Developing formulaicity: memorisation and production of formulaic expressions in L2 speakers of English

Stephen F. Cutler

Centre for Language and Communication Research
School of English, Communication and Philosophy
Cardiff University

Thesis submitted for the degree of PhD
Cardiff University
2020

Abstract

Research suggests that the use of formulaic language is a key feature of fluent and 'natural sounding' speech, but it is not used extensively by second language (L2) speakers, despite the potential benefits. To provide further insight into the acquisition and use of such language in L2 speakers, this thesis explores the operational and psycholinguistic processes by which targeted expressions become formulaic for the speaker. Building on ideas by Myles and Cordier (2017), formulaic expressions are here defined from a psycholinguistic perspective as multi-word strings that are either stored holistically in the mental lexicon or processed automatically as a single unit. This approach recognises that a target sequence, even if considered formulaic 'in the language', may not be immediately formulaic for the individual learner.

A series of experiments is undertaken, focussing on the speech of intermediate/advanced Japanese speakers of English. The first set of studies explores the use of formulaic expressions in speech, using samples collected from semi-structured interviews and a longitudinal case study. Subsequent studies concern the memorisation and reproduction of targeted multi-word expressions. The effect of various factors is investigated, including the memorisation process (whether it encourages a unitary or reconstructive approach to recall), the sentential context in which expressions are delivered and the degree and type of repetition and retrieval. Formulaicity in the target sequences is identified principally on the basis of fluency, but the analyses also incorporated measures of accuracy and ease of recall. Various methods (natural speech, cued responses and psycholinguistic methods) are used to extract and analyse output. The extent to which fluency is a sufficient indicator of processing advantage in formulaic expressions is also explored. Building on this, a further psycholinguistic method for assessing formulaicity is introduced for the final study and compared with fluency to give a fuller picture of the stages of acquisition.

The overall findings suggest that targeted expressions may become internally formulaic in different ways depending on the learner, specific characteristics of the expressions and the method of memorisation. Two particular routes are identified: the 'fusion' over time of component words and sub-sequences, and 'holistic acquisition', where an expression is internalised as a single unit from the start. In the former case, expressions appear to go through stages of initial reconstruction, via fluent (but not necessarily formulaic) production, to full holistic processing. With holistic acquisition, the development stages are more associated with strengthening semantic and contextual links to the expression and improving recall. Drawing on existing models of speech production, particularly the ‘superlemma’ model of Sprenger, Levelt, and Kempen (2006), a tentative model for how target sequences may be stored and processed at different stages of acquisition is proposed. Key implications of this for the acquisition of formulaic expressions in L2 speakers are discussed.

Acknowledgements

I would like to thank my supervisor Alison Wray for the great help and support she has offered me throughout the development of this thesis. Her insightful feedback, thoughtful discussion and constant encouragement helped me shape my ideas and maintain my motivation. Thanks also to Tess Fitzpatrick for her wise words and unfailing positivity throughout the study, and to all staff and colleagues on the Lexical Studies programme at Cardiff and Swansea universities for their interest and helpful suggestions.

In conducting the empirical studies, I would like to express gratitude to the companies in Tokyo who helped me recruit participants and organise rooms for the research. In particular, I would like to thank Steve Sherriff, Atsuko Yoshida and Aoi Shimizu for going above and beyond the call of duty with their generosity and support. Even more crucially, thanks to all my participants in Japan who gave up time in their busy lives to take part so willingly in the studies, and to friends and colleagues who provided ‘native speaker’ input. Without you all, none of this could have taken place.

Finally I would like thank Asako for all her help with Japanese translations and for being there throughout, and to all my family for keeping me going with their enthusiasm and humour.

Table of Contents

List of tables and figures

CHAPTER 1: Introduction

Overview, rationale and organisation

Although defined in many different ways, a common way to view formulaic expressions is as sequences of words that can in some sense be treated as a whole lexical unit. From this perspective, any of the following examples could be formulaic for a particular individual speaker:

(1) *by and large*
(2) *turn a blind eye to*
(3) *an accounts manager in Finance*

Suppose a second language (L2) speaker of English was presented with these to learn. How might they go about internalising them as single lexical units? It might seem that (1) would necessarily need to be memorised as a single unit because the component words give no clue to the meaning. On the other hand, (3) might be constructed initially and come to be treated as a single unitary expression after being used repeatedly over time (because it is the speaker's job perhaps). What of (2)? Either option seems possible: it could be learnt as single unit or might need to be broken down into parts first. These examples illustrate the kinds of question raised when considering the acquisition of formulaic expressions by L2 speakers. For example: How can we tell whether an expression is formulaic for a speaker? How does the internalisation to a single unit take place? How might the component parts of the expression influence this process?

This research seeks to address questions such as these. In particular, the aim of the thesis is to explore the identification of formulaic expressions in speech and the possible psycholinguistic processes involved in their acquisition by individual L2 speakers.

1.1. Why study the acquisition of L2 formulaic expressions?

Formulaic expressions[1] are widely used by native speakers and are considered to be a central feature in the language generally (e.g. Sinclair, 1991). Such expressions

---

[1] To avoid confusion with other popular terms for such strings (e.g. 'formulaic sequence', 'conventional expression', or 'multiword expression') which have their own specific connotations, I have adopted the neutral term 'formulaic expression'. Terms and definitions are discussed in more detail in Chapter 2.

are typically thought to include collocations (*young adult*), multi-word verbs (*run the risk of*), idioms (*turn over a new leaf*), conventional expressions (*nice to meet you*), discourse markers (*by the way*), exclamations (*I don't believe it),* sentence starters *(I think that…),* grammatical constructions (*not only X but also Y)* and others. Their use by speakers is thought to bring benefits in terms of fluency and speed of processing (Siyanova-Chanturia & Van Lancker Sidtis, 2019; Towell, Hawkins, & Bazergui, 1996; Wray, 2002a). The underlying view is that formulaic expressions are stored and processed as whole units without the need to construct them on-line. It is argued that this holistic storage and processing provides the efficiencies necessary to enable the fluent, connected multi-clause discourse of native speakers (Pawley & Syder, 1983; Tremblay & Baayen, 2010). As well as the potential fluency and processing benefits, using formulaic expressions provides the opportunity to convey standard (specific, nuanced) meanings within a speech community in a clear and understandable way (Schmitt & Carter, 2004; Siyanova-Chanturia, Conklin, Caffarra, Kaan, & van Heuven, 2017).

However, a variety of research suggests that, despite these benefits, L2 speakers of English tend not to use formulaic expressions to anything like the extent of native speakers (Granger, 2019; Meunier, 2012; Paquot & Granger, 2012)[2]. Reasons given include a lack of sufficient exposure and a failure to notice that expressions may have a holistic nature. Other explanations (Wray, 2019) are related to the different ways that native and L2 speakers approach language learning. For example, while children are thought to learn such expressions in a holistic way in their L1, Wray and Perkins (2000) have suggested that there may be a tendency for adult L2 learners to explicitly analyse any new expression in terms of its component parts. This may be part of a natural tendency of an adult learner to abstract and generalise, or it may be an education-induced strategy to help remember the expression and support subsequent recall and accuracy.

There has been some research around the acquisition of formulaic expressions by L2 speakers, although the question of how individual learners break down expressions for learning is rarely addressed. Some research has shown that focussing learning on formulaic expressions does seems to increase usage and awareness of them in L2 learners which in turn leads to general increases in fluency of production (X.

[2] Of course, there are many fully proficient L2 speakers of English who do use formulaic language in a nativelike way. In this thesis, 'L2 speakers of English' should be taken to refer to learners who are not yet fully proficient.

Chen, 2009; Wood, 2009). Other research has explored the benefit of drawing learners' attention to particular prosodic features (such as alliteration or assonance) of the expression (Boers & Lindstromberg, 2012), or to semantic features such as L1 congruence (Conklin & Carrol, 2019) or imageability (Steinel, Hulstijn, & Steinel, 2007), as a support in committing formulaic expressions to memory. Other studies have focussed more on the targeted learning of specific sequences. For example, Wray (2004) showed that an individual could learn specific expressions in a new language (Welsh) and use these in the context for which they were intended. Wray and Fitzpatrick (2008) had L2 English learners memorise target utterances which they then used in real transactional situations in the UK. They found that learners varied considerably in how accurately they delivered the utterances they had learnt (and in their propensity to attempt to use them). Evidence from deviations in the output suggest that reconstruction of the learned expression on-line may be a feature of its recall and production, at least during some stages of the acquisition process. A number of researchers have proposed that formulaic expressions are not always learnt in an all-or-nothing way (Bardovi-Harlig, 2019; Schmitt & Carter, 2004) but may be learnt with some phonological elements, component words or syntactic structures initially not known but filled in later. There has also been a study on early learners of French (Myles, Hooper, & Mitchell, 1998; Myles, Mitchell, & Hooper, 1999) which showed that expressions learnt as whole units may subsequently be broken down over time. While none of the studies considered above focusses precisely on the process by which individual L2 speakers acquire particular expressions, they do suggest some important points about individual acquisition. Firstly, there is support for the contention (e.g. Meunier, 2012) that the specific and targeted learning of formulaic language by L2 learners is an important means for facilitating its acquisition. Secondly, even when targeting specific expressions, there are different possible routes to acquiring formulaicity. Further study of these routes and how (and when) expressions are analysed during learning will be valuable for developing our understanding of how formulaic expressions are acquired by L2 speakers.

1.2. Challenges in studying L2 acquisition

One particular challenge in drawing together the results of research into formulaic language is the wide variety of ways in which formulaic expressions have been labelled and defined, reflecting the different requirements of different schools of enquiry. Different accounts have focussed on their frequency in the language as recurring multiword expressions (MWEs) (Ellis, 2012; Siyanova-Chanturia, 2015), their function (Nattinger & DeCarrico, 1992), their idiomaticity (Van Lancker Sidtis, 2009) or their conventionalised usage (Bardovi-Harlig, 2019). Wray (2012) highlights the dangers of over-generalisation and circularity when pooling results based on different types of conceptualisation. For example, empirical results for MWEs do not necessarily apply to idioms as these are generally not frequent in corpora (e.g. Moon, 1992).

In the case of exploring the acquisition of formulaic expressions by individual speakers, formulaicity needs to be defined in a way that takes the individual's relationship with the expression into account. An important distinction noted by Wray (2008a) is between externally-defined sequences that are considered to be formulaic 'in the language' (such as idioms and high frequency multiword units) and those which may be 'psycholinguistic' units in the lexicon of the individual speaker. Some researchers (e.g. Dahlmann, 2009; Erman, 2007) have shown that these are not necessarily the same, particularly for L2 speakers. For example, an L2 speaker may know of a particular idiom (which is formulaic in the language) but not be able to use it smoothly. At the same time, a specific non-idiomatic expression (such as *I'm an accounts manager in Finance*) may become internalised as a (psycholinguist) formulaic expression for that speaker (because it is relevant and often-repeated) while not being considered generally formulaic.

A useful definition for formulaicity which takes an internal perspective is that proposed by Myles and Cordier (2017). They define a 'processing unit' as being an expression which is used and processed as a single unit by the individual speaker, so that it gives a processing advantage (dues to holistic storage or automaticity) compared to the same expression being constructed on-line from its component elements. Although it is not possible to observe internal features such as holistic storage or automatic processing directly, a common approach for identifying formulaicity has been to use sequence fluency. For example, in studies by Erman (2007) and Dahlmann (2009), the absence of dysfluency markers (such as pauses, hesitation, and repetition) was used as a criterion for formulaicity. Myles and Cordier (2017) developed a set of criteria for the internal formulaicity of a sequence whereby fluency (indicating phonological coherence) along with evidence of its unitary nature (such grammatical irregularity or semantic opacity) are the two necessary conditions. This approach was applied to a group of British speakers of L2 French at an advanced level in a study by Cordier (2013) and shown to be a practical method for identifying internally formulaic expressions in speech. While legitimate questions remain about what constitutes a processing advantage and how this relates to fluency and automaticity in expressions, the approach of Myles and Cordier is a good starting point for exploring acquisition and use.

1.3. Aims and scope

The aim of the current research is to explore how L2 speakers acquire formulaic expressions and to describe the possible routes by which it occurs. This work seeks to build on the existing L2 acquisition research by providing a more detailed account of the possible mechanics of acquisition, storage and usage of formulaic expressions in L2 speakers. In particular, it explores the psycholinguistic processes involved in internalising the target and how different factors associated with the memorisation process might influence these processes. The research also looks at ways of modelling holistic storage and processing, building on relevant models and theoretical approaches. Through this study, it is hoped that, as well as contributing to the theoretical understanding of formulaic expressions and how their acquisition and storage might be represented in the mind, insights can be gained that will support the future learning of formulaic expressions by L2 speakers.

While these are the broad aims, it is important to acknowledge the boundaries of what this thesis covers. The topic of L2 acquisition of formulaic expressions could potentially be explored from a wide variety of perspectives within applied linguistics. As highlighted by Siyanova-Chanturia and Pellicer-Sanchez (2019), as well as the cognitive and psychological perspectives on formulaicity, there are socio-cultural and pragmatic aspects and pedagogical issues that are relevant. While drawing on findings and ideas from those areas, the current research has a very specific focus. It is concerned with ‘psycholinguistic’ formulaic expressions, that is, ones that are formulaic for the individual speaker (internal) rather being specified as formulaic ‘in the language’ (external). It takes the approach of Myles and Cordier (2017) in defining such expressions in terms of holisticity and automation of processing and storage, and identifying them primarily in terms of their phonological coherence and unitary form. The target language for acquisition is English and the outcome of acquisition is mainly assessed in terms of spoken production. The emphasis on spoken use of formulaic expressions by L2 speakers is necessary from an operational perspective. At the same time, this focus is valuable because speech is a fundamental aspect of language use, and fluency and the ‘natural’ delivery of speech are frequently a major aim for learners.

There are also limits in the scope of the empirical studies with respect to the nature of the acquisition studied, the type of formulaic expression and the participants who take part. Regarding acquisition, while formulaic expressions may potentially be acquired through incidental, semi-incidental or intentional learning (Pellicer-Sánchez & Boers, 2019), this thesis is primarily concerned with certain types of intentional learning. In particular, in the acquisition studies, participants are presented with novel target sequences which are learnt in different ways and their output is monitored over time. While recognising the importance of implicit learning in second language acquisition (Ellis, 2015), this research focuses on particular aspects of learning that can be made explicit (such as how a target sequence is structured or its phonological form) and how such explicit knowledge can lead to fluent production.

The type of formulaic expression targeted for acquisition is also restricted. It is widely accepted that formulaic expressions (whether defined internally or externally) can be fully fixed (e.g. *by and large*, *as sure as eggs is eggs*), allow for some variation in terms of inflection or expansion (e.g. *[very] nice to meet you*, *come/came/is coming to a head*), or be structural frames with slots for open class items (*as X as a Y*). The latter types are very important and may have significance in language development, for example as abstractions from fixed exemplars (Ellis, 2012; Woolard, 2013). However, for simplicity and in order to maintain a clear focus on the initial acquisition, the focus of the thesis is on fixed or semi-fixed expressions.

Finally, the L2 speakers tested in the studies are all Japanese speakers of English at intermediate/advanced level. This is mainly for pragmatic reasons (as this is the group of speakers I have most access to). However, fluency is thought to be a particular challenge for Japanese speakers of English (Cornwell, Simon-Maeda, & Churchill, 2007; Seki, 2004). Issues regarding the acquisition of formulaic expressions may therefore be particularly pertinent for such learners. It is also recognised that the role of formulaic language for L2 speakers at an earlier stage of learning may be differ from those at a more advanced level (Myles, 2004). Nevertheless, the nature of the findings from the studies suggests a much broader range of L2 learners than those tested may also benefit from deeper insights into the different possible routes to achieving formulaicity and fluency in the production of useful target sequences.

1.4. Organisation of the thesis

The thesis begins in Chapter 2 by reviewing in more detail the existing research on the L2 acquisition of formulaic language. This review considers the wider perspective on formulaic language and how it has been conceptualised before moving to a focus on the idea of internal formulaic expressions. Appropriate means for identifying these expressions in speech are considered, leading to a more detailed exploration of the hierarchical approach of Myles and Cordier (2017). There is also a closer look at empirical research on L2 acquisition, with a particular focus on studies of what happens when L2 speakers memorise targeted sequences for future use (Wray, 2004; Wray & Fitzpatrick, 2008). Finally, the many factors that could influence the targeted acquisition of expressions are considered, including features both of the expressions themselves (e.g. length, structure, prosodic features) and of the way in which they are learnt (e.g. analytic or holistic approaches). As well as raising questions about the different routes to acquiring formulaic expressions that may be possible, the review also highlights useful methodological approaches for exploring them. These include different ways of selecting and presenting target sequences, methods for assessing formulaicity, and methods for collecting and analysing sample speech.

Building on the questions and suggested methodological approaches arising from the review, the first two empirical studies (S1 and S2) explore how some Japanese speakers of English use formulaic expressions in their speech. Chapter 3 describes the first study (S1) in which the identification approach of Myles and Cordier (2017) is applied to speech data collected from eight intermediate and advanced Japanese speakers of English (JSE), along with two native speakers (to provide a comparison). The results are compared with those of Cordier (2013) and the study provides a useful way of exploring the practicalities of this approach. The second study (Chapter 4) addresses some issues associated with identifying formulaic acquisition in a single sample of text and explores the extent to which the acquisition process can be observed over time. For this, the speech of an individual JSE speaker was sampled on multiple occasions (15) over the course of 9 months and the fluency of recurring unitary expressions analysed.

The next phase of the research investigates aspects related to the memorising of target utterances in general. The main experimental work (Chapter 5) is a replication of the study by Wray and Fitzpatrick (Fitzpatrick & Wray, 2006; Wray & Fitzpatrick, 2008) which explored the targeted memorisation of utterances by L2 speakers for subsequent reproduction. The aim of the replication is to explore the original methodology and verify particular findings related to the types of deviations introduced and the participants' approach. In Chapter 6, the study is extended to also measure the phonological coherence of the learnt utterances in order to better understand what kinds of analysis or segmentation take place when target sequences are memorised and how that affects what becomes formulaic for the speaker. These studies provide some useful insight into the identification of expressions that may be formulaic for the speaker at the time of delivery and demonstrate stages in the evolution of that formulaicity through practice and repetition.

Building on this, an approach is developed in Chapter 7 via a small exploratory study (S4) in which specific target sequences are chosen and introduced to a small group of L2 speakers. This study investigates the potential effects of segmenting sequences into shorter elements or of embedding them into longer sequences. While the results are inconclusive, it is useful for testing the method and fine-tuning it for the subsequent studies. Taking stock of the findings up to this point, Chapter 8 returns to the research literature in order to explore ways of modelling the acquisition and processing of formulaic expressions, homing in on a cognitive model describing the representation of formulaic expressions proposed by Sprenger, Levelt, and Kempen (2006). This provides a basis for describing how formulaic expressions may be represented in the mind of the individual. At the same time, combined with earlier research, the review provides some insights into how to explore possible routes to acquisition in the following studies.

The final experimental phase of the research explores two broad routes by which a targeted expression could become formulaic. In the first study (S5), described in Chapter 9, a group of 10 Japanese speakers of English learn novel target sequences using two different methods for the initial memorisation. These are specifically designed to induce either 'holistic acquisition' (where a speaker takes on the expression as single whole unit) or 'fusion' (where the expression is initially reconstructed but later becomes unitary). Their ability to reproduce these expressions is assessed immediately and after one and three weeks. Chapter 10 describes a follow-up study (S6) in which the same participants and targets are assessed a few months later. In addition to fluency, psycholinguistic methods are used to assess the formulaicity of the targets, based on the idea that formulaicity represents a qualitatively different state than just being a fast, fluent on-line reconstruction of the expression.

The results of the studies and the research are brought together in Chapter 11 which discusses ideas raised regarding the identification of formulaic expressions in speech and the way expressions achieve formulaicity over time. A tentative model describing two routes to formulaic acquisition, based on the model of Sprenger et al. (2006), is proposed and discussed in the light of the findings and examples from the studies. The thesis concludes in Chapter 12 with a brief summary of the limitations and general implications of the research along with suggestions for future research.

CHAPTER 2: Understanding formulaicity and L2 acquisition

A review of what is known about the acquisition and use of formulaic expressions by L2 speakers

2.1. Introduction

This research aims to explore the following central question: how do L2 speakers acquire, store and process formulaic expressions for spoken production? In order to do this, five key review questions have been identified:

- What are formulaic expressions and how have they been defined from the perspective of individual speakers?
- What are the key issues regarding the acquisition and use of formulaic expressions by L2 speakers?
- What features could be used to identify (internal) formulaic expressions in speech?
- What factors may influence the targeted acquisition of formulaic expressions (in L2 speakers)?
- What are the cognitive processes involved in the acquisition and use of formulaic expressions, and how have they been modelled?

This chapter presents research from the literature that addresses the first four of these questions. The final question on processing and cognitive models is postponed until Chapter 9 as it builds on the experimental work that precedes it and leads more naturally into the final experimental work later. While the main focus overall is on the acquisition and usage of targeted sequences by L2 speakers of English, the review will refer to issues of formulaic language more generally (e.g. usage by native speakers) and other languages where these are appropriate.

For each review question, a broad range of research is examined to establish the main existing ideas, and some key studies are covered in more detail where appropriate. At the end of the chapter, there is an overall conclusion drawing together the main substantive, theoretical and methodological issues raised and the key areas for further enquiry. These will shape the main research questions and the structure of the subsequent experimental studies.

2.2. Formulaic language and the individual speaker

2.2.1. *What do we mean by formulaic language?*

Research in corpus linguistics, psycholinguistics and second language acquisition (SLA) has provided a variety of evidence that language is highly ‘formulaic’. That is, people tend to use familiar, standardised ways of expressing thoughts and ideas rather than exploit the creative potential of the language offered by the free combination of smaller units into grammatical patterns. For example, Pawley and Syder (1983), drawing upon studies of English conversational talk, observe that a key feature of nativelike speech is the use of lexical phrases (prefabricated phrases which are stored as single wholes). They identified many such lexical units “whose grammatical form and lexical content is wholly or largely fixed” (p.191). They suggest that it is the linking of such units that gives rise to fluent idiomatic speech. Sinclair (1991) proposed that speakers tend to adopt the ‘idiom principle’ in that most texts include “a large number of semi-pre-constructed phrases that constitute single choices, even though they may appear analysable into segments” (p.110) and speakers tend to use these whenever they can. In many cases, the formulaicity of expressions (as a ‘single choice’) can be recognised by their syntactic irregularity (e.g. *by and large*) or semantic opacity (e.g. *beat about the bush*) since such expressions could not have been generated by the language grammar at the time of use (Sinclair, 1991). However, as Wray (2000) notes, many regular expressions (such as *It was lovely to see you*) may also be formulaic due to their conventionalised usage.

A variety of research has tended to support the idea that formulaicity is ‘ubiquitous’ in language and that normal discourse, both written and spoken, contains a large percentage of language which is formulaic (Ellis, Simpson-Vlach, & Maynard, 2008; Granger & Meunier, 2008; Wray, 2002a). For example, Jackendoff (1995) makes the claim that there are as many formulaic expressions in American English as there are single words. Erman and Warren (2000) estimate that about half of fluent native text is formulaic (in that expressions are constructed according to the idiom principle). Research also suggests that formulaic expressions are more frequent in spoken language than in written. For example, Biber, Johansson, Leech, Conrad, and Finegan (1999) analysed a conversation corpus and an academic prose corpus to determine the frequency of lexical bundles[3], which they define as “sequences of word

[3] Lexical bundles and the extent to which they are classed as formulaic expressions is discussed further in Section 11.2.1

forms that commonly go together in natural discourse" (p. 990). They found that these accounted for around 30 percent of the words in the spoken corpus and about 21 percent of the written corpus. Overall, these corpus studies suggest that formulaic language makes up at least one third and possibly as much as one half of all discourse. Such prevalence clearly makes it an essential component of adult linguistic competence.

2.2.2. *How has formulaic language been conceptualised?*

Formulaicity has been investigated from a variety of perspectives, and a wide range of terminology has been used to describe it. According to Wray (2002), at that time over 40 terms had been used to describe formulaic expressions. Since then, while efforts have been made to standardise the approach and terminology, there is still variation, reflecting different ways that formulaicity is conceptualised. Formulaicity has been described in terms of formal features of expressions (such as idiomaticity), functional qualities, context of usage and frequency within corpora. For example, Nattinger and DeCarrico (1992) used the term 'lexical phrases' to describe form/function composites which have a well-defined form that is used for specific functions in context. Aijmer (1996) also proposed functional categories for classifying formulaic expressions, including discourse devices (for orienting the hearer or organising content) and social interaction (e.g. thanking, greeting, apologising, offering, etc.). The term 'formulaic sequence' was introduced by Wray (2002) and defines formulaicity in terms of multiword sequences that appear to be stored as whole units in the speaker's mental lexicon. This has been widely used as an umbrella term, although often without a clear connection to the original definition. (This will be discussed further in Section 2.2.5). The term 'formula' has also been used as an umbrella term referring to multiword units that follow the idiom principle of Sinclair (1991). Erman and Warren (2000, p. 31) use the term 'prefab' which they define as "a combination of at least two words favoured by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalisation".

The idea of conventionalised usage is also captured in the proposal (e.g. Van Lancker Sidtis, 2009) that formulaic language consists of those expressions that can be categorised as automatically produced. That is, they are not newly created from the operation of grammatical rules on lexical items. Such expressions (including idioms, proverbs, speech formulas, conventional expressions, expletives, etc.) have stereo-typed form (specific words and word order and intonation contour) and conventionalised meanings. Corrigan, Moravcsik, Oulali, & Wheatley (2009, p. xiv) suggest two distinctive characteristics that differentiate expressions that are formulaic from those that are not. They have restricted form (they are not amenable to lexical or structural reformulations) and restricted distribution (they are confined to particular communicative situations or contexts). However, Corrigan et al. recognise that these restrictions on form and distribution are probabilistic rather than absolute, and some variation may be allowed. Indeed, as Wray (2000, p. 466) points out, many formulaic expressions will be regular and canonical and therefore either literal in their own right or identical in form to a literal meaning (e.g. *it was lovely to see you*). Similarly, while many expressions will be preferred ways of expressing meaning or function in a given context (Wray, 2002a, p. 87), other expressions may be formulaic without being uniquely keyed to a specific situation (e.g. *I think* ..., *it seems to me that...*).

The term 'multi-word expression' (MWE) has also been used by many researchers (e.g. Dahlmann & Adolphs, 2007; Siyanova-Chanturia et al., 2017; Strik, Hulsbosch, & Cucchiarini, 2010), particularly in corpus research, where MWEs are identified on the basis of frequency-based statistics. Some researchers have combined different features in their definitions. For example, Boers and Lindstromberg (2012, p. 83) define formulaicity as "the use of word strings that have become conventionalised in a given language as attested by native-speaker judgment and/or corpus data".

These examples illustrate both the multiplicity of terms and also that they may actually refer to different sets of expressions. For example, Grant (2005) showed, through analysis of idioms in a corpus, that idiomatic expressions are not necessarily frequent. She identified 104 core idioms in English, defining them as multiword units that are non-compositional (i.e. the meaning of the sequence is not retained if all the lexical words are replaced by their definitions), are non-figurative, and have more than one word which is non-compositional or non-literal (Grant, 2005, p. 430). These criteria exclude examples such as *divide and conquer* (compositional), *within spitting distance* (figurative), and *drive someone to distraction* (only one non-literal word: *distraction*), but include such quintessential idioms as *kick the bucket*, *by and large*, and *red herring.* She found that not one of these idioms occurs frequently enough in the British National Corpus to merit inclusion in the 5,000 most frequent words of English (which occur at least 19 times per 1 million words).

While there are many ways that formulaicity could be conceptualised, Wray (2009, pp. 32-34), in a chapter specifically focussing on this question, suggests that restrictions on form, distribution or compositionality are not defining features but represent different dimensions upon which formulaic expressions may vary. She argues that certain typical (but non-essential) features of formulaic expressions (such as non-canonical form, semantic opacity and specific pragmatic function) can be explained as resulting from their patterns of use over time. The sequence is passed around within a community and becomes easier to select and use (as it requires less processing than a novel string of similar size). Component words gradually lose phonological precision and morphological immediacy through holistic form-meaning mapping. Sequences attached to agreed meanings and functions may adopt new connotations and associations, or become phonologically distinctive and protected from grammatical and other change in the language. The longer an item is insulated in this way, the more it will stand out from novel material until it is viewed as irregular in form or opaque in meaning.

2.2.3. *What are the benefits of using formulaic expressions?*

There are a number of potential benefits in using formulaic language in spoken language. The principal one highlighted in the literature is its role in facilitating greater fluency and automation of the processing in speech production (Ejzenberg, 2000; Towell et al., 1996; Wray, 2002a). The suggestion is that this is achieved through the retrieval of longer chunks of language which are stored whole in the lexicon and do not therefore need to be constructed consciously in working memory. This follows because estimates for the number of items that can be held in the mind at one time is around four pieces (Cowan, 2010). So, in order to construct a novel sentence with many more than four words (as native speakers often can do fluently), some of those words would need to be packaged within a larger, holistically managed, unit. This suggests that the use of prefabricated chunks is necessary for fluent speech, and should be faster and easier than creating novel combinations of words.

2.2.3.1. Research into processing advantages

There has been some research which supports the idea of a processing advantage for certain types of idiom (e.g. Siyanova-Chanturia & Lin, 2017; Underwood, Schmitt, & Galpin, 2004; Van Lancker, Canter, & Terbeek, 1981). For example, Underwood et al. (2004) found that terminal words in idioms are read more quickly than in non-idioms. Siyanova-Chanturia and Lin (2017) explored the acoustic differences between idiomatic expressions. They collected recordings of 66 native speakers of British English reading aloud three types of expression: idioms used figuratively, idioms used literally, and matched novel control phrases. Each was embedded within a paragraph appropriate to the context. They found that figurative idioms had shorter duration than their literal versions and controls.

In addition, a variety of studies have shown evidence that multiword expressions (MWEs) tend to be processed more quickly. This has been shown for comprehension (Jiang & Nekrasova, 2007; Kim & Kim, 2012) and production (Arnon & Cohen Priva, 2014; Ellis et al., 2008). These studies principally rely on comparing certain aspects of processing between higher and lower frequency expressions. For example, Kim & Kim (2012) investigated the self-paced reading of two-word verbs (e.g. *find out*, s*ort out*) and found that reading times were longer for lower frequency items. Arnon and Cohen Priva (2014) analysed a large corpus of spontaneous speech (containing over 8000 trigrams from 40 different speakers) and found there was a significant effect of trigram frequency on the duration of the middle word of the trigram, even when controlling for frequencies of the constituent words and bigrams within the trigram. Ellis et al. (2008) found that more frequent phrases were delivered faster in native speakers. They also found that phrases with a higher Mutual Information (MI) score[4] had shorter voice onset latency.

