' est la responsibilité des: parents (.) euh plus parce que euh si on (.) [/] si on sait enfant la différence entre les: [//] la nourriture bon et la nourriture grasse euh on peut (.) faire le choix .
*Ros: mais euh je pense que euh c' est une bonne idée de [//] d' augmenter les prix des aliments trop sucrés parce que c' est euh [//] ce n' est pas (.) trop évident .
*Ros: et euh les enfants ne sont pas forcés euh à euh (.) à manger la nourriture grasse .
*Ros: je pense que c' est trop sévère d'interdire les [/] les distributeurs [/] distributeurs automatiques de chips parce que euh quand on euh on essaye de forcer un enfant spécialement de <ne faire> [//] ne faire pas un chose euh ils (.) veulent faire le chose plus .
*Ros: et aussi c' est une bonne idée de [//] d' avoir un plus grand choix de plats qui sont bons la santé .
*Ros: et la publicité aussi .
*Ros: euh (.) on peut (.) faire les pubs pour (.) la nourriture qui est euh mieux pour la santé (.) dans une (.) moyen positif .
*Ros: mais je pense que: euh (.) c' est un problème différent quand les enfants euh euh <a plus> [//] ont plus âgés (.) parce que l' idée et l' image du corps est trop [/] est trop important .
*Ros: et les problèmes sont plus (.) l' anorexie et le boulimie je pense.
*car: qu' est ce que tu penses de la nourriture qui est disponible sur le campus ?
*Ros: je pense que les prix <sont (.) agréables> [//] sont (.) bons mais la nourriture (.) je pense que ce n' est pas le mieux .
*Ros: le bâtiment de droit est euh est euh (.) plus loin .
*Ros: et il y a une [//] un café près du bâtiment de droit euh qui est mieux .
@End

Food discussion task, Rose, time 2 Text format

il me semble que
les écoles puissent
prendre beaucoup de mesures
pour
combattre l'obésité
mais

aussi il faut
qu'on
qu'on rappelle que
il y a un problème avec les
2 extrêmes, l'obésité et l'anorexie
mais
je pense que
c'est
ce n'est pas une bonne idée vraiment d'interdire les distributeurs automatiques de chips de boissons gazeuses aussi etc parce que si on
si on interdit quelque chose
les
les
les enfants et même les
les gens
justement
veulent ces choses plus
et si
si ils veulent
si ils veulent
acheter et manger ces choses ils peuvent
aller à une magasin de
différent
pour l'acheter
les acheter
mais je pense que c'est
c'est une bonne idée d'augmenter les prix
mais aussi les
les
les
les gens peuvent
aller à une différent
aller
aller à une
magasin différent pour les
pour les acheter
je pense que la
la publicité
de
de1
d'être en bonne santé
peut-être c'est une bonne idée
et aussi
si la
l'école
a une plus grand choix de plats
qui sont bons pour la santé
mais
je pense que
le problème d'obésité
base
se base mais

le problème d'anorexie et
de boulimie
va augmenter alors je pense que dans l'avenir
on peut voir un changement
dans les tendances
des gens
la
la culture en général
il y a une grand
une grand habitude de
de rencontrer des amis
dans un café ou
dans un restaurant
et les choses comme ça
c'est une
c'est une aspect de
de la culture
très
très important je pense
et
la
la culture française est connue pour la nourriture et
les
et
je pense que
la
la nourriture est très
très importante dans la
dans la vie des Français et je pense que c'est une
c'est une rapport
de bonne santé ils
je pense
il me semble qu'il y a
moins de personnes
qui ont le problème de
de l'obésité
en France
que
qu'en Angleterre
oui je suis d'accord mais
je sais qu'il y a une grande différence d'entre les étudiants
les étudiants en Angleterre
spéc
en particulier à NCL
les
les filles
sont
un peu moins
intéressées dans l'apparence
je sais qu'
je
je pense que

ils s'inquiètent aussi mais
dans une
moyen différent
les
les étudiants français
les filles en particulier
se maquiller parfaitement et le
et leurs cheveux
toujours parfaits
mais
je pense que c'est
c'est une différence entre les cultures
oui
je ne sais pas exactement si
les
les étudiants français sont obsé
obsédés
mais
c'est vrai que il y a moins de
de
des étudiants
avec des problèmes de l'obs
l'obésité
généralement
en France on
mange plus tard
le
les choses que
les Français
mangent sont un peu différents beaucoup de pain et beaucoup d' fromage et
comme ça mais
oui je sors dans les restaurants mais
je pense que généralement sont
plus ou moins les mêmes je pense
non
beaucoup de
de mes amis
mangent
à la restaurant de
universitaire resto U
et c'est
c'est pas cher pas du tout mais
la nourriture n'est pas
très bon tout le temps
et je ne
je ne veux pas
prendre le risque
qu'on ne
on ne sait pas
comment
qu'est ce qu'on va
manger

Story-retelling, Iris, time 1 CHAT format

@Begin
@Languages: fr
@Participants: Iri Subject, Car Investigator
@ID: fr|ncl|Iri|20;00.00|female|Y2||Subject||
@ID: fr|ncl|Car||female|||Investigator||
@Date: 25-JUN-2009
@Location: NCL
@Situation: Loch Ness 1
*Iri: une femme est allée en vacances avec sa mère et ses trois enfants .
*Iri: euh et ils ont loué une maison près euh du lac du Loch Ness monster .
*Iri: euh et au [/] au long du vacances les enfants euh ont fait de pêche .
*Iri: et euh la grand_mère euh elle a fait la peinture .
*Iri: et aussi un des enfants l' a fait .
*Iri: et la mère elle a fait la lecture .
*Iri: mais la grand_mère a peinté [/] peinté ?
*Car: peint .
*Iri: une image de [//] du Loch Ness Monster .
*Car: du monstre du Loch Ness .
*Iri: du monstre du Loch Ness .
*Iri: quand la mère est allée en ville pour faire du shopping .
*Iri: la grand_mère et les enfants a décidé euh de créer euh le monstre du Loch Ness .
*Iri: avec les bouées euh ils ont peinté +/.
*Car: peint .
*Iri: +, ils ont peint euh les bouées euh pour recréer le monster .
*Iri: et puis ils se [//] ils euh ont allé euh sur le Loch pour [/] pour laisser les bouées euh dans l'eau .
*Iri: et puis quand la mère euh s' est [//] quand elle est rentrée euh les enfants euh [/] les enfants euh l' a montré.
*Car: lui ont montré .
*Iri: lui ont montré euh le monster .
*Iri: la mère était très choquée .
*Iri: et puis euh la mère a téléphoné à la presse .
*Iri: et tous les journalistes du village euh a [/] a visité le Loch euh pour voir le monster .
*Iri: et euh le monstre a été aussi sur la télévision .
*Iri: mais euh la grand-mère euh elle a admis que le monstre n' était pas euh le vrai [/] le vrai monstre .
*Iri: et la mère elle était euh un peu (.) cross ?
*Car: en colère ?
*Iri: en colère .
*Iri: mais les journalistes et les locaux euh l' a trouvé euh très amusant .
*Iri: et puis après euh l' épisode euh le vrai monster était dans le Loch mais la famille euh n' as pas la vu .
@End

Story-retelling, Iris, time 2 Text format

j'pense qu'ils ont
y sont en vacances
y viennent d'arriver en voiture
à la maison
voilà y sont en train de
prend les valises
d'les
d'les apporter
dans la maison
là ils sont dans le jardin
le
la mère elle est en train de lire les deux enfants sont en train de
faire la pêche
et les deux autres sont en train de
en train de peindre
la grand-mère a peint une
un dessin
d'un monstre
c'est le LN monstre
et le petit enfant il a
ouais il a dessiné un
une image du lac
et là le
la mère elle
elle part avec l'un des enfants elle est en train d'faire des courses en ville
et en c' moment là le
la
la grand-mère elle décide de créer le
le monstre qu'elle a dessiné avec les enfants
ils prend des
avec des pierres pour
pour construire le monstre
il va sur le lac
pour le mettre dans l'eau
dans un petit bateau et
après ça le
la mère elle revient avec l'enfant et
les deux autres enfants disent ah voilà y a le monstre dans le lac
la grand-mère elle est sur le ponton
avec
un fil pour contrôler le monstre
ils savent pas
quoi dire
parce qu'y pensaient que
qu'il n'existait pas
donc ils appellent le
le média
tout l'monde
tout l' monde vienne en fait c'est pas que les médias c'est juste des gens qui sont
intéressés

y viennent pour voir le monstre
ça passe à la télé
et après la grand-mère elle avoue que c'était juste
des pneus et des pierres
c'était une blague
et
ah le
la majorité du monde
semble assez content
mais la mère elle est un peu énervée
et
ouais à la fin y a
ouais y a le vrai monstre
dans le lac

Interview, Lola, time 1, CHAT format

@Begin
@Languages: fr
@Participants: Lol Subject, Car Investigator
@ID: fr|ncl|Lol|20;00.00|female|Y2||Subject||
@ID: fr|ncl|Car||female|||Investigator||
@Date: 15-JUN-2009
@Location: NCL
@Situation: Oral interview
*Lol: d' accord je m' appelle Lola et j' ai vingt ans .
*Car: est ce que tu pourrais te décrire physiquement ?
*Lol: moi ?
*Lol: d' accord .
*Lol: je suis assez grande avec les che(veux) [//] cheveux bruns
les yeux verts .
*Lol: et maintenant je porte un jean et une chemise bleue .
*Car: est ce que tu peux un petit peu décrire ta personnalité .
*Lol: je suis une personne assez diligent .
*Lol: euh et euh je suis responsable aussi .
*Lol: mais aussi j' aime beaucoup rire .
*Lol: oui euh j' aime beaucoup rester avec mes amis et aller au
cinéma voir les films etc .
*Car: tu es d' où en Angleterre ?
*Lol: j' habite à B oui dans le sud ouest .
*Lol: oui .
*Car: est ce que tu pourrais me parler de ta vie avant d' être ici à l'
université ?
*Lol: je suis allée à l' école euh pour pour les filles
pour [//] pendant sept ans .
*Lol: euh oui euh j 'ai j' ai habité avec mes parents et mon frère
.
*Lol: euh maintenant il a seize ans .
*Lol: euh oui je suis allée à l' école .
*Lol: j' ai travaillé dans un supermarché des samedis .
*Lol: je n'ai pas fait grand chose [//] beaucoup [//] grand
chose .
*Lol: je suis allée au cinéma .
*Lol: je suis allée dans les boîtes avec
mes amis .
*Lol: j' ai fait mes études .
*Lol: j' ai joué la guitare aussi .
*Lol: je sais pas je je préfère la vie actuelle parce que
j' ai plus d'ind(ependance) [//] indépendance .
*Lol: euh euh mais aussi maintenant um je crois que la vie euh à
Bath <était plus facile> [//] était plus simple .
*Lol: je suis allée à l' école .
*Lol: j' ai fait mes devoirs .
*Lol: c' était tout .
*Lol: mais maitenant je dois me préparer pour l'année prochaine .
*Lol: je dois faire beaucoup de choses comme euh l' électricité à
la maison et les choses comme ça qui sont un peu ennuyeux .

*Car: tu peux un peu m' expliquer ta vie étudiante ?
*Lol: j' habite à West Jesmond avec sept étudiantes cinq garçons et deux [/] deux filles .
*Lol: notre maison est assez dégoûtante parce qu' il y a huit personnes qui habitent là .
*Lol: mais oui j' ai une chambre assez grande et il est près des autres étudiants et de Tesco .
*Lol: donc pour faire des courses et aller à l' université .
*Lol: il fait vingt minutes pour marcher à l' université .
*Lol: comme ci comme ça .
*Lol: j' entends très bien avec une fille et l' autre euh habite à l'université je crois ou dans la bibliothèque .
*Lol: mais ils font beaucoup de bruit donc j' ai des problèmes avec euh le couchage ?
*Lol: j' ai des problèmes de dormir .
*Lol: mais non ça marche .
*Car: et donc quelles sont les matières que tu étudies cette année ?
*Lol: je fais le français parce que dans mon premier an j' ai commencé avec la géographie .
*Lol: donc cette année ce n' était pas possible de faire une autre langue .
*Lol: euh donc je fais la littérature l' histoire de [/] de Vichy le grammaire le film français et aussi les valeurs républicains .
*Car: et tu fais aussi de la géographie ?
*Lol: non c' était seulement l' année dernière .
*Car: et quel est l' aspect du français que tu préfères ?
*Lol: la matière ?
*Lol: j' aime beaucoup le film français parce que j' adore le cinéma .
*Lol: je vais très souvent avec <mes amis> [//] mes amis français euh au cinéma à Tyneside .
*Lol: aussi la littérature est assez intéressante .
*Lol: et les valeurs républicains ça m' interesse beaucoup parce qu' il intéressant de voir comment les français pensent et leurs valeurs .
*Car: est ce que tu peux me parler de films que tu as bien aimés ?
*Lol: euh j' ai vu il y a longtemps que je t' aime .
*Lol: oui c' était un très bon film .
*Lol: aussi j' ai vu euh au bout des autres au goût des autres ?
*Lol: c 'était un film assez ancien .
*Lol: mais c' était un très bon film .
*Lol: qu' est ce que j' ai vu aussi ?
*Lol: je ne peux pas souvenir maintenant .
*Car: quel serait ton film préféré ?
*Lol: j' aime beaucoup les choristes .
*Lol: oui parce que j' adore .
*Lol: c' est un film très (.) c' est un film très joli .
*Lol: j' adore aussi les chansons dans le film .
*Lol: j' ai le C_D .
*Lol: je suis allée avec ma famille quand j' étais petite .
*Lol: euh je suis allée quand j' avais douze ans pour faire euh une