Duration effects have also been found for children. For example, in a repetition study, Bannard and Matthews (2008) found that young children articulated frequent phrases (e.g. *sit in your chair*) more quickly than similar but less frequent expressions (*sit in your truck*). They also repeated them more accurately. In addition to these psycholinguistic studies, processing differences between multiword expressions and 'novel' expressions have been investigated via neurolinguistic studies that compare brain activity in the processing of such sequences. These are based on the idea that brain activity associated with the composition of phrases is reflected in particular event-related potentials (ERPs) which can be measured during on-line processing of the expressions. Studies have found reduced ERPs in the processing of more frequent MWEs (Siyanova-Chanturia *et al*., 2017; Tremblay *et al*., 2016; Tremblay & Baayen, 2010; Tremblay & Tucker, 2011). Further discussion of this kind of study is given in Chapter 8.

The general finding then is that frequent expressions (MWEs) are processed faster than less frequent ones. This does not prove that formulaic expressions in general

[4] MI is a measure that takes into account the frequency of the component words within an expression. It is thought to indicate a degree of cohesiveness of the expression based on statistical information

have a processing advantage since frequency is not necessarily a reliable indicator of formulaicity (and excludes the many formulaic expressions which are not frequent in the language). However, the general observation of faster delivery and phonological reductions in idioms and MWEs may be relevant to individual acquisition based on salience and frequency of hearing. For example, hearers (including children acquiring the language) might spot that there is something special about these expressions, and earmark them for holistic learning (because that is how others seem to be processing them). Alternatively, they may *have* to be made holistic because the hearer cannot actually work out the components accurately due to the speed and phonological reductions.

2.2.3.2. Other potential benefits of formulaicity

A further benefit of formulaic expressions suggested by many researchers (e.g. Siyanova-Chanturia et al., 2017) is that, as highly familiar, conventional ways of expressing thoughts and ideas, they represent a meaning or function that is easily recognised, thus helping to make our discourse natural and easily comprehensible. Schmitt and Carter (2004) suggest a number of pragmatic and functional benefits in using standard prefabricated expressions in interactions including: transacting specific information in a precise and understandable way *(cleared for take-off*); realising practical functions (*I'm just looking*); and signalling discourse organisation (*on the other hand*). A further possible benefit suggested by Wray (2002a) is that the use of formulaic language is an important linguistic tool for socially aligning ourselves with other people in our community. Wray (2002a, pp. 88-92) argues that using standardised, possibly idiomatic, expressions that are recognised ways of saying something may help signal group or personal identity within a target speech community that has a shared way of speaking. Since these fixed sequences and frames are retrieved whole from the lexicon, they require less decoding and so also minimise the risk of misunderstandings, which is beneficial to both speaker and hearer.

As well as delivering benefits during spoken interactions, formulaic language may play an important role in the structure and acquisition of language more generally. Various psycholinguistic research has shown that native language users are sensitive to the frequencies with which words co-occur within expressions (Ellis *et al*., 2008; Ellis & Sinclair, 1996) and that children learn many sequences as whole units if they encounter them frequently (Dabrowska & Lieven, 2005). It has been suggested that language development may involve these formulaic expressions first being learnt as exemplars which are combined together by simple rules to form new utterances (Lieven, Behrens, Speares, & Tomasello, 2003). Many researchers such as Ellis (2012) therefore highlight the importance of formulaic language within usage-based theories of language. These theories contend that the “cognitive organisation of language is based directly on experience with language” (Beckner et al., 2009, p. 5) and that the basic units of language are constructions. These are form-meaning pairings which vary along dimensions of size (from words to clause level) and abstraction (from fully fixed words or phrases to empty frames) (Beckner *et al*., 2009; Ellis, 2012; Goldberg, 2006; Hoey, 2005). In this approach therefore, formulaic expressions are an integral part of language in that they may be basic lexical units in their own right, as well as staging posts on the way to acquiring more abstract constructions. Woolard (2013) suggests there may be benefits for L2 speakers in learning formulaic expressions as exemplars which can be broken down by the speaker into more frame-like constructions later, if necessary.

2.2.4. ***Formulaicity and the individual speaker***

As noted by Wray (2008a), there is an important difference between sequences that are considered formulaic ‘in the language’ and those that may be formulaic for the individual speaker. Wray (2012) suggests that a failure to fully appreciate this difference may underlie some ambiguity in the terminology used and a lack of consistency in the way that conclusions have been drawn about formulaic language from empirical studies. Expressions that are considered formulaic in the language are generally defined as such by criteria external to the speaker such as frequency of occurrence in large corpora, idiomaticity (e.g. irregularity of syntax or opacity of meaning) or their pragmatic function (e.g. expressions such as ‘*How do you do’*). Formulaicity defined with reference to the individual speaker, on the other hand, draws on features of the internal processing and storage of the expressions by the speaker (e.g. faster processing or holistic storage). Much of the research into formulaic language deals with speaker-external sequences such as corpus-derived frequent multiword expressions or idiomatic expressions and collocations (Conklin & Schmitt, 2008; Underwood *et al*., 2004).

There has been an implicit assumption that these two conceptions of formulaic expressions are the same (Pawley & Syder, 1983; Sinclair, 1991). However, evidence shows that while this might be broadly true for native speakers (Underwood *et al*., 2004) it is not always so (Schmitt, Grandage, & Adolphs, 2004) and for L2 learners it is generally not the case (Dahlmann, 2009; Erman, 2007; Siyanova-

Chanturia, Conklin, & Schmitt, 2011) unless the sequences are common and transparent (Jiang & Nekrasova, 2007).

A useful study which shows that externally defined formulaic expressions are not necessarily processed holistically by individual speakers is that of Schmitt et al. (2004). They derived potential sequences from reference works, including Nattinger and DeCarrico (1992), and checked them against both general and specialised corpora (including BNC and CANCODE[5]) for frequency of occurrence. Formulaicity of the resultant target sequences was measured using a special dictation task done at speed to ensure that participants could not memorise and repeat the spoken text using rote memory. The procedure involved incorporating the target sequences into sentences of about 20–24 words which together formed a short story. The sentence sequence was then spoken out loud, with gaps in between for the participant to repeat each sentence as exactly as possible.

Because this is a challenging task where the participant is unlikely to be able to reformulate the whole sentence on a word-by-word basis, Schmitt et al. (2004, p. 137) postulated that any sequence that is formulaic is likely to be reproduced fully intact (or not at all), with no hesitation pauses or transformations. Partial, amended or dysfluent sequences would suggest that these targets are not in fact processed holistically but constructed out of the individual words. They also stipulated that conclusions about formulaicity could not be drawn on targets that were fully absent, since the speaker may have just failed to register the chunk (akin to forgetting an individual word). Based on this criterion, many of the target sequences were found not to be formulaic, and for others the formulaicity varied from individual to individual. Their results show that a sequence deemed formulaic on the basis of conventionality and frequency in the language is not necessarily formulaic (i.e. holistically stored or processed) for the individual. This and other studies, then, lend support to the idea that, while there may be extensive overlap between them, externally and internally defined formulaicity are not the same.

For internally defined formulaicity, Tabossi, Fanari, & Wolf (2009) suggest that the speaker's knowledge of the sequence is an essential factor in whether it is formulaic for them. This knowledge will depend on the degree of familiarity or experience with the sequence and the way it has been learned. Formulaic expressions are not

[5] BNC is the British National Corpus; CANCODE is the Cambridge and Nottingham Corpus of Discourse in English

thought to always be learnt in an all-or-nothing way, even by native speakers. For example, Schmitt and Carter (2004) suggest they may be learnt with some phonological elements, component words or syntactic structure initially not known but filled in later, or they may be created by a process of “fusion” whereby an often used expression, initially constructed, becomes formulaic (i.e. stored whole in the lexicon) by regular usage. The latter process may be particularly important for L2 speakers as fused expressions could constitute items that the individual has consciously or subconsciously identified as likely to be useful and relevant. The key point, then, is that formulaicity is a dynamic feature of the idiolect. For any L2 speaker, different expressions may be at different stages in the process of becoming fused together. This process may work in the opposite direction too, with some initially holistic expressions starting to be broken down (Ellis, 2012). Thus, for L2 learners in particular, there may be ‘degrees of formulaicity’ which depend on the stage of acquisition of the sequence in question.

2.2.5. ***Defining ‘internal’ formulaicity***

While the internal and external conceptualisations of formulaicity may be different, Myles and Cordier (2017) suggest that the distinction has frequently not been taken into account in the research. A particular case in point is the term ‘formulaic sequence’. This was introduced by Wray (2002a) as an inclusive way to capture the broad range of possible formulaic expressions. She defined the formulaic sequence (p.9) as:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, pre-fabricated; that it, stored and retrieved whole from memory at the time of use, rather than being subject to generation by the language grammar.

This clearly defines formulaicity from an internal perspective. However, this term has been widely used subsequently in a broad range of research. Often the term has been used to refer to expressions such as idioms and collocations (Conklin & Schmitt, 2012; Underwood *et al*., 2004) or frequent multiword expressions (Brooke *et al*., 2015; Lin, 2010) all of which are actually formulaic by external criteria. Wray (2012) suggests that these different applications of the terminology are potentially problematic if findings are applied to formulaic expressions in general while the approach taken only deals with one type of formulaicity.

An advantage of considering formulaicity from the internal perspective is that it provides a clear theoretical basis for describing the phenomenon and for distinguishing features that are fundamental and features that arise as a consequence. A disadvantage is that it is difficult to operationalise the concept in the absence of direct access to the systems of storage and processing in the mind of the individual speaker. Myles and Cordier (2017) therefore suggested creating a definition for formulaicity that could be used more directly. They proposed a new term 'Processing Unit' (PU) and defined this 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 (Myles & Cordier, 2017, p. 10)

This is proposed as a pragmatic definition which allows researchers to focus on the desired effects of formulaicity (which can be observed) rather than stipulating the manner in which they are achieved. The definition combines two conceptually different qualities that could be applied to a sequence of words for a given speaker. The first is that the sequence has a unitary quality based on its meaning or function (as they are 'semantic or functional units'). The second is that they afford a 'processing advantage' which is defined to necessarily arise from one of two sources: either the unit is stored holistically in the mental lexicon or it is highly automatised. The term 'advantage' entails a comparison and, although not stated specifically, the implication is that PUs must be processed more quickly or easily than a similar 'normal' (non-PU) sequence of words in the speech stream. Myles and Cordier (2017, pp. 17-18) propose that 'phonological coherence', as represented by fluent delivery of the expression during delivery, is a directly observable indicator of this processing advantage.

2.3. Some key issues in L2 acquisition and use

2.3.1. *To what extent do L2 speakers make use of formulaic language?*

Since the use of formulaic language seems to offer many advantages to the speaker of a language, the acquisition of an appropriate stock of formulaic expressions would seem to be a useful goal for L2 learners wishing to develop their fluency. Paquot and Granger (2012) suggest that formulaic language should play an important role in foreign language learning and teaching. In particular, the use of formulaic expressions such as collocations, phrasal verbs, compounds, and idioms may impact positively or negatively on the complexity, accuracy, and fluency of the L2 speaker's output (Housen & Kuiken, 2009). Ellis (2012) also highlights the significance of formulaic language in L2 acquisition.

However, the existing research on use of formulaic expressions in L2 speakers suggests that such usage is limited and inconsistent at best and that most intermediate and advanced L2 speakers tend not to learn or use it to the same extent as native speakers (Forsberg, 2009; Foster, 2001; Paquot & Granger, 2012). The nature of this inconsistency has been characterised in several ways. For example, Wray (2008b) suggests that L2 learners vary significantly in their use of formulaic sequences due to the diversity of their learning situations. Ellis et al. (2008) highlight the difficulty that L2 speakers have with idiomaticity and collocations, suggesting that it often results from transfer of first language (L1) combinatorial restrictions and idiomaticity. Granger (2009) found that one-third of L2 collocations have errors in them. Paquot and Granger (2012) reviewed the use of formulaic expressions (based on standardised lists or frequency-based methods) in L2 learner corpora and found an underuse of referential collocations, multiword verb phrases, and idiomatic usage, and an overuse of some meta-discursive expressions even in more advanced speakers.

Qualitative explorations of individual cases also suggest limited usage of formulaic expressions in L2 speakers. For example, in a case study on the speech of an intermediate-level Japanese speaker of English, Wood (2009) found that about 12% of her speech consisted of formulaic expressions. Of course any measure of L2 usage of formulaic expressions depends on the way that they are identified. In particular, using external criteria such as frequency or idiomaticity of the expression may under-estimate individual usage. For example, Cordier (2013) used a different set of criteria (to be discussed in Section 2.4.4) and found that, for her sample of advanced learners of French, formulaic expressions represented 27.8% of their speech. As discussed in Section 2.2.1, estimates for native speakers range from 30% to 50% and these are not based on the (less restrictive) criteria of Cordier (which have not yet been applied to native speaker speech).

Wray (2019) suggests a number of possible explanations for why L2 learners may not target formulaic expressions despite the obvious benefits of doing so. One idea is that, for a number of reasons, L2 speakers generally are just not able to target them. Reasons include that: they are not exposed to any expression sufficiently often; the common mode of input is written and this does not highlight them sufficiently; or, the expression is overlooked as formulaic because all aspects (component words/structure) are already known. These claims are difficult to test because the level of exposure of any L2 speaker to any particular expression over the lifetime of their learning is hard to gauge. However, research which explores the effect of frequency (of expressions in the language) on aspects of processing of those expression by L2 speakers may be indirectly relevant. For example, in the reading out loud study by Ellis, Simpson-Vlach and Maynard (2008) mentioned in Section 2.2.3.1, production duration and voice onset latency times were measured for non-native speakers as well as native speakers. Ellis et al. found that, in contrast to native speakers, frequency did not have a significant effect on production times although it did have an effect on voice onset latency. This suggests that frequent expressions may be more easily recognised by L2 speakers, but they are not necessarily produced more quickly.

Wolter and Gyllstad (2013) used a reaction time (RT) approach to compare the processing of collocations of different frequencies. In this experiment, L1 speakers of English and advanced level L2 speakers (with Swedish L1) undertook an acceptability judgement task on lists of collocations. They found that collocational frequency significantly predicted RTs for both L1 & L2 speakers. However, word-level frequency did not. In a similar experiment with L1 Japanese speakers, Wolter and Yamashita (2017) found that, for L1 & advanced L2 speakers, collocational frequency was a better predictor of RTs than word-level frequency, while the opposite was true for intermediate L2 speakers. They found that RT performance was influenced by both word and collocation-level frequency in all groups, but the relative importance of these factors changed. This suggests that, as the L2 proficiency increases, the ability to recognise collocations on the basis of their frequency also increases. It is not unreasonable to suppose that this effect is due to increased exposure, at least insofar as exposure to frequent expressions is related to L2 proficiency. There is also research suggesting that, given sufficient exposure, L2 speakers do have the ability to recognise formulaic expressions in the language. For example, in a lab-based study exploring incidental collocation learning by L2 speakers, Durrant and Schmitt (2010) found that adult second language learners retained information about what words appear together in their input.

Boers, Lindstromberg, & Eyckmans (2014) reviewed a number of explanations for why even many of the most advanced learners fail to use nativelike collocations. Their conclusion was that it was largely due to a lack of attention to their occurrence.

So, while they may recognise frequent multiword units, they do not apply the attentional resources to learn them for productive use. This may be linked to motivation. For example, Meunier (2012) suggests that, since much of the language is formulaic and there may be many different expressions for a particular situation, it is demotivating to have to learn each of these along with their subtly different meanings and uses. Indeed, the findings from Paquot and Granger (2012) suggest that L2 speakers tend to learn one expression for a type of situation and are then disinclined to learn other more nuanced variations. Studies have demonstrated that L2 language learners find formulaicity particularly challenging because it is impossible for them to use the native intuition usually associated with formulaic language use (Granger, 1998; Howarth, 1998; Nesselhauf, 2003) as L2 formulaic expressions do not reliably map on to those in the L1. Further, in the case where English is being learnt as a lingua franca (e.g. for international business), learning formulaic expressions may be considered counter-productive due to the possibility that their use is limited to one speech community only.

Another possible explanation for the relative failure of L2 speakers to utilise formulaic expressions is related to the different ways that native and L2 speakers tend to approach language learning. Firstly, as Wray (2019) points out, there are differences in the acquisition sequence for L2 speakers compared to L1. In L1 learning, the holistic expression is learnt simultaneously with the new complex concept or idea it expresses. In L2 learning, the idea is usually already known and the acquisition of the expression (from idea to expression) may be less instinctive. Secondly, it has been shown (Dabrowska & Lieven, 2005) that native speaking children will often accept a whole multi-word expression as the representation of a particular meaning based on context. They may later break this down on the basis of further examples to form more abstract constructions or generalisations. So, while they do analyse expressions, it is done implicitly, conservatively and on a needs-only basis.

Meanwhile, Wray (2002a) suggests that there may be a tendency for adult learners to explicitly analyse any new expression in terms of its component parts. This is particularly likely when the learning culture emphasises the importance of understanding and applying grammatical rules. However, the extent to which such overt analysis during acquisition actually undermines the learning of expressions as formulaic is unclear. As DeKeyser (1994) points out, it is possible for implicit knowledge (such as that which is required to produce a multiword sequence fluently and automatically) to result from explicit learning. Vocabulary learning research suggests that the effects of analytical learning approaches on eventual storage and production are in fact quite complex. This is an important issue that will be explored in more detail in Section 2.4

2.3.2. ***Do formulaic expressions actually benefit L2 speaker fluency?***

While there seem to be clear fluency and processing benefits for native speakers in using formula expressions, it is useful to check the extent to which those benefits are achievable for L2 speakers. Of course, if internal formulaicity is defined in terms of processing advantage and identified in terms of fluency (as suggested in Section 2.2.5), overall fluency would be expected to improve with increased use of internally formulaic expressions. The question being asked here, however, is whether using more external formulaic expressions (i.e. unitary expressions considered formulaic due to frequency in the language, idiomaticity or other external criteria) tends to provide benefits to L2 speakers in terms of fluency. While the research in this area is limited, an influential study which looked at qualitative factors associated with fluency in L2 speakers is that of Towell et al. (1996). There have also been some corpus-based and longitudinal studies investigating the links between use of formulaic expressions and oral fluency.

2.3.2.1. Towell, Hawkins and Bazergui (1996)

Towell et al. (1996) were interested in L2 oral fluency and the factors that influence its development. An interesting feature of their approach is that they focussed on the cognitive processes underlying fluency and interpreted their findings from the perspective of Anderson's Adaptive Control of Thought (ACT-R) theory of skill development (Anderson, 1996). The idea is that skilled performance involves a conversion of declarative knowledge into procedural knowledge. Declarative knowledge is knowledge about something and is usually conscious and able to be described in words. For example, it could apply to the conscious process of constructing a sequence of words (using grammatical rules) to express something. Procedural knowledge is the ability to do something and is usually unconscious, automatic and not easily describable (e.g. saying an expression automatically without thinking about the component words). It could be argued therefore that conversion to procedural knowledge (proceduralisation) confers a form of processing advantage, since it facilitates faster and more automatic production or comprehension of speech.

Towell et al. tested 12 English-speaking university students studying French before and after a year abroad in a French-speaking country. The task (both before and after the year abroad) was to re-tell a short film story which had been presented without words. They found a significant increase in certain temporal indicators of fluency: speech rate, articulation rate (AR) and the mean length of fluent runs between pauses (MLR). On the other hand there was no increase in the phonation time ratio (the proportion of time spent actually speaking) or in average pause length. They concluded that, since the increased speech rate was down to increases in MLR rather than decreases in pausing, these longer speech runs pointed to greater proceduralisation in the putting together of speech.

Most relevant, however, is the detailed qualitative analyses they did to establish the source of increased fluency and automatisation of speech. They analysed samples of more and less fluent participant speech and found that a key difference was in the use of lexical phrases, as defined by Nattinger and DeCarrico (1992), especially sentence builders (e.g. *c'est* __ *qui* __ = 'it is __which __') and institutionalised expressions (e.g. *qu'est-ce que c'est* = 'what is it'). These tended to be used more efficiently and without pauses in the more fluent speech samples. In the less fluent samples, on the other hand, speakers appeared to "laboriously put together sentences" (Towell et al., 1996, p. 109) with conscious effort. Where expressions considered formulaic 'in the language' were used, these were mostly interrupted by pauses, suggesting that they were not (yet) psycholinguistically formulaic for those speakers. Towell et al. suggested that increased fluency came from greater proceduralisation in the use of known syntactic constructions and in the use of lexical phrases. This reduced dysfluency in the delivery of lines of speech, leading to longer runs and a higher speech rate. In particular, comparing language used before and after the year abroad, they concluded that increased fluency was less the result of changes to conceptual knowledge or the ability to articulate speech but a change in the way that linguistic knowledge is stored – as procedural knowledge.

2.3.2.2. Other studies

There have also been a limited number of longitudinal studies exploring the links between use of formulaic expressions and oral fluency and these have focussed on the co-development of both over time. For example, Wood (2007) investigated how four intermediate students on a full-time intensive ESL course in Canada developed fluency and formulaicity over a 6 month period. He had them do a narrative speaking task whereby they were shown short films (without preparation) once every month and then asked to talk about them. Temporal measures including speech rate (SR) and the mean length of runs (MLR) were used to assess fluency. Note: a 'run' is defined as an unbroken length of speech between two pauses. Formulaicity was measured by the formula to run ratio (FRR). This is the ratio of 'number of formulaic expressions used' to 'number of runs', expressed as a percentage. Formulaic expressions were identified by native speaking linguistics graduates who were trained to apply five diagnostic criteria (as discussed in Section 2.4.1). He found that fluency (on all temporal measures) and formulaicity (FRR) increased for all participants. However, the trend was not linear and individual variations were quite marked.

In a larger-scale study by X. Chen (2009), seventy intermediate level ESL students at a Chinese university were divided equally into control and experimental groups. The experimental group only was provided with specific input and practice of new phrases and sequences including formulaic expressions over a 3 month period. All participants were assessed using a narration task based on retelling a short film clip, followed by a free talk on a related subject. Fluency was measured using SR and MLR, and formulaicity by FRR. Formulaic expressions were identified using the taxonomy of Nattinger and DeCarrico (1992). Chen found that the experimental group increased in all three measures significantly more than the control group implying that the input did support fluency. Also, fluency increased significantly for both groups over the 3 months, but formulaicity only increased for the experimental group. A potential issue with the studies of Wood and Chen is that the measures of fluency and formulaicity may be conflated. This can happen in two ways: firstly, in identifying formulaic expressions through listening to audio, there is the chance that cues related to fluency are used in the identification process. Secondly, since FRR (Formula to Run Ratio) and the fluency measure MLR (the Mean Length of Runs) both depend on the number of runs in a speech sample, they are not independent measures.

A number of studies have utilised spoken learner corpora to investigate the relationship between multiword expressions (based on n-gram frequency and strength of association) and indicators of proficiency. For example, Kyle and Crossley (2015) found that use of multiword expressions positively correlated with holistic oral proficiency scores (based on human rating). However, this could be because human raters unconsciously favour spoken output that contains multiword expressions. Garner and Crossley (2018) found that L2 speakers produce a gradually increasing proportion of high-frequency bigrams and trigrams as their language proficiency increases over time.

In general, the research does suggest a relationship between the use of formulaic expressions and oral proficiency for L2 speakers. However, in the studies mentioned, formulaicity is identified via either diagnostic criteria (in the longitudinal studies) or frequency measures (in the corpora studies). In the former case, the challenge is that fluency is not fully separated from the formulaicity measure so any relationship may be circular. In the latter case, frequency may not be a good indicator of formulaicity on its own, and even if there is a relationship between use of MWEs and fluency, this may be a due to an associated variable such as proficiency rather than a direct causal relationship. It seems that, rather than the number of externally defined formulaic expressions an L2 speaker uses, it is the way that they are used that is the crucial factor in contributing to fluency. This then is a good argument for focussing on internal factors (processing advantage, fluency, automaticity) in defining and identifying unitary expressions as formulaic, as suggested in Section 2.2.5. In this way, increased fluency and proceduralisation of speech as a result of using more formulaic expressions is a given, while the attainment of formulaicity in any given expression becomes the focus. Nevertheless, expressions considered formulaic in the language are still important because these are often (but not always) examples of expressions that the L2 speaker would do well to acquire as internally formulaic.

2.3.3. *What happens when L2 speakers memorise targeted expressions?*

Because of the potential benefits and challenges of acquisition highlighted above, Meunier (2012) suggests that the specific and targeted learning of formulaic language by L2 learners is necessary. Boers & Lindstromberg (2012) have reviewed a number of experimental and intervention studies on formulaic expressions in a second language (L2). Based on these, they propose some interventions to help L2 speakers acquire more formulaic expressions, including drawing learners' attention to formulaic expressions as they are encountered, and stimulating lookups in dictionaries. They also highlight that particular semantic or prosodic features (such as alliteration or assonance) can help learners commit certain sequences to memory. Such features will be explored further in Section 2.5.

However, these approaches are less suitable for sequences without such stand-out features. As highlighted earlier, many formulaic expressions useful for an individual to learn are canonical. This section will therefore focus on some studies that have targeted the learning of useful multiword expressions for L2 speakers without necessarily relying on features pertaining to specific types of formulaic expression in the language. These are useful to consider both from a methodological perspective and also for the insights they give on the acquisition and processing of the targeted sequences in the individual L2 speaker.

2.3.3.1. Wood (2009)

Wood (2009) described the focussed instruction of formulaic expressions in a case study focussing on an individual Japanese student. The student was part of a group taking a weekly fluency class over six weeks. The class involved a variety of interventions, including: highlighting formulaic expressions in native speaker text; shadowing (where students read aloud a written text while simultaneously listening to it being spoken); dictogloss (where students listen to a text, jot down notes, and then try to re-construct it in groups); and, mingle jigsaw (where each student memorises a single sequence and then goes round a class sharing with others, while learning the ones from other students). Wood reported a significant increase in the use of formulaic expressions by the student, many of which were from the examples presented to her, and a corresponding increase in the fluency of her speech when giving a monologue on personal topics. In this case, the focus was on the general use of formulaic expressions identified via diagnostic criteria. While there was a phonological criterion related to the delivery of the expression, it was not a necessary one, meaning that not all expressions deemed formulaic were necessarily fluently delivered. The study showed that the interventions increased the general use of formulaic expressions. However, it did not focus on the acquisition of specific sequences and the recall, accuracy and fluency of their subsequent usage.

2.3.3.2. Wray (2004)

This study by Wray (2004) follows a single individual learning Welsh from scratch for a specific purpose (doing a presentation in a TV programme) under controlled learning and production circumstances. The advantage of this study is that the learner had little previous knowledge of the language, while the specific pragmatic aims and time constraints of the situation meant that learning a script as a series of formulaic expressions was the most appropriate approach. Another advantage was that much of the learning (everything except her private study time) was filmed and made available for analysis. The situation was to present a cookery demonstration in

Welsh as part of the TV programme ‘Welsh in a Week’ on the Welsh language television channel S4C.

The individual, Margaret, prepared for the demonstration with three days of tutorials, each of which was repeated three times (for filming purposes). The challenge itself was then performed twice (P1 and P2) in front of a live audience. For the study, there was a follow-up after 5 months (R5) and 9 months (R9) in which recall of the script was attempted. In all, 60 different utterances were learnt, with six repetitions of each plus the four different attempts at recall. The percentage of successfully recalled (completely error-free) items repeated (of those attempted) was 79% at P1 but reduced significantly for P2 and again for the follow-up recalls. However, there was not a significant difference in the success or recall of items between the two follow-up (R5 and R9) despite them being 4 months apart. During that time, one section of the challenge was broadcast giving the opportunity for Margaret to review her performance. However, this review did not appear to affect recall and success at R9. Wray suggests that any loss of accuracy had plateaued at R5 with the best-learned items firmly placed in memory and the others forgotten. It may also be that successful retrieval at R5 facilitated recall of the same items at R9. (See Section 2.5.3.3 for more on the importance of retrieval in memorisation).

The study also offered the opportunity to explore whether errors and pauses are more likely to fall at the boundaries of formulaic expressions than within them. In this case, formulaic expressions were defined purely on the basis of how they were learned, rather than on any formal or grammatical basis. Some of the 60 items were not learnt whole, being broken into smaller chunks. So, formulaic expression boundaries were defined on the basis of input; specifically, if there was a sub-sequence within an item which was also part of another item, that sub-sequence was taken to be formulaic itself. Using the recall data from the challenges and reviews, the percentage of sequence boundaries containing pauses (36.3%) was significantly higher than the percentage of possible-non-boundary points that contained pauses (7.4%). A similar result was found for errors suggesting that pauses and errors are more likely to fall at the boundaries of formulaic expressions. That said, some pauses and errors did fall within the expressions. This confirms the point that ‘learnt as a single chunk’ does not necessarily imply ‘fluently and accurately reproduced’ and these two potential criteria for formulaicity are not the same.

While the expressions were learnt as a whole without much scope for analysis (since the learner was a beginner and the material to memorise was well beyond her knowledge of the language grammar and lexis), errors were nevertheless introduced. In some cases these suggested internal editing. For example, in the case of front-mutation (a pattern within Welsh involving phonological changes to the beginnings of some words based on complex rules), errors seemed to be induced when multiple forms of a word had been encountered in different chunks of input, but not when just one form was ever used in the entire enterprise or when the form variation was specifically taught. Other morphological errors such as the loss of grammatical particles occurred, even after initial success. Suggested explanations for this were that formulaic material is phonologically under-specific in general (in that expressions can be reproduced convincingly without all the unstressed elements being fully accurate or present). Wray speculates that certain segments of a new utterance may be informally matched by the learner to elements of its meaning, while other segments may have no apparent semantic function. Unstressed forms within those unaligned segments may be particularly susceptible to omission. Other examples suggesting a degree of internal editing (or segmentation) were also present. For example, Margaret sometime replaced a Welsh word within a learnt sequence with the English version or with a Welsh synonym.

This was a single case study, but it did demonstrate that an individual speaker is able to effectively learn and deliver targeted formulaic language for a specialised function and occasion. It also provided a rich source of information due to the highly controlled nature of the data collection and content, and the lack of interference with existing knowledge of Welsh.

2.3.3.3. Wray and Fitzpatrick (2008)

To understand better how L2 speakers are able to memorise specific spoken sentences and use these later in real conversations, Wray and Fitzpatrick conducted a study with six intermediate/advanced learners of English (three Japanese and three Chinese) based in the UK (Fitzpatrick & Wray, 2006; Wray & Fitzpatrick, 2008). The aim was to explore individual differences in L2 memorisation (Fitzpatrick & Wray, 2006) and to analyse the kinds of deviations that learners introduced when trying to reproduce the target utterances (Wray & Fitzpatrick 2008). In their study, the participants identified messages they individually wanted to convey in specific future situations and were then given model utterances (MU) to memorise. They then used these first in a practice performance (PP) and later in a real conversation (RP). In total there were 21 conversation cycles with around 10-12 target utterances in each cycle. The utterances had a mean length of 10.5 words. The participants' performances were measured according to the proportion of target utterances they attempted and the closeness of their attempts to the original. In addition, the kinds of deviations they made were analysed in detail. Deviations were categorised as Grammatical, Lexical or Phrasal, and assessed as to whether they were 'nativelike' or 'non-nativelike' changes to the original material.

They found that the six participants varied considerably in the proportion of targets they attempted and the closeness of these to the original. However, there was no clear relationship between these variables and the measures of proficiency and language-learning aptitude taken beforehand. Analysis of deviations by category showed that 43% were Phrasal, 33% were Lexical and 24% were Grammatical. Phrasal and lexical deviations were generally nativelike (i.e. a deviation that a native speaker might also make if they had forgotten, or chose to change, the model*)* whereas grammatical deviations were non-nativelike. This reflects the relative lack of flexibility of morphological choices compared to those of lexical and phrasal items.

Wray and Fitzpatrick suggest that the overall lack of correlation between the various deviation measures and proficiency may be due to individual differences in the propensity to take risks during memorisation (i.e. pay less attention to the detail) or in attitudes to making changes during reproduction (as some participants indicated that they changed utterances deliberately). They argued that risk-taking may seem a legitimate strategy for more proficient speakers because they are more able to complete partially memorised utterances appropriately using their knowledge of English. However, some speakers (e.g. those with a much higher receptive than productive vocabulary) may be over-optimistic in their ability to do this, therefore producing more non-nativelike deviations. Wray and Fitzpatrick (2008) also found that the model utterances most likely to be used in the real performance "used pronominal forms to avoid over-specificity, and did not begin with an adjunct or conjunction that might constrain its cohesion to previous text" (p.13). In other words, memorised utterances were most useful (in terms of fitness for purpose) when cohesion was left implicit.

The study highlighted that individual L2 learners vary in the way that they memorise utterances and in the way they reproduce them, and that this is not necessarily related to proficiency. During practice performances, accuracy (especially morphological components) and fluency were often compromised. This suggests a tendency to segment or analyse utterances for learning and then reconstruct them for production. Interviews with participants also suggested that focussing on key component words and using these to build the utterances was a common strategy.

2.4. What features might identify formulaicity in speech?

When investigating the processing of formulaic expressions, it is clearly important to have a way of determining which word-strings produced in speech are internally formulaic for the speaker. The discussion in this section looks at research related to this identification process. It starts with the use of diagnostic criteria as a general approach to identifying formulaicity in individual speech and then looks in more detail at phonological aspects of different types of formulaic expression. This includes the positioning of formulaic expressions within intonation units and the use of fluency. It concludes by exploring an approach suggested by Myles and Cordier (2017) which combines a number of the features discussed.

2.4.1. *Identifying formulaicity using diagnostic criteria*

Due to the variety of possible identification methods and the fallibility of each, Read and Nation (2004, p. 33) argue that no one method is appropriate on its own and a mixture of factors should therefore be considered. In keeping with this, such factors have been developed into diagnostic criteria by a number of researchers (Wood, 2007, 2009; Wray, 2008a; Wray & Namba, 2003) and used to identify formulaic expressions in the speech of experimental participants. For example, Wray (2008a, pp.116-121) developed a set of 11 diagnostic criteria, adapted from a study by Wray and Namba (2003). These are based on a detailed analysis of the various possible features of formulaic expressions, and are summarised in Table 2.1. The criteria provide a list of possible features to check to see if an expression is likely to be formulaic. Since there is a diversity of types of formulaic expression, it is not the case that every criterion needs to be satisfied, and absence of evidence for one criterion does not preclude the expression being formulaic. As can be seen, the criteria include features associated with the delivery of the expression (F, G), features related to the form, function or meaning of the expressions (A, B, C, D, I, J) and features related to the specific speaker and their experience with the expression (E, H, K).

Table 2.1: Eleven diagnostic criteria for assessing formulaicity

By my judgement A. there is something grammatically unusual about this word string B. part or all of the word string lacks semantic transparency C. this word string is associated with a specific situation and/or register D. the word string as a whole performs a function in communication or discourse other than, or in addition to, conveying the meaning of the words themselves E. this precise formulation is the one most commonly used by this speaker when conveying this idea F. the speaker has accompanied this word string with an action or phonological pattern that gives it special status as a unit, and/or the speaker is repeating something just heard or read G. the speaker, or someone else, has marked this word string grammatically or lexically, in a way that gives it special status as a unit H. based on direct evidence or my intuition, there is a greater than chance-level probability that the speaker will have encountered the precise formulation before in communication from other people I. although the word string is novel, it is a clear derivation, deliberate or otherwise, of something that can be demonstrated to be formulaic in its own right J. this word string is formulaic, but has been unintentionally applied inappropriately K. the word string contains linguistic material this is too sophisticated, or not sophisticated enough, to match the speaker's general grammatical and lexical competence

From 'Eleven diagnostic criteria for assessing intuitive judgements about formulaicity' (Wray, 2008a, pp.116-121)

It is important to note that the intention behind such criteria is to help the researcher pin down the reasons for what they already believe about an expression and to reflect the range of ways in which intuition about what is formulaic is used. They explicitly involve the judgement of the researcher and, in some cases, require particular knowledge of the speaker's level, experience and patterns of use if this is available. For example, to apply criterion K, the researcher needs to know the speaker's general level of proficiency; to apply criterion H requires some understanding of the speaker's learning experience. Since, as Sinclair (1991) notes, subjective judgement can be highly fallible when applied to a native-speaker's perceptions of language use, they should not be relied upon to identify formulaic expressions outside the specific research situations for which they were designed. However, such criteria are very useful in highlighting features that other researchers have explored and which may be developed for identifying formulaicity in a more objective way. For example, the criteria relating to form, meaning and function are helpful in establishing the string as a holistic unit, in the sense that it is non-canonical and unlikely to have been constructed on-line from constituent words. Criterion F is specifically related to the speaker's delivery and may be relevant to how the individual processes the expression. This aspect of identification is considered further in the next section.

2.4.2. ***Phonological features of formulaic expressions***

2.4.2.1. Phonological features of idioms

Van Lancker and Canter (1981) explored the ability of native English speakers to recognise idiomatic and literal meanings of ditropically ambiguous sentences (e.g. *It broke the ice* or *David spilled the beans*) when they heard recordings of them spoken out loud. They found that participants could reliably distinguish between the meanings, and that idiomatic versions tended to have fewer pauses and shorter duration that their literal counterparts. In this experiment however, the speakers who read the sentences out loud were asked to convey the contrasting meanings as clearly as they could, thereby emphasising the differences. In versions of the experiments where the sentences were produced naturally (read from within disambiguating paragraphs and then extracted), listeners were not able to distinguish them. There have been other experiments where the idioms have been produced naturally. For example, Yang, Ahn, and Van Lancker Sidtis (2015) showed that Korean speakers were able to distinguish between literal and non-literal versions of Korean idioms based on prosodic cues. While native speakers may be able to notice such cues in familiar idioms, it appears to be difficult for non-native speakers. For example, using a similar approach to her earlier study, Van Lancker Sidtis (2003) found that advanced L2 students of English performed no better than chance in distinguishing between literal and idiomatic versions of ditropic expressions.

Further work on the prosodic patterns of idioms has been done by Ashby (2006). He identified a number of different possible prosodic patterns for idioms. For example, one type has an accentual patterns that is the same as that of the literal version (e.g. *have a chip on one's shoulder*, *have eyes in the back of one's head*) and another type has an accentual pattern different from the corresponding literal expression (e.g. *pour down, be rolling in money*). Although the patterns appear to be arbitrary and fixed (in the sense that changes to the pattern lose or disrupt the idiomatic meaning), this is not always the case. For example, *he has eyes in back of his head* retains its idiomatic meaning, as does *she has eyes in the back of her head too* (if describing a second person with the same uncanny ability to notice what's going on). Ashby argues that the key feature for idioms is that they cannot change the focus within the non-compositional part of the idiom. For example, shifting focus within the non-compositional part of the previous example (i.e. from *head* to *back*) would disrupt the idiomatic meaning (i.e. *she has eyes in the* *back* *of her head*). While this appears a useful way of understanding the prosodic patterns of idioms, it does not really offer an applicable way of identifying idiom usage in speech (due to the variety of patterns and their seemingly arbitrary nature). However, it could provide a diagnostic criterion for checking that known idioms were being used formulaically by L2 speakers.

2.4.2.2. Prosody and intonation units

Lin (2010, 2013) has explored the prosodic features of formulaic expressions more generally, especially in regard to their relationship with intonation units (IUs)[6] . She suggests that formulaic expressions are aligned to IUs and have restricted accentuation patterns (Lin, 2013; Lin & Adolphs, 2009). Lin (2013) analysed a set of formulaic expressions extracted from the prosodically annotated IBM/Lancaster Spoken English Corpus. This corpus contains approximately 53,000 words of British spoken English, mainly taken from radio broadcasts dating from the mid1980s. Formulaic expressions were defined as prefabricated multiword sequences stored holistically in the mental lexicon (following Wray, 2002) and extracted from the Corpus using the tool 'Wmatrix' (Rayson, 2009) which utilises a large number of formulaic expression templates. Short expressions (of two words), proper nouns and time and number expressions were excluded, leaving 218 expression types (339 tokens) in the analysis. For each expression, its position within the IU and the location of the nucleus (the accented syllable with the highest prominence in the unit) within the expression were identified.

The study confirmed previous research in that most formulaic expressions did indeed occur within IUs. Analysing further, Lin found that around half the formulaic expressions occurred at the end of the IU (including those that occupied the whole unit). For these, about 70% received the nucleus on the final lexical word of the expression, mirroring the typical ("broad focus") situation in general (non-formulaic) speech. In general speech, the nucleus is thought to fall outside the final lexical word

[6] The intonation unit (or tone unit) is a prosodic unit in natural speech, defined as a speech segment that falls into a single coherent intonation contour (Chafe, 1987). Although there are different approaches, intonation units are generally identified according to prosodic features (such as pitch level and movement on prominent syllables) within the unit, and pauses, pitch re-sets, and syllable lengthening around the boundaries. IUs are also linked to syntactic units, with the clause being the prototypical intonation unit type from which most other types are derived (Chafe, 1987, p. 38). Accordingly, before major syntactic boundaries, final syllable of syntactic units tend to be lengthened, and there is a global decline in pitch (Cruttenden, 1997; Wichmann, House, & Rietveld, 2000).

in the IU only 5-15% of the time and this "narrow focus" is thought to provide emphasis or contrast in speech (Crystal, 1969). So, the 30% of formulaic expressions in IU ending positions which did not have the nucleus at the end represents a high figure, particularly since these examples did not involve emphasis or contrast. One possible way to account for this finding is that the formulaic expression is behaving as a single word-like unit and, as such, the nucleus falling anywhere within that unit (or not) is in line with the expected broad focus case. For example, in Lin's data, *if you like* was an IU ending expression with no accent (the nucleus fell on the preceding word). This follows the expected pattern provided the expression (a 'semantically empty' filler) is treated as a non-lexical unit at the end of the IU. In other cases, the position of the nucleus within the formulaic expression is determined by the nature of that expression (as accent within a word is determined by the nature of the word). For example, IU ending cases such as *more or less* or *to a certain extent* have stress on words that are prominent within the expression (such as words expressing degree). Formulaic frames such as *far as X is concerned* have the nucleus on the flexible slot variable (rather than the final lexical word) since this represents the key (least predictable) information in the unit.

The word-like nature of the formulaic expression therefore may have an influence on the prosody within the IU and so could potentially provide clues for identifying formulaicity in speech. For example, the absence of a nucleus in the final lexical word of the IU might indicate that a formulaic expression is present. However, it is important to note that other factors affect canonical prosody in speech too, such as emphasis, contrast or whether something is given rather than new. For example, compare *I told him it was a fait accompli* and But *it wasn't a fait accompli*. In the first case, the prosody is canonical and in the second it is not. This illustrates the complexity in using prosodic clues in the identification of formulaicity; firstly, the prosody of the expression may change as a result of discourse factors, and secondly, not every use of formulaic expressions within an IU is non-canonical.

One aspect of the relationship between IUs and formulaic expressions which can be applied more easily however, is the finding that formulaic expressions tend to fall within IUs. This offers the possibility of using IU boundaries as a means for dividing up text as a first step in analysis. Formulaic expressions can then be searched for within the IU segmented text and clues from prosody may help this process (with the caveats mentioned above). Such an approach would require a clear indication of how to segment text into IU in practice, as there are different views of this in the literature.

For example, in the Discourse Intonation approach (Brazil, 1997), IUs are identified according to the tone on prominent syllables (the onset and tone units). However, other researchers (Barth-Weingarten, 2011) suggest focusing instead on the boundaries. In either case, there is a recognition that these boundaries may be 'fuzzy' (i.e. subject to different interpretation by different transcribers). Indeed, in the study by Lin (2013), the two expert phonologists analysing the spoken English corpus had only 69% agreement on where IU boundaries should be placed. So, in practical terms, the criteria for specifying boundaries would need to be laid out very clearly by the researcher.

2.4.3. *What is the relationship between pausing and formulaic delivery?*

While the use of prosodic features and IU boundaries may be possible for identification, another feature of delivery that is perhaps more easily observable is the presence or absence of pauses. As noted in Section 2.3.3.2, the study by Wray (2004) suggested that pauses tend to occur at boundaries of formulaicity rather than within them. This general idea is supported indirectly by a variety of research. For example, studies have found that pause length reflects hierarchical organisation in phrases (Gee & Grosjean, 1983), that pauses are more likely to occur in newly organised propositional speech than old automatic speech (Goldman-Eisler, 1964) and that pauses are more likely to occur at syntactic boundaries than within phrases (Goldman-Eisler, 1972). For formulaic expressions, it is commonly thought that pauses do not generally fall within the expression. For example, Dahlmann (2009) studied the distribution of pauses around automatically extracted multiword units in her two corpora, the English Native Speaker Interview Corpus (ENSIC) for adult native speaker speech and the Nottingham International Corpus of Learner English (NICLE) for second/foreign language learner speech. Drawing on research by Raupach (1984) and Wray (2004), she makes the assumption that pauses should not normally be found within multiword units. This concept then forms the main basis of her phonological criterion when she attempts to develop an inventory of spoken formulaic expressions especially for second/foreign language learners. She extracted lists of multiword units of various lengths automatically from her two corpora and then used the phonological criterion as a filter to select psycholinguistically valid formulaic expressions for inclusion in the inventory. Tavakoli (2010) in a comparative study of 40 native and 40 L2 speakers narrating picture stories found that L2 speakers tended to pause more in mid-clause than native speakers did, but did not tend to pause within formulaic expressions. However, the way in which formulaic expression were identified was not specified.

Erman (2007) explored the distribution of ‘production pauses’ within “prefabricated and non-prefabricated” language extracted from two corpora of spoken English. Production pauses are those considered to involve cognitive processing effort in lexical retrieval. They are involuntary and interrupt the flow of speech mid-structure. In the study, they were identified on the basis that they occurred either within the boundaries of a phrase, after a function word or before a content word. The corpora were the Bergen Corpus of London Teenage Language (COLT) and the London-Lund Corpus of Spoken English (LLC) and the material study consisted of 60,000 and 30,000 words from each respectively. Prefabs were taken to be multiword units with conventional meanings, and the main identification criterion was ‘restricted exchangeability’ (i.e. there is a word in the prefab that cannot be replaced with a synonym without causing a change in meaning or function). Prefabs were considered fixed if the constituent words were fixed or had restricted variability. The theoretical assertion was that prefabs are stored as a single unit in declarative memory and that lexical retrieval of the unit includes retrieval of all constituent words within the unit automatically. On the other hand, words in open slot positions in the text are thought to require cognitive “focussing” to choose and retrieve, which is effortful. Any position in non-prefabricated language was taken to be open along with any variable slot in a prefabricated frame structure. Erman hypothesised that production pauses would occur much more in open slot positions than restricted ones and this is what was found clearly in the results of her analysis. The results were taken to support the idea that prefabs are stored and processed as a whole and that production (involuntary or planning) pauses do not occur within formulaic expressions.

In addition, pauses before prefabricated material were shorter than before ‘creative language’. Erman suggested that pause duration was an indicator of cognitive processing load which was reduced when choosing a prefabricated expression compared to preparing for the construction of a novel expression. Before prefabricated expressions, pauses can occur for reasons other than processing load. For example, a short pause could be a citation marker (e.g. consider the likely difference between: ‘He saw twelve angry men outside the court house’ and ‘He saw ‘Twelve Angry Men’ at the cinema’). Such a usage satisfies criterion F in the list of diagnostic criteria for formulaic expressions given by Wray (2008a) in Section 2.4.1. Thus, pause duration could help distinguish between pauses that reflect the planning of subsequent constructed text and those that demarcate what is formulaic for semantic reasons.

While the cause or meaning of the pause may not always be the same, the research overall does suggest that pauses, in general, do not fall within formulaic expressions. However, an important point to note is that in Erman's study, the number of production pauses overall was actually very small (9 per 1000). So, in most cases, there is no (production) pausing at all, even in non-prefabricated expressions. This is consistent with the idea that planning can also take place at the same time that speech is being produced (Ford & Holmes, 1978) and usually does not require pauses. Therefore, the absence of pauses alone may not be sufficient to identify formulaic expressions in the speech stream, and other criteria would be important to establish the unitary nature of the expression. However, it does seem a valuable tool for segmenting utterances to narrow down potential formulaicity.

2.4.4. *Phonological coherence and the hierarchical approach*

Based on the research considered so far, formulaic expressions in speech seem likely to display a combination of shorter duration, phonetic reduction, a unitary intonation pattern and the absence of pauses. Myles and Cordier (2017) use the term 'phonological coherence' to describe such features and suggest that such coherence may be a good indicator of the processing advantage required for speaker-internal formulaic expressions. They introduce a method of identification based on the use of hierarchical conditions (p.19). In this, the first necessary criterion is phonological coherence, operationalised as the absence of dysfluency. In their method, indicators of dysfluency within a sequence were defined as any of the following: a pause longer than 0.2s; fillers (e.g. *er, um*); reformulations or repetitions of words; elongated syllables (>0.4s). Recognising that not all fluent runs represent processing units (Raupach, 1984), a second necessary criterion is that the sequence possesses a "holistic quality". That is, the potential sequence should display at least one typical condition showing a holistic dimension: either semantic/functional unity or an indication that the sequence was learnt holistically. A third, graded criterion based on frequency of occurrence in the sample is included to add to the reliability of the identification. Myles and Cordier give an example of applying these criteria on the samples of speech of advanced learners of French. Using the methodology, they found that formulaic expressions represented 27.8% of the speech, a higher figure than previous research suggests for formulaicity in L2 learners (as mentioned in Section 2.3.1).

The hierarchical application of phonological coherence followed by an indicator of a sequence's holisticity is a practical method closely aligned to the theoretical definition of internal formulaicity. Using absence of dysfluency for the first condition provides a simple way of segmenting the speech for further analysis and reducing the number of potential sequences that need to be assessed on their unitary status. However, it is worth noting that there may be some occasions when pauses do occur within formulaic expressions. One possible case is when a speaker attempts to adapt or generalise an already established formulaic expression. As discussed in Section 2.2.3.2, there is a view of language acquisition (Ellis, 2012; Ellis & Sinclair, 1996) whereby an expression which is fixed and formulaic for a speaker starts to be analysed, leading to the acquisition of a frame or a more abstract construction. Ellis (1996, p. 98) gives the example of a fixed expression such as '*I'm sorry to keep you waiting'* being expanded to '*Mr Brown is sorry to have kept you waiting*', creating a model for a lexicalised sentence stem '*NP be*-tense *sorry to keep*-tense *you waiting'*. During this process, dysfluency within the established formulaic expression may arise due to the extra potential for analysis or choice as the less fixed versions of it become available to the speaker.

Furthermore, common knowledge of the formulaicity of a sequence between speaker and hearer may encourage strategic dysfluency. For example, Wray (2019) suggests that pausing within a formulaic expression in order to plan a later segment may be effective for holding ones turn because the hearer knows the formula is not finished and so will not interrupt. Gruber (2009) found that, in law courts, delivering formulaic material less fluently (i.e. with pauses) helped defendants come across as more sincere. While these causes of dysfluency may not be relevant in most cases of L2 speech, they are important to keep in mind. It is also worth noting that the means suggested for assessing the second condition (holistic unity) draw on diagnostic criteria (see Section 2.4.1) and are therefore, in some cases, prone to the challenges of subjectivity and the need to know about the speaker's previous experience of a wordstring.

2.5. What factors influence acquisition of L2 formulaic expressions?

As suggested in Section 2.3, in later stages of acquisition, L2 speakers generally do not seem to target formulaic expressions for learning even though it may be in their interest to do so.[7] Many researchers therefore advocate specific instruction in this area (e.g. Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Boers, Lindstromberg, et al., 2014; Lewis, 1997; Nattinger & DeCarrico, 1992; Nesselhauf, 2003). For example, Lewis (1997) recommends a variety of classroom activities such as summarising a text on the basis of noted word combinations, categorising word combinations according to structural or semantic criteria, and reading passages aloud with emphasis on phonological chunking. Another approach is to give L2 speakers specific targeted expressions to memorise and practise for subsequent usage. While there are mixed views on the efficacy of rote memorisation as a way of learning a second language (Fitzpatrick and Wray, 2006, pp. 25-36), there are many cases (as highlighted earlier) where it has been shown to have beneficial effects (Au & Entwhistle, 1999; Ding, 2007; Wood, 2009; Wray, 2004; Wray & Fitzpatrick, 2008).

Exploring a targeted approach further may be particularly useful in the context of the current research as it provides a controlled way to investigate the acquisition process for formulaic expressions. In particular, it ensures that the provenance of the expression (how it was learnt) is known and provides a good focus for testing the formulaicity of the expression over time based on how it is used in subsequent speech. As part of this, it is important to understand what is known about factors that may influence acquisition and use in order that they can be controlled or explored further. The research by Wray and Fitzpatrick (2008) discussed in Section 2.4 demonstrated that individual differences between L2 speakers were an important factor in how they acquired and used targeted expressions. This section focuses on some other factors including features of the sequence itself (length and structure, congruence with L1, phonological features, existing knowledge) and aspects of the way an expression is memorised (types of processing, rehearsal, retrieval, generative or error-free approached, noticing, analytic or holistic approaches).

2.5.1. *What does memorisation 'success' mean?*

In order to understand what factors influence the memorisation of multiword strings, it is first helpful to consider what might be considered the goals of successful memorisation in terms of the knowledge and proficiency attained. In this section, relevant factors related to vocabulary research on single words will first be considered and then applied, where appropriate, to formulaic expressions.

---

[7] For early stages of acquisition, however, Myles (2004) suggests that formulaic expressions and unanalysed lexical chunks can form the majority of the L2 speaker's language.

A prevailing view in vocabulary research is that vocabulary knowledge is multi-dimensional with a number of different components contributing to it. Meara (2005a) and Daller, Milton, and Treffers-Daller (2007) suggest that knowledge of individual lexical items includes aspects of depth and fluency/accessibility. Depth has been described in terms of the structured, lexical network that makes up a learner's mental lexicon (Meara, 2005a) or in terms of various aspects of form, meaning and usage that may be associated with a particular lexical unit (Nation, 2001, p. 27). Some researchers contend that the knowledge of the form–meaning relation is the most important component of word knowledge (Laufer & Girsai, 2008) and suggest that this defines what it means to 'know' a word (Kersten, 2010). Vocabulary knowledge could therefore be summarised as the ability to retrieve the meaning of a given lexical form (recognition), and the ability to access the form for a given concept, idea or meaning (recall and accuracy). Depth of knowledge then allows the speaker to do these things consistently and accurately in the face of the many complexities around how words are used, and how they relate to other, semantically similar, words.

Regarding the fluency/accessibility aspect of knowledge, Daller et al. (2007, p. 8), using the term 'lexical fluency', state that it is intended to define "how readily and automatically a learner is able to use the words they know and the information they have on the use of these words". This may involve the speed and accuracy with which word forms can be recognised receptively or retrieved for expressing targeted meanings productively (in speech or writing). Henriksen (1999, p. 314) also describes an aspect of knowledge called *receptive to productive use ability*, which she argues is a continuum, describing "levels of access or use ability".

Insofar as formulaic expressions are thought to have their own holistic entries in the lexicon, this conceptualisation of vocabulary knowledge in terms of depth and fluency/accessibility may easily be extended. Although not addressed explicitly in the literature, the implication is that depth of knowledge for a formulaic expression would relate to knowledge of its own semantic and distributional characteristics as a unit. Similarly, fluency and accessibility of a formulaic expression would relate to its ease of access as a unit and the capacity to produce it fluently in its semantic context. Further aspects of knowledge may also be relevant. For example, research by Sprenger et al. (2006) suggests that associated knowledge related to component words is also linked to the formulaic expression. (This is explored in more depth in Section 8.3.3). In addition, as discussed in Section 2.4, the internal fluency or phonological coherence of the expression is an important and, in the approach of Myles and Cordier (2017), necessary feature of formulaicity.

Overall, 'successful learning' of a targeted (potentially) formulaic expression could be considered in terms of a variety of features associated with depth of knowledge (recognition of meaning/function; ability to recall; accuracy of form) and fluency/accessibility (automaticity of response; internal fluency). In terms of the speaker successfully acquiring the expression as a formulaic unit, the speed and fluency of processing and the degree of holisticity of the expression as a unit are particularly relevant, along with establishing a clear connection between the holistic form and its meaning or function. These are the aspects that will be the main focus of the current study. Nevertheless, it is important to acknowledge that other aspects of depth (such as understanding specific usages and restrictions or establishing a wide network of associated links) are clearly vital aspects of learning and using formulaic expressions in the longer term.

2.5.2. ***How do features of target sequences affect memorisation?***

Vocabulary research suggests that a number of factors influence how easily a word is learned by L2 speakers. For example Schmitt and McCarthy (1997) suggest that these include pronounceability, length, morphology, synformy[8], grammar, and semantic features of the word. Insofar as formulaic expressions are considered word-like, it is useful to consider how factors such as these may also affect the memorisation of sequences. In addition, complexities and variations associated with the presence of multiple component words in a sequence, may also play a part. This sub-section therefore briefly reviews research on a number of features that might be particularly relevant to the memorisation of formulaic expressions: length and structure, congruence with L1, phonological features, and existing knowledge of component words.

2.5.2.1. Sequence length and structure

Regarding sequence length, a natural intuition would be that longer sequences are in general more difficult to memorise and retain for subsequent recall and reproduction. Although there is little research on this for formulaic expressions, the study by Fitzpatrick and Wray (2006) investigating the learning of targeted sequences by L2

---

[8] Synformy is the phenomenon of form similarity between words in terms of sound, script or morphology e.g. *conceal* and *council; embrace* and *embarrass*

speakers generally found this to be the case. Support for the general principle also comes from traditional memory research. For example, Baddeley, Thomson, & Buchanan (1975) suggested that the immediate recall of verbal lists is limited by the overall length of the material to be recalled. This may be linked to the idea that storage in working memory (which is required to rehearse and repeat a sequence of words in order to store it in long-term memory) is limited (Cowan, 2010). Longer material would therefore be expected to be more greatly impacted by this limit.

However, as mentioned in Section 2.2.3, the chunking of information is also thought to be an important factor in the memorising of lexical material (Ellis, 1996). Chunking is the collecting together of individual words within a longer sequence so that they are treated as a single unit.[9] This complicates the effect of sequence length on learning because any effect also depends on the extent to which a learner is able to chunk components. More recent memory research by Chen and Cowan (2005) has shown that both overall length and number of chunks affect the ability to retain and recall verbal lists. They did a series of experiments where participants recalled lists of previously learned words (all short, frequent and with concrete meaning). For some lists, they taught paired associations between words to create two-word chunks. In free recall tasks, they found that chunking made a difference up to a limit. For example, 12-word lists consisting of six pairs were recalled much better than 12 single words (and at a similar level to six single words). On the other hand, for serially-ordered recall tasks, they found that the overall length of the list was the key factor. For example, in eight-word lists, four pairs had a similar recall to eight single words. Since multiword sequences require serial recall of component words, it suggests that overall sequence length is likely to be an important factor irrespective of any internal chunking of the components.

However, chunking is clearly highly relevant to the learning of formulaic expressions. For example, McCauley & Christiansen (2015) have shown that individual chunking ability predicts on-line sentence processing. They argue that the limit within working memory means that chunking is a necessary part of the ability both to produce fluent speech and to process and learn from it. Ultimately, the holistic view of formulaic expressions suggests they are processed in working memory as a single chunk. However, when memorising a new target sequence, some degree of intermediate

[9] Chunking is, of course, closely related to formulaicity in that a formulaic expression, as defined in this thesis, represents a fully chunked expression. Here, however, the focus is on the chunking of groups of words within a target expression as part of the process of acquiring formulaicity in that target.

chunking may also be required to hold the sequence in working memory for processing. Isbilen and Christiansen (2018) argue that, since chunking information is a fundamental human process, language may have evolved to favour linguistic structures that are more easily chunked. Thus, expressions which have become formulaic in the language may be those which are especially 'chunkable', enabling them to be learnt and produced as single holistic units. The authors do not specify the features that might make an expression 'chunkable'. However, one aspect of this might be that formulaic expressions are themselves easily chunked into four or fewer sub-chunks. Another could be phonological features that give the expression a certain cohesiveness as a unit.

2.5.2.2. Phonological features

Some research (Boers, Demecheleer, Coxhead, & Webb, 2014; Boers & Lindstromberg, 2005; Eyckmans & Lindstromberg, 2017) suggests that the particular lexical composition of many multi-word formulaic expressions (collocations, idioms, proverbs) may be due to a favouring of certain sound patterns such as alliteration (e.g. *wage war*; *from pillar to post*; *time will tell)* and assonance (e.g. *be left high and dry*). According to Eyckmans & Lindstromberg (2017), from 11% to 35% of English figurative idioms show either alliteration or assonance based on an analysis of learners' dictionaries of English idioms. They argue that the incidence of such patterns in formulaic expressions is far above chance and may explain the particular choice of certain words within the sequence over near synonyms. For example, the formulaic expression *time will tell* alliterates, whereas its plausible (but relatively unused) [10] alternatives *time will say/show/reveal* do not.