échange euh à Normandie .
*Lol: aussi je suis allée à Bordeaux trois fois pour visiter une amie quand j' ai fait l' échange Bristol Bordeaux .
*Lol: et euh l' année dernière je suis allée euh dans la région de Lot et aussi à TouLolse .
*Car: mais avec qui ?
*Lol: je suis allée euh pour être un au pair mais ce n' était pas très bien passé donc j' ai quitté je l' ai quitté .
*Lol: et je suis restée avec une amie à TouLolse pour une semaine .
*Car: une amie française ?
*Lol: les enfants étaient horribles terribles et euh non .
*Lol: et aussi euh ils ont voulu <un esclave> [//] une esclave .
*Lol: et ce n' était pas pour moi .
*Car: normalement tu n'es pas censée .
*Car: et l' année prochaine quels sont tes projets ?
*Lol: euh j' irai à Lyon, Lille ou Rouen pour faire un stage de huit mois d' être professeur de langue d' anglais euh pour des ados [//] des adolescents ouais ?
*Lol: et après je vais voyager dans le pays .
*Lol: je voudrais visiter mes amis et aussi je voudrais aller en Allemagne parce que j' ai euh j' ai la famille en Allemagne .
*Car: où ?
*Lol: dans le nord, c' est au bord de la mer .
*Car: pourquoi tu as choisi d' être assistante ?
*Lol: parce que euh j' ai pensé que euh si j' irais à l' université euh j' aurais de amis anglais .
*Lol: et aussi et je voudrais àméliorer mon français .
*Lol: donc j' ai choisi l' école pour parler plus de français j' espère .
*Lol: et aussi pour avoir un petit salaire et de faire [/] de faire un stage que je peux mettre sur mon C_V .
*Lol: non ce n' est pas beaucoup mais c' est plus euh c' est plus professionnel je crois que [//] qu' aller à l'université .
*Car: je comprends .
*Car: est ce que tu es inquiète par rapport à l'année prochaine ?
*Lol: un peu parce que je en sais pas euh je ne sais pas où où j' irai .
*Lol: mais euh non ça ne m'inquiète beaucoup parce que je suis assez adaptable .
*Lol: et oui [//] non .
*Car: est ce que tu peux me parler de tes loisirs ou passions ?
*Lol: euh j' ai déjà parlé de du cinéma .
*Lol: aussi je fais la guitare un peu .
*Lol: euh je fais aussi les scouts euh chaque semaine .
*Lol: euh qu' est ce que je fais aussi ?
*Lol: euh je sais pas j' aime beaucoup le shopping .
*Lol: et euh j' aime beaucoup lire .
*Lol: um j' aime beaucoup voyager .
*Car: est ce que tu pourrais me parler de très bonnes vacances que tu as passées ?
*Lol: d' accord en deux mille cinq je suis allée en Tunisie avec

ma famille pour une semaine .
*Lol: c' était très bien passé .
*Lol: j' ai mangé beaucoup .
*Lol: je je suis allée à la plage .
*Lol: j' ai lu le code da vinci .
*Lol: nous avons visité aussi euh des marchés et aussi un site de Carthage .
*Lol: et euh oui j' ai beaucoup aimé la semaine parce que (.) <c'était euh> [//] il faisait chaud .
*Lol: et normalement nous faisons des vacances euh éducatives .
*Lol: donc c' était très relaxant .
*Lol: c' était très bien passé .
*Lol: euh oui oui en juin euh je vais euh à Saint Pétersbourg en Russie euh pour trois semaines.
*Car: je reviens de Saint Pétersbourg !
*Lol: et oui je [/] m' attends avec impatience pour ça .
*Lol: <c' est un peu euh> [//] ça m' inquiète un peu parce que euh c' est assez loin d' Angleterre .
*Lol: et aussi je ne parle pas euh le russe .
*Lol: mais j' espère que ça va me préparer pour France en septembre .
*Car: oui tu verras ça sera plus facile en France !
*Car: est ce que tu peux me parler de quelqu' un que tu admires et pourquoi ?
*Lol: je sais pas .
*Car: ou le contraire quelqu' un que une personnalité que tu n'aimes pas et pourquoi ?
*Lol: personnalité que je n' aime pas ?
*Lol: une personne que j' admire est <la reine> [//] la reine d' Angleterre parce qu' elle a (.) plus de euh quatre quatre vingts ans.
*Lol: mais encore elle fait euh ses duties ?
*Lol: elle fait ses devoirs .
*Lol: elle a toujours l' esprit de passion pour son travail et ce n' était pas sa choix d' être la reine.
*Lol: et il n'y a pas beaucoup de personnes dans notre pays <qui travaillent> [//] qui font de travaux quand ils ont plus de soixante ans .
*Car: et au contraire quelqu' un que tu n'admires pas du tout ?
*Lol: euh je ne sais pas .
*Lol: peut être Jodie Marsh parce que je ne crois je ne pense pas que [//] qu' elle fasse beaucoup de choses .
*Lol: euh aussi je n' admire pas euh les personnes qui ont des bénéfices et qui sont paresseuses .
*Car: O_K euh quels sont à ton avis dans la société d' aujourd'hui les problèmes majeurs ?
*Lol: les problèmes majeurs je crois que notre gouvernement ont euh a fait [/] a fait des grandes erreurs quand ils ont euh décidé de de couper le le taxe pour les bénéfices pour des personnes qui se sont mariées parce que <dans notre société> [//] en Angleterre il est on gagne plus d' argent si on est une femme toute seule.
*Lol: et à mon avis ça cause des problèmes s' il n' y a pas deux

parents dans une famille .
*Lol: et euh donc oui c' est un problème .
*Lol: aussi euh les problèmes financiers .
*Lol: mais heureusement ça n' est pas un problème pour moi en ce moment .
*Lol: aussi je crois que nous avons un très grand problème avec l' alcool .
*Lol: ce n'est pas un problème majeur en France .
*Lol: par exemple quand je suis allée à TouLolse euh j' ai vu que il n' y avait pas la même culture de ça quand on sort dans le soir et on est bourré .
*Lol: donc c' est un problème ici et ça utilise des ressources de N_A_S [//] N_H_S .
*Car: la santé publique .
*Car: pour ton examen oral tu as choisi de parler des inégalités et de la discrimination envers les femmes .
*Car: pourquoi est ce que tu as choisi ce sujet ?
*Lol: parce que tous les autres sujets étaient très ennuyeux .
*Lol: d' accord alors j' ai choisi ce sujet parce que je suis une femme donc et je voudrais travailler après l' université .
*Lol: et je sais que il y a (.) des problèmes avec la discrimination en ce par exemple les salaires et aussi la matérnité .
*Lol: et donc euh j' ai pensé que c' était un sujet assez intéressant de rechercher .
*Car: par exemple concrètement est ce que tu pourrais donner des exemples de problèmes ?
*Car: quels sont les problèmes auxquels elles sont confrontées ?
*Lol: il y a le plafond du verre .
*Lol: c 'est euh c' est imaginaire c' est euh oui c'est qui arrête les femmes d' avoir des postes avec plus de responsabilités .
*Lol: c' est [/] <c' est causé non> [//] la raison pour ça est si les enfants peut être a les enfants euh elle prend le temps pour aller chez elle donc ça arrête le progrès .
*Lol: aussi et donc il y a des femmes qui travaillent de temps partiel .
*Lol: elle aura des salaires plus petits que le hommes .
*Lol: mais je ne sais pas s' il y a solution pour ça parce que nous sommes biologiquement différents .
*Car: alors par exemple mais si un jour tu as des enfants .
*Car: est ce que tu voudrais soit t'arrêter de travailler ou travailler ensuite à temps partiel ?
*Car: quelle est à ton avis la meilleure chose à faire ?
*Lol: euh je ne sais pas mais <je voudrais> [//] si j' aurais le support financière je voudrais rester avec mes enfants .
*Lol: parce que ma mère a arrêté de travailler euh quand j' étais née .
*Lol: et euh c' était c' était très positif pour mon frère et moi parce qu' elle était toujours là et même aujourd'hui .
*Lol: euh et mais je sais que <nous sommes assez> [//] nous sommes dans une situation assez privilégiée pour faciliter ça .
*Lol: donc peut être que je vais travailler quand si j' aurais euh des enfants si quand ils auront plus de cinq ans peut-être ?

*Lol: mais je ne sais pas encore .
*Car: il y a des gens qui disent que les femmes au foyer devraient avoir un salaire ?
*Lol: euh c' est [//] je sais qu' il y a une grande division dans les femmes avec euh l' idée de travailler ou de ne pas travailler après avoir un enfant .
*Lol: mais à mon avis euh ma mère travaille très dur .
*Lol: elle [//] parce qu' elle n' a pas un travail euh notre maison est un travail pour elle .
*Lol: et elle doit avoir toute la maison complètement euh euh nettoyée et complètement euh bien rangée.
*Lol: non je ne pense pas qu' elle a besoin d' avoir un salaire et ou les autres femmes si [/] si elles ont le support financier .
*Lol: donc c' est leur choix .
*Car: est ce que tu penses que pour certaines professions euh les hommes soit les hommes soit les femmes sont plus qualifiés en raison de leur sexe ?
*Lol: oui je crois que qu' il y a des métiers qui sont euh qui sont <qui sont mieux> [//] qui sont meilleurs pour chaque euh chaque sexe .
*Lol: euh non oui parce que nous sommes nous sommes différents.
*Lol: nous nous pouvons être égal mais aussi nous sommes différents et nous avons des qualités différentes .
*Lol: euh par exemple en général les femmes sont plus euh .
*Lol: euh je ne connais pas le mot .
*Lol: euh gentle ?
*Car: douces .
*Lol: douces sont plus douces .
*Lol: et et elles font des bons chefs à mon avis .
*Lol: les hommes peut être ils sont plus euh euh occupés avec leur Euh l'argent et le succès de l' entreprise .
*Lol: oui mais peut être les femmes ont ont euh euh plus de euh .
*Lol: they care about .
*Lol: oui plus de respect .
*Lol: et oui mais ça dépend des hommes et des femmes .
*Lol: mais aussi s' il y a un travail qui est assez physique peut être <c' est> [//] ce serait meilleur pour un homme mais ça dépend des femmes .
@End

Interview, Lola, time 2, Text format

je m'appelle Lola
j'ai 21 ans
depuis la semaine dernière
j'habite à B dans le sud ouest d'Angleterre
je vais à la fac à Newcastle
à la fac je fais français la littérature la langue l'histoire
j'ai fait un peu d'la géographie dans mon premier an
j'adore aller au cinéma
aussi j'aime beaucoup faire du shopping

j'aime beaucoup faire du vélo
d'la natation
j'aime beaucoup lire
j'sais pas j'aime beaucoup voyager
voir d'autres pays différents villes
des choses comme ça
je travaille dans une lycée
qui est à
3 minutes
d'ici
c'est une lycée scientifique et je
j'enseigne l'anglais
avec des groupes de
12 étudiants
ils ont entre 16 et 22 ans
donc
le but est que
ils parlent
ça ne marche pas toujours comme ça mais
j'enseigne l'anglais
les autres profs sont très sympas
il y a
4 profs d'anglais
un qui est un peu bizarre mais les autres sont très sympas
je
je connais les autres profs parce que je suis avec eux
au salle des profs
ils étaient très
accueillants
avec moi
ce
ça passe bien avec eux
j'ai
je n'sais pas j'ai environ
10 groupes
non plus peut-être parce que
j'ai des secondes premières terminales et autres classes comme BTS
cet âge
ça passe bien avec les 2 groupes qui sont des groupes oraux
parce que ils ont un bon niveau d'anglais mais les autres
les secondes par exemple sont
ne sont pas très motivés ils sont
c'est un vrai défi
avec eux
et
c'est assez épuisant
leur
leur cours
avec eux
oui mais les premières et les terminales sont plus
plus murs que les autres je crois
lundi

le lundi le jeudi et le vendredi je travaille au lycée
c'est 4 heures chaque jour mais
je reste là toute la journée
et puis
mardi
je dors normalement et puis je fais cours de français le soir
mercredi je sais pas je vais au cinéma ou quelque chose avec les amis et le week-end

je fais des
je n'sais pas je fais des choses avec des amis
j'ai fait
quelques voyages en France
Paris
Reims Amiens
des villes
assez proches d'ici
j'ai beaucoup aimé Paris bien sur c'était
j'ai visité 3 fois
cette année
aussi j'ai beaucoup aimé Lyon
j'y suis allée en février et aussi à Grenoble
mais j'ai préféré Lyon c'est une très belle ville
aussi bien sur Paris j'aime beaucoup
et aussi je suis allée à Bruxelles et Bruges
j'ai beaucoup aimé B c'est une
c'est une
ville
belle
j'habite dans un foyer
il y a
100
120 chambres
avec d'autres étudiants d'autres assistantes et aussi ceux qui cherchent du travail
et j'ai trouvé un loge
le logement parce que j'avais une amie anglaise qui habite ici
et
elle m'a raconté qu'il y a ce logement ici
j'ai essayé de trouver quelque chose par Internet mais
j'avais pensé je voudrais voir
la chambre avant d'habiter ici
donc ma mère m'accompagnait
à L
pour 3 jours pour chercher un logement
et nous avons trouvé
ce foyer ici
parce que à cause de la
la CAF
donc je paye la moitié
de loyer
non c'est pas trop cher c'est
je crois que je paye moins qu'en Angleterre à cause du CAF
j'ai rencontré bien sur beaucoup d'autres assistantes anglaises