Eyckmans, Boers, & Lindstromberg (2016) argue that such features have strong mnemonic potential provided that, in line with the Noticing Hypothesis (Schmidt, 2001), they are recognised as such by the L2 learner. They showed this in a study which looked at the effect of both alliteration and L1 congruence (see Section 2.5.2.3) on acquisition of new lexical phrases. In the study, 65 EFL students (L1 Dutch) from a school in Belgium were first given a pre-test and then divided into three groups and each given 15 minutes to learn 32 lexical phrases. The first group was simply asked to memorise the phrases for subsequent testing. The second group was additionally instructed to mark all phrases which showed alliteration (e.g. *gain*

[10] In the iWeb corpus of 14 billion words (Davies, 2018), there are 28827 instances of *time will tell*, 413 of *time will show*, 183 of *time will reveal*, and 35 of *time will say*.

ground), the third to mark all phrases which were non-congruent (i.e. the verb used was not a direct translation from the Dutch). The phrases were in four groups, covering each combination of presence or absence of the two features. Following a short break after the learning, participants were tested on their recall (post-test) and again 10 days later (delayed test). The tests consisted of 32 gapped sentences containing the target phrases in meaningful contexts. In the pre- and delayed test participants needed to supply only the missing verb; in the post-test, the whole phrase was missing. They found that the group asked to attend to alliteration recalled significantly more phrases on the delayed test (compared to pre-test) than the no-intervention group. This was the case whether or not the phrases themselves contained alliteration. From these results, they concluded that attending to alliteration in the deliberate learning of phrases can have a positive effect on recall and may help to over-ride the influence of semantic factors on memorability.

This general feature is supported by earlier research on the learning of lists of (arbitrarily) paired associates by L1 speakers (e.g. *cat/hat*) which has shown that 'phonological similarity' between the pairs significantly can enhance memory of the items when the similarity is highlighted (Rubin, 1995). Other studies focussing on lexical phrases confirm the mnemonic potential of alliteration (Boers, Lindstromberg, et al., 2014; Lindstromberg & Boers, 2008a) and also of assonance (Lindstromberg & Boers, 2008b) provided there has been a pedagogical intervention likely to induce the noticing of phonological form. Boers, Lindstromberg, and Eyckmans (2012) suggest that the amount of intervention required is small. They found that there was an advantage (as measured on a free recall test) for alliterative word pairs (e.g. *private property*) over matched non-alliterative controls (e.g. *private collection*) when the learning process consisted of simple oral repetition and writing down without explicit attention being drawn to alliteration.

2.5.2.3. Congruence with L1

It is widely accepted that L1 knowledge has a strong influence on the learning of new L2 vocabulary and this has been shown for the learning of collocations too. In particular, it is suggested that incongruent L2 collocations (i.e. those for which there is no corresponding L1 collocation that can be easily translated in terms of the core meanings of each word component) are more difficult to process or learn. This has been shown both in production (Laufer & Waldman, 2011) and reception (Wolter & Gyllstad, 2013). A key study proposing the use of contrastive analysis and translation in the learning of L2 collocations is that of Laufer and Girsai (2008). They had three groups of EFL students read a text containing 10 single words (e.g., *relish*) and 10 collocations (e.g., *hit the headlines*), which pretesting had shown to be unknown to them. The meaning of the unfamiliar vocabulary was clarified to the students. One group was then asked to discuss the contents of the text and to debate a moral issue it raised. The second group focussed on the meaning and the form of the target vocabulary contained in the text via multiple-choice exercises and completion exercises, respectively. The third group, the contrastive analysis group, engaged with the target vocabulary through translation exercises from L2 to L1 and vice versa. One day after the treatment, all the learners were given an unannounced post-test where they were required to explain the meaning and reproduce the form of the target words and collocations, prompted by the translation equivalents. One week later, a delayed post-test was administered. The contrastive analysis group outperformed both other groups significantly in both tests. As expected, the group which had not done vocabulary-focussed exercises obtained the lowest scores. The authors conclude that a particular type of vocabulary-focussed engagement, that is, comparing and contrasting with L1, seems particularly apt for collocation learning.

It should be noted though, that the choice of post-test format (producing translations) may have favoured the group who had learned the items through a translation route. If the outcome tested had been more production oriented, such as free recall or gap-filling in the L2, then the benefits of contrastive analysis might not have been found. Indeed, this was the finding of Eyckmans et al. (2016) in their study described in 2.5.2.2. They found that attending to the congruence of lexical phrase during learning had no beneficial effect on recall. Eyckmans et al. suggest that the poor performance of the congruency-noticing group may be because the process required more cognitive effort or that it induced participants to focus on L1 phrases at expense of L2 target forms. A simpler explanation may be that, due to the challenging nature of the task, the participants in this group simply had less time to actually memorise the phrases in the 15 mins they had to do the task and learn the phrases or that their strategy of L1 transfer over-rode the one exposure they'd had to the (non-congruent) items. Also, deciding that a phrase was non-congruent (in the intervention) required learners to pay attention to the incorrect (L2 translation of the L1) phrase, so there was considerable competition for resources.

The research then is not clear about the effects of L1 congruence or the possible beneficial effects of contrastive analysis of L1 and L2 word choices. However, it does suggest that noticing a feature (such as congruence) which is thought to be associated with its memorability does not necessarily enhance learning if the time and effort required to do so detracts from other aspects of the memorisation process.

2.5.2.4. Existing knowledge

A distinctive feature of learning new sequences is that L2 speaker may already know some or all of the component words even though the sequence as a whole is new. This existing knowledge may potentially influence how the sequence is memorised, for example by helping (no new words to learn) or hindering (interference from similar expressions, lower salience of known words, etc.). Associative learning approaches (Ellis, 2006) suggest a number of effects that may be relevant to the learning of formulaic expressions. For example, an effect known as latent inhibition suggests that words which are well-known but have not in the past been pertinent cues to meaning generally (e.g. function words such as *the*, or verbs such *get* and *have)* are not easy to associate with the expression's meaning and may be overshadowed by constituent words that are more salient. However, there may be interference effects if a salient component word (e.g. *hang* in *get the hang of*) is usually associated with a different meaning. This suggests that expressions containing a word which is unfamiliar to the speaker may be easier to recall (with that word as cue to the meaning) because it is not subject to interference.[11] However, even in this case, accuracy may still suffer due overshadowing of the less salient components of the expression. Some research has been done on the learning of collocations in which new words are contained. However, this has mainly focussed on the recall of the component words rather than on the whole collocation. For example, studies by Hsu (2010) and Kasahara (2011) show that learning a collocation with a new word is as effective for recall as learning that word on its own. Boers and Lindstromberg (2012) suggest that learning the collocation containing the new word (e.g. *memory lapse*) can be helpful because the already familiar word (e.g. *memory*) may then help cue recall of its "newly learned syntagmatic partner" (e.g. *lapse*).

An associative learning effect relevant to the whole expression is a form of interference called proactive inhibition. This occurs when the meaning of an expression already has an existing cue (i.e. another word or phrase for that meaning). This effect implies that expressions for meanings that are not already represented by a word or phrase known to the speaker should be easier to learn

---

[11] An unfamiliar word may also be more easily noticed. This, according to the Noticing Hypothesis (Schmidt, 2001), should also support its recall.

(because form-meaning links are not subject to interference from the existing, non-relevant, associations). This matches general learning heuristics (e.g. Quine, 1960), such as the Mutual Exclusivity constraint (new words will more likely refer to something new), which suggest that it is easier to form a cue-response association when the response (new meaning) is as yet unlabelled.

There are also potential interference effects when learning groups of expressions. For example, it has been shown that, when learning a set of single words, the presence of semantically related words makes the learning more difficult (Waring, 1997). A study by Webb and Kagimoto (2011) suggests that a similar phenomenon occurs when learning expressions. In their study, EFL learners were asked to memorise sets of 12 unfamiliar adjective-noun collocations accompanied by their L1 translations. The learners had three minutes to learn the set and were then given a productive recall test (using the L1 translations as cues). The best recall was for the sets where several collocations had the same adjective (e.g. *deep respect*, *deep sleep*, *deep voice*, and *black sheep*, *black market*, *black eye*). Clearly, in this case, there were fewer words to learn overall, thus easing the learning burden. More significantly, the worst recall scores were for a set which contained collocations all made from different words but with adjectives that were semantically related (e.g. *narrow escape* and *slim chance*; *tall order* and *high spirits*). This suggest some interference across the semantically related words. While a drawback of the study is that it did not account for whether the component words were known or not, it does suggest that care should be taken in the selection of target sequences to be learnt in order to avoid interference effects.

2.5.2.5. Other features

There are other features and interventions that may influence the memorisation of formulaic expressions. One is associated with the ‘imageability” of a sequence. This is a semantic feature of the sequence related to how easy it is to create an image of a concrete object or scene relevant to the meaning. It has been found to positively influence the likelihood of lexical retention (Hamilton & Rajaram, 2001). In their study on the learning of lexical phrases (described in Section 2.5.2.2 above), Eyckmans et al. (2016) also investigated the effect of imageability. They found that imageability correlated with learning gains most strongly in the control group of their study, suggesting that it had an effect, but one that was overshadowed by other interventions such as noticing alliteration. Some studies have focussed on figurative idioms (Boers, Eyckmans, & Stengers, 2007), and with interventions involving how to make the sequences ‘imageable’, for example by exploring the context (e.g. boxing) of figurative idioms (e.g. *on the ropes*) (Steinel et al., 2007). These interventions are based on the idea that the meaning of many figurative idioms is connected in some sense (e.g. metaphorical or etymological) to their literal meanings. Drawing attention to such connections is thought to strengthen the form-meaning link, thus aiding recall.

Associative learning research (Ellis, 2006) has also demonstrated that expectation and surprise may be important elements in the learning of word combinations. For example, if a targeted idiom has a surprising lexical focus (e.g. *eat like a horse* – why a horse and not, say, a pig?) or unexpected grammar (e.g. *as sure as eggs is eggs* – why *is* and not *are*?), it is likely to be particularly salient for a learner. In some cases, this may be countered by interference from the expected form. It has also been shown that emotional aspects of lexical items have an impact on learning and memory (Kanazawa, 2016; Talmi & Moscovitch, 2004). For example, Kanazawa (2016) found that words with a positive emotional valence (e.g. *opportunity*) were learnt and recalled more easily by L2 speakers that than words with neutral or negative valence (e.g., *threat*). While little research has been done to explore such effects in the case of multiword sequences, it seems likely that these features would affect learning in a similar way.

2.5.2.6. Conclusion

The research suggests a wide variety of features of a target sequence that could potentially affect how easily it is learnt and recalled. In many cases, the effects are closely inter-linked with particular methods of learning (e.g. drawing attention to phonological or semantic features) and the individual’s relationship with the expression (e.g. knowledge of component words). It is important to note that the outcomes studied in most of the research are mainly associated with recall of the item (for example on being given the L1 translation as a cue). The effect of different features on the subsequent fluency or coherence of the learnt sequence has not in general been explored.

2.5.3. ***Methods of memorisation and their effects***

In the literature, there are few studies contrasting different methods for learning targeted formulaic expressions. However, there is wide variety of research looking at types of processing strategy from the perspective of memory and vocabulary research in general. These include the relative benefits of rehearsal (including both repetition and elaboration of the items to be learnt) and retrieval. These are reviewed briefly below along with some other aspects of learning that are more specifically related to formulaic expressions.

2.5.3.1. Types of processing strategy

Memory and vocabulary learning research suggests that there are different types of processing strategies for memorising lexical items and that these have different effects on which features of knowledge are learnt. For example, the levels of processing (LOP) theory (Craik & Lockhart, 1972) identifies perceptual, phonemic and semantic processing, with the last considered necessary for 'deep' processing. A variety of research has shown that semantic elaboration aids long term recall. However, more recent approaches (Roediger, Gallo, & Geraci, 2002) recognise the idea of Transfer Appropriate Processing (TAP). This suggests that different kinds of processing lead to different kinds of learning and that processing which matches the means of assessing the knowledge gained will be most effective. In other words, semantic processing may be most effective for establishing deep links to meaning, but other processing may be more effective for other types of outcome. For example, a study by Elgort et al. (2016) showed that, in the situation of reading new words in context, form-focussed processing (i.e. word-writing) had a better effect in their subsequent tests of word knowledge than meaning-focussed processing (i.e. actively deriving word meaning from context). The key point is that the tests they used involved reproducing word forms. So, the participants may have learnt the meanings either way, but the meaning-focussed approach put less focus on the detail of the form which, in this case, was the principal 'outcome' of the learning that was measured.

Research by Barcroft (2002, 2006) on processing specificity has shown that semantic, formal and mapping components are three separate and dissociable processes and that, under constraints, emphasising one may take resources from another. Subsequent studies featuring competing or different processes have tended to support this (e.g. Sommers & Barcroft, 2013). This type of processing dissociation is also implicated in research showing the benefits of task repetition (Birjandi & Ahangari, 2008; Bygate, 2001). The reasoning is that repeating the same tasks reduces the need for speakers to focus on meaning, thereby freeing them to attend to form. Related to TAP, research has also demonstrated the phenomenon of encoding specificity (Tulving & Thomson, 1973). This is where recall performance of a memorised item is enhanced by recreating the conditions of its original encoding.

The importance of these ideas to the memorisation of formulaic expressions is that the facilitative effect of any particular feature or method of learning depends (to some extent) on the particular learning outcome that is being measured. In addition, the research highlights the idea that attention is a limited resource for the learner and consideration should be taken regarding how it can be most profitably channelled through the instructional choices that are made (Schmidt, 2001). A further element of importance is that, when learning a new expression in a particular context (e.g. connected with a certain story, or embedded in a particular sentence), recall will be enhanced if the same context is given later.

2.5.3.2. Rehearsal: Repetition and Elaboration

In general, acquiring performance ability associated with fluency is associated with repeated rehearsal and meaningful practice as output (Swain, 1985). Rehearsal may be broadly taken to refer to the different mental techniques for helping learners remember information in preparation for subsequent usage. According to Baddeley (1997) two different kinds of rehearsal are maintenance rehearsal and elaborative rehearsal. He suggests that maintenance rehearsal is a means of remembering or maintaining information without any deeper encoding, and includes rote repetition. Its main purpose is to prevent forgetting rather than facilitate long-term learning (p.116). Elaborative rehearsal involves deep semantic processing (such as sentence writing), and is more likely to lead to long term memory than is maintenance rehearsal (Baddeley 1997, p.123). More broadly, elaboration refers to processing strategies that facilitate an increased evaluation of an item with respect to particular features such as its meaning or structure (orthographic or phonemic). This focussed attention is thought to help learners connect old knowledge with new knowledge. The beneficial effect of including context in the memorising of vocabulary is a particular example of this (Prince, 2012; Rodríguez & Sadoski, 2000).

As suggested above, the situation is complicated by the idea that the success of any elaborative approach depends on the type of knowledge that is being measured as a mark of that success. Craik (2002) states that encoding and retrieval are integrated in such a way that the initial processes determine the qualitative nature of the trait encoded, and deeper encodings through elaboration are associated with greater retrieval potential in an environment conducive to that kind of recall. So, while oral repetition may be thought of as a form of shallow processing in terms of semantic knowledge, it has been shown to be a key means of achieving fluency in a targeted sequence. For example, Nelson (1977, p. 151) demonstrated that repetition “at the phonemic depth of processing” facilitates memory for cued and uncued recall and for recognition. Yoshimura and MacWhinney (2007) demonstrated that oral production fluency for target sentences (using measures including read-aloud time and speech production time) shows gradual increases with the number of repetitions. Ellis and Sinclair (1996) found that the oral repetition of new expressions by L2 speakers improved the acquisition of forms, pronunciation, and aspects of productive grammatical fluency and accuracy, compared with silent controls and participants who were prevented from rehearsal by articulatory suppression. These studies also highlight the crucial role of the phonological loop (Baddeley, 2003) in working memory in facilitating the learning achieved through oral repetition.

While repetition therefore may be effective for acquisition and other aspects of learning, the way in which the repetition is conducted may also be important for recall and accuracy. Research into the effective learning processes of Chinese students (Au & Entwhistle, 1999) suggests that rote memorisation is more effective if it is accompanied by a link with meaning. Ding (2007) interviewed three winners of English speaking competitions and debate tournaments in China. These highly fluent speakers reported that memorisation (e.g. of a film script by copying a DVD) was an effective learning strategy provided they were fully attentive to an imitation process. Noice and Noice (2006) have researched on how actors are able to learn their lines. They showed (in a study on non-actors) that the strategy of ‘actively experiencing’ the line as they spoke it was more effective for accurate, fluent recall and reproduction than other memorising strategies. These kinds of repetition strategy may therefore be appropriate for achieving accurate acquisition of the complete phonological form while at the same time providing a strong automatic link to overall meaning and context.

2.5.3.3. Retrieval

A wide body of research has demonstrated that retrieval (i.e. actively generating or recalling the target item given a stimulus) is more effective for meaningful learning than just studying the association. This idea was called the Generation Effect by Hirschman and Bjork (1988) and has since been demonstrated by a variety of different studies (Karpicke, 2012; Kornell, Klein, & Rawson, 2015). Further, in experiments based on learning lists of items using flash-cards, repeated retrieval during learning has been shown to be a key factor in long-term retention (Karpicke & Roediger, 2007), and the spacing of retrieval attempts appears to be more effective than doing them all at once (Kornell, 2009). In addition, Kornell et al. (2015) found that retrieval success or failure does not matter provided that the correct version is given after the retrieval attempt. Kornell and Vaughn (2016) propose a two stage framework to explain the benefits of retrieval. The first stage is when one attempts to retrieve an answer. This causes a variety of knowledge related to the retrieval cue to be activated in a process of spreading activation in the network associated with the cue. When the answer becomes available (either by successful retrieval or by presentation of the answer), appropriate connections are strengthened, thus better linking the target to the cue. The retrieval research discussed focusses mainly on cue-response learning of associated pairs or question and answer pairs. However, it is relevant to an L2 learner memorising a new sequence that is to be recalled given an appropriate semantic cue (L1 translation or context), and suggests that spaced retrieval attempts will be beneficial in this situation too.

2.5.3.4. Generative v error-free approaches to memorisation

Generative approaches to learning are where a learner attempts to derive form (for a given meaning) or guess meaning (e.g. from context) for a new expression before being given the answer. Such approaches are thought to support the link between form and meaning and help long-term recall. However, in learning form, initial errors during generation do seem to have a long-term negative effect on accuracy. Strong (2016) showed that, in the learning of new phrasal verbs in an L2, an error-free approach was better for accurate recall than trial-and-error (a generative approach). A similar result has been shown for collocations (Stengers & Boers, 2015). In this study, an errorful (generative) learning approach using trial and error with corrective feedback was compared with errorless practice (prior consultation of exemplars). The approaches were found to be equally effective for recall. However, when students under the trial-and-error procedure supplied a wrong response in the exercises, the corrective feedback seldom had a remedial effect. In another study, Cyr and Andersen (2014) compared errorless and errorful learning. They found that recall performance for conceptual (semantic) information was highest with errorful learning, but for lexical/structural performance (including formal accuracy) errorless was better. These findings suggest that, in the learning of new sequences, generative learning may be beneficial in linking the sequence to its conceptual meaning (for example, guessing the meaning of a new sequence). However, in a trial and error approach to producing new sequences, initial errors may be detrimental to subsequent recall accuracy, despite corrective feedback after producing the error.

These results seem to contrast with the findings of Kornell et al. (2015) on corrective feedback in the previous section. Indeed, Kornell and Vaughn (2016) also cite a variety of studies in which ‘pre-testing’ (giving learners a cue-response test on something they do not yet know and therefore have to guess) seems to enhance learning. However, the errors in these studies amount to choosing the wrong recall response to a particular cue. In the generative approach discussed in this section, it is errors of form that are the focus (i.e. when the correct expression is linked to the cue, but it has errors within it).

2.5.3.5. Noticing / focussed attention

As suggested by the research reviewed in Section 2.5.2, many features of collocations and idioms (such as alliteration, assonance or figurative meaning) may have a mnemonic effect but only if the feature in question is actively ‘noticed’ by the learner. This also applies to the noticing of formulaicity in itself. In the ‘lexical approach’ of Lewis (1993, 1997), learners are systematically encouraged to notice recurring lexical chunks in authentic L2 language. The suggestion is that raising awareness of the sequences triggers acquisition through imitation. Lewis (1997; 2000) suggests that students respond positively to these type of activities, but he does not provide empirical evidence of learning gains. Boers et al. (2006) demonstrated that an instructional method that had learners notice formulaic expressions in text had a positive outcome in terms of their subsequent use of such expressions in conversation. Although it was only a small study, the increased usage was significant compared to a control group. However, only about a third of the sequences encountered were actually used. So, Boers et al. acknowledge that it is not the case that every chunk ‘noticed’ will automatically be added to one’s linguistic repertoire for active use even under favourable experimental conditions. An important aspect of conversations is that they involve attending to a variety of different things related to the message, and these may distract speakers from thinking of a target sequence or indeed may obviate the need for it

Some forms of noticing may also be thought of as a type of elaboration. For example, the interventions suggested by Boers et al. (2007) for the learning of figurative idioms involve learners trying to make sense of idiomatic meanings in terms of the original, literal meaning of the idiom. For such interventions, the relatively deep mental processing involved may be an important factor in helping the meaning and form to be remembered.

2.5.3.6. Analytical v holistic learning approaches

There is some debate in the literature on the extent to which the learning of formulaic expressions is (or should be) associated with analysis of the individual words and structure of the expression to be learnt. According to Wray (2000), a number of the classroom approaches to the learning for formulaic expressions involve such 'internal inspection' (p.463). For example, Nattinger and DeCarrico (1992) advocate pattern drills of formulaic forms followed by controlled variation (leading to an appreciation of patterns with open slots), followed by increasing variation of the form to encourage analysis of the patterns. The 'lexical approach' of Lewis (1993, 1997) views lexical phrases as word-like units, but highlights aspects of word grammar and the individual words within collocations. Willis (1990) suggests a focus on words (such as '*way*') that can be used to create different formulaic expressions (e.g. *by the way*, *by way of*).

Two possible purposes for such approaches are that they are generative (i.e. help learners acquire new frames and constructions) and that they aid deeper learning through a form of elaboration. The generative aspect may be thought to follow in part from a variety of research (e.g. Dabrowska & Lieven, 2005) that shows that children learn multiword chunks whole in their native language and then analyse them (unconsciously) to extract patterns and generalisations. Ellis (2012) suggests that this derives from an inherent ability to draw conclusions on the basis of statistical information from the input, and that this ability is retained by L2 speakers. Wray and Perkins (2000) argue that children only do this up to the age of eight and after that "the organisation of the language system becomes progressively more formulaic" (p.21). They argue that this is due to children learning in a 'socio-interactional' bubble that limits the number of interactional situations and concepts, and which is conducive to the effective analysis of the expressions they initially learn as chunks. Longitudinal research from Myles et al. (Myles et al., 1998; Myles et al., 1999) shows that, to a limited extent, beginner L2 learners may use holistic expressions for generative purposes in a similar way. Myles et al. studied the use of certain basic rote-learned expressions (e.g. *je m'appelle*... or *où habites-tu?)* by beginner level L2 students of French as a foreign language. They found that, over the two years of the study, some learners attempted to break down and adapt these expressions for other situations based on communicative needs arising in the classroom. The key point about all these studies is that useful expressions are first learnt and used holistically and only later broken down. For children and L2 learners holistic expressions may provide a relevant source of linguistic material which can be analysed once the need and the necessary linguistic tools or data are there to do so. Wray (2002) provides examples suggesting that native speakers tend to acquire formulaic expressions without analysing their component parts, and resort to analysis on a needs only basis. Clearly there is value in breaking down formulaic expressions for further learning, but in the research studied, this comes after the expression is already established for the speaker in its holistic form. Wray proposes that, for adult L2 learners, analysis happens more readily, but may be counter-productive to the aims of achieving native-like formulaicity due to the difficulty in deciding which sequences are regular (and therefore generalisable) and which are not. She suggests that the value of formulaic expressions (as whole forms that are processed more easily) lies in by-passing analysis.

While the generative potential of learning formulaic expressions through an analytic approach is not clear, such an approach may be beneficial in supporting subsequent recall and accuracy of the learned expression. Analysis of the expression may be thought of as a type of elaboration, drawing attention to the form and the component words, which may otherwise not be salient. For example, the pedagogic approaches suggested by Boers & Lindstromberg (2012) are based on the research highlighting the benefit (in terms of recall) of drawing attention to specific features of target sequences such as alliteration and assonance (although the efficacy of this for highlighting cognates is less clear). Further, regarding accuracy, Schmidt (2010) argues that attention to aspects of form is necessary in order to accurately learn and correct errors. This is the basis of his Noticing Hypothesis which drew on a number of case studies. In particular he highlights the case of 'Wes', a Japanese speaker who had lived in Hawaii for many years and learnt English incidentally through interaction. Despite having good communicative ability, he had failed to develop appropriate use of grammatical form, with consistent fossilised errors in morphology, articles, etc. Schmidt put this down to his lack of attention to the form of the expressions and a failure to compare (or even notice) errors. His inability to accurately reproduce common formulaic expressions does suggest that some attention to form is important for accurate reproduction.

In studies where L2 learners have specifically learnt memorised sequences as whole units, there is evidence that in many cases they do engage in some form of analysis even when they have not been instructed to do so. For example, Ellis and Sinclair (1996) showed that English speakers learning L2 Welsh acquired some implicit knowledge of grammatical form through the oral repetition of a series of new expressions (without explicit analysis). In the study by Wray and Fitzpatrick (2008) described in Section 2.3, participants often introduced phrasal or lexical changes to the target sequences (in some cases deliberately and in others without realising) despite being instructed to learn them exactly as given. These changes were generally ‘nativelike’ suggesting the expressions had been analysed at the component or structural level to some degree.

If L2 learners do break up targeted formulaic expressions (rather than accept them as an L1 speaker might) during learning, this may be part of a natural tendency in adult learners to abstract and generalise, or it may be a strategy to help remember the expression and support subsequent recall and accuracy. This distinction seems important in understanding the acquisition process(es) of formulaic expressions, but does not seem to be addressed in existing research. An important aim in learning a targeted sequence as formulaic (the focus of this thesis) is that it should ultimately be processed as a holistic unit, meaning that in production it is delivered fluently and automatically. The effect on this of analysing the sequence in terms of its component words is not clear. On the one hand, analysing the component words and form may encourage a reconstructive approach which is a form of ‘controlled’ processing subject to errors and dysfluency. In particular, such an approach appears counter to the idea of viewing the sequence as a single holistic unit. On the other hand, the goal of fluent automatic production does not necessarily preclude an analytic approach to the learning. For example, there may be potential mnemonic effects from breaking down a longer expression into smaller segments or words that supports both recall and fluency.

2.6. Summary and implications

This section draws together what has been discovered from the literature and the implications of those findings for the research moving forward. In particular, the summary provides the catalyst for the empirical studies that follow.

2.6.1. ***What is known about the acquisition of L2 formulaic expressions?***

As shown in this chapter, formulaic expressions have been well studied over the past 20 years and appear to be an important feature of English (and other languages). A range of research has shown that, as well as other benefits, these expressions are delivered faster and more fluently than novel expressions constructed during the speech process. Due to these benefits, it might be expected that formulaic expressions should be a target for L2 learners as a means for improving the fluency and appropriacy of their language. Indeed, research via longitudinal studies (e.g. Towell et al., 1996; Wood, 2007) and L2 corpora studies (e.g. Granger, 2019; Kyle & Crossley, 2015) suggests that increased use of formulaic language in L2 speakers is associated with higher overall fluency and increased proceduralisation of speech. Research on different kinds of pedagogic support for developing L2 formulaicity (e.g. Boers & Lindstromberg, 2012; Pellicer-Sánchez & Boers, 2019) highlights the value of drawing attention to formulaic expressions (or aspects of them) in text and of targeting specific expressions to learn. There is some evidence that such focussed attention does increase usage and awareness of in L2 learners which in turn leads to general increases in fluency of production (X. Chen, 2009; Wood, 2009). Regarding the learning of target sequences, the studies by Wray (2004) and Wray and Fitzpatrick (2008) described in Section 2.3.3 showed that useful expressions can be effectively memorised and reproduced in practice and real situations.

In these studies, the focus was on whether the learnt expressions were subsequently attempted and their accuracy of reproduction. However, that might only tell part of the story when it comes to identifying whether an expression is (or has become) internally formulaic for an individual speaker. Formulaicity in individual speakers has tended to be identified via diagnostic criteria (Wood, 2009; Wray, 2008a) which are primarily concerned with establishing the holistic nature of the expression as a semantic or functional unit. As well as being somewhat subjective and difficult to apply, diagnostic criteria may not adequately address the psycholinguistic aspects of internal formulaicity. As described in Section 2.4, research on pausing and other phonological aspects of formulaic expressions (Dahlmann & Adolphs, 2007; Erman, 2007; Lin, 2010, 2013) suggests these may be useful identifying features. Drawing on this research, the hierarchical approach of Myles and Cordier (2017) for identifying formulaic expressions in speech seems a good basis for exploring L2 acquisition. In this approach, the ‘processing advantage’ aspect of internal formulaicity is identified primarily by the fluency of the expression when it is delivered in speech, and this provides a practical and quantitative means of establishing formulaicity.

The various research explored in Section 2.5, suggests that, when it comes to acquiring targeted formulaic expressions, there are aspects of the individual, of the expression and of the memorising approach that are likely to make a difference to the process. For example, the study by Wray and Fitzpatrick (2008) showed individual differences possibly associated with attitudes to risk taking and aural memory proficiency. In addition, individual effects such as previous knowledge of component words, usefulness of the expression, chunking ability and/short-term memory span may also affect how the sequence is learnt. Aspects of the expressions themselves which may affect how they are learnt include: length, structure, category of sequence, prosodic features, imageability and congruence with L1. In addition, the effects are likely to be complex. For example, the effects of length and structure may be complicated by intra-lexical features of component words (familiarity, word class/function, relative salience). Also, many features such as alliteration, assonance, congruence and imageability of the sequence seem to depend on that feature being noticed.

The way in which an expression is memorised was also thought likely to have differential effects on the acquisition outcome. The research explored in Section 2.5.3 suggests that activities such as repetition, elaboration, retrieval and focussed attention can benefit the learning process. Further, work on transfer appropriate processing and processing specificity (e.g. Barcroft, 2002) show that learning outcomes are generally related to the type of approach taken to learning. For example, elaborative approaches focusing on the deeper processing of meaning have been shown to increase recall, while highlighting aspects of form and errorless processing may enhance accuracy. On the other hand, for achieving increased fluency and automaticity, approaches using repetition may be more conducive. A key question raised, but not fully answered by the research, concerns the extent to which analytic approaches (focussing on the component words and structure) and holistic approaches (which treat the expression as a single unit) influence the process of an expression becoming formulaic for a speaker.