mais aussi je suis allée à la fac ici pour
pour trouver des amis français
français
donc j'ai quelques amis français ils
ils ont beaucoup de travail
donc j'les vois je n'sais pas
2 ou 3 fois par mois c'est pas
c'est pas beaucoup mais
c'est
non c'est sympa
je suis allée
dans la salle de
de langue
et j'ai rencontré quelqu'un qui m'a dit ah j'ai un
un ami qui voudrait apprendre l'anglais
qui voudrait améliorer son anglais
oui donc j'ai pris son numéro de
de portable
et
aussi il y a
une assistante allemande
ave qui
qui je travaille et
elle est bilingue en français donc on parle en français tout le temps et ça c'est bien aussi

oui il y a des
des filles espagnoles ici qui sont très sympa
et l'année dernière il y avait aussi des Italiens
qui ont travaillé dans un hôpital ici
oui les Espagnoles sont très sympas aussi
et aussi je vais à l'Eglise internationale ici donc j'ai rencontré
des gens
un peu partout du monde
j'ai beaucoup aimé la ville parce que il faisait chaud
ou il était
sympa j'ai pas travaillé le der
le 1er mois j'avais
2 semaines au lycée quand
j'observais des cours
j'sais pas c'était
j'crois qu' c'était
excitant
pour moi parce que ah oui j'étais en France
c'était très cool
j'ai rencontré beaucoup de
de personnes ici
le premier mois je suis allée à Paris pour voir un match du foot
au stade de Paris donc ça c'est
c'était
très cool aussi
donc j'ai fait plein de choses
c'était très cool j'ai beaucoup aimé mon temps ici

mon temps ici
je crois que c'était une bonne expérience en lycée parce que même
même si
j'n'ai pas aimé travailler au lycée
je n'ai pas
c'est
c'est une
une ville assez
déprimée
déprimante
c'est pas à Lille c'est
c'est
c'est une autre ville
et donc
les élèves de temps en temps ils sont très malpolis ils sont
très différents de
parce que je suis allée à une école pour des filles
en Angleterre donc c'est très différent c'est
70% sont des immigrés
et donc ça
je sais pas
c'est
c'est une atmosphère
différent mais
c'est une bonne expérience quand même
avant Noel j'étais très nostalgique
pour mon pays
j'ai voulu retourner chez moi parce que
j'ai
j'habitais ici pour 3 mois
sans voir ma famille
mais après Noel ça améliorait
mais j'ai
ma famille est allée à
en février ici
je suis allée à Grenoble et Lyon
et maintenant j'ai que 2 semaines au lycée donc
ça a passé très vite c'était une très bonne expérience pour moi je crois
j'ai 2 semaines de vacances maintenant donc je vais à C
pour une semaine parce que le parrain
de mon frère
a un appartement là bas
et sa femme est française
ils habitent normalement à Londres
donc
je vais rester avec eux pour une semaine
puis
j'ai une rédaction à faire
en français
pour
N
la fac à N

je vais parler un peu d'la
le tourisme à L
parce que c'était le
capitale de culture en 2004
et je vais faire quelque chose sur ça
c'est pas
c'est pas
quelque chose qui m'intéresse beaucoup
c'est un sujet parce que j'avais des problèmes de trouver quelque chose
oui donc je vais faire ça
puis
j'ai 2 semaine au lycée
et puis
je vais retourner le premier week-end de mai parce que c'est un jour férié
en Angleterre
donc mes parents vont venir ici
pour me chercher en voiture
puis je vais finir mon rédaction
je vais
donc
je vais
je sais pas reconnecter avec l'Angleterre
et puis j'espère d'aller en Suisse
au fin de mai
pour voir une amie qui fait son Erasmus là bas
oui en Suisse
à L
oui donc c'est une très belle ville j'ai jamais
visité Suisse
donc j'espère de faire ça
et puis juin et juillet
je suis pas encore sure si je vais
retourner en France ou si je vais
faire quelque chose d'autre
puis en Aout je vais au Pays de Galle parce que nous avons
une 2ème maison au bord de la mer
pour rester avec ma famille avant de partir
non j'avais pensé peut-être d'être au pair
jeune fille au pair
sinon je vais retourner en Russie
pour 1 mois
donc
je n'sais pas
2 choix très différents
je dois décider
chaque semaine je crois
pour moi je crois premièrement c'est le système
scolaire c'est très différent le relation entre le prof
el les élèves est très différent que chez nous
c'est plus
c'est beaucoup plus proche en Angleterre
on connait je n'sais pas on connait des faits de leur famille on connait

quelques trucs ici c'est vraiment
madame le prof et les élèves
ça c'est différent
aussi le
un petit truc mais pour moi c'est
c'est amusant
que tout le monde dit toujours bon appétit bonne fin d'après midi c'est toujours bon
quelque chose c'est toujours
c'est
c'est amusant
oui quand on mange à la cantine c'est quelque chose très important
pour nous c'est que on mange quelque chose et on parte mais non
on mange ensemble
3 fois par semaine
oui ça va
de temps en temps puis la chose
que j'aime pas trop mais
aussi il y a
des choses qui sont
sont bien
il n'y a pas
beaucoup de choses
très sucrés
il y a toujours
yaourt ou fruit
chez nous on avait
des gâteaux avec
je sais pas
beaucoup de frites
choses comme ça
mais
et aussi on peut prendre
une entrée un plat principal un dessert
un yaourt et un
fromage
c'est
ouais c'est beaucoup
mais
c'est
oui la nourriture ici les repas sont plus importants pour les Français je crois que
que chez nous
d'autres choses
parce que je vais au cinéma
beaucoup je n'aime pas que les films sont toujours en version française
ça c'est
quelque chose qui m'énerve un peu
j'aime beaucoup voir les films français mais je n'aime pas voir des acteurs américains
avec des voix français
et aussi quand
il y a
j'ai des élèves qui
qui pensent que ah ce film est français j'ai dit

non c'est américain
d'originalité américain mais
ils ne connait pas
je crois aussi quelque chose qui m'a frappée que
ici il n'y a
il n'y a
pas beaucoup de gens qui parlent anglais
à L
donc c'est bien pour moi
mais aussi
au lycée il n'y a pas beaucoup
il y a ceux qui pensent ah Angleterre c'est
c'est un pays
je n'veux pas
y aller
parce que avant
tous les Français que j'ai rencontrés en Angleterre
étaient là parce que ils ont aimé l'anglais
et j'avais pas pensé de ça
donc c'était un vrai
surprise au lycée de découvrir des gens qui
qui détestent les Anglais
même moi juste parce que je suis anglaise
les Français ne font pas beaucoup de shopping peut-être
ils n'ont pas la même obsessions
qu'en Angleterre
par exemple le dimanche
tout est fermé ça c'est
quelque chose que
j'aime pas trop c'est
j'aime beaucoup faire mes courses le dimanche quand il y a moins du monde
mais ici le samedi
il y a
il y a du monde dans les magasins mais
aussi
oui je crois en Angleterre on peut dire ah je vais faire du shopping c'est mon hobby

mais pas en France c'est
on achète un petit truc c'est tous
tous en modération
on pourrait dire
on mange en modération on fait
on shop en modération
oui mais
ça c'est quelque chose qui m'a frappée aussi
non pas vraiment parce que c'est une église internationale donc c'est un style
assez similaire
de
celui en Angleterre mais je suis allée aussi à une église française
mais
c'est très moderne
mais

j'ai
je préfère les églises en anglais parce que c'est ma langue maternelle
non c'est pas très différent c'est un style up peu
parce que je vais
à une église baptiste
à NCL
je ne vais pas à l'église à Bath
pas encore
donc c'est un peu moins
américain
on pourrait dire ici mais non c'est bien aussi j'ai rencontré un grou
groupe de gens qui
qui vient de
tous les pays du monde
oui c'était bien
c'est bien oui
j'ai eu des visiteurs
qui étaient à B
et aussi en France je les ai rencontrés pour
un jour 2 jours
à L
maintenant
la plupart de mes amis sont en Espagne ou
Argentine en Chine
donc ils sont
moins proches
j'ai
j'ai fait quelques visites à Reims pour un ami
Paris
Grenoble pour voir des amis donc
Tours non Liz est allée ici
elle a restée avec une amie française donc
j'l'ai vue
un jour à Lille
non j'avais pensé d'y aller mais
aussi à Nancy
mais finalement
j'ai préféré
d'aller à G et aussi que je vais
à C
la semaine prochaine
hou ça c'est une question difficile je sais pas parce que
c'était le choix de ma mère
oui parce que j'avais pensé que ça serait trop difficile de trouver un
un stage de travail
et
je vais à la fac en Angleterre je n'ai pas besoin d'aller ici
aussi
donc ma mère m'a dit ah tu devrais être
peux être assistante ce serait bien
j'ai dit pourquoi pas
c'était plus facile je crois parce que

ouais
j'avais pas un problème de trouver quelque chose
je crois que oui je voudrais répéter l'expérien
si
si j'avais le choix de rechoisir
parce que oui c'était une bonne expérience
même
s'il y avait quelques
novembre l'année dernière j'étais
vraiment
j'avais vraiment pensé je voudrais retourner en Angleterre
France m'énerve tous les choses français
mais maintenant
mais c'était
c'était bien quand même c'était
une expérience en lycée c'était
aussi d'être un plus
peu plus indépend
pen
pendante
d'habiter en France
oui j'ai appris plein de choses donc
oui je voudrais
si
si j'avais le choix
je fais le même choix je crois
oui ça dépend des gens j'ai
je crois que mes amis ici aiment beaucoup des écoles primaires et des éc
des collèges mais pas les lycées c'est ceux qui sont dans des lycées
à L ou près de L qui
qui avaient des problèmes
non pas vraiment
peut-être travailler pour une
une organisation caritative de faire quelque chose comme ça
je ne sais pas ou maintenant je voudrais
je voudrais aller aux Etats Unis
pour faire Camp America
c'est
c'est pour des
pour des ados
c'est pour 8
8 semaines
ils font des choses assez
je sais pas comment expliquer ils font des choses sportives ils font des choses
dessins trucs comme ça
je voudrais faire ça pour 2 mois et puis je voudrais
voyager un peu aux Etats Unis
puis je ne sais pas si je vais rester ici
rester en Angleterre
ou peut-être aller en Allemagne parce que
j'ai d'la famille là bas
donc

je sais pas encore mais
j'ai beaucoup
j'ai
beaucoup d'options
peut-être prof mais
aux Etats Unis ou en Angleterre
je
je n'aime pas le système ici
c'est
à mon avis les jours sont trop longs
pour des élèves
et aussi
oui je préfère les relations
entre les profs
et l'élève en Angleterre
mais
c'est vrai qu'il y a des problèmes dans notre système éducatif aussi
donc
je ne sais pas
essayer de rencontrer ceux qui ne sont pas anglais
mais aussi d'avoir des amis qui sont anglais
avoir
un mélange entre
les différentes nationalités
parce que
sinon c'est
c'est bien de pratiquer
français
votre français mais aussi
c'est bien d'avoir des amis qui comprend la situation qui
qui comprend les choses culturelles que
que
tu te manques
aussi d'autres choses je crois
je ne sais pas
de décou
découvrir les villes de voyager le plus que
que possible
aussi de retourner
tous les
tous les 2 mois peut-être en Angleterre
parce que j'avais fait 3 mois ici et c'était trop long pour moi
de télécharger Skype
ça c'est bien aussi
au lycée
je ne sais pas
de
d'avoir
de
d'avoir beaucoup d'options pour chaque cours
parce que ça change de temps en temps
de temps en temps

il y a une chose qui ne marche pas
d'avoir toujours des jeux
dans ta tête pour faire à la fin du cours si
s'il y a les
des élèves qui sont plus vite

Appendix C. FS List

This appendix gathers the lists of FS identified in the transcripts regrouped in appendix B. These lists were obtained by transferring the FS identified on the Praat files onto text files. Each line represents the formulaic material in the same fluent run.

FS identified for French 1, Sally

pour moi
j'aime beaucoup
j'aime beaucoup
c'est pas
je vais à l'étranger
on va
c' est vraiment difficile de
je pense que
dépend de
je pense que ça pourrait être intéressant
c'est pas
c'est pas
c'est
je sais pas comment on peut
c'est ça
ça va
on est quelqu'un qui
pour moi
je vais je veux
je pense que ça c' est
c' est vraiment difficile de
c' est difficile de
c'est le même
je suis d'accord
ça dépend
ça n' importe pas c' est difficile
je pense que
c'est facile
ça c' est pas une bonne raison c' est
je sais pas
je peux pas
c'est utile c'est facile
tout le monde
c'est pas vrai
c'est impoli
c'est
c'est impoli c'est pas
s'il y avait
tout le monde
je pense que
c'est
on peut

FS identified for French 2, Sally

pour moi
c'est
j'aime bien
j'aime bien
c'est très intéressant de
tu peux pas
il y a
t'essayes de
ça marche pas
j'aimerais
j'aime bien aller à l'étranger
pour moi
je sais que le fait que c'est pas pour ça que
ça c'est
c'est cool c'est pas pour moi
j'aimerais bien
je sais que c'est dur
je sais même pas si
je vais
je vais
à Newcastle
ce que je veux faire
il faut faire un master
fais un master en Angleterre
je vais en Angleterre
tout l' temps en cours
par contre à Paris en France en Espagne je sais pas
faire un master
j'sais pas si En Espagne
en France
décide de
c'est pas pour moi
c'est sûr
c'est juste
ça m'a beaucoup plu
ça dépend
c'est amusant
j'aime bien
je dirais pas qu' c'est facile
c'est pas que c'est difficile
il faut juste il faut juste
plein d'
chez moi
j'sais pas chez moi
c'est
j'avais pas peur de faire des erreurs
quand même
je suis beaucoup plus à l'aise
je sais que de temps en temps je vais avoir du mal à
je crois que c'est grâce à
il y a aux Etats-Unis
vient de vient de

c'est pas toujourss facile de
de temps en temps je fais des erreurs
je peux pas
et tout
c'est plus facile
j'ai
en français
c'est
c'est que
j' comprends pas
en fait dans le train
et tout ça
j'vais pas quand-même je sais qu'
chez moi l'année prochaine
c'est que je peux
continuer à
je suis allée j'vais pas trop au cinéma en fait
en VO
 qui s'appelait
y avait pas beaucoup d'
c'était intéressant
pas grand chose
pas vraiment
j'aime bien
c'est à l'heure où je rentre chez moi
je peux
avant d'aller à l'école
de temps en temps il y a
séries américaines
en français
c'est juste
et tout ça

FS identified for Alcohol 1, Lily

j'aime
je ne pense pas que
je ne pense pas c'est
à la fac
par conséquent
c'est un problème
ce n'est pas un problème
je pense que
c'est
c'est difficile
c'est important
continuent à
à cause de
face à face
ça serait
c'est
c'est difficile c'est
c'est
c'est
c'est
par exemple
pour la plupart du temps par exemple
c'est
c'est
aller au restaurant
on va en ville
c'est la plupart
il y a
à l'université
c'est choquant
c'est
c'est difficile
la plupart des (gens)
il y a
c'est normal
je n'ai pas beaucoup d'argent
je vais en ville
de temps en temps
c'est facile
c'est
c'est

FS identified for alcohol 2, Lily

je pense que
je pense que ça va marcher
je pense que ça va
ça va (vraiment) marcher
il faut qu'
ça va marcher
on peut pas on peut pas
je pense que
au niveau
je pense que c'est pas juste
ça va marcher
ça va
ça va marcher
il y a
ça va marcher
faut que
c'est mal c'est pas bon
il y a
je pense que ça va marcher
ça va marcher
la plupart
je pense
ça
je pense que
commencent à
ça va
ça va marcher par exemple
commencé
c'est cool
c'est
c'est normal
je pense au niveau de
un verre de vin
c'était pas
c'était pas
c'était ça va
il y avait
il y a
faut que
c'est je pense
je pense qu'au fond il y a
(d'un côté) il y a beaucoup de gens qui
c'est
c'est être bourré
j'aime pas ça du tout j'aime pas
j'aime
j'aime pas
être bourré c'est
c'est
t'as pas besoin de

d'avoir une bonne soirée
je pense qu'au fond c'est
c'est (plutôt) comme ça
c'est
c'est pas que
les jeunes les jeunes
je pense que c'est
ça marche
il y a les gens qui
en Angleterre je pense
je pense que
en France
c'est pas (d)'être bourré c'est (de)
avoir une bonne soirée
en Angleterre
c'est
c'est
c'est
les émissions télévisées les choses comme ça
c'est
c'est vrai que
il faut que
avoir une bonne soirée
ça marche
je sais pas si c'est je pense que c'est
c'est c'est normal
tout le monde
c'est
c'est-à dire
le soir
ça c'est intéressant
en Angleterre
le soir c'est
c'est plus important
je pense que
j'aime
il y a
en Angleterre
il y a plus de
 pour la plupart
il y a
c'est bon
comme ça en France j'aime
j'aime bien
c'est bon
c'est assez cher c'est
à l'intérieur
il y a
on peut
on peut c'est sympa
on peut
en Angleterre

c'est pas c'est pas

FS identified for Alcohol 1, Lola

c'était
plus de
je ne pense pas qu'il
qu'il y a
c'est
c'est
en Europe
en Grande Bretagne
je n'sais pas
je sais qu'il y a
moins de
sans abri
ça pourrait
je crois que nous avons besoin de plus de
il y a il est
ce n'est pas
de moins de
plus de
je crois que c'est
c'est
nous sommes
plus de
je ne crois pas qu'il y a en Europe
je crois que ça change un peu en ce moment
je suis allée à Toulouse l'année dernière
je suis allée
c'était
c'est
c'est plus de
ça va changer en Europe
par exemple
au début de l'année
(chaque nuit) il y a
au centre
par exemple je n'aime pas
je n'aime pas
j'ai des problèmes est-ce que
est-ce que
ce n'est pas mon problème
à Newcastle par exemple il y a
beaucoup beaucoup de
c'est
plus de
j'ai
j'ai
j'ai
il y a
des choses comme ça
en France
par exemple

j'aime
et caetera
c'est c'est normal
par exemple
nous avons
beaucoup d'
c'est
ça c'est
je crois que c'est un problème beaucoup d'
j'ai
à mon avis ce n'est pas
c'est assez cher

FS identified in Alcohol 2, Lola

je crois que
par exemple
quelque chose comme ça
avait des problèmes
je ne sais pas si ça pourrait marcher je crois que
c'est
c'est quelque chose que
par exemple
la semaine dernière il y avait
plus de
est allé en Espagne
il y avait beaucoup de
(ils étaient)
c'est
ça c'est
ça c'est
ça c'est
je ne sais pas s'il y a (quelque chose qu') on peut (faire) c'est
c'est
je crois que
c'est a des problèmes
est-ce que
je n'sais pas
je sais que
il a
c'est pas
il y avait
sans abri
ils ont
je n'sais pas je crois que
je n'sais pas s'il y a
en Angleterre
je n'sais pas
par exemple
je ne sais pas
ils ont
ils ont je sais pas

il n'y a pas
je ne sais pas si je crois
de plus en plus de
je crois que ça va
on peut on va à la fac
aux Etats-Unis c'est on est
on est à la fac
ça ne marche pas
je n'sais pas
c'est
qu'est-ce que c'est
est-ce que
il y a
est-ce que ça c'est
je n'sais pas je crois qu'
il y a
je crois qu'en France en ce moment
ça change
peu à peu ça change
je suis allé
quelque chose comme ça
je crois que ça change
il y a
en France
ça change
c'était
un peu
un peu de c'était tout je crois que ça change
c'est pas
c'est pas en Angleterre
c'est pas
 ça change en France
ça dépend à Lille
je suis allée à Reims
je crois ça dépend des
en fait
je suis allée à il y avait
ça c'est normal ça c'est
ça dépend des gens je crois
je sais que
c'est différent
c'est différent
c'est
je suis allée à la fac
c'est
c'est
c'est très différent
je sais pas c'est pas il y a
près de
par exemple
chez moi
c'est

en Angleterre c'est
chez nous
j'ai
j'ai
à la bibliothèque
c'est
à la fac
chez nous c'est
on a besoin de
c'est différent
je n'aime pas
je crois que c'est bien
on a besoin de
on peut
on a
je n'sais pas
j'ai
à la fac
il va
je n'sais pas
quelque chose comme ça
j'aime que
c'est pas
c'est pas
j'ai

FS identified in Food 1, Rose

je pense que c'est important
je crois que
c'est
on peut (NC)
je pense que
c'est une bonne idée de
c'est
2
je pense que
c'est
on essaye de
c'est une bonne idée de
je pense que
c'est
je pense
je pense que
je pense que
il y a
près de
près du (NC)

FS identified in Food 2, Rose

il me semble que
beaucoup de
il faut
il y a
je pense que
c'est
une bonne idée
je pense que c'est
c'est une bonne idée d'
je pense que
en bonne santé
c'est une bonne idée
bons pour la santé
je pense que
je pense que
on peut voir
en général
il y a
les choses comme ça
c'est
c'est
je pense
je pense que
je pense que c'est
c'est
je pense
il me semble qu'il y a

en France
en Angleterre
je suis d'accord
je sais qu'il y a
en Angleterre
en particulier
je sais que
je pense que
en particulier
je pense que c'est
c'est
je ne sais pas (exactement) si
c'est vrai que il y a moins de
en France
beaucoup de beaucoup de
comme ça
je pense que
plus ou moins je pense
beaucoup de
c'est
c'est pas cher pas du tout
tout le temps

FS identified in LN1 (story-retelling task), Iris

allée en vacances
est allée en ville faire du shopping
sur la télévision
elle était
un peu
et puis

FS identified in LN2 (story-retelling task), Iris

j'pense qu'
en vacances
viennent d' en voiture
à la maison
y sont en train de
y sont
elle est en train de sont en train de
sont en train de
en train de
c'est
il a
elle est en train d'faire des courses en ville
en c' moment là
décide de
après ça
pensaient que
tout l'monde
tout l' monde en fait c'est pas que c'est juste
(ça passe) à la télé
c'était juste
c'était une blague
la majorité du
elle est un peu
à la fin y a
y a

FS identified for Inter 1, Lola

je m'appelle j'ai
je suis
je suis
je suis
j'aime beaucoup
j'aime beaucoup
aller au cinéma etc
j'habite à B
(dans le sud ouest)
je suis allée à l'école
j'ai
12
je suis allée à l'école
je n'ai pas
grand chose)
grand chose
je suis allée au cinéma je suis allée
j'ai
je sais pas
plus de
je crois que
à Bath
je suis allée à l'école j'ai fait mes devoirs c'était tout
l'année prochaine
beaucoup de choses
à la maison
les choses comme ça
j'habite
il y a
faire les courses à l'université
à l'université
comme ci comme ça
à l'université je crois
beaucoup de
j'ai des problèmes
j'ai des problèmes
ça marche
le français
la géographie
cette année
ce n'était pas possible de
la littérature
l'histoire
c'était l'année dernière
j'aime beaucoup j'adore le cinéma je vais
au cinéma
la littérature
c'est
ça m'intéresse beaucoup
c'était

c'était (X2)
c'était
qu'est ce que
j'ai
j'aime beaucoup
j'adore
c'est
c'est
j'adore
j'ai
je suis allée
je suis allée
je suis allée à B
l'année dernière
je suis allée
à T
je suis allée
ce n'était pas très bien passé
à T
ce n'était pas
à L
faire un stage
je vais + inf
je voudrais
je voudrais
en Allemagne j'ai
j'ai en Allemagne
c'est au bord de la mer
à l'université
je voudrais
faire un stage
ce n'est pas beaucoup
c'est
c'est je crois
à l'université
un peu je ne sais pas
je ne sais pas
je suis
parlé de
un peu
chaque semaine
qu'est ce que
j'sais pas j'aime beaucoup le shopping
j'aime beaucoup
j'aime beaucoup
je suis allée
en Tunisie
(pour une semaine)
c'était très bien passé
je suis allée à la plage
j'ai beaucoup aimé
c'était

faisait chaud
c'était
c'était très bien passé
je vais
en Russie
pour 3 semaines
c'est un peu
un peu
c'est
ça va
en septembre en juin
je n'sais pas
elle a
plus de
ce n'était pas
il n'y a pas beaucoup de
plus de
je ne sais pas
je ne pense pas
beaucoup de choses
je crois que
décidé de
en Angleterre
plus d'
toute seule
à mon avis
il n'y a pas
c'est un problème
ça c'est un problème
c'est
ce n'est pas un problème
je crois que nous avons
ce n'est pas un problème
ce n'est pas un problème en France
par exemple je suis allée à T
il n'y avait pas
on est
c'est un problème
je suis
je voudrais
je sais que il y a
par exemple
j'ai
c'était
il y a le plafond de verre
c'est
c'est
c'est
plus de
c'est
c'est
il y a

je ne sais pas
il y a
je n'sais pas
je voudrais
arrêter de
c'était
c'était très positif
je sais que
je vais
plus de
je n'sais pas encore
c'est
je sais qu'il y a
l'idée de
à mon avis
je ne pense pas qu'elle a besoin de
c'est
je crois que
il y a
par exemple en général
à mon avis
plus de
ça dépend des
il y a c'est
ce serait
ça dépend des

FS identified for Inter 2, Lola

je m'appelle
j'ai
la semaine dernière
j'habite à Bath
je vais à la fac à Newcastle
à la fac
un peu d'
j'adore aller au cinéma
j'aime beaucoup faire du shopping
j'aime beaucoup faire du vélo
j'aime beaucoup
j'sais pas j'aime beaucoup
des choses comme ça
c'est
ça ne marche pas toujours comme ça
il y a
un peu
ça passe bien
j'ai
je n'sais pas j'ai
ça passe bien
par exemple
c'est

c'est
je crois
le lundi le jeudi le vendredi au lycée
c'est chaque jour
toute la journée
cours de français le soir
je sais pas je vais au cinéma le week-end
je n'sais pas
en France
j'ai beaucoup aimé c'était
cette année
j'ai beaucoup aimé
en février à Grenoble
c'est
j'aime beaucoup
je suis allée à Bruxelles
j'ai beaucoup aimé
j'habite
il y a
il y a
essayé de
je voudrais
avant de
pour 3 jours
à cause de
c'est pas trop cher c'est
je crois que moins qu'en Angleterre à cause du
beaucoup d'autres
je suis allée à la fac
j'ai
je n'sais pas
3 fois par mois c'est pas
c'est pas beaucoup
c'est
c'est sympa
je suis allée
il y a
en français en français tout le temps ça c'est bien aussi
il y a
l'année dernière il y avait
je vais à l'Eglise
un peu partout
j'ai beaucoup aimé il faisait chaud
au lycée
j'sais pas c'était
j'crois qu' c'était
en France
c'était très cool
beaucoup de
je suis allée à Paris
ça c'est
c'était

plein de choses
c'était très cool j'ai beaucoup aimé
je crois que c'était une bonne expérience
j'n'ai pas aimé au lycée
c'est
c'est
c'est pas à Lille c'est
c'est
c'est
de temps en temps
différents de
je suis allée
en Angleterre c'est très différent c'est
je sais pas
c'est
c'est
c'est une bonne expérience
chez moi
j'ai
pour 3 mois
en février
je suis allée à Grenoble
j'ai que au lycée
c'était une très bonne expérience je crois
j'ai 2 semaines de vacances
pour une semaine
là bas
à Londres
je vais pour une semaine
j'ai à faire
en français
je vais un peu
c'était
en 2004
je vais
c'est pas
c'est pas
c'est
je vais faire ça
au lycée
et puis
c'est un jour férié
en Angleterre
en voiture
je vais
je vais
je vais
je sais pas
et puis
là bas
c'est
j'espère de