2.6.2. ***Implications for current research***

The current research is concerned with the process by which any sequence that is useful for or targeted by an individual L2 speaker becomes formulaic. Of particular interest is understanding the extent to which expressions may either be ‘fused together’ to form a whole or start whole and remain so for that speaker. In either case, there is a question as to the prevalence and effect of analysing component words or sub-sequences within the expression. Associated with these ideas, is the possibility raised in the study by Wray and Fitzpatrick (2008) that on-line reconstruction of the learned is a regular feature of its recall and production, at least during some stages of the acquisition process. Memory researchers (Tulving, 2001) believe that memory recall is generally reconstructive in nature and involves rebuilding a target memory from salient components. In the case of elicited imitation (EI), where a learner repeats a sentence which has just been read out to them, the premise is that the sentence is being reconstructed by the learner (Yan, Maeda, Lv, & Ginther, 2015). This is a necessary feature for the claims that EI provides an effective means of testing grammatical knowledge.

While reconstruction may be a natural or default strategy for recall and reproduction of memorised sequences, this is not the way that psycholinguistic formulaic expressions are produced in the end. Reconstruction is associated with non-automatic processing and is likely to impact accuracy (especially of non-salient components) and fluency. Moving to psycholinguistic formulaicity through repeated reconstruction may be challenging due to the lack of sufficient repetitions and the perpetuation of errors and dysfluency. A key question then is the extent to which a segmentation stage is typical, necessary or desirable in the targeted learning of formulaic expressions by L2 learners: Does it support or detract from (e.g. induce errors or dysfluency) the learning goal? What can be done to accelerate (or by-pass) this stage?

Detailed investigation of the possible mechanics of acquisition, storage and usage of formulaic expressions in L2 speakers has been limited. Following the approach of Myles and Cordier (2017) for defining and identifying internal formulaic expressions, this research seeks to address this matter by exploring certain aspects of acquisition and use by L2 speakers. In particular, their approach recognises the importance of delivery features such as fluency and automaticity as well as recall and accuracy. The aim is to build a picture of the possible paths to formulaicity and apply these to potential psycholinguistic models of storage and processing.

Another target outcome is to determine approaches to the memorisation process that may help facilitate acquisition. As the research to date demonstrates however, there are many factors that may possibly influence acquisition and it is necessary to home in on those that are of most interest and seek to control the other factors. Thus, features such as the effects of expression length and structure on acquisition, the extent to which expressions are segmented and reconstructed and comparisons between holistic and analytical methods of learning expressions are useful areas to explore further. On the other hand, factors such as the prosodic features of expressions and individual differences, while not part of the focus of the research, will be important to be aware of and to control.

The key research questions are:

1. What are the psycholinguistic processes by which targeted expressions become formulaic for L2 speakers?
2. How do different approaches to memorisation (operational/learning process/sequence of actions) influence the acquisition process? What approaches may enhance it?

Associated with these are four sets of subsidiary questions arising from the literature review which will serve to frame the experimental research:

- What kind of formulaic expressions do L2 speakers use in their speech? Can we observe those expressions going through the process of becoming formulaic?
- What happens when L2 speakers memorise new expressions and use them later? Which expressions become formulaic and why? To what extent do L2 speakers reconstruct learnt expressions or learn them in one go?
- Can we influence the way that targeted sequences are memorised and processed from the beginning? Does analysis actually affect the way that expressions are learnt as formulaic?
- What is the end-point of formulaic acquisition? What are the paths to formulaicity and how may these be modelled cognitively?

In order to answer the first of these subsidiary questions, two empirical studies (S1 and S2) are conducted and described in Chapters 3 and 4. The first study (S1) applies the approach of Myles and Cordier (2017) to understand how some Japanese speakers of English use formulaic expressions in their speech. It is a way of exploring the practicalities of this approach and how it could be adapted to observe expressions in the process of becoming formulaic. Chapter 4 describes a case study where samples of the speech of an individual Japanese speaker of English, Kensuke, are collected over time. This provides the potential to observe the natural development of formulaicity in certain expressions.

The remaining questions are addressed through a series of studies (S4-S6) which focus on the targeted memorisation of particular expressions by Japanese speakers of English. These studies begin with a replication of Wray and Fitzpatrick (2008) which is described in Chapter 5 and extended in Chapter 6 to analyse fluency behaviour and formulaicity in the target utterances. Chapter 7 describes a small study (S4) which explores an experimental method designed to influence how participants memorise a specific set of target sequences. In Chapter 8 there is a return to the literature in order to review approaches to understanding and representing formulaic expressions in the mind. Combining the theoretical and empirical understanding from the previous chapters, the final two experimental studies (S5 and S6) compare two different approaches to memorising the set of targets in order to explore different routes to formulaicity (Chapter 9), and better understand the end-point of formulaic acquisition (Chapter 10). The final discussion in Chapter 11 returns to the two key research questions. In seeking to answer these, the chapter draws on findings from the studies to discuss the identification of formulaicity in speech, describe ways in which acquisition of formulaicity takes place, and highlight some key factors that influence this process. A simple model for representing two routes to formulaicity fusion and holistic acquisition) is then proposed, and pedagogic and other implications are also discussed.

CHAPTER 3: Formulaicity in L2 speech samples (Study S1)

Exploring psycholinguistic formulaic expressions in Japanese speakers of English

3.1. Introduction

In order to explore formulaicity in L2 speakers, it is important to have both a clear definition for formulaic expressions and a means of identifying their occurrence in speech. As discussed in Chapter 2, a useful definition of internal formulaicity and one that will be adopted in this thesis is that offered by Myles and Cordier (2017) for what they term a 'processing unit'. Specifically, an internal formulaic expression is here defined 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.

Closely aligned to this definition, Myles and Cordier propose a method for identifying formulaic expressions in speech based on a set of hierarchical criteria. This chapter describes an empirical study which utilises this approach to explore psycholinguistic formulaicity in a group of intermediate/advanced Japanese speakers of English.[12] The methodology follows that used by Cordier (Cordier, 2013; Myles & Cordier, 2017) who applied the approach to five advanced level L2 French speakers before and after a period of study abroad in order to determine how their use of formulaic expressions changed. In her study, the participants undertook a set of different speaking tasks consisting of interviews, discussions and story narrations. Formulaic expressions were identified by applying the fluency and diagnostic criteria on a hierarchical basis (following Hickey, 1993), meaning that conditions were applied in a strict order. For a sequence to be declared formulaic, it first had to satisfy the necessary condition of fluency and then also had to satisfy at least one diagnostic criterion to indicate that it showed signs of being a holistic unit. The main results using this method were that on average 27.7% of her participants' speech was formulaic overall (a mean of 25.1% before study abroad and 30.5% after). In addition, she found significant differences in observed formulaicity across different tasks, with

[12] A version of this study has been published in the journal *Vocabulary Learning and Instruction* (Cutler, 2017)

the story-telling task producing fewer formulaic expressions than the interview or discussion tasks.

The current study adopts the same identification process and hierarchical criteria but, unlike the Cordier study, only focuses on samples from one time period (utilising two different task types). This difference is because Cordier sought to compare formulaicity before and after a period of study abroad, whereas the current study focuses on the use of formulaic expressions (by Japanese speakers of English) at a single point in time. In addition, the study also samples and analyses the speech of two native English speakers in order to provide a point of comparison.[13] One aim of the study is to get a feel for the amount and type of formulaic speech used by a particular group of intermediate and advanced Japanese speakers of English and to check how this compares with the previous research. The three main research questions associated with this aim are as follows:

- RQ1: To what extent do psycholinguistic formulaic expressions feature in the speech of these intermediate/advanced Japanese speakers of English, and how does this compare with results from previous research?
- RQ2: How does the nature of the task affect the number of formulaic expressions used?
- RQ3: What types of formulaic expressions are used by the speakers and how do they contribute to overall formulaicity in these speakers?

A second key aim is to explore the methodology for sampling and analysing text to identify internal formulaic expressions in these L2 speakers. The practical and theoretical issues associated with investigating formulaicity in this way will therefore also be discussed, along with any enhancements that may need to be applied for the purposes of the overall research. This study, being the first to apply these hierarchical criteria to Japanese speakers of English, provides an opportunity for testing the methodology as well as giving insight into the prevalence of formulaic expressions in this group of speakers.

[13] The study uses data from speech samples originally collected for an earlier study (Cutler, 2013). The analyses applied and described in the current chapter, however, are quite different from those of the original study. In the original, formulaicity of expressions was determined via the diagnostic criteria of Wray (2008a) and did not involve segmentation according to dysfluency.

3.2. Method

3.2.1. *Participants*

The participants were eight Japanese speakers of English (JSE), all of whom were volunteer office workers recruited from companies in Japan that were known to the researcher. The participants were chosen on the basis of availability and to provide a mix of background (in terms of experience and opportunities to use English) and proficiency levels. There were seven females and one male and their ages ranged from 32 to 55. Four of the participants were from the same company, and three of the participants had similar jobs (associated with book-keeping and accountancy). To provide a point of reference, two native speakers from the UK also undertook the identical process. Both were working adults with occupations unrelated to teaching English or linguistics.

On joining the study, participants were given a brief questionnaire asking about any English tests taken and their experience in using and learning English. Participants were also given a vocabulary check based on the Productive Vocabulary Levels (PVL) test developed by Laufer and Nation (1999).

Table 3.1: Details of participants (using pseudonyms)

Participant	Sex/Age	TOEIC*	PVL	CEFR
Junko	F-40+	650	27	B1
Eri	F-50+	735	27	B1
Mami	F-30+	-	41	B2
Sachi	F-40+	865	42	B2
Kanae	F-30+	940	44	B2
Yoko	F-40+	940	44	B2
Wataru	M-40+	-	47	B2
Yayoi	F-40+	975	51	C1

** TOEIC is the Test of English for International Communication (ETS, 2019)*

Details of the participants and their estimated level on the Common European Framework of Reference for languages (CEFR) (Council of Europe, 2001) are given in Table 3.1.[14] Four participants (Eri, Kanae, Yoko and Yayoi) used spoken English regularly in their current work. All other participants used English at work occasionally, but mainly in written form. Three participants (Sachi, Yoko and Yayoi) had experience (over a year) of living or studying abroad in an English-speaking country.

3.2.2. ***Procedure***

Full University ethical procedures were followed in the collection of data for the study (and for all subsequent studies in the research). The participants each undertook two speaking tasks: a structured interview about their work lasting 4-5 minutes, and a story narration based on a picture sequence (around 3-4 minutes). For the story, they had a choice of three picture sequences (shown in Appendix 3.1) and were given two minutes to prepare. Participants were told the nature and timings of the tasks but not the focus of the research.

All tasks were recorded and transcribed, with pauses and other relevant dysfluency marked. Formulaic expressions were then identified according to a set of hierarchical conditions, following the methodology described by Cordier (2013). These conditions were applied in three stages to provide a progressive filtering of the transcribed speech.

3.2.2.1. Condition 1: Phonological coherence (Necessary)

The first necessary condition was that of phonological coherence, here operationalised as fluent pronunciation. As highlighted in Section 2.4, this has been used as a validation measure in the identification process before (e.g. Dahlmann, 2009; Erman, 2007; Raupach, 1984) but not (other than by Cordier) as an initial necessary condition in a hierarchy of criteria. Signs of dysfluency were defined to be:

- unfilled pauses > 0.25s
- filled pauses (e.g. *er*, *umm*, *ah*)
- syllable lengthening > 0.4s
- repetition or repair/retracing

---

14 Proficiency levels for each participant were estimated on the basis of their most recent score on TOEIC (ETS, 2019), where available. This was converted to a CEFR proficiency level on the basis of score comparison research undertaken by the publishers of the test (ETS, 2007). The vocabulary test scores were consistent with the TOEIC scores and, on this basis, were used to estimate the proficiency band of the participants without other test scores.

The 0.25s cut-off for unfilled pauses follows a standard used frequently in fluency research (e.g. Kormos & Dénes, 2004; Lennon, 2000).[15] Filled pauses were those featuring non-words (i.e. not containing semantic information). Lexical fillers (e.g. *you know, yeah*) were not taken as filled pauses since they have a function and may themselves be examples of formulaic expressions. The identification of syllable lengthening follows Dahlmann (2009) and was taken to indicate the end of a run. An oblique line (/) was inserted into the text wherever a dysfluency occurred (and at the beginning and end of a speaking turn). In this way, the lines segmented the speech stream into fluent runs. For example:

> SACHI: *it's / funny because he / I'm working in the office / and it / it's just he and me / so / when he went on business overseas / I just...*

3.2.2.2. Condition 2: At least one typical condition showing a holistic dimension (Necessary)

Fluent runs can potentially be quite long stretches of speech and are not necessarily formulaic in themselves. Indeed there may be several formulaic expressions along with individual words within a fluent run. Therefore, a further way of identifying the formulaic expressions from within the runs was required. The second necessary criterion defined by Cordier (2013) was that there should be at least one typical condition showing a holistic dimension. The diagnostic criteria used here were adapted from those used by (Wray, 2008a) and Wood (2009) and are as follows:

a) Grammatical or semantic irregularity: The meaning of the sequence is not given by its parts, or the grammar of the sequence is not that typically used to express the meaning. Examples from the current study included: '*they lived happily ever after*', '*bits and pieces*', '*can't handle*'.

b) Regular sequences with semantic or functional unity: These are typically grammatical units, common collocations, proper names or other sequences with a clear holistic mapping of form to meaning or function. Examples from the current study included: '*in charge of*', '*of course*', '*on the other hand*', '*typical day*', '*Toshima Ward*'.

[15] According to Myles and Cordier (2017, p. 19), using a low value (between 0.2 and 0.3 seconds) for what constitutes a pause ensures that all processing dysfluencies are picked up, even for advanced speakers of English. It also means that any possible cultural differences in what length of pause is acceptable in speech should not influence the analysis.

c) Sequences likely to have been learnt or used as a whole by the speaker: This was based one of the diagnostic criteria from Wray (2008a, p. 116): “based on direct evidence or my intuition, there is a greater than chance level probability that this speaker will have used this precise formulation before in communication with other people”. Examples from the current study included expressions from the speaker’s work experience (e.g. ‘*total administration time*’, ‘*TOEIC essay contest*’) or ones that they were likely to have learnt before (‘*on the other hand*’).

It should be noted that the above criteria are by no means mutually exclusive, and a sequences may satisfy more than one criterion (e.g. ‘*on the other hand*’ above). This is not surprising since there are a number of potential causal or theoretical links between the criteria. For example, most irregular sequences known to a speaker are likely to have been learnt or experienced as a whole. However, evidence of holisticity only requires the satisfaction of one criterion. So, for the purposes of this procedure, no special significance is attached to sequences satisfying multiple criteria.

3.2.2.3. Condition 3: Frequency (Graded)

A further graded condition used was that of intra-speaker frequency (i.e. does the speaker use the same term repeatedly). In a small speech sample, it is not possible or desirable to use the repetition of an expression as a necessary criterion. However, when expressions are repeated by a speaker, it adds to the likelihood that they are formulaic (assuming the other conditions are also satisfied). For example, one participant said ‘*I’m very surprised*’ on three different occasions (even when narrating the past).

3.2.3. ***Measures***

Two main measures of ‘formulaicity’ were used. For comparative purposes these were identical to the ones used by Cordier (2013):

1. FS% (Percentage of formulaic syllables): the number of syllables in the speech sample that were part of a formulaic expression divided by the total number of syllables in that sample.
2. ANR (Average number of formulaic syllables per fluent run): the number of syllables that were part of a formulaic expression divided by the number of fluent runs in the speech sample.

The FS% measure gives an overall sense of how much of speech is part of a formulaic expression, while ANR gives a sense of how they divide up the speech stream. In addition to the formulaicity measures, some standard temporal measures of speech fluency were calculated for each sample in order to explore how formulaicity may vary with fluency. These were the Speech Rate (SR) in syllables per minute, and Mean Length of Runs (MLR) which measures the average length in syllables of a fluent runs between dysfluency markers (see Kormos & Dénes, 2004; Lennon, 2000).

3.3. Results

Overall, 4798 words (6340 syllables) were spoken by the eight JSE participants over the two tasks and 663 formulaic expression tokens were identified (449 types). These contained 1685 words (2285 syllables). The average number of words (syllables) per formulaic expression was 2.54 (3.56). There were 214 repetitions (22.2%) overall, with 67 tokens (40 types) being repetitions across two or more participants. The most repeated sequences were '*for example*' (12 tokens across 5 participants), '*you know*' (11 tokens / 2 participants) and '*I think*' (9 tokens / 6 participants).

3.3.1. *Types of formulaic expression used*

To explore the different types of formulaic expression that participants used, sequences were categorised according to a broad typology developed by Cordier (2013). This was chosen to provide a direct comparison with her study.

In this typology, 'referential sequences' are defined as those predominantly used to refer to entities such as objects, places, times or ideas. 'Meta-discursive expressions' are sequences used to structure, comment on or engage with the discourse or message, and 'sentence builders' (from Nattinger & DeCarrico, 1992) are the fixed parts of patterns used to build sentences and phrases. The relative distribution of sequences across each category type is given in Table 3.2 along with examples from the study for each category and sub-category.

Table 3.2: Distribution of formulaic expressions by category

Category	Sub-category and examples	No. (%)
Referential	Verb phrase – *have to deal with* Noun phrase – *book stores* Time/place complements – *last year* Adverbials – *on behalf of* Whole clause – *they lived happily ever after*	486 (74%)
Meta-discursive	Hedges - *some kind of* Fillers - *you know* Asides – *what do I do?* Discourse structure – *for example*	100 (15%)
Sentence builders	*I think __* *I'm not good at __.* *It's nothing to do with __*	77 (12%)

3.3.2. *Use of formulaic expressions by task*

In order to explore differences in the usage of formulaic expressions across the two tasks, mean values of each formulaicity measure across the participants were calculated. Table 3.3 shows these values (along with the range for each) for each task and in total.

Table 3.3: Mean values (and ranges) for both formulaicity measures

	Task 1 (Work interview)	Task 2 (Picture story)	TOTAL
FS %	38.2 % (33.2 – 48.1)	31.0 % (26.0 – 38.4)	34.6 % (29.6 - 40.3)
ANR	1.89 (1.03-2.79)	1.39 (0.53-2.64)	1.64 (0.82 - 2.63)

Comparing the two tasks, the results show that more formulaic expressions were used in the first task (the interview about their job) than in the second (picture narration task). Using a paired t-test (two-tailed), these differences were found to be significant for both of the formulaicity measures, FS% and ANR, $t(7)=3.14$, $p=0.016$ and $t(7)=3.62$, $p=0.009$ respectively. For the combined samples, the mean FS% was 34.6% and mean ANR was 1.64. These mean figures are substantially higher than those found by Cordier (2013) whose five advanced French learners had mean

FS%=27.7% (range 22.1–31.0) and mean ANR=1.50 (range 0.83–1.90) over the 5 tasks they undertook. This is discussed further in Section 3.4.1.

3.3.3. *Use of formulaic expressions by participant*

A summary of the quantitative measures of formulaic expression usage and fluency for each participant are given in Table 3.4, arranged in order of fluency (SR).

Table 3.4: Summary fluency and formulaicity of participants over both tasks

Participant	FS %	ANR	SR (syll/min)	MLR (syll)
Junko	30.9%	0.81	70.9	2.54
Eri	29.6%	0.84	83.6	2.82
Wataru	40.3%	1.44	97.0	3.50
Sachi	36.0%	1.78	115.7	4.96
Kanae	35.6%	1.58	123.4	4.44
Mami	33.8%	1.81	127.3	5.34
Yayoi	31.9%	2.21	148.3	6.80
Yoko	38.5%	2.63	175.9	6.85

As can be seen from the data, formulaicity as measured by ANR (the average number of formulaic syllables per fluent run) increases consistently in line with fluency (SR). In particular, the two participants (Yayoi and Yoko) who had considerable experience (2 years or more) of living overseas also had the highest fluency and ANR scores. On the other hand, the FS% measure does not show a clear pattern with respect to fluency. For example, the participant Wataru has the highest FS% score (40.3%) but was one of the less fluent speakers (SR=97.0) on the tasks. The two native speakers who did the same tasks and followed the same procedure had considerably higher usage of formulaic expressions than all of the participant (FS%=46.4% and 48.1%, ANR= 3.74 and 4.81) and they were also more fluent (SR=182.0 and 195.7). This provides a good validation of the procedure.

3.4. Discussion

3.4.1. *Use of psycholinguistic formulaic expressions*

Insofar as they can be reliably measured on the basis of the criteria used here, the FS% figures show that psycholinguistic formulaic expressions are a significant part (e.g. 30-40%) of the speech of the Japanese Speakers of English (JSE) participating in the study. The sequences used were mainly referential (verb phrases, noun phrases, time /place complements), accounting for 74% of all sequences. Within this category there were few repetitions between or within the individual participant samples and, as Cordier (2013) found, there were few examples of expressions that are grammatically or semantically irregular (like *by and large*). Meta-discursive and sentence building sequences accounted for a smaller proportion of the sequences overall (15% and 12% respectively), but the majority of repeated expressions (e.g. '*I think*', '*for example*', '*you know*') were from these two categories. The distribution of sequences by category, and the mostly standard nature of these, matches what Cordier (2013) found with her advanced French learners. Overall, as in her study, the picture of psycholinguistic formulaic expression usage that emerges is that of the speakers using a breadth of canonical (transparent and grammatical) referential sequences, each being used only once or twice with almost no overlap across participants. These are then supplemented by a number of useful meta-discursive or sentence building expressions which tend to be repeated more, particularly by the participants with higher degrees of formulaicity in their speech.

As noted in Section 3.3.2, there was a significant difference between the two task types for both formulaicity measures FS% and ANR, with the interview task producing more sequences than the story-telling in each case. This supports the findings of Cordier (2013) who found significant differences between all the task types used, with the more interactive interview and discussion tasks yielding more formulaic usage than the narrative task. In the current study, this may be thought to reflect the familiarity of the topics as much as the tasks themselves. In the work interview task, participants tended to use expressions specifically related to their work and experience (e.g. '*procedures for foreigners*', '*put the cheque in*', '*test administration*', '*month end*' etc.) which they have likely used frequently before. In the story narration however, the content was not so familiar to the participants and there were likely to be fewer referential sequences easily available to them. On the other hand, when narrating in general, there are potential opportunities to use common sequences for organising discourse (e.g. expressions for sequencing time and events such as '*last year*' or '*after that*') that the participants could have usefully employed. However, apart from a few examples (e.g. '*the next day*', '*ten years later*'), these were not used extensively by most speakers in this study.

While the distribution of sequences by task, category and regularity is similar to that found in Cordier's study, the formulaicity figures in the current study (for intermediate/advanced Japanese speakers of English) are higher than those found for her advanced British speakers of French. Despite the obvious difference that the texts were in different languages in the two studies, the size and direction of the difference in the FS% scores is perhaps surprising. A possible contributory factor is a small difference in the pause cut-off length used (0.25s compared to the 0.2s used by Cordier). However, a follow-up analysis on a sample of the sequences identified as formulaic in the study found that none would have been rejected even if a 0.2s cut-off was applied. A further possibility is that, due to the essentially probabilistic and contextual nature of diagnostic criteria, there may be systematic differences in applying the criteria in the second stage of the identification process. This point is explored further in the next section.

3.4.2. ***Identification challenges***

Although a consistent and well-defined process was used, the actual application of the method highlighted particular challenges inherent in identification arising from the nature of formulaic expressions themselves and the necessarily interpretative nature of diagnostic criteria. Three particular challenges were illustrated in the study. These will be described in the sections below, and the implications for subsequent studies will then be addressed in Section 3.5.

3.4.2.1. Degree of 'fixedness' within the sequence

Formulaic expressions may be either fixed or constructed as frames with slots for variables (Wray, 2002a). In addition, they may be subject to expansion (e.g. adding an intensifier within the sequence) or nesting (placing one sequence in the variable slot of another). Deciding which of these options is applicable in individual cases can be challenging, and use of the conditions and criteria may not always be able to resolve this. Such decisions are important however since they may affect which words within the string are taken to be part of the formulaic expression, thereby affecting the quantitative measures of formulaicity. The following example from the study illustrates this challenge.

(a) YAYOI: *it's partially the subcontractor's job to train proctors*

The expression in (a) was delivered fluently by the participant and therefore satisfies the first condition for being a formulaic expression. For the second (holistic) condition, either criterion 2b ('has functional or semantic unity') or 2c ('has been used in the same form to convey the same meaning') may be applicable. However, they may potentially be applied at different levels of abstraction. For example, it is possible that the whole expression is formulaic as this is a work related topic which has been discussed before. On the other hand, it could be that the frame 'it's someone's job to do something' is formulaic for this speaker, with the (familiar) variables slotted in appropriately and the qualifier '*partially*' added as an (optional) expansion. To be able to choose between the options would require more data, such as a follow-up interview or a larger sample of the participant's speech. However, even then it may not be possible to resolve. This issue is further explored in Chapter 4.

3.4.2.2. Dynamic nature of formulaicity

The study also provided examples illustrating the potentially dynamic and context-based nature of formulaicity in the individual speaker (e.g. Ellis 2012). For example, Junko in her interview initially appeared to construct the phrase '*PR unit*' (as the English translation of her department name) and then subsequently used it in a formulaic way.

(b) JUNKO: *My job is a PR- (1) unit? (..) I am in PR unit. […] I think (...) PR unit is very conservative*

The phrase '*PR unit*' does have a semantic unity (Criterion 2b) and is repeated (Criterion 3). So, the two fluent cases of the phrase in the example are taken to be formulaic expressions in the current procedure. However, the evidence of earlier dysfluency of the sequence also seems important. For example, here it seems to indicate that the sequence is newly formed and, as such, may only be temporarily available in a holistic form. Other potential indicators of such 'temporary formulaicity' may include mixtures of fluent and non-fluent usage of a sequence, or the repetition of a formulaic expression taken from the interviewer's question. Indeed, examples of both indicators were observed in the current study. The extent to which this kind of contextual information should be applied will depend on the needs of the research and how one views the status of newly formed or temporary formulaic expressions.

However, since such decision will affect the count of formulaic expressions observed, it is important for any identification process to be explicit about how it deals with these cases. This will be explored further in Chapter 4.

3.4.2.3. Use of 'multi-word' as a defining feature

In most approaches, formulaic expressions are explicitly taken to be multi-word units (i.e. contain more than one word). However, as Wray (2014) argues, the concept of the 'word' is not always clear, due to the existence of contractions, polywords, compound nouns, hyphenated words, etc. While explicit clarifications can be made at the definitional stage (e.g. in this study, contractions, polywords, hyphenated words are all taken to be multiple words), there were examples from the study that reveal the slightly arbitrary nature of using the word as a defining feature for identification. For example, '*test takers*' and '*a lot of*' were included as formulaic but not the single words '*examinees*' or '*many*', even though on definitional criteria they are essentially equivalent. This highlights a challenge in applying a multi-word criterion as a definitional feature of formulaic expressions, and is another potential source of difference in the identification process.

3.4.3. ***The formulaicity measures***

Two variables, ANR and FS%, were used in this study to provide a measure of the 'formulaicity' of the participants' speech samples, and the results show a different pattern across the participants for each. ANR (the average number of formulaic syllables per fluent run) seems to have a close association with fluency, with ANR values increasing in line with increasing speech rate (SR). However, for FS% (the proportion of syllables that were part of a formulaic expression), there is not such a clear pattern. As already noted, one participant, Wataru, had a high value for FS% even though he spoke quite hesitantly (as shown by his fluency measures). At the same time, one of the most fluent speakers, Yayoi, had a comparatively low FS% over her two samples. One way to interpret this is to acknowledge that different measures indicate different aspects of performance and processing. For example, as discussed in Section 2.3.2.1, Towell et al. (1996) argued that fluency as measured by MLR (i.e. a greater ability to formulate runs) may be due to greater proceduralisation in processing within the formulator (in Levelt's 1993 model of speech production), and that such proceduralisation is facilitated by the use of formulaic expressions. However, how such usage is measured is also important and the results here suggest a possible differentiation of the roles of FS% and ANR.

A case such as that of Wataru demonstrates that the proportion of syllables that are formulaic (FS%) is not necessarily a useful measure of formulaicity to associate with aspects of speech processing such as proceduralisation. The FS% figure represents the proportion of speech that is part of a formulaic expression, but it does not indicate the number and length of sequences or how they fit together into fluent runs (for which ANR may be more appropriate). What this highlights is that although the FS% variable may have intuitive appeal as an apparent measure of how formulaic a speech sample is, it may not be the most appropriate measure for this purpose.

3.5. Conclusion

This study shows that psycholinguistic formulaic expressions, defined as fluent, semantically or functionally coherent multi-word units, may be a significant feature in the speech of intermediate / advanced Japanese speakers of English. The results of this first study to use these particular identification criteria on such speakers broadly agree with the main findings of the previous research by Cordier (2013) using the same method, and give some further insight into the prevalence of psycholinguistic formulaic expressions in L2 speakers as well as the practical challenges of identifying them. The study also adds further weight to the finding that use of formulaic expressions is sensitive to the kind of task that is used to elicit speech. Overall, the study demonstrates how a systematic hierarchical procedure can be used to identify formulaic expressions in a useful way. In particular, the use of dysfluency as an initial criterion provided a clearly quantifiable starting point for identification that can be consistently applied. Particular examples of expressions used by participants also highlighted some theoretical and definitional aspects of formulaicity that will be helpful in making the diagnostic criteria more robust and in interpreting the meaning of formulaicity measures such as FS% and ANR.

Undertaking the procedure highlighted a number of the inherent challenges in identifying formulaic expressions in spoken output. These centred on the dynamic and graded nature of formulaicity and the interpretative nature of diagnostic criteria. Two recommendations for making the process more robust therefore can be proposed. Firstly, ensure that there are explicit, theoretically justified ‘rules’ to cover ambiguous cases (such as when there is a mix of fluent and dysfluent examples of the same sequences or when there are multiple options for breaking a fluent segment into formulaic parts). These will help in further standardising the process. It is also particularly important to use contextual information from the task and from the individual's speech sample as a whole, and to specify how to apply it. However, even with such refinements, it should be recognised that the diagnostic criteria are based on likelihoods and are not always strictly quantifiable on the evidence available. So, a second important recommendation might be to utilise multiple judges to make the diagnostic assessments and to have explicit rules and procedures to deal with disputed cases when pooling the results.