et puis
je suis pas encore sûre je vais
en France je vais
quelque chose d'autre
en Aout je vais au Pays de Galle
au bord de la mer
avant de
jeune fille au pair
en Russie
chaque semaine je crois
pour moi je crois c'est
c'est très différent
c'est beaucoup plus en Angleterre
je n'sais pas
ça c'est différent
pour moi c'est
c'est amusant
tout le monde bon appétit bonne fin d'après midi c'est c'est
c'est amusant
on mange à la cantine c'est quelque chose très important
c'est que
3 fois par semaine
ça va
de temps en temps
il y a
il n'y a pas
bcp de choses
il y a
chez nous
je sais pas
beaucoup de
choses comme ça
on peut
c'est
c'est beaucoup
je crois
chez nous
d'autres choses
je vais au cinéma
je n'aime pas que
ça c'est
j'aime beaucoup je n'aime pas
il y a
c'est
je crois
pas beaucoup de
à L
c'est bien
au lycée il n'y a pas beaucoup
il y a c'est
c'est
en Angleterre

c'était
au lycée
je suis anglaise
pas beaucoup de
en Angleterre
par exemple le dimanche
c'est fermé ça c'est
j'aime pas c'est
j'aime beaucoup faire mes courses le dimanche il y a (moins) du monde
le samedi
il y a
il y a du monde
je crois en Angleterre on peut dire je vais faire du shopping c'est
en France c'est
c'est
en modération
on pourrait dire
en modération
en modération
ça c'est
c'est c’est
en Angleterre je suis allée
c'est très moderne
en anglais c'est
c'est pas très différent c'est un peu
je vais
à NCL
je ne vais pas à l'église à B
c'est un peu
on pourrait dire c'est bien
c'était bien
c'est bien
à B
en France
à L
la plupart de en Espagne
en Chine
ils ont
j'ai
à R
à L
à N
à C
la s’maine prochaine
ça c'est je sais pas
c'était
je vais à la fac en Angleterre je n'ai pas besoin d'
pourquoi pas
c'était plus facile je crois
je crois que je voudrais
c'était une bonne expérience
il y avait

l'année dernière j'étais
je voudrais en Angleterre
c'était
c'était bien c'était
c'était
en France
plein de choses
je voudrais
je crois
ça dépend des
c'est
à L près de
avaient des problèmes
une organisation caritative quelque chose comme ça
je ne sais pas je voudrais
je voudrais aux Etats Unis
c'est
c'est
c'est
je sais pas
j'voudrais je voudrais
un peu aux Etats Unis
je n’ sais pas j’ vais
en Angleterre
j'ai d'la famille là bas
je sais pas
beaucoup d'
aux Etats Unis en Angleterre
je n'aime pas
c'est
à mon avis
en Angleterre
c'est vrai qu'il y a
je n’ sais pas
essayer de
c'est
c'est bien de
c'est bien d'avoir des amis
je crois
je n’ sais pas
en Angleterre
c'était trop long
ça c'est bien
au lycée
je ne sais pas
beaucoup de
de temps en temps
de temps en temps
il y a
à la fin du
il y a

Appendix D. Combo

Example of FS frequency search for the sequence ‘à mon avis’, using the Combo command of CLAN.

```
combo +u +f +sà^mon^avis *.cha
Fri Feb 15 20:35:08 2013
combo (25-Feb-2010) is conducting analyses on:
  ALL speaker tiers
****************************************
From file <FrenchIri.cha>
From file <FrenchLil1.cha>
From file <FrenchLol1.cha>
From file <FrenchRos.cha>
From file <FrenchSal.cha>
From file <InterIri.cha>
From file <InterLil.cha>
From file <InterLola.cha>
----------------------------------------
*** File "InterLola.cha": line 174.
*Lol:   et (1)à (1)mon (1)avis ça cause des problèmes s' il n' y a pas deux
        parents dans une famille .
----------------------------------------
*** File "InterLola.cha": line 220.
*Lol:   mais (1)à (1)mon (1)avis euh ma mère travaille très dur .
----------------------------------------
*** File "InterLola.cha": line 239.
*Lol:   et et elles font des bons chefs (1)à (1)mon (1)avis .
From file <InterRos.cha>
----------------------------------------
*** File "InterRos.cha": line 152.
*Ros:   et (.) (1)à (1)mon (1)avis c' est une situation très différente en Angleterre
        .
From file <InterSally.cha>
From file <IrisAlcool2trans.cha>
From file <IrisFrench2.cha>
From file <IrisInter2.cha>
From file <IrisLN2.cha>
From file <IrisObes2.cha>
From file <LNIri.cha>
From file <LNLil.cha>
From file <LNLol.cha>
From file <LNRos.cha>
From file <LNSally.cha>
From file <LilAlcol2Trans.cha>
From file <LilFood2Trans.cha>
From file <LilFrench2.cha>
----------------------------------------
*** File "LilFrench2.cha": line 6.
*TXT: donc (1)à (1)mon (1)avis c' est c' est oui c' est c' est c' est pour c' est
```

utile d' apprendre (1)à parler le français pour avoir un meilleur
emploi mais parce que ça va ça va montre les compétences pour tu
peux parler avec tout le monde tu peux déménager dans un autre pays
tu tu es facile (1)à (1)à bouger comme ça .
From file <LilInter2Trans.cha>
From file <LilLN2.cha>
From file <LolALcol2Trans.cha>
From file <LolFood2Trans.cha>
From file <LolFrench2Trans.cha>
From file <LolInter2Trans.cha>

*** File "LolInter2Trans.cha": line 236.
*TXT: peut-être prof mais aux Etats Unis ou en Angleterre je je n' aime
pas le système ici c' est (1)à (1)mon (1)avis les jours sont trop longs pour
des élèves et aussi oui je préfère les relations entre les profs et
l' élève en Angleterre .
From file <LolaLN2.cha>
From file <RosFrench2Trans.cha>
From file <RosInter2Trans.cha>
From file <RosLN2Trans.cha>
From file <RosObes2Trans.cha>
From file <Rosalcool2trans.cha>
From file <SalAlcohol2.cha>
From file <SalFrench2.cha>
From file <SalInter2.cha>
From file <SalLN2.cha>
From file <SalObes2.cha>
From file <alcoholIriKar.cha>

*** File "alcoholIri.cha": line 35.
*Iri: (1)à (1)mon (1)avis ça c' est la plus important (1)à faire chose je pense chose
(1)à faire (2)à (2)mon (2)avis .
From file <alcoholLil1.cha>
From file <alcoholLol1.cha>

*** File "alcoholLol1.cha": line 75.
*Lol: et (1)à (1)mon (1)avis ce n' est pas une très bonne chose pour lui .
From file <alcoholRos1.cha>
From file <alcoholSally.cha>
From file <obesityIriAmy.cha>
From file <obesityLol1.cha>
From file <obesityLolLil.cha>

*** File "obesityLolLil.cha": line 8.
*Lil: (1)à (1)mon (1)avis c' est euh effectif d' intedire les distributeurs (.)
automatiques euh des euh des malbouffes parce que euh (1)à mon école
euh euh ces distributeurs étaient interdits et euh ça euh ça
marchait euh parce que euh les étudiants sont trop paresseux de
venir euh euh au dehors dehors de l' école .
From file <obesityRos.cha>
From file <obesitySally.cha>

*** File "obesitySally.cha": line 9.
*Sal: d' accord je pense que (.) introduire plus de cours de sport (.) (1)à (1)mon (1)avis c' est pas utile parce que (.) en ce moment quand moi j' étais au collège il y avait beaucoup (.) [//] assez beaucoup de cours de sports .

*** File "obesitySally.cha": line 30.
*Sal: et (1)à (1)mon (1)avis ça c' est stupide si on veut encourager euh les jeunes et les gens en général de manger (.) euh bien .

Strings matched 12 times

Example of FS frequency search for the sequence ‘ça va marcher’, using the Combo command of CLAN.

combo +u +f +sça^va^marcher *.cha
Fri Feb 15 20:37:46 2013
combo (25-Feb-2010) is conducting analyses on:
ALL speaker tiers
****
From file <FrenchIriKar.cha>
From file <FrenchLil1.cha>
From file <FrenchLol1.cha>
From file <FrenchRos.cha>
From file <FrenchSal.cha>
From file <InterIri.cha>
From file <InterLil.cha>
From file <InterLola.cha>
From file <InterRos.cha>
From file <InterSally.cha>
From file <IrisAlcool2trans.cha>
From file <IrisFrench2.cha>
From file <IrisInter2.cha>
From file <IrisLN2.cha>
From file <IrisObes2.cha>
From file <LNIri.cha>
From file <LNLil.cha>
From file <LNLol.cha>
From file <LNRos.cha>
From file <LNSally.cha>
From file <LilAlcol2Trans.cha>

*** File "LilAlcol2Trans.cha": line 6.
*TXT: donc pour réduire la consommation d' alcool chez les mineurs oui je pense que si il y avait une amende plus sévère aux magasins et aux bars qui vendent de l' alcool aux adolescents je pense que (1)ça (1)va (1)marcher parce que les les bars et les magasins peuvent oublier le le droit de de vendre de l' alcool je pense que (1)ça va (1)ça va vraiment marcher quand .

*** File "LilAlcol2Trans.cha": line 12.
*TXT: quand j' avais seize ans j' ai pu aller au au bar facilement mais parce

que les les lois ont a changé quand j' avais dix-sept ans et tous
les bars ont dit non il faut qu' on on vérifier et donc oui (1)ça (1)va
(1)marcher parce que si on peut pas acheter d' alcool on peut pas le
boire .

*** File "LilAlcol2Trans.cha": line 20.
*TXT: peut-être (1)ça (1)va (1)marcher un peu mais si les gens veulent boire ils
vont boire boire (1)ça va peut-être (2)ça (2)va (2)marcher un peu parce que tu
peux pas boire boire beaucoup si t' as pas l' argent .

*** File "LilAlcol2Trans.cha": line 23.
*TXT: et aussi si il y a les les peut-être les alcooliques qui peut vont
dans les écoles pour parler aux jeunes peut-être (1)ça (1)va (1)marcher pour
parce que les tout le temps les les professeurs qui disent ah faut
que tu tu boives pas beaucoup parce que c' est mal et c' est et s'
il y a vraiment quelqu'un qui est pre presque mort à cause d'
alcool je pense que (2)ça (2)va (2)marcher si tu as quelqu'un à côté face à
toi qui dit ah non tu vas tu vas mourir si tu fais (2)ça .

*** File "LilAlcol2Trans.cha": line 30.
*TXT: (1)ça (1)va (1)marcher la plupart non je pense peut-être si l' âge légal
était abass(é) abaissé jusqu'à seize ans peut-être (1)ça va je pense
que les les les jeunes qui ont dix-huit ans qui commencent à aller
au au bar ils ils boivent trop .

*** File "LilAlcol2Trans.cha": line 34.
*TXT: ils vomit ils fait n'importe-quoi peut-être si ils le font quand
ils avaient seize ans peut-être ça va passer trop ra(pidement) plus
rapidement mais ça va si c' est pas si (1)ça (1)va (1)marcher mais par
exemple quand quand j' ai commencé à boire peut-être je faisais oh
c' est cool .

From file <LilFood2Trans.cha>

*** File "LilFood2Trans.cha": line 12.
*TXT: mais si c' était possible oui je pense que (1)ça (1)va (1)marcher si les
cours étaient s' il y avait beaucoup de choix parce que si il faut
que tout le monde fasse quelque chose fasse de rugby ou de d' hockey
de hockey et si tu n' aimes pas (1)ça tu es t' as pas hate de faire
d' l' exercise et tu penses que toute l' exercice c' est pareil et
c' est pas comme (1)ça qu' on quand j' étais à l' école .

*** File "LilFood2Trans.cha": line 24.
*TXT: à la cantine je pense que ça serait fantastique parce que si la
plupart des élèves mangent à la cantine (1)ça (1)va (1)marcher par parce que
mon ex mon expérience quand j' étais au lycée (1)ça (1)ça s' est passé il
n' y avait plus de chips et donc on a on a dû manger les choses qui
étaient bons pour la santé .

*** File "LilFood2Trans.cha": line 29.
*TXT: sauf si on est allé au café au resto pour déjeuner c' est ce qu' on
peut pas faire tous les jours parce que ca coûte beaucoup plus que à
l' école donc oui je pense que (1)ça (1)va (1)marcher .

*** File "LilFood2Trans.cha": line 32.
*TXT: et oui si si tu interdis les distributeurs automatiques peut-être
(1)ça (1)va (1)marcher un peu parce que c' est pas aussi facile d' ache de
prendre les produits qui sont pas bons pour la santé mais enfin si
s' il y a un élève qui veut manger quelque chose il va le manger et
je pense c' est important de de de montre que ces ces produits sont
affreux pour la santé mais à la fin c' est le choix de chacun de de
manger ce qu' on veut .

*** File "LilFood2Trans.cha": line 39.
*TXT: et c' est difficile d' en de forcer quelque chose comme ça si les
prix des aliments trop sucrés étaient augmentés ça va énerver les
gens qui aiment qui elles qui les aiment manger un peu comme comme
moi je le j' adore les gâteaux je je les mange un peu assez
fréquemment mais pas tout le temps et si les prix étaient très chers
ça me ça va m' énerver parce que c' est pas juste si je suis en bonne
santé mais si je veux avoir quelque chose comme ça c' est c' est pas
vraiment juste je pense mais je sais pas si (1)ça (1)va (1)marcher parce que
en encore s' il y a quelqu'un qui veut manger quelque chose comme (1)ça
et il va le manger .
From file <LilFrench2.cha>
From file <LilInter2Trans.cha>
From file <LilLN2.cha>
From file <LolALcol2Trans.cha>
From file <LolFood2Trans.cha>
From file <LolFrench2Trans.cha>
From file <LolInter2Trans.cha>
From file <LolaLN2.cha>
From file <RosFrench2Trans.cha>
From file <RosInter2Trans.cha>
From file <RosLN2Trans.cha>
From file <RosObes2Trans.cha>
From file <Rosalcool2trans.cha>
From file <SalAlcohol2.cha>

*** File "SalAlcohol2.cha": line 25.
*TXT: et finalement abaisser l' âge légal de la consommation d' alcool à
seize ans peut être (1)ça (1)va (1)marcher .
From file <SalFrench2.cha>
From file <SalInter2.cha>
From file <SalLN2.cha>
From file <SalObes2.cha>
From file <alcoholIri.cha>
From file <alcoholLil1.cha>
From file <alcoholLol1.cha>
From file <alcoholRos1.cha>
From file <alcoholSally.cha>
From file <obesityIri.cha>
From file <obesityLol1.cha>
From file <obesityLolLil.cha>
From file <obesityRos.cha>

From file <obesitySally.cha>

Strings matched 14 times

Appendix E. Typology of FS used in the study

This typology is a more complete version than the one presented in chapter 7. Like the one in chapter 7, each FS is accompanied by its English translation and number of occurrences.

A. Referential FS

a. Whole sentence

i. Whole sentence expressing an opinion or a commonplace idea

1. Personal constructions

Je suis (pas) d'accord ('I am (not) of agreement' I agree / I (don't) agree) 6
Je comprends pas / *j'comprends pas* / *je ne comprends pas* 8 (I don't understand). This FS has various forms depending on the level of its grammatical and/or phonetic reduction.
Je n'ai pas le temps (I don't have time)

2. Impersonal constructions

Ça m'est égal ('this me is equal'=I don't mind)
Ça marche (pas) 16, *Ça va marcher* 14: (it works, 'it goes work'=it is going to work)
Ça craint 2 ('it fears'= it sucks)
Ça me gêne 2 (this me hinders/bothers'=I mind)
Ca me plait (beaucoup) 6 ('it me pleases'+I like it)
Ça ne change rien 2, *Ça change* 9 (it doesn't change anything, it changes)
Ça va ('it goes'=it's ok) 15
C'est une bonne idée (Rose 7 times): it's a good idea
C'est un problème 9, *ce n'est pas un problème* 3, *C'est un vrai problème* 3 (Lo) : it's a problem, it's not a problem, it's a real problem
C'est le même: it's the same. This sequence is an example of a non-target FS : in correct French, one should say *c'est la même chose* (it's the same thing)

ii. Whole sentence expressing an opinion and following the structural pattern: C'est/ c'était + adjective

C'est difficile (it's difficult) 32, *C'était difficile* (it was difficult) 4 , *C'est vrai* (it's true) 17, *C'est important* (it's important) 16, *C'est différent* (it's different)16, C'est intéressant (it's interesting) 14, *C'est super* (it's great) 13, *C'était super* 10 (Sar and Isy),

C'est normal 12, *C'est facile* (it's easy)10, C'est bon (it's good) 10, *C'est mieux* (it's better) 10, *C'est (pas) pareil* (literally it's (not) identical= it's (not) the same) 9, *C'était bien* (it was good) 8, *C'est bizarre* 10 (it's strange), *C'est sûr* 5, *C'est utile* (it's useful) 5, *C'est pas grave* (literally it's not serious= it doesn't matter) 4 (only Sally and Iris), *C'était sympa* (it was friendly/ nice) 3 (Only Sally), *C'est cool* 3

iii. Ready-made whole sentences to express a 'necessary topic'

Il fait/ faisait beau, il fait du soleil, Il faisait chaud, il faisait froid (literally 'it does' beautiful /some sun/ hot/cold=the weather is nice/ it's sunny/it's hot/it's cold)
je joue au hockey (I play hockey)*, j'aime aller au cinéma* (I like going to the cinema)

b. Time, space and other referential FS complements

i. Place complements, usually prepositional phrases

Countries: *En France* 88, *en Angleterre* 108, *aux Etats-Unis* 7, en Espagne
Cities: *A Newcastle, à Paris*
Various Places: *A l'université* (at the university= at university) 35, *au collège* (at the highschool= in highschool) 9, *à l'école* (at school) 20, *à la maison* (at the home=at home) 13, *en ville* (in/to town) 12, *à la fac* (at the faculty (abbreviated form) = at uni) 11, *à la cantine* (at the canteen) 5
Spatial organisation: *Au fond* (at the back/ in the background) 4, *à l'intérieur* (at the interior=inside) 3, *pas loin d'ici* (not far from here=nearby)
Chez + pronoun (strong form): *Chez moi* (at me=at/to mine) 23, *chez elle* (at her= at/to hers) 6, *chez eux* (at them= at/to theirs) 5

ii. Time complements (nominal phrases and prepositional phrases)

NPs
Tout le/ l' temps (all the/definite article with elision time) 31, *tous les jours* (all the days= every day) 7, *toute la journée* (all the day= all day) 3
Le + weekday e.g. *le dimanche* (the Sunday=on Sundays) 18
Chaque semaine (every week) 5
La plupart du temps ('the most of time'= most of the time) 6
L'année dernière ('the year last'= last year) 11, *l'année prochaine* ('the year next'=next year) 6
Cette année ('this year') 14, *ce soir* ('this evening'=tonight) 4

Le matin (‘the morning’=in the morning) 4, *la nuit* (‘the night’=at night) 8 *le soir* (‘the evening’= in the evening) 13

Tout de suite (straight away) 2, *N’importe quand* (whenever) 3

PPs

De temps en temps (from time to time) 24

En ce moment (‘in this moment’= at the moment) 18

Par jour / semaine / mois (per day/week/month) 21

En + month e.g. *en septembre* (in September) 18

iii. Other types of complements (means etc….)

En vacances (on holiday) 13

En avion (by plane), *en voiture* (by car) 8

En moderation (in moderation) 5

En bonne santé (in good health=healthy) 2

c. Multiword NPs referring to a single entity

Different patterns:

Noun + de + noun : *Lunettes de soleil* (glasses of sun=sunglasses), *verre de vin* (glass of wine), *salle des profs* (room of the teachers (familiar abbreviation)= staff room, *boîte de nuit* (box of night=nightclub), *bruit de fond* (noise of background= background noise)

Noun + prep + noun : *transports en commun, année à l’étranger, choses à faire*

***Un/ une bon(ne)* + noun** : *un bon début* (a good start), *une bonne expérience* (a good experience), *un bon choix* (a good choice)

Noun+adjective : année scolaire (school year), vie étudiante (student life), *boisson alcoolisée* (alcoholic drink), comédie romantique (romantic comedy)

Definite article + noun : e.g. school subjects: le droit (definite article+ law), *le français, la linguistique* or common groups of people e.g. *les gens* 69, *tout le monde* (all the world=everybody) 45, *les jeunes* (the young=young people) 20

d. Multiword VPs

Parler anglais (speak English) 27, *parler français* (speak French) 18

Faire du shopping (to do some shopping) 7, *faire du ski* (‘to do some ski=to ski) 3, *faire ses études* 3 (to do one’s studies=to study (only Sally), *faire un / des cours* 8, *faire un master* 3, *faire attention* 3, *faire pareil* 2, *faire des erreurs* 4, *faire du sport* 5

Avoir l’air (to have the air= to look/ seem) 4, *avoir honte* (to have shame=to be ashamed) 4, *avoir des problèmes* (to have problems) 11

Etre à l'aise (to be at the ease= to be comfortable) 3, *se sentir à l'aise* ('reflexive pronoun+ feel at the ease'=to feel comfortable 2, *être bouche bée (*'to be mouth open'=to be gob-smacked 1, *être en colère* ('to be in anger'=to be angry) 6
Prendre un verre (literally 'take a glass= have a drink) 4
Prendre le train 2 / *l'avion* 1 (take the train / the plane)
aller 5 / être 3 / partir 5 en vacances (to go/be/go on holiday)
Aller à l'étranger ('to go to the stranger'=to go abroad 7, *Aller au cinéma* (to go to the cinema) 9
(super) bien s'entendre ('(very) well reflexive pronoun hear'=to get on (very well)) 8
Dire bonjour (say hello) 2

B. Meta-discursive FS

a. Fillers

Fillers are the most important subcategory of this category.
The most two common fillers are:
je sais pas (often reduced as *chais pas)* 90
je pense 79
Other fillers include:
Je crois (I believe= I think) 46
C'est ça (it is that=yes/exactly) 38
Et tout 36 ('and everything', difficult to translate, filler used at the end of utterances)
Et tout ça 12 ('and all this', difficult to translate, filler used at the end of utterances)
Quelque chose comme ça (something like that) 21
Des / les choses comme ça (things like that) 34
Ça dépend (it depends) 14 (used as a filler/ useful expression rather than sentence builder)
Je dirais (I would say) 3
Quoi d'autre (what of other=what else) 4 (only Sally)

b. Prepositional phrases introducing one's opinion or structuring one's discourse

- introduce one's opinion: *Pour moi* (for me) 72, *A mon avis* (in my opinion) 12
-structure one's discourse e.g. *En fait* (in fact) 83, *Par exemple* (for example) 57, *En général* (in general) 8, *Par contre* (literally 'by against= on the opposite) 4 (only Sally).
Etre honnête 3 (Rose): this is an example of non-target like FS resulting from a literal translation of 'to be honest'

c. Temporal expressions to structure narration

(Et) après ça 35 ((and) after that)

Et puis 38 (and then)

Au début (at the beginning) 10, *à la fin* 22

C. Sentence-builders

a. Express one's opinion

i. FS followed by a clause

Je pense que (I think that) + clause 186

Je crois que (I believe that) + clause 54

Je sais que (I know that) 30

Je trouve que (I find that) 14

Je suppose que (I suppose that) 3 (only Sally)

Il me semble que ('it me seems that =it seems to me that) 3

Il semble que (it seems that) 3 (only Rose)

Avoir l'impression que (to have the impression that) 2

Il faut que 9, il fallait que (3) (impersonal contruction to express obligation)

C'est vrai que (it's true that) 7

Je dirais que (I would say that) 3 (only Sally)

Le problème c'est que (the problem it is that= the problem is that) 5

ii. FS followed by an infinitive verb

Il faut 47

C'est important de 7

iii. Expressions of likes and dislikes + NP/Infinitive verb

J'aime bien (I like well= I like) 38

J'aime beaucoup (I really like) 38

J'aime (I like) 122

J'adore (I love) 17

J'ai beaucoup aimé 8

b. Existential expressions : il y a + NP

Il y a 262 (there is)

Il y avait 31 (there was)

c. Presentative constructions

Il y a + NP + qui (there is + NP + subject relative pronoun) and more specifically Il y a +beaucoup de+ NP + qui (there are a lot of + NP+ subject relative pronoun)
Il y a + NP + que (there is + NP+ object relative pronoun)
C'est + NP + que (it is + NP + object relative pronoun) (at least 5)
NP c'est que (NP it is that): (only avec le problème ?) (I and S)
J'ai + NP+ qui (I have + NP + subject relative pronoun)
C'est quelque chose qui

d. Necessary topics

je m'appelle (I me call= my name is) + name (10)

j'habite (I live) + place (17)

e. le fait que (the fact that) + clause

f. après avoir + past participle

g. Sentence builders for description

On voit (one sees= we can see) (17)

On peut voir (one can see=we can see) (8)

h. Frequent subject+ verb units

je suis allée (I am gone= I have been/ I went) / *on est allé* (one is gone= we have gone/ we went) + place

je peux (I can) / *on peut* / *on pourrait* + infinitive verb

je voudrais (I would like) + infinitive verb

je vais (I go=I'm going) / *on va* (one goes=we're going) / *ça va* (it goes=it's going) + infinitive verb to express the future

ça prend (it takes) + time

je vais (I go) / *on va* (one goes=we go) + place

j'ai (I have) / *on a* (one has=we have)

i. Multiword prepositional phrases + NP or infinitive

au bord de +NP 7 (at the edge of)

à cause de/du/des + noun 24 (because of)

au niveau de +NP 10 : at the level of

près de +NP 12 : ('near of'=near)

avant de +infinitive 13 ('before of'=before)

au lieu de +infinitive 4 ('at place of'=instead of)

en train de +infinitive 18 ('in train of'=marker of an action in progress')

j. Multiword quantifiers + noun

La plupart de / du / des 19 (most +noun), *la majorité des* (the majority of) 7, *beaucoup de* (a lot of) 117, *plus de* (‘more of’=more) 62, *moins de* (‘less of’=less) 17, *plein de* (plenty of) 30

k. Verb or phrasal verb + preposition+ infinitive verb

Venir de 6 (to come from)

Venir de 3 (different meaning : to have just)

Essayer de 17 (‘to try of’=to try)

Commencer à 9 (to start to)

Décider de 5 (to decide to)

Avoir envie de 23 (‘to have envy of’=to feel like) Iris, pas tros envie de 5

Avoir besoin de 20 (‘to have need of’= to need)

Etre obligé(e) de 12 (‘to be obliged to’=to have to) (only Rose) : on est obligé de 7, j’étais (2), je suis/ ne suis pas 2, elle est obligée 1

Avoir peur de 5 (‘to have fear of’= to be scared of)

Avoir du mal à 2 (‘to have some difficulty to’=to find it difficult to)

Se rendre compte de/que 5(‘reflexive pronoun+give account that’=to realise that)

Appendix F. Case study of *je pense que*

There is a total of 215 occurrences of *je pense que* (I think that) in the corpus. Despite the very high frequency of this FS, it is never used by Lola. The 215 occurrences are spread evenly amongst the 4 other learners with about 50 occurrences for each of them. *je pense que* is an interesting sequence as it encapsulates many of the phenomena observed in this study:

- Differences in the learners' formulalects because *je pense que* is frequent for all the learners but not used at all by Lola who uses *je crois que* (I believe that) instead.
- Presence of phonetic reductions of highly frequent FS as *je pense que* gets reduced as *j' pense que as* je undergoes schwa deletion.
- More phonetic reductions at time 2: for example, most of the occurrences of *je pense que* used by Iris at time 2 are reduced (17 reduced forms out of 22 occurences of *je pense que* at time 2).
- Influence of the type of task on the type of FS used: *je pense que* is not used in the story-retelling task but is extremely frequent in all the discussion tasks in which the learners are asked to offer opinions.
- Over-reliance of L2 learners on certain FS: *je pense que* is a good example of overused FS.

  je pense que is a good illustration of the difference between time 1 and time 2 in the distribution of certain FS. Lily and Rose use it much more frequently at time 2 (37 and 36 occurrences respectively) than at time 1. The reversed pattern, however, can be observed for Iris (39 occurrences at time 1 and only 22 at time 2) and even more strikingly for Sally, whose 47 occurrences of *je pense que* all appear at time 1. The fact that some of the learners use it much less at time 2 epitomises the development of their lexical diversity and the fact that they rely less on a limited set of sequences.