Overall, the study supports the suggestion that the use of psycholinguistic formulaic expressions (as measured by ANR for example) is associated with fluency of the speech overall. An observation from the study was that a principal area of difference in usage between participants with higher and lower ANR (and fluency) was in the use of meta-discursive and sentence starter expressions and their repetition. In particular, higher fluency participants tended to use (and repeat) a greater number of general discursive expressions (sequencers, hedges and fillers) and longer types of sentence building patterns. This suggests that a useful focus, even for higher-level Japanese speakers of English such as those in this study, would be to support them in becoming fluent in the production of a prioritised set of such formulaic expressions in order to enhance their output delivery.

CHAPTER 4: Observing formulaicity over time (Study S2)

A case study exploring formulaicity and fluency in an individual L2 speaker over time

4.1. Introduction

4.1.1. *Background and context*

Study S1 (Chapter 3) explored the extent to which some intermediate Japanese English speakers of English (JSE) used formulaic expressions in their speech. That study utilised an identification method based on hierarchical criteria (Cordier, 2013; Myles & Cordier, 2017). While the method followed a clear procedure, one of the questions it raised was about how to deal with sequences that recur in the individual's speech sample but which are delivered fluently only some of the time.

Exploration of such expressions may be useful in understanding and observing the acquisition of formulaicity. Inconsistent fluency could potentially indicate some kind of 'partial' formulaicity where the expression is in the process of becoming formulaic at the time the sample is taken. On the other hand, it could be that the expression is in fact formulaic but is on some occasions disrupted for a legitimate reason (such as planning ahead), or that the expression is not formulaic (i.e. there is no holistic storage or processing) but has been constructed fluently on some occasions. If inconsistent fluency were indicative of a trajectory of increasing formulaicity, it might be predicted that the change in fluency would be in a consistent direction from less fluent to more fluent. On the other hand, other explanations would predict random oscillation between fluent and dysfluent renderings, in any order.

To explore the nature and prevalence of this phenomenon therefore, it is necessary to look at multiple samples of an individual's speech in order to track potentially formulaic expressions recurring over time. While variations on such an approach have been used for native speaking children (Dabrowska & Lieven, 2005; Peters, 1983), no equivalent investigation for adult L2 speakers appears to have been carried out. There have been some analyses of individual L2 speech over time, such as those detailed in Schmidt (2001), Wood (2009) and Cordier (2013). However, these focussed on how overall formulaicity of speech developed between two time periods rather than how the delivery behaviour of individual expressions changes over multiple samples.

4.1.2. ***Overview of study S2***

In the current study, a series of speech samples based on a recurring conversation topic were obtained from an individual Japanese speaker of English over a period of about 9 months. The speech samples then formed a small individual corpus that could be analysed to explore the delivery features (fluency) of expressions that recurred across samples. Since the aim was to explore the changing fluency of expressions that could be (or become) formulaic, it was necessary to have a clear method for selecting them which avoided any potential circularity. In particular, since fluency was the quality being explored, it could not be used as criterion for selecting the target sequences in the first place.

Many approaches to formulaicity 'in the language' utilise frequency in some way for identifying potential formulaic expressions, and this is also a graded element in the internal approach of Myles and Cordier (2017). Since the current study is concerned with expressions that are repeated across the samples, it was feasible to utilise frequency (within the corpus) as a criterion for the selection of the target sequences. In order to do this, a review of frequency-based methods for extracting potential formulaic expressions from corpora was undertaken (see Appendix 4.1). Based on this review, an algorithmic approach from Brooke, Tsang, Hirst, and Shein (2014) was selected and adapted for the specific needs of the current study.[16] This will be described in detail in Section 4.2. In essence, it is an approach which segments the sample into recurring lexical units and resolves issues (such as two possible expressions that overlap) in a consistent way. In addition, as most approaches to formulaicity require expressions to have a unitary holistic dimension, this criterion was also included in the process. The extracted expressions were then termed 'recurrent unitary expressions' (RUEs).

The set of RUEs was explored in terms of the fluency of each expression as it occurred in the multiple samples over time. RUEs that were consistently fluent could be considered fully formulaic for this speaker (according to the criteria for formulaicity used in this research). Similarly, RUEs that were rarely or never fluent could be considered non-formulaic. This group are interesting in that they raise questions of why these expressions might resist formulaicity. The key group, however, was those RUEs which were fluent some (but not all) of the time. A main aim of the study was to

[16] It should be emphasised that while the Brooke et al. (2014) method claims to extract formulaic sequences 'in the language' from corpora, the expressions extracted here were not (necessarily) psycholinguistically formulaic since their fluency was yet to be established.

focus on those expressions in particular and explore what they may show about evolving formulaicity and the acquisition process. This involved firstly seeing whether such inconsistency was actually a common feature, and secondly exploring the extent to which it could be associated with expressions being on a path towards formulaicity. The key research questions are:

- What proportion of RUEs have inconsistent fluency – and is this a typical feature of the output?
- Of the RUEs with inconsistent fluency, is the pattern of changing fluency consistent with them becoming formulaic (i.e. moving from non-fluent to fluent over time)?

4.2. Method

4.2.1. *Participant*

Kensuke (a pseudonym) was a male Japanese speaker of English in his 30s. He worked for an American-owned trading company based in Japan. Kensuke occasionally spoke English with suppliers in the US and other countries and he read English reports and wrote some e-mails in English. However, most of his work with colleagues and clients was conducted in Japanese. He studied conversational English once a week for 90 minutes, and was at an intermediate level of English with a TOEIC score of 840. He was enthusiastic about improving his English but highly committed to his job which took up most of his time.

4.2.2. *The speech samples*

The speech samples were in the form of 5-10 minute conversations that the researcher had with Kensuke during his 1:1 English lessons. These took place early in the lessons after he had warmed up. He was asked to speak naturally and as fluently as possible with a focus on communicating what he wanted to say. The conversations took place in a coffee shop and the atmosphere was relaxed and friendly. The idea was for Kensuke to talk about his work and what had happened that week or what was concerning him. The researcher started the conversation with an open prompt like "Tell me about your work this week" and followed up with further prompts ("What happened?", "Really, why did he say that?", "What does that mean for you?") to keep the conversation moving. By design, most of the speaking was done by Kensuke.

There were 15 speech samples recorded in this way over a period of 9 months. Each sample was recorded on a Sony IC Recorder, transferred to computer and transcribed using the software Transcriber AG (DGA, 2014). This provides a visual display of the sound waves alongside the transcription, enabling pauses and other speech phenomena to be timed accurately and noted. In order to facilitate subsequent analysis, Kensuke's speech was divided into lines as indicated by turn taking and major clause boundaries. Researcher speech was retained for reference, but excluded from counts and quantitative analysis. For each sample, an annotated transcription was saved into a text file.

For the annotation, dysfluencies were marked as follows within each clause line:

/ = pause > 0.2s; repetition dysfluency (false starts where repetitions precede a fluent continuation)

// = pause > 0.5s; short filled pause or elongated previous syllable (>0.3s)

/// = pause > 1.0s, long filled pause, any other major pause phenomenon

The following were separated out by the use of tag markers (< >): speech by the researcher; repetitions of single words (indicating dysfluency); unfinished words. This allowed them to be excluded from the quantitative analysis by the concordance software, but retained for reference in the qualitative analysis. Contractions were treated as separate words (e.g. *don't = do n't*; *it's = it 's*). Although, as noted, single word repetitions were excluded, reformulations and repetitions of multiword expressions were retained since the purpose of the analysis was to observe repeated uses of multiword expressions.

4.2.3. *Producing the list of potential formulaic expressions*

The first stage of the analysis was to extract all the recurrent unitary expressions (RUEs) in a systematic and reasoned way without reference to their fluency. For this purpose, an adapted version of the Brooke et al. (2014) segmentation process was used. The steps and calculations for this approach were performed in Excel using custom VBA[17] macros coded specifically for the purpose.

[17] VBA (Visual Basic for Applications) is a programming language which can be used for manipulating data in Excel.

4.2.3.1. Initial list of n-grams

The first stage was to obtain a list of all n-grams (frequency ≥ 3; range ≥ 2) for n ≥ 2. Here, frequency is the number of times the n-gram occurs over the whole corpus, and range is the number of different samples it appears in. To create the list, the text files were loaded into the concordance software AntConc (Anthony, 2018), set to exclude text marked with < >. A frequency of three occurrences was considered the minimum to constitute a clearly repeated sequence, and matches a common threshold suggested by O'Donnell (2011). A range of at least two samples was required so that change across time could be observed. This meant that some multiword sequences that were good candidates for formulaicity on other grounds (e.g. *New Orleans)* were excluded. Word frequency information was also obtained for use in subsequent calculations.

4.2.3.2. Initial segmentation

The next stage was to divide the corpus into lines of text and then segment each line using the initial list of n-grams extracted. In the Brooke et al. approach, each line is a sentence. However, Carter and McCarthy (2017) suggest that, when dealing with spoken text, it is more useful to focus on the clause as the basic unit. It has also been shown that formulaic expressions (at least those that are fixed and contiguous) fall within clauses or intonation units (Lin, 2013). The main purpose of the division into lines was to simplify the process and exclude any recurring n-grams that should not be included because they crossed obvious breaks in text. For this reason, the division was done on the basis of turns (i.e. when Kensuke resumed speaking after a question or break) and finite clauses.[18]

For each line, a set of maximal n-grams was obtained. A maximal n-gram is one that satisfies the minimum frequency and range criteria (i.e. one that is in the list of candidate n-grams) and for which the *n+1*-grams that contain it do not satisfy the criteria (i.e. are not in the list).

> Example 4.1; *but most of our rival companies they already purchased some portion in September and October from Brazil*

In the clause given in Example 4.1, the maximal n-grams were:

- *most of*

[18] Non-finite clauses, embedded clauses and any cases which where division was unclear were not separated into separate lines. The aim at this stage was to maximise the number of recurring n-grams that could potentially be formulaic.

- *of our*
- *our rival companies*
- *already purchased*
- *in September*
- *September and October*

Expressions such as *September and* and *rival companies* also satisfied the frequency and range criteria, but as these were contained in longer candidate *n*-grams (*September and October*, *our rival companies*) these were not maximal.

4.2.3.3. Removing overlaps

The next step was to deal with the cases where a maximal n-gram overlapped with another. For example, in clause above (Example 4.1), there are two sets of overlaps:

Example 4.2:	*most of; of our ; our rival companies*
Example 4.3:	*in September; September and October*

The aim is to segment the clauses without overlaps and this means selecting from possible competing segmentations. For example, in the first set of overlaps above (Example 4.2), there are two possible segmentations.[19]

Example 4.4	*most of* | *our rival companies*
Example 4.5	*most* | *of our* | *rival companies*

The method of choosing between segmentations was a variation on Brooke et al. (2014, p. 755) and involved conditional probabilities calculated from frequency data within the corpus. The idea was to calculate a ‘sequence predictability’ for each sequence within the segmentation and sum these to give a segmentation score. The segmentation with the highest score was then chosen.

Table 4.1: Example of segmentation comparison

Segmentation	**Sequences**	**Sequence predictability**	**Segmentation score**
most of | *our rival companies*	*most of*	0.0638	0.7638
	our rival companies	0.7000	
most | *of our* | *rival companies*	*most*	0.0012	0.7162
	of our	0.0150	
	rival companies	0.7000	

[19] These are the minimal segmentations in that they are the only options that resolve the overlap without any unnecessary additional boundary points. For example, in a possible segmentation such as *most of* | *our* | *rival companies*, the second boundary point could be removed

For example, Table 4.1 shows the sequence predictabilities and segmentation scores for the two competing segmentations given in Examples 4.4 and 4.5. In this case, the first segmentation (Example 4.4) was chosen because it had the higher segmentation score. Details about how the sequence predictabilities were calculated are given in Appendix 4.2.

4.2.3.4. Creating the revised list

Resolving all overlaps (185 in total) resulted in a revised list of sequences based only on the segmented lines of the sample and subject to the same frequency and range criteria. This was a reduced list compared to the original set of n-grams because the numbers of some n-grams were reduced to below the minimal frequency as a result of the overlap resolution process.

In the Brooke et al. method, a further reduction of the list is proposed by listing the sequences longest to shortest and then analysing each sequence to see if the sequences could be further broken down in a way which increased the total predictability of that group of words. However, since the sequences in the revised list were short (mostly 2 or 3 words), this step was not considered necessary for this study.

4.2.3.5. Checking that the sequences were unitary

The final part of the process (not part of the Brooke et al. method) was to remove any sequences from the revised list that did not have a holistic dimension. In particular, the sequences were required to have some semantic or functional unity, or to be assessed as being likely to have been learnt or used whole by the speaker.

This kind of assessment is not straight forward for two main reasons. Firstly, the individual speaker will have their own experience and idiolect, and particular word combinations may have taken on a unitary meaning or function for them even if they are not thought to be so in the wider language. Secondly, many formulaic expressions are frames with slots for variables, and the semantic or functional unity comes when the variable is filled. For this reason, a very broad interpretation of the unitary criterion was taken. In particular, the only sequences that were rejected as being non-unitary were:

- sequences starting with conjunctions/discourse markers (e.g. *but they, so I*, *yeah so*)

- sequences consisting only of functional/closed class words (e.g. *and the*, *on this*, *for that*, *us to*)

There were 23 sequences removed according to the above two criteria. The remaining 142 expressions were considered to form the set of RUEs.

4.2.4. ***Analysis***

The recurring unitary expressions (RUEs) were then analysed both quantitatively and qualitatively. The quantitative analysis recorded the fluency information for each sequence for every occasion it was spoken in the samples. This included the extent of the dysfluency, its location (within the sequence) and which of the occurrences it was in.

4.2.4.1. Categorisation

The percentage of fluent repetitions (PFR = no. fluent repetitions / frequency) for each expression was calculated. The sequences were then grouped according to their PFR scores[20]:

- 100% (always fluent)
- 81 - 99% (mostly fluent)
- 61 - 80% (often fluent)
- 41 - 60% (sometimes fluent)
- 01 - 40% (mostly dysfluent)
- 0 % (always dysfluent)

This categorisation was helpful to explore the extent to which inconsistent fluency was a feature of recurring expressions. RUEs within the ‘mostly’, ‘often’ or ‘sometimes’ fluent groups were then analysed further to determine the time sequence of their dysfluencies.

[20] In some cases, categorisation of continuous data can result in a loss of experimental power and potential confounding effects due to the choice of boundaries. However, since most of the RUEs occurred 3-5 times, the boundaries coincided with discrete changes in the number of dysfluencies and therefore formed coherent groups. For example, the 61-80% category represented one dysfluency (i.e. the RUE was fluent 4/5, 3/4 or 2/3 of the time) in most cases. Further, inspection of the results at the item level, did not indicate any obvious discrepancies within or across categories.

4.2.4.2. *Tendency to go from dysfluent to fluent*

The main hypothesis was that if inconsistent fluency was associated with expressions being on the road to formulaicity, a necessary feature would be that the delivery of those expressions became more fluent over time.

In order to measure this tendency objectively, some measures and criteria were developed. These were based on the idea that, if an expression is becoming more fluent over repeated delivery, dysfluencies will tend to occur more in the earlier occurrences, and dysfluency should generally not increase in subsequent occurrences. The two measures created to evaluate these points for each RUE were:

- Dysfluency Centre of Gravity (DysCOG) = weighted average position of dysfluency in the ordered set of occurrences where dysfluency occurs
- Tendency to increase dysfluency (DysINC) = sum of dysfluency increase from one occurrence to the next (standardised to 'per occurrence')

An account of the way in which these were calculated and a justification for the criteria developed to measure increasing fluency are given below, followed by a worked example.

4.2.4.3. *Calculation of the measures*

DysCOG: For each RUE, its occurrences in the corpus were listed in chronological order (within and across samples) and numbered from n = 0 to F-1, (where F is the frequency of occurrence)[21]. For each occurrence n, the level of dysfluency D_n was then scored according to the pause marks (/) in the transcription: fluent=0; /=1, //=1.5; ///=2.

$$\text{DysCOG}, = [\ \Sigma\ (\ n \times D_n\ )\ /\ \Sigma\ D_n\ ] \times [1/(F-1)]$$

This calculation gives a standardised measure (from 0 to 1) representing how far along the occurrence points the average dysfluency would be.

DysINC: For the 'tendency to increase' measure, a change score C_n was calculated for each occurrence n=1 to F-1 as follows: if the level of dysfluency value ($D_{n)}$ had

[21] This numbering (from n = 0 to F-1) is chosen to simplify the formulae. The first occurrence of the expression is a point of reference and the n value represents the number of steps from that reference.

increased compared to the previous occurrence, it was scored by the amount of increase, otherwise it was scored as 0.

$$DysINC = \Sigma C_n / F$$

This measure gives a standardised measure (from 0 to 2) representing how much the average tendency to increase dysfluency would be. A value at or close to zero shows that the dysfluencies tend to remain constant or reduce as time passes.

4.2.4.4. Criteria for increasing fluency

A profile of DysCOG < 0.5 and DysINC < 0.25 was taken to indicate that an expression has ‘Increasing Fluency’ (i.e. a tendency to go from dysfluent to fluent delivery over time). The calculations ensure that these criteria are passed only when the dysfluencies occur before the half-way mark on average, and when dysfluency increase is minimal (a single small increase in dysfluency once every 4 occurrences or less). However, finding that an expression does not meet the criteria for increasingly fluency does not tell us whether it is actually decreasing in fluency (which would strongly challenge the experimental hypothesis) or whether it is relatively static, or perhaps full of data noise. To gain a clearer picture and to provide a balanced analysis, a similar approach was applied to determine whether an expressions had ‘Decreasing Fluency’. To do this, exactly the same calculation and criteria were applied, but to the set of occurrences listed in reverse chronological order.

To illustrate, Table 4.2 below gives the fluency behaviour and values for the five occurrences of the expression *we had some business*. In this example, DysCOG = [7.5/4] x [1/4] = 0.47. In other words, the centre of gravity for dysfluency is just under half way along the chronological time-line. However, DysINC = [1.5] / 4 = 0.3, meaning that the tendency to increase dysfluency is larger than the minimum. So, the criteria for ‘Increasing fluency’ are not satisfied.

Applying the process in reverse order: DysCOG(-) = [8.5/4]x[1/4] = 0.53; DysINC(-) = [1.5] / 4 = 0.3. This shows that the expression does not satisfy the criteria for ‘Decreasing fluency’ either.

Table 4.2: Calculating fluency tendency for 'we had some business'

S01: it was / a little / busy this week / because // **we had some business**.
S01: so / **we had // some business** this week.
S08: it was // relatively good /// because /// **we had // some business** this week.
S11: however /// **we had / some business** ///
S11: **we had some business** / in second week.

Sample no.	S01	S01	S08	S11	S11	Σ
n (n - *reverse order*)	0 (4)	1 (3)	2 (2)	3 (1)	4 (0)	
Dysfluency type	fluent	//	//	/	fluent	
D_n (Dysfluency value)	0	1.5	1.5	1	0	4
n x D_n (n x D_n - *reverse order*)	0 (0)	1.5 (4.5)	3 (3)	3 (1)	0 (0)	7.5 (8.5)
C_n (Dysfluency increase) (C_n – *reverse order*)	- (0)	1.5 (0)	0 (0.5)	0 (1)	0 -	1.5 (1.5)

4.3. Results

4.3.1. *Quantitative results*

The 15 samples contained 4964 word tokens (542 lemmas). As noted, 142 sequence types (857 sequence tokens) were identified as Recurring Unitary Expressions (RUEs) following the method of extraction. The least frequent RUE occurred three times across the total dataset, and the most frequent (*it was*) occurred 34 times. All RUEs occurred in at least two samples and some expressions (*it was*, *our clients)* occurred in as many as 14 different samples. There were 120 bigrams, 36 trigrams and nine sequences longer than three words. The longest sequences were:

I don't know how to say (frequency = 5, range = 3)

could you say that again (frequency = 4, range = 4)

Table 4.3 gives the number (and %) of sequences in each fluency category and the numbers (and %) of these which displayed increasing fluency (according to the criteria described earlier). For comparative purposes, the numbers of sequences with decreasing fluency are also shown. The mean value of the DysCOG measure is also shown for each category.

Table 4.3: Results by fluency category

Category	**# (%)**		**Fluency behaviour**	
		DysCOG	**Increasing #**	**Decreasing #**
always fluent (100% of time)	92 (65%)	-	-	-
mostly fluent (>80%)	10 (7%)	0.411	5 (50%)	4 (40%)
often fluent (>60%)	21 (15%)	0.375	11 (52%)	4 (19%)
sometimes fluent (>40%)	9 (6%)	0.372	5 (56%)	0 (0%)
rarely fluent (>0%)	7 (5%)	0.490	2 (29%)	2 (29%)
never fluent (0%)	3 (2%)	-	-	-
Total	142	0.399	23 (49%)	10 (21%)

By far the largest category was that of sequences that were always fluent. These included a variety of different expressions including: long sequences (*I don't know how to say*), standard phrases (*could you say that again*; *I don’t know*), discourse markers (*you know*; *to be honest*; *for example*; *in the end*; *of course*), hedges (*kind of*), sentence starters (*I think*; *it means*), time markers (*next week*; *last year*), multiword verbs (*have to*; *go down*; *couldn’t*), and examples specific to the work topic (*freight market*; *tough negotiation*; *our rival companies*; *fiscal year*; *loading port*; *Brazilian corn*; *in Japan*).

In the categories which featured a mix of fluent and non-fluent expressions, the mean value of the dysfluency ‘centre of gravity’ (DysCOG) was 0.399. This indicates that overall the dysfluency tended to occur in earlier occurrences of the expressions. Similarly, the tendency to increase fluency was much more prevalent than to decrease. This tendency was mainly focussed in the categories featuring expressions that were fluent between 40% and 80% of the time (‘often’ and ‘sometimes’ fluent). For these, DysCOG = 0.374 and the increasing fluency expressions (16) far outweigh the decreasing (4). That is, in this sample, expressions that are somewhat fluent showed a pattern of becoming more consistently fluent over time.

In the ‘mostly fluent’ category and ‘rarely fluent’ categories however, the fluency behaviour did not show a clear tendency either way. For ‘mostly fluent’ expressions, the average DysCOG= 0.411 (i.e. close to 0.5) and for ‘rarely fluent’ expressions, DysCOG= 0.490. In both cases, the number of increasing and decreasing fluency expressions is balanced. In other words, there is a more random distribution of dysfluencies in these categories; in the mostly fluent case, it is an occasional dysfluency that occurs randomly, while in the 'rarely fluent' case it is an occasional fluency.

4.3.2. ***Qualitative examples***

In exploring whether these expressions with increasing fluency could be considered as becoming more formulaic, it is instructive to look at the actual occurrences of each in more detail. Some illustrative examples are given below.

4.3.2.1. Increasing formulaicity?

Table 4.4 lists each occurrence of three example RUEs that showed increasing fluency (each a type of multiword verb).

Table 4.4: Details of occurrence for three sequences

Example1: decided to buy	
S01	They recognise // that situation // and /// they **decided / to buy** at higher prices
S01	They / didn't / **decided / to buy** more yet // but // they are thinking / now
S03	So // our clients // after all / our clients // **decided to buy** // but reluctantly
S11	So they / **decided to buy** /// they **decided to buy** / now.
Example 2: to explain	
S01	yeah // we /// needed / **to / explain** / the situation
S03	I had **to / explain** / our // boss // about / the situation // in here
S07	So /// we // had **to explain** much more than // usual year /// about …
S07	Yeah it's / difficult **to explain**
Example 3: started to	
S01	After long holidays /// our clients / had / **started / to** move / on their business
S01	which means I means /// they have / they had **started / to** buy corn
S06	but finally // our clients /// had **started / to** purchase \ corn
S06	So they thought that // corn / price // hit / the lowest // and just / **started / to** rally
S06	And // so / I / we think // prices /// has **started to** / rally
S12	finally // our clients /// **started to** / realise /// freight market /// is rallying high

These examples illustrate how an expression might be seen as becoming more formulaic for a speaker. For each, in the initial usages there is a short pause between two parts of the expression (e.g. *decided* and *to buy).* Then*,* in later samples, the expression is delivered without dysfluency. The fluent occurrences occur at different points for each example. Other aspects of delivery may also be relevant. For example, for *decided to buy*, the expression is repeated in sample S11 with identical rhythm and intonation, possibly for emphasis, and in S01, the past tense is used instead of the infinitive, which may indicate fixedness

It may be noted in passing that the examples from Table 4.4 are each associated with a verb + infinitive structure. They represent three different ways of splitting this structure: the verb + to (*started to*); the infinitive (*to explain*); the whole trigram (*decided to buy*). This exemplifies the difficulty there would be in ascertaining formulaicity via purely formal considerations.

4.3.2.2. Observing fusion?

In the process of fusion (increased formulaicity over time), the component words or sub-sequences within the expression are initially fitted together through some form of construction and then become ‘fused’ together as a result of repeated use and familiarity. To get a better idea of whether such a phenomenon could be occurring for an expression, it is useful to see whether there were subcomponents of that expression that had their own pattern of fluency within the corpus. In the set of 16 RUEs that showed increasing fluency, there were some which contained sub-sequences that also recurred in the corpus. These included: *decided to buy* (with *to buy*); *end of fiscal year* (with *end of* and *fiscal year*), *as I said* (with *as I _*) and *price of corn* (with *price of*). As an illustration, a list of the occurrences of *end of fiscal year* and its sub-sequences is given in Table 4.5.

Table 4.5: Occurrences of ‘end of fiscal year’ and its sub-sequences

S01	so /// from / **end of** April / to I think /// to September …
S07	So /// **end of** / November, // we /// had to / report
S07	in our company // our **fiscal** / in our company / **fiscal year** / start from / June.
S08	The vessel is arriving // **end of** December.
S10	In Japan, // March / is // **end of fiscal year** / for // many companies.
S10	some of our / clients /// requested us to // deliver corn by **end of** // March
S10	in our industry / January shipment means / finish delivery by **end of** March
S10	if /// shipment, /// is on **end of** / January /// the vessel should / arrive …
S10	our client /// they believe / it should / it should be delivered / **end of** March.
S11	As you know / March / is /// **end of / fiscal year** / for / almost all Japanese…
S11	because / they were / closing **fiscal year** / at the **end of** March.
S11	but / it was / beginning / at **fiscal year** // so not active.
S11	So /// **end of** / **end of** / March /// is important for / our industry,
S15	because /// May was / **end of** our **fiscal year** ///
S15	because it was / **end of fiscal year**, so we /// we will /// we have to review …
S15	Yeah maybe **end of** / **end of** / **end of** June // we will have // kind of interviews

Note: To highlight the sequences, ‘end of fiscal year’ is shown in purple type, ‘end of’ in blue type and ‘fiscal’ and ‘fiscal year’ in red type

As the ordered list of occurrences shows, the expression *end of* seems to be a formulaic frame for the speaker and is used regularly with different variable endings (e.g. *end of <u>March</u>*). The expression *fiscal year* also seems to have become formulaic in its own right, after initially being constructed in S07. It is used several times independently of the longer expression *end of fiscal year.* The joining of the two also seems to be supported by the pause between them (in S11) and the insertion of the word 'our' (in S15).

4.3.2.3. Temporary formulaicity?

A further interesting set of RUEs from the list are the three different time markers (*September and October*; *two or three weeks*; *three or four days*) shown in Table 4.6.

Table 4.6: Details of occurrence for three time-related sequences

Example 5: *September and October* (frequency=5; range=2)	
S13	they / already /// purchased / some portion /// in // **September and / October** from Brazil
S13	So this week we purchased Brazilian corn / in /// **September and October** / in advance.
S13	maybe // in **September and October** /// the US corn / will be // will be /// higher / than ...
S13	maybe / this year /// just in // **September and October** /// I don't know but ten or / fifteen vessels / from Brazil // will come to Japan
S15	we already purchased /// **September and October** shipment // Brazilian corn
Example 6: *two or three weeks* (frequency=4; range=3)	
S01	For this **two or / three weeks** // corn prices // has risen / very much
S01	So // that's why // corn prices / risen / for **two or three weeks**
S12	next couple of weeks /// it's / would fall /// to prices /// **two or three weeks** ago prices
S14	but / if /// it takes / **two or three weeks** / to become /// normal condition /// we ...
Example 7: *three or four days* (frequency=3; range=2)	
S03	it will / take / **three or four days** to fix it
S03	if it's /// if / the / **three or four days** / delay, / it will be big problem / in japan
S10	our vessel // arrived // at Kobe /// **three or four days** ago

As shown in the listing, the first two had only one dysfluency each (in their first occurrence) and the third was fully fluent. Their form and specificity suggest that they may be examples of a more general frame. However, for all three examples, there were no other case of similar form (e.g. such as *three or four weeks*, *November and December)* which would support this.

In addition to the fluency of the expressions, formulaicity is supported by other potential indicators of non-canonical unitary usage (e.g. 'September and October shipment', 'two or three weeks ago prices', 'the three or four days delay'). In the case of *September and October*, the limited range (four occurrences in S13 and one in S15) indicate that this was a localised use describing a particular time period where some things were due to happen (purchasing of corn, price rises, shipments arriving). This may indicate a type of 'temporary formulaicity'.

4.4. Discussion

This study started with the premise that unitary sequences which are consistently delivered fluently by a speaker during speech are formulaic for that individual, but the status of sequences that are sometimes (but not always) delivered fluently is less clear. The study was designed to explore the extent to which this inconsistent fluency occurs and what it might mean in terms of the acquisition of formulaicity within the sequence. Using the frequency-based algorithm, a set of 142 recurring unitary expressions (RUEs) was identified over the 15 speech samples.

4.4.1. ***Identifying formulaic expressions***

The results show that a high proportion (65%) of such expressions were delivered fluently on every occurrence and they were therefore deemed formulaic by the criteria used in this research. There were also a number of RUEs (7%) in the 'mostly fluent' category, meaning that these had a dysfluency less than once every five occurrences. While Myles and Cordier (2017) do not specify how strict the fluency requirement should be for multiple occurrences of an expression, there is a good argument for saying that these should also be deemed formulaic. As Wray (2009) argues, even in fully formulaic expressions, there may be an occasional dysfluency due to planning or for some discursive purpose (e.g. emphasis). If that were the case, it would be expected that such dysfluencies would occur randomly across occurrences. In fact, the dysfluencies in this category do appear to be random, because the average dysfluency centre of gravity (DysCOG) is 0.411 (i.e. close to 0.5) and the numbers of Increasing and Decreasing Fluency expressions (based in most cases on a single dysfluency) are balanced (4 and 5 cases, respectively). So, including the 'mostly fluent' expressions as formulaic seems to be supported here.