Appendix G. Transcription conventions for CHAT format

Summary of headers

@Begin	Indicates the start of a new transcript
@Languages	Language(s) used in the transcript
@Participants	Participants taking part in the task including investigator
@ID	Details about the participants: mother tongue, place where the task is taking place, name, age, gender, University year, Status
@Situation	Type of task: e.g. French Discussion Task
@End	Indicates the end of the transcript

Summary of abbreviations used in the headers

fr	French
en	English
NCL	Newcastle
Y	Year

Summary of symbols used in the transcripts

*****	Used at the start of a line to introduce a new speaker
[/]	Retracing without correction (repetition)
[//]	Retracing with correction
<text>	< and > indicate the start and the end of the part of the utterance which is retraced in the cases of retracing with correction
(.)	Pause between words
+/.	Interrupted or unfinished utterance
[/-]	False start without retracing
text **:**	Lengthened word

Bibliography

Abutalebi, J., Cappa, S.F. and Perani, D. (2001) 'The bilingual brain as a function of functional neuroimaging', *Bilingualism: Language and Cognition*, 4, pp. 179-190.

Altenberg, B. (1998) 'On the phraseology of spoken English: the evidence of recurrent word combinations', in Cowie, A.P. (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press.

Ambridge, B. and Lieven, E. (2011) *Child Language Acquisition*. Cambridge: Cambridge University Press.

Anderson, J.R. (1983) *The Architecture of Cognition*. Canbridge, MA: Harvard University Press.

Anderson, J.R. (1995) *Learning and Memory: an Integrated Approach*. New York: Wiley.

Bannard, C. and Lieven, E. (2009) 'Repetition and reuse in child language learning', in Corrigan, R., Moravcsik, E.A., Ouali, H. and Wheatley, K.M. (eds.) *Formulaic Language*. Philadelphia: John Benjamins, pp. 299-321.

Bannard, C. and Lieven, E. (2012) 'Formulaic Language in L1 Acquisition', *Annual Review of Applied Linguistics*, 32, pp. 3-16.

Bannard, C. and Matthews, D. (2008) 'Stored word sequences in language learning: The effect of familiarity on children's repetitions of four-word combinations', *Psychological Science*, 19, pp. 241-248.

Bates, E., Bretherton, I. and Snyder, L. (1988) *From first Words to Grammar: Individual Differences and Dissociable Mechanisms*. Cambridge: Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D. and Schoenemann, T. (2009) 'Language is a complex adaptive system: position paper', *Language Learning*, 59, pp. 1-26.

Biskup, D. (1992) 'L1 influence on learners' renderings of English collocations: a Polish / German empirical study', in Arnaud, P.J.L. and Béjoint, H. (eds.) *Vocabulary and Applied Linguistics*. Basingstoke: MacMillan, pp. 85-93.

Bolander, M. (1989) 'Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish', in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the Lifespan*. Cambridge: Cambridge University Press, pp. 73-86.

Butler, C.S. (1997) 'Enriching the functional grammar lexicon', in Butler, C.S., Hengeveld, K. and Garcia, J.S. (eds.) *The Structure of the Lexicon in Functional Grammar*. Amsterdam: IFOTT, pp. 60-77.

Bybee, J. (2002) 'Phonological evidence for exemplar storage of multiword sequences', *Studies in Second Language Acquisition*, 24(02), pp. 215-221.

Bybee, J. (2006) 'From usage to grammar: the mind's response to repetition', *Language & Communication*, 82, pp. 711-733.

Bybee, J. (2008) 'Usage-based grammar and second language acquisition', in Robinson, P. and Ellis, N.C. (eds.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, pp. 216-236.

Bybee, J. (2010) *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. and McClelland, J.L. (2005) 'Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition', *The Linguistic Review*, 22(2-4), pp. 381-410.

Bybee, J. and Scheibman, J. (1999) 'The effect of usage on degrees of constituency: the reduction of *don't* in English', *Linguistics*, 37, pp. 575-96.

Cacciari, C. and Tabossi, P. (1988) 'The comprehension of idioms', *Journal of Memory and Language*, 22, pp. 668-683.

Cameron-Faulkner, T., Lieven, E. and Tomasello, M. (2003) 'A construction-based analysis of child-directed speech', *Cognitive Science*, 27, pp. 843-873.

Chambers, F. (1997) 'What do we mean by fluency?', *System*, 25, pp. 535-544.

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clark, R. (1974) 'Performing without competence', *Journal of Child Language*, 1, pp. 1-10.

Conklin, K. and Schmitt, N. (2008) 'Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers', *Applied Linguistics*, 29(1), pp. 72-89.

Conklin, K. and Schmitt, N. (2012) 'The processing of formulaic language', *Annual Review of Applied Linguistics*, 32, pp. 45-61.

Corrigan, R., Moravcsik, E.A., Ouali, H. and Wheatley, K.M. (eds.) (2009) *Formulaic Language*. Philadelphia: John Benjamins.

Coulmas, F. (1994) 'Formulaic language', in Asher, R.E. (ed.) *Encyclopedia of Language and Linguistics*. Oxford: Pergamon, pp. 1292-1293.

Cutting, J.C. and Bock, K. (1997) 'That's the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blends', *Memory & Cognition*, 25(1), pp. 57-71.

Dabrowska, E. (1997) 'The LAD goes to school: A cautionary tale for nativists', *Linguistics*, 35, pp. 735-766.

Dahlmann, I. (2009) *Towards a multi-word unit inventory of spoken discourse*. University of Nottingham.

Daller, H., Milton, J. and Treffers-Daller, J. (eds.) (2007) *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.

De Bot, K. (1992) 'A bilingual production model: Levelt's 'speaking' model adapted', *Applied Linguistics*, 13, pp. 1-24.

Dechert, H. (1983) 'How a story is done in a second language', in C., F. and G., K. (eds.) *Strategies in Interlanguage Communication*. London: Longman.

Dell, G.S. (1986) 'A spreading activation theory of retrieval in sentence production', *Psychological Review*, 93, pp. 283-321.

Dewaele, J.-M. (2004) 'The acquisition of sociolinguistic competence in French as a foreign language: an overview', *Journal of French Language Studies*, 14, pp. 301-319.

Ejzenberg, R. (2000) 'The juggling act of oral fluency : a psycho-sociolinguistic metaphor ', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: University of Michigan Press.

Ellis, N.C. (2002) 'Frequency effects in language processing', *Studies in Second Language Acquisition*, 24(02), pp. 143-188.

Ellis, N.C. (2003) 'Constructions, chunking and connectionism', in Doughty, C.J. and Long, M. (eds.) *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell.

Erman, B. (2007) 'Cognitive processes as evidence of the idiom principle', *International Journal of Corpus Linguistics*, 12(1), pp. 25-55.

Erman, B. (2009) 'Formulaic language from a learner perspective: What the learner needs to know', in Corrigan, R., Moravcsik, E.A., Ouali, H. and Wheatley, K.M. (eds.) *Formulaic Language*. Philadelphia: John Benjamins, pp. 323-346.

Erman, B. and Warren, B. (2000) 'The idiom principle and the open choice principle', *Text*, 20(1), pp. 29-62.

Farghal, M. and Obiedat, H. (1995) 'Collocations: a neglected variable in EFL', *International Review of Applied Linguistics in Language Teaching*, 33(4), pp. 315-331.

Forsberg, F. (2009) 'Formulaic sequences: A distinctive feature at the advanced / very advanced levels of second language acquisition', in Labeau, E. and Myles, F. (eds.) *The advanced learner variety. The case of French*. Bern: Peter Lang.

Foster, P. (2001) 'Rules & routines: A consideration of their role in the task-based language production of native and non-native speakers', in Bygate, M., Skehan, P. and Swain, M. (eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching, Testing*. London, NY: Longman, pp. 75-94.

Foster, P. (2009) 'Lexical diversity and native-like selection: the bonus of studying abroad', in Richards, B., Daller, M.H., Malvern, D.D., Meara, P., Milton, J. and Treffers-Daller, J. (eds.) *Vocabulary Studies in First and Second Language Acquisition*. Basingstoke: Palgrave Macmillan.

Frank, V. (1997) 'Potential negative effects of homestay', *Middle Atlantic Conference of the American Association for the Advancement of Slavic Studies*. Albany, NY.

Freed, B., Dewey, D.P., Segalowitz, N. and Halter, R. (2004a) 'The language contact profile', *Studies in Second Language Acquisition*, 26, pp. 349-356.

Freed, B., Segalowitz, N. and Dewey, D.P. (2004b) 'Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs', *Studies in Second Language Acquisition*, 26, pp. 275-301.

Freed, B.F. (ed.) (1995) *Second Language Acquisition in a Study Abroad Context*. Amsterdam: John Benjamins.

Freed, B.F. (2000) 'Is fluency, like beauty, in the eyes (and ears) of the beholder', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: University of Michigan Press, pp. 243-265.

Goldberg, A.E. (2003) 'Constructions: a new theoretical approach to language', *Trends in Cognitive Sciences*, 7, pp. 219-224.

Goldman Eisler, F. (1964) 'Hesitation, information, and levels of speech production', in *Disorders of language* London: Churchill, pp. 96-111.

Goldman Eisler, F. (1968) *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Griffiths, R. (1991) 'Pausological research in an L2 context: A rationale and review of selected studies.', *Applied Linguistics* 12 (4), pp. 345-364.

Grosjean, F. (1980) 'Linguistic structures and performance strucutures: studies in pause distribution', in Dechert, H. and Raupach, M. (eds.) *Temporal Variables in Speech*. The Hague: Mouton de Gruyter, pp. 91-106.

Guiraud, P. (1954) *Les Caractéristiques Statistiques du Vocabulaire*. Paris: Presses Universitaires de France.

Hay, J. (2001) 'Lexical frequency in morphology. Is everything relative? ', *Linguistics*, 39, pp. 1041-1070.

Hay, J. (2002) 'From speech preception to morphology: affix ordering revisited', *Language & Communication*, 78, pp. 527-555.

Hickey, T. (1993) 'Identifying formulas in first language acquisition', *Journal of Child Language*, 20, pp. 27-41.

Hoey, M. (2005) *Lexical Priming: a New Theory of Words and Language*. London: Routledge.

Hopper, P. (1998) 'Emergent grammar', in Tomasello, M. (ed.) *The New Psychology of Language*. Mahwah, NJ: Lawrence Erlbaum, pp. 155-175.

Itoh, H. and Hatch, E. (1978) 'Second language acquisition: a case study', in Hatch, E. (ed.) *Second Language Acquisition: A Book of Readings*. Rowley, MA: Newbury House, pp. 76-88.

Jackendoff, R. (1983) *Semantic and Cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1997) *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jespersen, O. (1924/1976) 'Living grammar', in *The Philosophy of Grammar*. London.

Jiang, N. and Nekrasova, T.M. (2007) 'The processing of formulaic sequences by second language speakers', *Modern Language Journal*, 91(3), pp. 433-445.

Karniol, R. (1990) 'Second language acquisition via immersion in daycare', *Journal of Child Language*, 17, pp. 147-170.

Kormos, J. (2006) *Speech Production and Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Krashen, S. and Scarcella, R. (1978) 'On routines and patterns in language acquisition and performance', *Language Learning*, 28(2), pp. 283-300.

La Heij, W. (2005) 'Selection processes in monolingual and bilingual ', in Kroll, J. and de Groot, A.M.B. (eds.) *Handbook of Bilingualism. Psycholinguistic Approaches*. New York: Oxford University Press.

Langacker, R.W. (1987) *Foundations of Cognitive Grammar*. Stanford, CA: Stanford University Press.

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

Lennon, P. (2000) 'The lexical element in spoken second language fluency', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: The University of Michigan Press, pp. 43-60.

Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, W.J.M. (1999) 'Producing spoken language: a blueprint of the speaker', in Brown, C. and Hagoort, P. (eds.) *The Neurocognition of Language*. Oxford: Oxford University Press, pp. 83-122.

Lieven, E., Salomo, D. and Tomasello, M. (2009) 'Two-year old children's production of multiword utterances: A usage-based analysis', *Cognitive Linguistics*, 20, pp. 481-508.

Lieven, E.V.M., Pine, J.M. and Barnes, H.D. (1992) 'Individual differences in early vocabulary learning: redefining the referential-expressive distinction', *Journal of Child Language*, 19(2), pp. 287-310.