Overall then, most recurring unitary expressions in the corpus also appear to be internally formulaic for the speaker and the frequency-based and fluency criteria may be quite consistent with each other. The 'always fluent' expressions represented 24.1% of the total words spoken (or 30.5% if the 'mostly fluent' expressions are included). These figures are comparable with the percentages of formulaic sequences in L2 speech found in earlier studies using similar criteria. As noted, Cordier (2013) found an average of 27% of speech to be formulaic in her L2 samples, and in study S1 (described in Chapter 3), the figure ranged from 23% to 40% (average 34.6%).

4.4.2. ***Interpreting fluency behaviour***

While most RUEs were fluent or mainly fluent, there was still a sizeable proportion that had inconsistent fluency. For this behaviour to be associated with possible evolving formulaicity, the hypothesis was that these expressions should tend to go from dysfluent to fluent over time. The criteria and measures used to judge this did seem to indicate such a tendency for expressions in the 'often' and 'sometimes' categories (i.e. those with dysfluency between 40% and 80% of occurrences). For these expressions, the mean DysCOG was 0.374 and there were four times more Increasing Fluency expressions than Decreasing Fluency ones. That is, dysfluencies tended to come earlier and the favoured direction of fluency change was from dysfluent to fluent. This finding is consistent with the idea that, broadly speaking, fluency changes in recurring unitary expressions over time may be associated with them becoming more formulaic.

While a tendency towards increasing fluency may be a good indicator of developing formulaicity, it is important to consider other evidence which may help to support this. In particular, looking at the context of occurrences for candidate expressions is useful because it can also give some idea of how the formulaicity may be developing in each case. For example, the expression *decided to buy* showed a pattern of fluency entirely consistent with the fusion of *decided* and *to buy,* and this may be supported by the fact that the *to buy* is also an established RUE for this speaker, occurring 14 more times in the corpus (e.g. *want* *to buy*, *have* *to buy*). Similarly, in the analysis of the expression *end of fiscal year*, both *end of* and *fiscal year* appear to be separately formulaic for the speaker, suggesting the possibility that these have been joined to create the new expression. However, fusing component parts of an expression is not the only possible route to formulaicity. Many researchers (e.g. Wray, 2002) highlight the idea of formulaic expressions being learnt holistically from the start. This remains a possibility even if the expression has component parts that are themselves formulaic. In the case of *end of fiscal year*, the evidence from the samples could also be consistent with this direct form of acquisition. In the L1 Japanese, there is a set expression for the same meaning, and this would certainly have been familiar to the speaker. So, it is feasible that the whole expression was required and learnt earlier, and the observed dysfluency and later insertion of 'our' in the samples is actually a sign of subsequent analysis.

4.4.3. ***The dynamic lexicon***

Wray (2008, 2012) highlights the idea of the individual idiolect and a mental lexicon that is dynamic in its contents. The results of this study seem to support that idea. Kensuke seems to use formulaic expressions specific to his needs, circumstances and experience. While there are standard expressions (*for example*; *you know*; *I think*), there are many examples related to his work area (*freight market*; *tough negotiation*; *our rival companies*; *Brazilian corn*), to his learning of English (*I don't know how to say; could you say that again*), and to necessary components of narrative (*this week*; *we have to*). Many of them appear very specific to him as an individual (e.g. *we had some business*) and would not be formulaic for other people.

The time markers highlighted in Section 4.3.2.3 are also a good example of this. The expressions *three or four days* and *two or three weeks* do not appear obvious candidates for fixed formulaicity due to their numerical specificity. However, their usage suggests they have may have proven useful and acquired a holistic meaning for this speaker (e.g. for indicating a rough time period between a day and a week, or between a week and a month). A further variation is the idea of temporary formulaicity as suggested by the example *September and October*. Its repeated fluent usage over a short period of time suggests that it may have become usefully formulaic for the speaker in order to facilitate describing this particular time period where a lot of things were happening. Such an expression may suffer rapid attrition (as a unit in the lexicon) and need to be re-constructed if it were required again (e.g. a year later). Such an effect could reverse the expected direction of any fluency change.

4.5. Conclusion

For this study, internal formulaicity was defined with respect to the holistic nature of an expression in the individual speaker's lexicon and the processing advantage in production. The formulaicity of any recurring unitary expression (RUE) was indicated by its consistent fluency over multiple occurrences. The quantitative analysis of a set of RUEs for one L2 speaker over 15 samples showed that the majority of these expressions were also formulaic. For RUEs with variable fluency, the tendency was for them to show increasing fluency over time. This is consistent with the possibility that such fluency change is associated with developing formulaicity. Although only a single case study, the multiple sampling methodology does demonstrate how fluency behaviour of expressions can be monitored over time, and highlights the importance of looking at individual cases when exploring the acquisition of formulaicity.

However, the qualitative analysis and examples also highlight that the relationship between acquisition of formulaicity and changing fluency is complex. There may be multiple routes to acquisition and the evidence of delivery may be open to different possible interpretations. In addition, even with multiple samples, the initial usage of any expression is not usually observed, and the likely route to acquisition can be hard to determine without this. One way of addressing this challenge would be to introduce new expressions to the speaker and then see how they developed subsequently. Further study is therefore indicated using the same approach to identification of developing formulaicity but with multiple participants and targeted expressions.

CHAPTER 5: Memorising model utterances (Study S3)

Investigating the memorisation and use of multiword utterances by Japanese Speakers of English

5.1. Introduction

5.1.1. *Background*

As the studies S1 and S2 have shown, phonological and unitary coherence are criteria that can be applied practically for identifying expressions that are 'internally' formulaic for a speaker (with fluency as an indicator of phonological coherence). In these studies, the criteria were applied to samples of L2 speech to gain useful information about the numbers and types of formulaic expressions used by the speakers, along with examples of expressions that may be at different points on the way to becoming formulaic. However, in looking at snapshots of current speech (even with multiple samples over time), it is not possible know how these expressions were originally learned and therefore how they came to be at their current state of formulaicity. Each individual's experience of learning and usage is likely to be different and, for any potentially formulaic expression, how it was originally acquired could be unique to that person.

In order to get a better idea of the process by which expressions become formulaic for individual speakers, a useful approach is to introduce novel expressions to L2 learners in a planned manner, and monitor their use over subsequent practice and delivery. This ensures that the provenance of the expression is known and it provides a point of focus for subsequent analysis. It also eliminates the need to apply the 'unitary' criterion when assessing formulaicity. A further benefit of such an approach is that it offers a way of exploring how memorisation of multi-word utterances might be used by learners to develop their stock of formulaic language. As Fitzpatrick and Wray (2006) suggest, there are mixed views on the efficacy of rote memorisation as a way of learning a second language. However, it is an approach used extensively by Chinese students (Dahlin & Watkins, 2000), and research suggests that when used in certain ways, such as combined with understanding (Au & Entwhistle, 1999) or repetition out loud (Ellis & Sinclair, 1996), it can be an effective way of learning a language. Furthermore, being able to remember and reproduce is clearly a feature of learning generally and in specific situations (e.g. giving a presentation, explaining something complex, undertaking a highly predictable transaction) memorising specific utterances verbatim has direct practical benefits for any speaker.

A number of studies have investigated what happens when L2 speakers are given spoken text to memorise and the opportunity to use it later (e.g. Wray, 2004; Wray & Fitzpatrick, 2008). The study by Wray and Fitzpatrick (2008), described in Chapter 2, is particularly instructive in this context because it provides a method for introducing nativelike expressions to learners and monitoring their reproduction over time. While their study did not pose the same range of questions about the acquisition of internal formulaic expressions as asked in this research, it is nevertheless highly relevant. When a learner makes a concerted effort to memorise a useful, well-formed utterance, their subsequent attempt at reproducing it can provide a useful insight into how the words have been stored in the individual's lexicon, as demonstrated in the study by Wray (2004). Furthermore, analysing variations in the use of memorised material, and in the types of deviation from target forms that each speaker introduces, highlights factors that influence the approach to, and success of, memorisation. This chapter reports a replication of Wray and Fitzpatrick's study along with some extensions of the original. Chapter 6 provides some additional analysis of the results with respect to fluency and internal formulaicity.

5.1.2. *Overview of study S3*

The current study replicates Wray and Fitzpatrick (2008), applying their approach to a homogenous group of Japanese Speakers of English (JSE). Some minor variations were introduced (see Section 5.2.2.1) due to the location of the study (Japan rather than the UK). The aim of the study was to gain a detailed insight into how Japanese learners based in Japan were able to memorise and reproduce nativelike text that expressed idea they wanted to say. In particular, the study sought to understand how the individuals differed in their approach, the kinds of deviations that were made and features of the utterances themselves that made them susceptible to change.

In addition to these aims (which mirror those of the original), the current study also sought to explore certain aspects of the memorisation and reproduction that related to the acquisition of formulaic expressions.

The research questions were:

1. What variables between learners might account for any differences in their success in memorising and recalling the nativelike utterances they were given?

2. How does the context of reproduction (i.e. real or practice conversation) affect the recall and accuracy of the target utterances?
3. What sort of changes are made when memorised material was not fully reproduced?
4. What features of the nativelike utterances influence the nature and extent of changes that occurred?
5. How do participants segment their target utterances when reproducing them?

5.2. Method

5.2.1. *Participants*

Fourteen adult Japanese learners of English living in Japan were recruited from an English school and several companies in Tokyo known to the researcher. There were 4 males and 10 females, aged from 32 to 64. Participants were volunteers interested in further developing their English and each completed a questionnaire and took a series of proficiency and language aptitude tests before starting the main study. A list of the participants (pseudonyms) and their basic test results is given in Table 5.1.

Proficiency was measured by two simple vocabulary tests: V_YesNo (Meara & Miralpeix, 2016) and Lex30 (Fitzpatrick & Clenton, 2010; Meara & Fitzpatrick, 2000) both accessed from the Lognostics website (Meara, 2014). The YesNo test gives an estimate of the test-taker's receptive vocabulary size. It presents a list of words each from a given frequency band. Test-takers indicate whether or not they know each word and from this an estimate of their vocabulary size (up to 10,000 words) is calculated. Non-words are also included in the list in order to check the reliability of the participant's judgements. Lex30 is a test of productive vocabulary which presents a list of 30 words and invites test-takers to write down up to four other words associated with each. The total number of different infrequent words (i.e. those outside of the 1000 most frequent words) written down gives an indication of the taker's productive vocabulary. Both test types have been shown to reliably correlate with other measures of proficiency (Fitzpatrick & Meara, 2004; Meara & Jones, 1988).

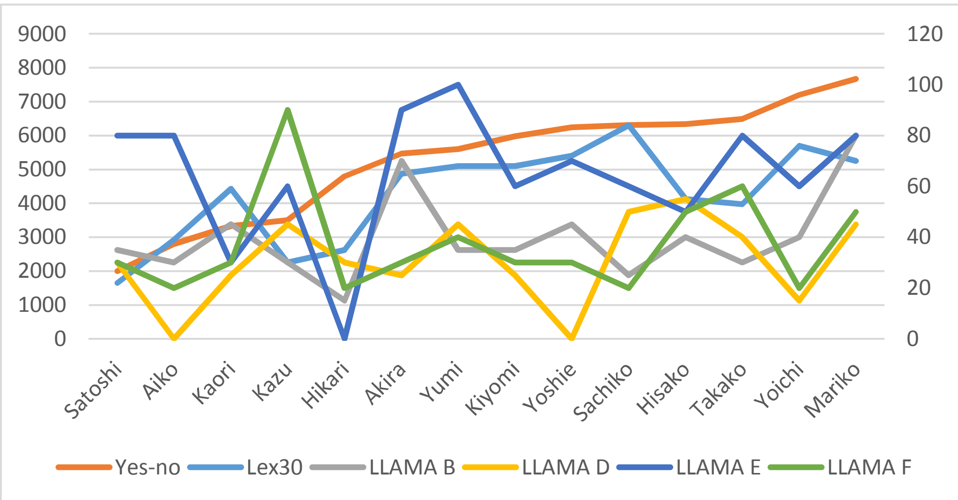
Table 5.1: List of participants and test scores (ordered by YesNo score)

Name	Sex	Age	Yes No	Lex 30	LLAMA B	LLAMA D	LLAMA E	LLAMA F
Satoshi	M	40-49	2000	22	35	30	80	30
Aiko	F	40-49	2800	39	30	0	80	20
Kaori	F	40-49	3333	59	45	25	30	30
Kazu	M	30-39	3500	30	30	45	60	90
Hikari	F	50-59	4800	35	15	30	0	20
Akira	M	30-39	5466	65	70	25	90	30
Yumi	F	40-49	5600	68	35	45	100	40
Kiyomi	F	30-39	5976	68	35	25	60	30
Yoshie	F	40-49	6240	72	45	0	70	30
Sachiko	F	50-59	6310	84	25	50	60	20
Hisako	F	60+	6340	55	40	55	50	50
Takako	F	40-49	6486	53	30	40	80	60
Yoichi	M	50-59	7202	76	40	15	60	20
Mariko	F	30-39	7674	70	80	45	80	50

In addition, the LLAMA tests of language aptitude (Meara, 2005b) were administered to all participants. These four tests were:

- Llama B: a simple test of memory for novel vocabulary items
- Llama D: a test of aural memory for sounds from a novel language
- Llama E: a test of ability to infer relationship between the sounds and symbols.
- Llama F: a simple grammatical inferencing test

Figure 5.1: Participant scores on the tests (ordered by YesNo score)



As can be seen from Table 5.1 (and illustrated graphically in Figure 5.1), the participants represented quite a wide range of proficiencies (as measured by YesNo and Lex30) and there was also considerable variation in their language learning aptitude (as measured by the LLAMA tests).

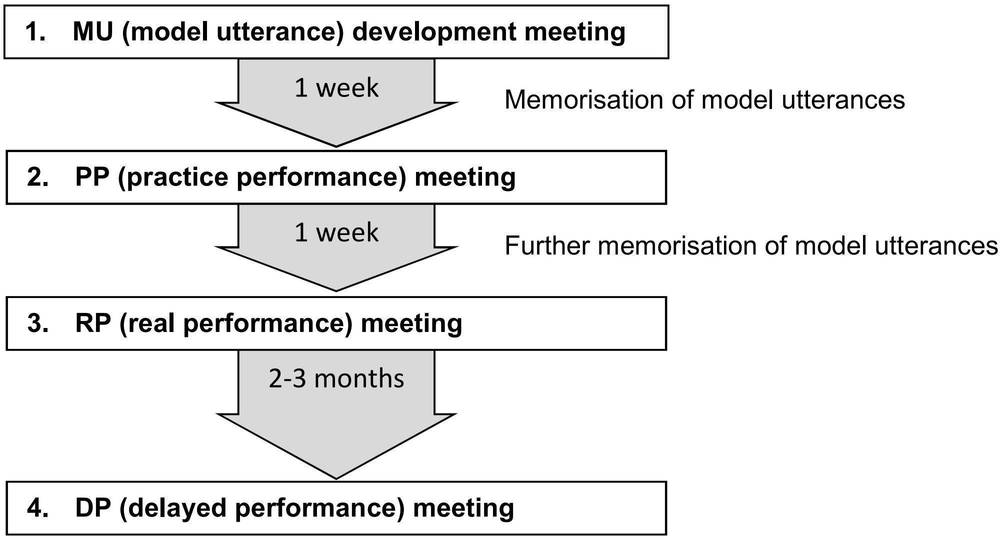
In addition to the tests, a questionnaire adapted from the original study was given in order to collect data on English learning experience and daily usage, and preferred learning methods and goals. Five of the participants (Sachiko, Yoshie, Hisako, Yumi and Mari) used English fairly frequently as part of their work. Five of the participants (Takako, Sachiko, Akira, Aiko and Yumi) had experience of living in an English-speaking country for a year or longer. All participants were well motivated to improve their English and saw the research as a good opportunity to practise and use their spoken English.

The questionnaire also asked about participants' beliefs and attitudes towards language learning. This included questions adapted from inventories on language learning beliefs and attitudes (Beebe, 1983; Ely, 1986; Horwitz, 1987). The questions covered four main areas: confidence (belief that they would improve), fit (how closely their learning beliefs corresponded to the procedure to be used in the study), risk (inclination to take linguistic risks, such as using a word without fully understanding it) and native-orientation (inclination to sound like a native English speaker). Overall, participants responded at a level of medium to high in all of these areas. The exceptions were Aiko with a slightly lower fit and attitude to risk, and Yoichi with a lower native-orientation. Overall, the responses did not indicate any great differences between the participants. So, apart from adding support to the observation that all participants were well-motivated and engaged in the study, the questionnaire results were not used further.

5.2.2. ***Procedure***

The overall procedure was for each participant to memorise a set of nativelike utterances that they would subsequently use in a real conversation with a native speaker. For each participant, a conversation cycle consisted of them taking part in each of the stages shown in Figure 5.2.

Figure 5.2: The conversation cycle



In the MU meeting, the researcher explained that the participant would, at a later stage, be meeting a native English speaker (recruited by the researcher) to have a conversation. The participant was prompted to think about specific things they wanted or expected to talk about in that conversation, and the researcher then provided a natural, nativelike way of expressing them[22]. These model utterances were personal to the participant and constituted new ways of expressing ideas, thoughts or questions that they wished to convey. Some examples include:

- *My bakery teacher and I are planning to start an NPO to make bread using plastic bags.*
- *Being stared at all the time was a strange feeling but probably a good experience to have at least once in your life.*
- *Have you ever come across a person who you found really difficult to communicate with?*

The conversation was recorded and, after the meeting, the researcher made a spoken recording of each of the model utterances, creating an mp3 file with each utterance separated by 3 seconds of silence. This was then sent to the participant by e-mail. The participant was instructed to repeat and memorise all of the utterances

[22] Despite the apparent limitations of only preparing one half of a conversation, this design has been shown effective as a conversation fluency aid when used by a non-speaking person with cerebral palsy (Wray, 2002b; Wray & Fitzpatrick, 2009).

exactly as recorded in order that they could use them in conversation. No written version was provided and they were specifically asked not to write anything down.

The next stage was the PP meeting. In this, the participant met the researcher again and had a practice conversation in which they attempted to include all the model utterances appropriately. At the end of the meeting, the researcher reviewed the utterances with the participant and answered any questions. Following this practice performance, participants had a further opportunity to learn the set of model utterances before having the real performance (RP) conversation with the native speaker. In most cases, the interval between each meeting was one week.

Each participant did two or three different conversation cycles. In the MU meeting of the first cycle, the participants were fully briefed, given written instructions (in Japanese and English) and informed consent was obtained. At the start of a new cycle, the researcher got feedback from the participant about how they found the previous real performance. After the final real performance, participants were sent a questionnaire to find out their thoughts on the process. Two or three months after the final real performance, the delayed performance (DP) meeting took place. In this, the researcher again met each participant to see, without prompting, how well they could reproduce the model utterances from their two or three cycles.

5.2.2.1. Real Performance

An important feature of the study was that participants should be highly motivated to memorise the model utterances as completely and accurately as possible. In addition, when they performed these utterances in the real performance (RP), the situation needed to be as pressurised and unpredictable as possible in order to test how the utterance would actually be used in a real situation. In the original, where participants were L2 English learners based in the UK, motivation, relevance and authenticity were provided by having them prepare for an interaction that they were going to have anyway as part of their daily life (e.g. taking a pet to the vet). For participants based in Japan, the opportunities for everyday transactions in English were felt to be too limited. So, the participants instead took part in a conversation with a native speaker organised by the researcher.

This RP conversation was designed to be as authentic as possible. For the first cycle, the native English speaker was someone the participant had never met before. For the second cycle, participants met with the same native speaker i.e. they had met them precisely once before. This allowed two different types of conversation and enabled participants to generate new utterances for the second cycle based on their knowledge of him/her from the first meeting. For the four participants who did three cycles, the final meeting was with a new native speaker.

Five different native speakers were used, four male and one female. Three were British, one American and one Canadian. Four of the speakers lived in Japan and one was in the UK; conversations with the UK speaker took place via Skype. The native speakers were not familiar with the details of the study and were simply told "I'd like you to meet this person and have a conversation. Try to get to know them better". Feedback from the native speakers at the end of the study confirmed that they remained unaware of the memorisation process taking place and that they continued the conversation in whatever way they felt was natural at the time.

While the RP process represented a significant difference from the original, it was felt that meeting an unfamiliar native speaker and engaging them in conversation in an interesting, informed way would be sufficiently motivating and challenging for these participants. The ability to talk about family, work, holidays, interests or recent news items in a natural and sophisticated way is a genuine communicative outcome, and opportunities for learners in Japan to do this are generally quite limited. Feedback from the participants and their obvious commitment to the task suggested that they saw this as a good opportunity for conversation which was relevant to their communicative aims for English.

5.2.3. ***Analysis***

5.2.3.1. Transcription and recording

All stages of each conversation cycle were recorded on a Sony IC recorder and the relevant parts of the conversations were transcribed with the support of software Transcriber AG (DGA, 2014). The output of the different stages is illustrated in Table 5.2. Stage I shows the participant's initial attempt at expressing an idea they wanted to communicate with the native speaker. MU represents the Model utterance, and PP, RP and DP what was actually said in the Practice, Real and Delayed performances.

Table 5.2: Takako-2-14 (14th utterance of her second cycle)

I	I want to say it is normal thing in British er for families to gather around in you know Christmas time – but he is in Japan (.) so I don't know ...
MU	I understand that British people normally like to get together with their families at Christmas.
PP	You know English people usually get together with their families at Christmas.
RP	I understand English people like to get together with family in the- [Yeah I think…]
DP	You celebrate Christmas with all your family.

In practice, the participants' actual speech included many fillers, pauses, repetitions, false starts and reformulations. In the original study, these were removed to produce a 'cleaned up' version of the reproduced utterance for the purpose of the analysis. A similar process was applied in the current study except that an indicator of each dysfluency point was also retained. This was to facilitate the additional analysis of fluency and segmentation (described further in Chapter 6).

5.2.3.2. Quantification

In order to compare individual differences in the extent to which utterances were attempted during the real and practice performances and how close they were, two measures were defined, identical to those in the original study.

$$\text{Propensity to attempt} = \frac{\text{total no. of utterances attempted}}{\text{total no. of utterances}}$$

$$\text{Closeness of utterance} = \frac{\text{no. of words with same form \& function as MU}}{\text{total no. of words in utterance}}$$

The propensity to attempt gives an indication of the proportion of model utterances that were attempted at practice or real performance. The closeness measure indicates how close an attempted utterance was to the original model. For example, in the case of Takako-2-14 given in Table 5.2, nine out of 15 words were correct at PP (closeness = 0.60) and eight at RP (closeness = 0.53).

5.2.3.3. Classification of deviations

Each attempted utterance was also analysed to determine what deviations occurred. As in the original study, deviations were classified into three main categories:

Morphological (i.e. involved a particle, function word, article or inflection), Lexical (involving a content word or adjunct) or Phrasal (involving a phrase or change in word order). Each main category was further divided into a number of sub-categories.

Broadly, the morphological deviations consisted of inflections (e.g. changing *families* to *family*, *get* to *gets*) and insertions, omissions or substitutions of grammatical words (i.e. particles, articles and function words, including prepositions). Lexical deviations were any non-morphological changes involving a single word. These were sub-classified as adjuncts or as content words, and they could be either inserted, omitted or substituted. Adjuncts are defined according to Eggins (2004) as words that contribute some additional (but non-essential) information to a clause either by providing a mood or comment (e.g. *perhaps*, *honestly*) or by linking it to another clause (e.g. *but*, *because*). Phrasal deviations were any multi-word sequences that are inserted, omitted or substituted, or any change in word or phrase order within the utterance.

Table 5.3: Summary of deviations for Kazu-1-5 (RP)

MU: I also manage the vessels that bring corn from the US and other countries.
RP: We also manage vessels bring in corn from the US

Deviation	Category	Sub-category	Type
I -> we	LEXICAL	Content word substitution	Nativelike
the vessels -> vessels	MORPHOLOGICAL	Article omission	Non-nativelike
that	MORPHOLOGICAL	Function word omission	Non-nativelike
bring corn -> bring in corn	MORPHOLOGICAL	Function word insertion	Nativelike
and other countries	PHRASAL	Phrase omission	Nativelike

Table 5.3 illustrates the deviation analysis for one RP response (Kazu-1-5). This shows one lexical substitution (*We* for *I*), two morphological omissions (the article before *vessels,* and the function word *that)*; a morphological insertion (of the function word *in)*; and a phrasal omission (of *and other countries*).

Each deviation was also classified as to whether it was nativelike or non-nativelike. A nativelike change is one that might have been made by a native or highly competent speaker of English. For example, the substitution of '*We*' for '*I*' is appropriate in the context, whereas the omission of the article *the* is not. The judgements were made by the researcher and in most cases were unproblematic as they involved clear grammatical errors. In cases where there was doubt, a second opinion from a native speaker was sought.

5.3. Results

In total, there were 32 cycles and 404 model utterances (average 12.6 utterances per cycle). Of the 404 statements, 340 were statements while 64 were questions. All but 12 of the model utterances were attempted at least once in the practice or real performance. The 12 not attempted in either were still included in the totals since participants had been instructed to learn all the utterances and, in a free conversation, it was deemed that all utterances were potentially relevant and usable.

The total number of words was 5632 (average 13.27 words per utterance). Table 5.4 gives the number of model utterances memorised by each participant and the total number of words contained within these.

Table 5.4: Profile of the data set

Name	Takako	Hikari	Kaori	Sachiko	Kiyomi	Kazu	Yoichi
# cycles	2	3	2	3	2	2	2
# MUs	27	39	26	39	26	21	26
# words	393	530	351	548	323	280	342

Name	Satoshi	Akira	Yoshie	Aiko	Hisako	Yumi	Mari
# cycles	2	2	2	3	2	3	2
# MUs	27	24	23	41	24	37	24
# words	301	284	317	531	310	498	354

5.3.1. *Individual differences in use of memorised material*

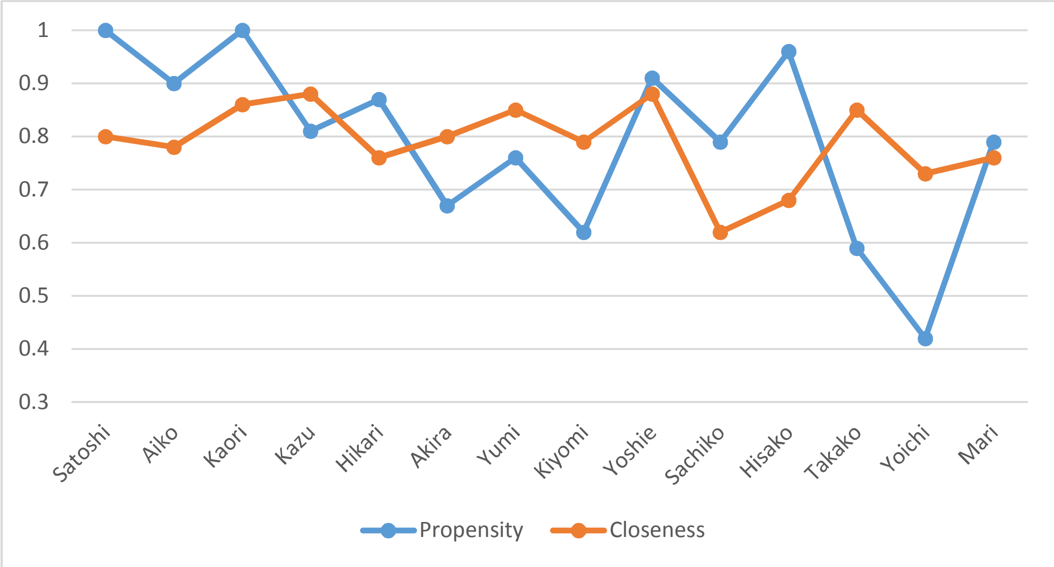
Table 5.5 gives the overall proportion of utterances attempted and the average closeness scores for each participant at the practice (PP), real (RP) and delayed (DP) performances. Although most of their utterances were attempted during the practice performance, the 14 participants varied considerably in their propensity to attempt during the real performance, as illustrated in Figure 5.3. A test for correlation using Pearson's coefficient indicated no correlation between the proportion of model utterances attempted at RP and the Lex30 scores, $r(12)=-0.471$, $p=0.089$, but there was a significant inverse correlation with the YesNo vocabulary scores, r(12)=-0.588, p=0.027. Using Spearman rank correlation tests, no significant correlations were found between the proportion of utterances attempted by an individual at RP and any of the language aptitude tests.

For closeness of the attempts to the model utterance, there was some variation between participants, but no significant correlation with either YesNo or Lex30, nor was there any significant correlation between the closeness measure and any of the language aptitude tests.

Table 5.5: Propensity to attempt and average closeness by participant

	Propensity to attempt			**Closeness**		
	Practice	**Real**	**Delayed**	**Practice**	**Real**	**Delayed**
Satoshi	1.00	1.00	-	0.76	0.80	-
Aiko	0.95	0.90	0.37	0.77	0.78	0.54
Kaori	0.88	1.00	0.46	0.73	0.86	0.38
Kazu	0.95	0.81	0.48	0.93	0.88	0.32
Hikari	0.87	0.87	0.31	0.70	0.76	0.49
Akira	0.96	0.67	-	0.84	0.80	-
Yumi	0.89	0.76	0.19	0.71	0.85	0.39
Kiyomi	0.88	0.62	0.77	0.83	0.79	0.56
Yoshie	0.96	0.91	0.48	0.80	0.88	0.38
Sachiko	0.90	0.79	-	0.68	0.62	-
Hisako	0.96	0.96	0.67	0.71	0.68	0.31
Takako	0.96	0.59	0.67	0.71	0.85	0.34
Yoichi	0.88	0.42	-	0.83	0.73	-
Mari	0.96	0.79	0.71	0.70	0.76	0.51
Mean	0.93	0.79	0.51	0.76	0.78	0.42
St Dev	0.041	0.169	0.192	0.077	0.082	0.099
Range	0.87 - 1.00	0.42 - 1.00	0.19 - 0.77	0.68 - 0.93	0.62 - 0.88	0.31 - 0.56

Figure 5.3: Propensity to attempt and average closeness at RP by participant



5.3.2. ***Relationship between performance stages***

5.3.2.1. Practice and real performance

As shown in Table 5.5, there were clear differences between performances at PP and RP. As predicted by the results of the original study, the overall mean propensity to attempt at RP (0.79) was lower than at PP (0.93) and a paired samples t-test revealed that this was a significant difference, $t(31)=3.195$, $p=0.003$. There was also far greater variation at RP. In particular, there were two exceptional conversations (Yoichi-1 and Takako-1) where utterances were hardly attempted at all (1/12 and 2/12 respectively). However, even without these, there was still a significant difference between RP and PP. This suggests that, as in the original study, it was more challenging for participants to find opportunities to use the model utterances in the real conversation with the unfamiliar native speaker than in the practice.