Lin, P. (2010) 'The phonology of formulaic sequences: a review', in Wood, D. (ed.) *Perspectives on Formulaic Language*. London: Continuum, pp. 174-193.

Lin, P. and Adolphs, S. (2009) 'Sound evidence: phraseological units in spoken corpora', in *Researching Collocations in Another Language: Multiple Interpretations*. Basingstoke: Palgrave Macmillan, pp. 34-48.

Logan, G.D. (1988) 'Toward an instance theory of automatisation', *Psychological Review*, 95, pp. 492-527.

MacKay, D.G. (1982) 'The problem of flexibility, fluency and speed accuracy trade-off in skilled behaviour', *Psychological Review*, 89, pp. 483-506.

MacWhinney, B. (2008) 'A unified model', in Robinson, P. and Ellis, N.C. (eds.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, pp. 341-371.

Malvern, D., Richards, B., Chipere, N. and Dur´an, P. (2004) *Lexical Diversity and Language Development: Quantification and Assessment*

Basingstoke: Palgrave Macmillan.

Marriott, H. (1995) 'Acquisition of politeness patterns by exchange students in Japan', in Freed, B.F. (ed.) *Second Language Acquisition in a Study Abroad Context*. Amsterdam: John Benjamins, pp. 197-227.

McCarthy, P. and Jarvis, S. (2007) 'Vocd: a theoretical and empirical evaluation', *Language Testing*, 24(4), pp. 459-488.

McKee, G., Malvern, D. and Richards, B. (2000) 'Measuring vocabulary diversity using dedicated software', *Literary and Linguistic Computing*, 15(3), pp. 323–337.

Miller, G.A. (1956) 'The nagical number seven, plus or minor two: some limits on our capacity for processing information', *Psychological Review*, 63, pp. 81-97.

Miller, J.E. and Weinert, R. (1998) *Spontaneous Spoken Language: Syntax and Discourse*. Clarendon: Clarendon Press.

Mitchell, R. and Martin, C. (1997) 'Rote learning, creativity and 'understanding in classroom foreign language teaching'', *Language Teaching Research*, 1, pp. 1-27.

Möhle, D. and Raupach, M. (1987) 'The representation problem in interlanguage theory', in Lörscher, W. and Schülze, R. (eds.) *Perpectives on Language in Performance*. Tübingen: Gunter Narr, pp. 1158-1173.

Myles, F. (2004) 'From data to theory: The over-representation of linguistic knowledge in SLA', *Transactions of the Philological Society*, 102(2), pp. 139-168

Myles, F., Hooper, J. and Mitchell, R. (1998) 'Rote or rule? Exploring the role of formulaic language in classroom foreign language learning', *Language Learning*, 48(3), pp. 323-362.

Myles, F., Mitchell, R. and Hooper, J. (1999) 'Interrogative chunks in French L2': A basis for creative construction?', *Studies in Second Language Acquisition*, 21, pp. 49-80.

Nattinger, J.R. and DeCarrico, J.S. (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nelson, K. (1973) 'Structure and strategy in learning to talk', *Monographs of the Society for Research in Child Development*, 38, pp. 1-2.

Nelson, K. (1975) 'The nominal shift in semantic-syntactic development', *Cognitive Psychology*, 7, pp. 461-479.

Newell, A. (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Oppenheim, N. (2000) 'The importance of recurrent sequences for nonnative speaker fluency and cognition', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: University of Michigan Press, pp. 220-240.

Paradis, M. (2000) 'Cerebral representation of bilingual concepts', *Bilingualism: Language and Cognition*, 3, pp. 22-24.

Pavlenko, A. (1999) 'New approaches to concepts in bilingual memory', *Bilingualism: Language and Cognition*, 2, pp. 209-230.

Pawley, A. and Syder, F.H. (1983) 'Two puzzles for linguistic theory: Nativelike selection and nativelike fluency', in Richards, J.C. and Schmidt, R.W. (eds.) *Language and Communication*. New York: Longman, pp. 191-226.

Pawley, A. and Syder, F.H. (2000) 'The one-clause-at-a-time hypothesis ', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: University of Michigan Press.

Perkins, M.R. (1999) 'Productivity and formulaicity in language development', in Garman, M., Letts, C., Richards, B., Schelletter, C. and Edwards, S. (eds.) *Issues in Normal and Disordered Child Language: From Phonology to Narrative*. Reading: University of Reading, pp. 51-67.

Peters, A.M. (1976) 'Language learning strategies: Does the whole equal the sum of the parts? ', *Language*, 53, pp. 560-573.

Peters, A.M. (1983) *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Peterson, R.R., Dell, G.S., Burgess, C. and Eberhard, K.M. (2001) 'Dissociation between syntactic and semantic processing during idiom comprehension', *Journal of Experimental Psychology / Learning, Memory & Cognition*, 27(5), p. 1223.

Pierrehumbert, J. (2001) 'Exemplar dynamics: word frequency, lenition and contrast', in Bybee, J. and Hopper, P. (eds.) *Freqeuncy and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, pp. 137-157.

Pine, J.M. and Lieven, E. (1993) 'Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech', *Journal of Child Language*, 20, pp. 551-572.

Pinker, S. (1994) *The Language Instinct: The New Science of Language and Mind*. London: Allen Lane.

Plunkett, K. (1993) 'Lexical segmentation and vocabulary growth in early language acquisition', *Journal of Child Language*, 20, pp. 325-343.

Poulisse, N. and Bongaerts, T. (1994) 'First language use in second language production', *Applied Linguistics*, 15, pp. 15-48.

Raupach, M. (1984) 'Formulae in second language speech production', in Dechert, H., Möhle, D. and Raupach, M. (eds.) *Second Language Productions*. Tübingen: Gunter Narr, pp. 114-137.

Regan, V., Howard, M. and Lemée, I. (2009) *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Bristol: Multilingual Matters.

Rehbein, J. (1987) 'On fluency in second language speech', in Dechert, H. and Raupach, M. (eds.) *Psycholinguistic Models of Production*. Norwood, NJ: Ablex, pp. 97-105.

Riggenbach, H. (1991) 'Toward an understanding of fluency: a microanalysis of nonnative speaker conversations', *Discourse Processes*, 14, pp. 423-441.

Riggenbach, H. (ed.) (2000) *Perspectives on Fluency*. Michigan: The University of Michigan Press.

Schmidt, R. (1992) 'Psychological mechanisms underlying second language fluency', *Studies in Second Language Acquisition*, 14, pp. 357-385.

Schmidt, R. (1995) *Attention and Awareness in Foreign Language Learning*. Honolulu: University of Hawaii Press.

Schmitt, N. (ed.) (2004) *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins.

Schmitt, N., Grandage, S. and Adolphs, S. (2004) 'Are corpus-relevant clusters psycholinguistically valid?', in Schmitt, N. (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins.

Schmitt, N. and Underwood, G. (2004) 'Exploring the processing of formulaic sequences through a self-paced reading task', in Schmitt, N. (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins, pp. 173-189.

Schumann, J.H. (1978) 'Second language acquisition: the pidginization hypothesis', in Hatch, E. (ed.) *Second Language Acquisition: A Book of Rreadings*. Rowley, MA: Newbury House, pp. 256-271.

Segalowitz, N. (2000) 'Automaticity and attentional skill in fluent performance ', in Riggenbach, H. (ed.) *Perspectives on Fluency*. Ann Arbor: University of Michigan Press.

Segalowitz, N. (2003) 'Automaticity and second languages', in Doughty, C.J. and Long, M. (eds.) *Handbook of Second Language Acquisition*. Malden, MA: Blackwell, pp. 383-408.

Segalowitz, N. (2010) *Cognitive Bases of Second Language Fluency*. New York: Routledge.

Segalowitz, N. and Freed, B.F. (2004) 'Context, contact, and cognition in oral fluency acquisition - Learning Spanish in at home and study abroad contexts', *Studies in Second Language Acquisition*, 26(2), pp. 173-199.

Segalowitz, N. and segalowitz, S. (1993) 'Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition', *Applied Psycholinguistics*, 14, pp. 369-385.

Segalowitz, N., Segalowitz, S. and Wood, A. (1998) 'Assessing the development of automaticity in second language word recognition', *Applied Psycholinguistics*, 19, pp. 53-67.

Shapira, R.G. (1978) 'The non-learning of English: case study of an adult', in Hatch, E. (ed.) *Second Language Acquisition: A Book of Readings*. Rowley, MA: Newbury House, pp. 246-255.

Siegal, M. (1995) 'Individual differenes and study abroad: women learning Japanese in Japan', in Freed, B.F. (ed.) *Second Language Acquisition in a Study Abroad Context*. Amsterdam: John Benjamins, pp. 225-244.

Sinclair, J.M. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A., Conklin, K. and Schmitt, N. (2011a) 'Adding more fuel to the fire: an eye-tracking study of idiom processing by native and non-native speakers', *Second Language Research*, 27, pp. 1-22.

Siyanova-Chanturia, A., Conklin, K. and Van Heuven, W.J.B. (2011b) 'Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), pp. 776-784.

Siyanova, A. and Schmitt, N. (2008) 'L2 learner production and processing of collocation: A multi-study perspective', *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 64(3), pp. 429-458.

Skehan, P. (2009) 'Lexical performance by native and non-native speakers on language-learning tasks', in Richards, B., Daller, H., Malvern, D., Meara, P., Milton, J. and Treffers-Daller, J. (eds.) *Vocabulary Studies in First and Second Language Acquisition*. Basingstoke: Palgrave Macmillan.

Swinney, D. and Cutler, A. (1979) 'The access and processing of idiomatic expressions', *Journal of Verbal Learning and Verbal Behaviour*, 18, pp. 523-534.

Tabossi, P., Fanari, R. and Wolf, K. (2009) 'Why are idioms recognized fast?', *Memory and Cognition*, 37, pp. 529-540.

Temple, L. (2000) 'Second language learner speech production', *Studia Linguistica*, 54(2), pp. 288-297.

Temple, L. (2005) 'Investigating cognitive and processing mechanisms in instructed SLA instructed learners' fluency and implicit/explicit language processes ', in Housen, A. and Pierrard, M. (eds.) *Investigations in Instructed Second Language Acquisition*. New York: Mouton de Gruyter.

Towell, R. (2002) 'Relative degrees of fluency: A comparative case study of advanced learners of French', *IRAL*, 40(2), pp. 117-150.

Towell, R., Hawkins, R. and Bazergui, N. (1996) 'The development of fluency in advanced learners of French', *Applied Linguistics*, 17, pp. 84-119.

Tremblay, A. (2011) 'Proficiency assessment standards in second language acquisition research', *Studies in Second Language Acquisition*, 33, pp. 339-372.

Tremblay, A. and Baayen, R.H. (2010) 'Holistic processing of regular four-word sequences: a behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall', in Wood, D. (ed.) *Perspectives on Formulaic Language: Acquisition and Communication*. London: Continuum, pp. 151-173.

Underwood, G., Schmitt, N. and Galpin, A. (2004) 'An eye-movement study into the processing of formulaic sequences', in Schmitt, N. (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins, pp. 153-172.

Weinert, R. (1995) 'The Role of formulaic language in second language acquisition: A review', *Applied Linguistics*, 16(2), pp. 180-205.

Weinert, R. (2010) 'Formulaicity and usage-based language: linguistic, psycholinguistic and acquisitional manifestations', in Wood, D. (ed.) *Perspectives on Formulaic Language*. London: Continuum, pp. 1-20.

Weinreich, U., Labov, W. and Herzog, M.I. (1968) 'Empirical foundations for a theory of language change', in Lehmann, W.P. and Y., M. (eds.) *Directions for Historical Linguistics*. Austin: University of Texas Press, pp. 95-195.

Wilkinson, S. (1998) 'Study abroad from the participants' perspective: A challenge to common beliefs', *Foreign Language Annals*, 31, pp. 23-39.

Wong-Fillmore, L. (1976) *The second time around: cognitive and social strategies in second language acquisition.* Stanford University.

Wong-Fillmore, L. (1979) 'Individual differences in second language acquisition', in Fillmore, C.J., Kempler, D. and S-Y.W., W. (eds.) *Individual Differences in*

Language Ability and Language Behaviour. New York: Academic Press, pp. 203-228.

Wood, D. (2009) 'Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study', *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée*, 12(1), pp. 39-57.

Wood, D. (2010) *Formulaic Language and Second Language Speech Fluency : Background, Evidence and Classroom Applications*. London: Continuum.

Wray, A. (2000) 'Formulaic sequences in second language teaching: Principle and practice', *Applied Linguistics*, 21(4), pp. 463-489.

Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008) *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. (2009) 'Idntifying formulaic language', in Corrigan, R., Moravcsik, E.A., Ouali, H. and Wheatley, K.M. (eds.) *Formulaic Language*. Amsterdam: John Benjamins, pp. 27-51.

Wray, A. (2012) 'What do we (think we) know about formulaic language? An evaluation of the current state of play', *Annual Review of Applied Linguistics*, 32, pp. 231-254.

Wray, A. and Namba, K. (2003) 'Formulaic language in a Japanese-English bilingual child: A practical approach to data analysis', *Japan Journal for Multilingualism and Multiculturalism*, 9, pp. 24-51.

Wray, A. and Perkins, M.R. (2000) 'The functions of formulaic language: An integrated model', *Language and Communication*, 20(1), pp. 1-28.

Wulff, S. ((in press)) 'Words and idioms', in Trousdale, G. and Hoffmann, T. (eds.) *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Yorio, C.A. (1989) 'Idiomaticity as an indicator of second language proficiency', in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the Lifespan*. Cambridge: Cambridge University Press, pp. 55-72.

Zellner, B. (1994) 'Pauses and the Temporal Structure of Speech', in Keller, E. (ed.) *Fundamentals of Speech Synthesis and Speech Recognition* Chichester: John Wiley.