Regarding Closeness to the MU, there was little overall difference between the two performance types. This was so whether considering all utterances attempted (as shown in Table 5.5) or just the 302 utterances that were attempted at both RP and PP. So, contrary to the results of the original study, RP utterances were no less accurately reproduced. Indeed, looking at deviations, it was found that the mean number of deviations per utterance at PP (2.09) was higher than at RP (1.87) and this difference was significant, $t(301)=2.350$, $p=0.019$. According to their feedback, all participants continued memorising the utterances after the PP meeting. A number commented that they felt the utterances were not yet fully memorised at PP and further time was required. Examples such as that of Kiyomi below suggest that further consolidation of learning took place after the practice.

Table 5.6: Example from Kiyomi-2-10

MU	Putting on a kimono is a surprisingly time-consuming process.
PP	Mm the kimono is (0.3) um a surprisingly mm (2) consum- <laugh> (4) [OK, so-] time consum- time con (1) time time-spending time-consump- spending [Right, so it takes a long time]
RP	Yes, putting on my kimono is su- (0.7) very- surprisingly time-consuming process. I need one hour

Note: (n) = pause of n seconds. [Square brackets] indicate native speaker interlocutor's speech

Based on the Pearson correlation coefficient, there was a moderate correlation between individual performance at PP and at RP in both mean closeness and mean deviations/word, and in the latter case this was significant, $r(13)=0.541$, $p=0.045$. However, contrary to the findings in the original study, there was no evidence of a correlation between PP and RP in the propensity to attempt.

Overall the results suggest that participants attempted fewer utterances at RP than PP in general, but that the extent of the effect was not consistent across the set of participants. For utterances that were attempted, however, the closeness and degree of deviation remained similar at PP and RP.

5.3.2.2. Delayed Performance

The delayed performance meeting between the participant and the researcher took place 2-3 months after their final real performance. This was possible for 10 of the participants. Overall, the propensity to attempt an utterance and the closeness of the attempts to the original model were much lower than in either the practice or real performance, as shown in Table 5.5. However, two participants (Takako and Kiyomi) were able to attempt more utterances in the delayed performance compared to the real performance (although not compared to the practice performance), despite the delay. Both indicated that they had in fact continued to use some of the expressions outside of the study.

Looking more closely at the utterances that were attempted at DP (139 out of a possible 288 for the 10 participants), there were just 22 that had a closeness score above 0.7. The nature of these utterances is explored further in the discussion, Section 5.4.

5.3.3. ***Types of deviations***

Overall, there were 1698 deviations across the practice and real performances where utterances were attempted (694 attempts). The overall mean deviation per utterance was 4.20 and the mean deviation/word was 0.183. Deviations classified as Morphological accounted for 31.7% of the total while 39.9% were Lexical and 28.4% were Phrasal. This represents a difference from the original study in which Phrasal deviations were most common.

Deviations were also assessed as to whether they appeared to be nativelike changes or not. Of the total number of deviations, 55.3% were classified as nativelike. However, within the three categories of deviation, the percentages of nativelike deviations varied considerably. For Morphological deviations only 25% (136/538) were nativelike while for Lexical deviations the figure was 77% (524/677), while Phrasal deviations were somewhere in the middle (275/483 =57%). The low proportion of nativelike deviations was a similar finding to the original study and reflects the relative lack of flexibility of morphological choices compared to those of lexical and phrasal items.

5.3.3.1. Sub-categories of deviations

Each category of deviation was divided further into sub-categories. Table 5.7 gives the sub-categories for the morphological deviations. This shows that a large number of these were insertions and substitutions, the latter tending to be somewhat more nativelike than the other morphological deviations.

Table 5.7: Sub-categories of morphological deviations

	Total	**% nativelike**
Inflections	99	25%
Insertions	195	23%
Omissions	81	26%
Substitutions	164	36%

Table 5.8: Sub-categories of lexical deviations

		#	% nativelike
Content word	Insertion	86	71%
	Omission	113	73%
	Substitution	160	70%
	Total	359	71%
Adjunct	Insertion	256	85%
	Omission	33	82%
	Substitution	30	67%
	Total	310	78%

For lexical deviations, given in Table 5.8, adjuncts represent nearly half (46%) of the deviations and the majority of these are insertions. This again is a similar result to the original study. Most of the adjunct insertions were nativelike and a large proportion of these (235/256) were added at the beginning of the utterance.

Finally, for phrasal deviations given in Table 5.9, there was a comparatively large number of substitutions (178) and this was the third largest subcategory of deviations overall (after adjunct insertions and morphological insertions).

Table 5.9: Sub-categories of phrasal deviations

	Total	% nativelike
Word order	56	52%
Insertions	156	66%
Omissions	89	58%
Substitutions	178	48%

An interesting group of deviations were those involving adjuncts or adjunct phrases at the beginning of the utterance. There were 313 such insertions (18% of all deviations) which involve the participant adding a word or phrase at the beginning in order to integrate the utterance into the conversation. These insertions were generally nativelike (85%) and 75% involved just a single word.

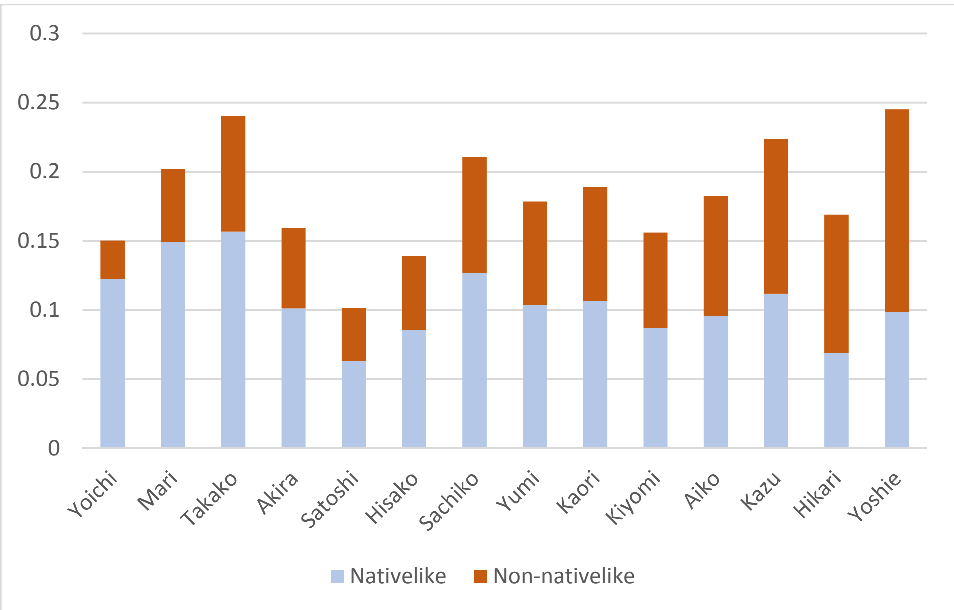
Some of the model utterances (50/404) had an initial adjunct. Around half of these were lexical (e.g. *maybe*) and half phrasal (*to be honest*, *at the moment*). Of the 87 attempts at using these in the practice and real performances, 39 were repeated exactly, and there were 48 deviations. The deviations were generally nativelike (71%) and were mainly single word adjunct omissions (18 attempted, 17 nativelike) or substitutions (12, 7). Where phrasal substitutions were attempted, these were the least successful (9, 2).

Overall then, adding an initial adjunct or amending an existing one occurred for over half (363/694 = 52%) of all the attempts at repeating an utterance during the performances.

5.3.4. ***Deviations and individual differences***

Figure 5.4 shows the average number of deviations per word over the utterances that were attempted for each participant and the proportion of these that were deemed 'nativelike'. As the chart shows, there was considerable variation in the number of deviations that different participants introduced and in the proportion of those that were nativelike. However, there were no obvious correlations between these variables and any of the proficiency or aptitude measures. Further, there was also no correlation between the number of deviations per word and the proportion of deviations that were nativelike. Overall, this suggests that participants have an individual propensity to deviate from model utterances which is not related to their ability to do this successfully, or to their level of proficiency or language aptitude.

Figure 5.4: Deviation profiles for each participant



Participants ordered from highest proportion of nativelike deviations to lowest

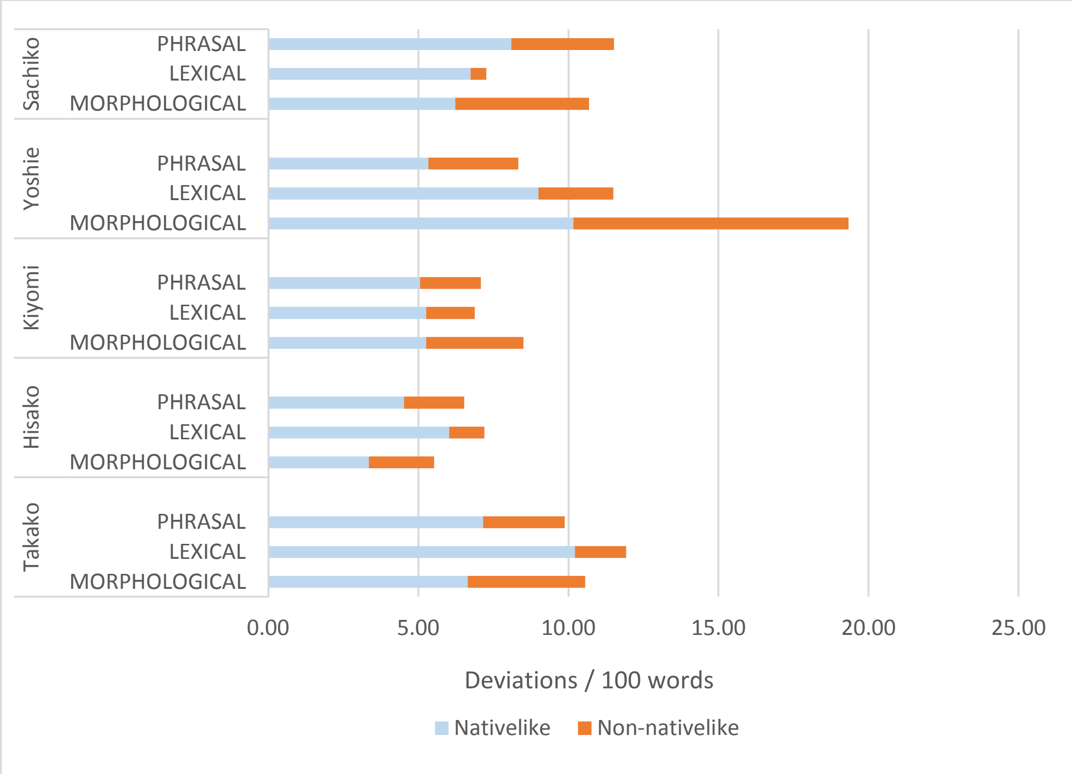
For example, the participants with the highest number of deviations/word (Takako, Kazu and Yoshie) varied considerably in the proportion of these that were nativelike (64%, 49%, 38%, respectively). These individual variations in deviation profile support results in the original and suggest that people memorising utterances vary in their overall approach to how they learn and use them in real conversation.

5.3.4.1. Deviations and risk-taking

A further idea in the original study was that deviation level may be predicted by some kind of relationship between proficiency and the tendency to take ‘linguistic risks’. In particular, Wray and Fitzpatrick (2008) suggested that the gap between receptive and productive knowledge represents what a person believes they know but cannot reproduce reliably. They conjectured that, for participants at a similar level of proficiency, this difference might predict their tendency to deviate from the original utterances in a non-nativelike way.

To check this possibility, a measure of the difference between receptive vocabulary size (YesNo) and productive vocabulary (Lex30) was calculated using standardised versions of each measure. This difference measure (R-P difference) was found to have no significant correlation with a participant’s propensity to produce non-nativelike deviations (as measured by mean number of non-nativelike deviations/word on the real performance). This finding is further supported by comparing the detailed deviation profiles for a sub-group of 5 participants with similar receptive vocabulary (YesNo scores between 5500 and 6500). These are shown in Figure 5.5, in order of increasing R-P difference. These illustrate that there is considerable variation between the participants at a similar receptive proficiency, but this does not seem to be in any way related to the gap between their receptive and productive abilities.

Figure 5.5: Detailed deviation profiles for 5 participants with similar proficiency



5.3.5. ***Characteristics of targets***

Model utterances were developed according to what the participants wanted to convey in their conversation. No particular attention was given to controlling the kind of structures or vocabulary used. The main criterion was that it was a natural way of expressing what the participant wanted to say and was not something they had been able to say themselves before being told it. The utterances were not graded according to a person's levels or restricted in other ways. In some cases there were new words or phrases or constructions, in others it was a new way of organising words that were already familiar.

Utterances therefore varied in terms of type (statement or question), length (in words) and complexity. It was possible therefore to see the extent to which these features influenced the number and type of deviations. Using Spearman rank correlation tests, length of the utterance (in words) was found to correlate significantly with deviations/word (r_s=0.187, $p<0.01$) and inversely with closeness (r_s=-0.158, $p<0.05$). In other words, longer utterances tended to have a higher rate of deviation per word and were less accurately recalled. Utterance length did not seem to have any effect on the propensity to attempt, however. Questions had significantly fewer deviations and were more closely recalled than statements, but this may have been influenced by the fact that questions were generally slightly shorter than statements (with means of 11.4 words and 13.6 words per utterance respectively).

Table 5.10: Analysis of utterances by number of clauses

Clauses	Count	Mean word length	Propensity to attempt	Closeness	Deviations per word
1	167	10.9	0.85	0.810	0.130
2	186	14.2	0.86	0.802	0.139
3	47	17.2	0.86	0.784	0.132
4	4	20.5	1.00	0.764	0.161

The number of clauses for each utterance was recorded as a means of analysing the effect of utterance complexity, and these ranged from one to four clauses. As Table 5.10 shows, there was little difference between the different types in terms of propensity to attempt, closeness, and deviations per word. This was despite the fact that utterances with more clauses tended to be longer. So, there is a suggestion that having more clauses may offset the effects of word length on memorisation success. This may be because clauses are easier to learn as sub-parts and this facilitates ‘chunking’.

5.4. Discussion

The main aims of this study were to explore what happens when learners are given natural, relevant utterances to memorise and repeat. There were two main areas of focus. The first was to see how individuals varied in their performance and approach to the memorisation and to consider the factors that might influence this. The second was to explore features of the utterances themselves and how these may have impacted on performance. A third outcome of the study, to be addressed in the next chapter, concerns what the results can tell us about how multi-word sequences are committed to memory.

5.4.1. *Memorisation and individual differences*

Overall, the results support the findings in the original study in suggesting that the way memorised phrases are used (in terms of propensity to attempt, closeness, amount and type of deviation) does not seem to be directly related to proficiency or aptitude. However, there are clearly individual differences in participant performance with regard to memorising and reproducing utterances, and it is likely that there are a number of factors involved in this. In particular, they include the time dedicated to the memorisation, the approach taken and performance factors.

5.4.1.1. Time spent on memorisation

In order to accurately memorise the phrases and be able to reproduce them in conversation, the participants needed to spend sufficient time on each utterance, with its level of difficulty a factor determining the time needed. For this study (and the original), the assumption was that the participants would have sufficient time to learn the utterances. However, there were examples where utterances were not reproduced accurately in the practice performance (PP) but were in the real performance (RP), suggesting that insufficient time was given initially by some participants. Indeed, it appeared that the time put in to memorising the utterances varied widely among the participants. According to the nine participants who gave feedback on this, a third said they spent around 2-3 hours in the week on learning all the utterances for a cycle (e.g. 20 mins / day or 10-15 minutes / phrase), another third said they spent a good part of the week on it (20 - 30 hours) while the remainder were somewhere in between. One factor that may be thought to affect the time put in is motivation. However, feedback from the participants and other evidence of their level of engagement suggest that they were all highly motivated to do their best and the task was relevant and worthwhile to them. So, the variations in time input appear more to do with availability and each individual's approach to the task. It is also worth noting that the participants who indicated that they spent a lot of time (Aiko, Kiyomi and Yumi) did not perform appreciably better. So, time spent (beyond a certain minimum) may not be such an important factor on its own.

5.4.1.2. Approach to the memorisation

The way each participant approached the memorisation is a potentially important factor in their success. Most participants indicated that they followed a basic method whereby they listened to each utterance a number of times and then repeated it out loud. Within this, some mentioned specifically focussing on getting the right speed (e.g. by shadowing) or rhythm and intonation (e.g. by using their hands and gestures). Some visualised the model sentences in written form or used a specific word to mentally 'tag' the utterance. Beyond the basic method, some inter-related factors that may be significant are the participant's attention to detail, their ability to gauge when they have memorised something sufficiently and the degree of certainty they require.

Attention to detail is important when memorising an utterance for completely accurate reproduction. This is particularly so for ‘weak forms’ within the utterance such as function words and inflectional morphemes which have low salience (Ellis, 2006), as discussed in Section 2.5.2. Learners may be more prone to making errors in reproducing these forms because they do not notice them. This may be especially so when learning utterances through listening and repetition only (as in this study). Analysis of the morphological deviations may therefore provide useful insights into aspects of attention. As well as low salience, these deviations have less variation in terms of correct usage compared to phrasal or lexical changes. Indeed, for all participants the proportion of morphological deviations that were non-nativelike was much higher than for other deviation types. However, as it turns out, there was no relationship between these morphological deviations and proficiency, with some higher proficiency participants (e.g. Yoshie) also having a high level of morphological errors. It seems likely therefore that attention to detail is an individual trait that varies across participants and is not directly related to proficiency.

A possible factor in this, raised in the original study and described in Section 5.3.4.1, may be each participant’s approach to linguistic risk-taking. This is the idea that participants may vary in terms of the point at which they decide they have done enough to memorise the utterance. This is partly about what they consider to be ‘good enough’ in terms of the probability that they will be able to reproduce the utterance accurately (risk aversion), and partly the degree to which they can accurately judge whether they have reached that point (self-awareness). In the original study, the hypothesis was that this self-awareness may be linked to a difference between productive and receptive English knowledge. The reasoning was that participants use their recognition and familiarity with the components of an utterance to judge whether they have memorised it sufficiently. Participants with high receptive knowledge compared to productive (R-P difference) may be overconfident (i.e. have lower self-awareness) in their ability to reproduce familiar words and structures, and so apply less time and attention to the utterance than they need to. However, the results did not show that the gap between productive and receptive ability is related to the deviation profiles observed in their real performance, even when differences in proficiency were taken into account.

It is not possible to be sure why this predicted pattern wasn't observed, but one possibility that could leave the underlying risk hypothesis intact is that the estimate of receptive knowledge was inaccurate. The measure used was the YesNo test, and some participants scored lower on this test than the level of their spoken performance and results in other published tests would suggest. For example, Kazu had a predicted vocabulary of 3500 on the YesNo but had a TOEIC score of 840 (upper intermediate) while Sachiko with a similar TOEIC score had a predicted vocabulary size of 6310. The YesNo test relies on test-takers making an appraisal of whether they know a word and this may be sensitive to individual features such as risk-aversion just as much as an individual's approach to memorisation is. Indeed, Kazu marked some items on the test as unknown but used the same words spontaneously and appropriately in his recorded conversations in the study, suggesting he did in fact know those words. The way that a potential under-estimate of receptive knowledge would affect the analysis is complex, however. On the one hand, without the under-estimate, there would be a higher R-P difference (assuming the productive score was accurate) meaning that Kazu's relatively high propensity for non-nativelike deviations would be more in line with the hypothesised relationship. On the other hand, if the cause of the under-estimate is related to Kazu's risk aversion, this is at odds with the relatively high number of deviations he produced generally.

5.4.1.3. Individual differences and performance

In the real performance (RP), participants met a native speaker for the first (or second) time to have a genuine conversation. At the same time, they were trying to weave in particular utterances exactly as they had memorised them (and without the native speaker knowing what they were trying to do). It might be expected that the additional stress of this situation (compared with the practice) would affect their performance on the task, and that the extent of this effect might be different for different individuals. In the original study, Fitzpatrick and Wray (2006) did find an effect of performance type, with lower propensity to attempt and lower closeness at RP compared with PP. However, they also found a high level of correlation between PP and RP results across their participants, suggesting that the effect of the real situation on performance was fairly uniform.

The results in the current study were slightly different in two respects. Firstly, the average closeness at RP was not significantly lower than that at PP. Indeed, if anything, there were fewer deviations/word at RP. For the RP, there was an additional week of practice compared to PP, and the beneficial effect of attempting to retrieve the expressions at PP (Kornell & Vaughn, 2016). While this was also true of the original study, the extra practice may have had a greater effect in the current study because participants (as working adults in Japan with limited free time) were somehow less prepared for the PP than those in the original (as university students in the UK). Another possible difference is that participants in the current study had more flexibility in choosing which utterances to use in the real performance (as part of a relaxed conversation rather than a specific real-life situation). This would allow them to select utterances they were more confident of at the time, resulting in fewer deviations.

A second difference in the results compared to the original study relates to the propensity to attempt. While both studies found that propensity to attempt an utterance was lower at RP compared with PP, the current study did not find any degree of correlation between the PP and RP results in this measure. There was evidence of individual differences in the effect of the real performance situation on the propensity to attempt. For example, four participants (Hisako, Hikari, Kaori and Satoshi) attempted the same number or more utterances at RP, while others (such as Takako and Yoichi) attempted considerably fewer, suggesting that they were more affected by the real situation (in terms of being able to access utterances) than other participants were. This effect was particularly pronounced in the first cycle as it was the first time they had met the native speaker and taken part in the real performance. Indeed, some participants did indicate that they were nervous about the real performance, and most said they preferred the second occasion they met the native speaker (as they already knew them and understood the process better). So, overall the results and feedback support the idea that the performance situation affects recall of memorised utterances and the extent of this effect varies across individuals.

5.4.2. *Where do the deviations come from?*

In order to better understand how the approach of the participants and the nature of the utterance influence the memorisation, it is useful to consider the range of deviations and how they may have arisen. In this study, participants were explicitly instructed to memorise and use all the utterances exactly as presented. At the same time, they were told to have a natural conversation with the native speaker. So, one source of deviations may relate to how participants chose to resolve these potentially conflicting requirements.[23] Two relevant aspects of this are how the participants managed the conversation and how they chose to link the utterance to the conversation flow. Another important source of deviations, is the utterances themselves and certain features that may have made memorisation and reproduction more difficult.

5.4.2.1. Conversation management

In order to successfully use the utterances in their real conversations, the importance of conversation management became obvious to many participants as the study progressed. From feedback at the end of the first cycle, most participants talked about the need for them to steer the conversation so that they had the chance to use all their model utterances as given. This was recognised as a challenge (Yumi: "I have two persons in my head – I have to follow the conversation and I also have to lead") and led some participants to take more control in the second meeting (Yoshie: "For the second session, I positively controlled the conversation").

For some participants, as illustrated by Satoshi in Table 5.11, this meant 'forcing' the utterances in verbatim even if the context or the form was not exactly appropriate. For others, like Aiko, it meant announcing up front the topics they wanted to talk about. This is perhaps not entirely natural in an ordinary conversation, but it appeared an effective strategy for her[24]. These kinds of approaches helped some participants to maximise the number of utterances attempted: for example, Satoshi was able to use all his utterances in every conversation. For such participants, adopting an effective conversation management approach also helped them follow the instruction to reproduce the utterances verbatim, thus reducing the need to make changes.

---

[23] This issue touches on a much more fundamental one concerning the extent to which memorised material can be fully integrated into spontaneous usage. In particular, there is a question as to how (and whether) explicit knowledge (as a result of memorised learning) can be converted into implicit productive knowledge (e.g. see Hulstijn & De Graaff, 1994). This experiment is one of the few that explores whether this process can take place, though it cannot in itself shed much light on how. (However, see the later discussions on automatic processing in Section 8.2.4 and Chapter 10).

[24] This kind of strategy was also used effectively by the non-speaking person with cerebral palsy using computer-supported communication in Wray (2002b). Despite it seeming unusual, blind judges could not distinguish her conversations from ones between two regular students.

Table 5.11: Examples of conversation management

Satoshi-2-10	
MU	A lot of the Cuban people assumed I was Chinese.
RP	[And did you dance?] No, no [But were there lots of people dancing though, I guess] Er yeah. Some people dancing, some people sitting uh listening to music. Ah yeah mm a lot of Cuban people (.) assumed I was Chinese. I walked in the street they said to me 'ni hao'
Aiko-2-1	
MU	MU: I remember last time you said you liked jazz - and that you'd been to that jazz bar called "Sometimes".
RP	RP: I have some- a couple of topics I want to talk about with you. The first is the music um music [mm OK] It's a common interest (.) for us, isn't it? [Yes it is, it's a common interest yeah] So, I remember last time you said you liked jazz [mm hm, I do] uh huh, and you been to the jazz bar called "Sometime".

Note: [ ] = native speaker speech

These observations resonate with reflections on the original study by Wray and Fitzpatrick (2009). Conversation management is an important idea for learners in that it can help them utilise and practice the language they know. So, an unexpected benefit of the study was to bring this to the attention of the participants. On the other hand, strict reproduction of some of the utterances exactly as memorised could render them unnatural within the conversation.

5.4.2.2. Linking the utterance to the conversation flow

However well target utterances have been chosen, in a real conversation there is the possibility that some may need to be amended to fit in with flow of the conversation or the context. This is an important factor for any speaker when learning new material in advance and the ability to do this well is part of speaking proficiency. While two participants, Takako and Kiyomi, acknowledged that they did change a few utterances partially in order to fit them in, most of the participants maintained that they did not deliberately do so during the performance. However, looking at the results, a large number were in fact changed in this way. For example, adding or changing a lexical or phrasal adjunct at the beginning of the utterance occurred for over half (363/694 = 52%) of the attempts at repeating an utterance during the performances.

As the first two examples in Table 5.12 illustrate, insertions or substitutions such as these are a natural part of fitting the utterance into the conversation flow, and the vast majority of adjunct-related changes were nativelike and appropriate. As noted earlier, there were 50 cases (out of 404) where model utterances already included an initial adjunct. These had an average deviation rate of 0.343 devs/word, which is almost twice as high as the average (0.183 dev/word). In other words, initial adjuncts and conjunctions were much more prone to change. This finding also complements the observation in Wray and Fitzpatrick (2008) that model utterances which do not begin with an adjunct or conjunction are more likely to be used without the need for change.

Table 5.12: Examples of nativelike changes

Kiyomi 1-3	
MU	The food culture there is similar to Japan.
RP	**And I think** the food culture there is similar to Japan
Kaori-1-10	
MU	**When travelling,** I feel like an explorer who's discovering many new things.
RP	**So,** I feel like an explorer who's discovering many new things
Yumi-2-3	
MU	I remember you mentioned that you play the guitar
RP	[I like to play the guitar] wow oh yeah I remember you mentioned that **last time**
Kaori-1-6	
MU	I check what they're reading and then I try to strike up a conversation with them
RP	I **just** check what they are reading and then I try to strike up a conversation with them

Although less common than adjunct changes, there were other types of deviation introduced as a natural part of the conversation, as illustrated in the second two examples of Table 5.12. For example, in Yumi-2-3, the context of the conversation provoked an appropriate phrasal substitution. In Kaori-1-6, a lexical insertion added extra nuance to the utterance. Changes such as these may have felt more natural than the original, and in most cases appear to have been made quite unconsciously.

5.4.2.3. *Difficulty of the utterances*

It is hard to quantify how difficult an utterance is to memorise and reproduce because it depends on the individual participant and their degree of familiarity with the components of the utterance. However, it is clear that difficulty did vary (e.g. Hikari said her first cycle was much easier because the utterances were easier) and general features that may be relevant in this include: length, structural complexity, ease of pronunciation and frequency of the vocabulary in the language at large. The creation of the model utterances was done in a collaborative way with the only criteria being that the utterance be the most natural way of saying something that the participant wanted to express, and that it was new to them. There was no conscious taking into account of the 'difficulty' of the utterance. However, a basic degree of monitoring was done by recording the length (in words) and the number of clauses in each utterance. Overall, the intention was that difficulty would vary randomly across the sample and, at least for length and number of clauses, this did indeed seem to be the case.

In general, longer utterances were found to be more 'difficult' in that length (in words) correlated significantly with the number of deviations/word in the performance. There was also a significant inverse correlation between length and closeness. On the other hand, length did not appear to have any effect on the propensity to attempt. It appears therefore that the existential awareness of an utterance and its relevance to the conversation (or motivation to use it) are not affected by its length. However, the ability to memorise it accurately or to piece it together fully in real time are. It is worth noting that using the deviations/word measure does take into account the fact that with more words there are more opportunities for deviation or error. So, it seems that there is something about memorising the longer utterance as whole (rather than, say, as two separate parts) that makes it more difficult. Memory research (e.g. Ellis, 2012) suggests that the chunking of text into smaller parts is essential when there is a lot of new information to remember. In particular, it provides a way of integrating new and existing knowledge in a way that falls within the limits of what short term memory can hold: around four chunks of information according to Cowan (2010). Thus utterances with features that make them easier to chunk for the learner should be easier to memorise. One such feature may be the existence of multiple clauses within the utterance. This possibility is supported by the results in this study which show that having more clauses in an utterance does appear to offset the disadvantageous effects of length, suggesting that participants found utterances with multiple clauses easier to chunk.

A second factor that may support chunking is having sub-sequences or phrases within the model utterance that are already familiar to the participant. For example, the utterance could consist of known and unknown expressions concatenated, or of known frames with new inserted completions. A useful way of classifying 'difficulty' therefore might be in terms of the number of new things a participant had to contend with. While the information for making such a classification was not specifically recorded within this study, some clues to it may be gleaned from examining the types and locations of deviations, as an indicator of the sub-sequences that they were processing as single items. This will be addressed in more detail in the next section.

5.4.3. *What deviations might tell us about memorisation and storage*

Deviations suggest that a word or phrase within the utterance has not been integrated in the whole. This may be because that particular part was not noticed or because it was not learnt sufficiently during memorisation. In either case, it was not available during the performance and an alternative has been supplied (which may or may not be appropriate). Deviations can therefore give clues as to how an utterance has been segmented and stored in memory.

The deviations were spread across morphological, lexical and phrasal types. In general, all such deviations (whether nativelike or not) can be thought of as possible points at which the utterance was segmented in term of the different chunks that were stored. However, there are different implications for different types of deviation.

5.4.3.1. Morphological deviations

The majority of morphological deviations were non-nativelike which is to be expected since there is less scope for variation in the choice of function words, determiners and inflections etc. Ellis (2006) suggests that these elements are generally less salient than lexical words due to their familiarity, lack of stress (as weak forms) and lack of contingency (in that they do not appear to contribute much to meaning). So, unless they are memorised (or recognised) specifically as being part of an expression within the utterance, they will be subject to reconstruction during reproduction. Generally, morphological deviations signal that the phrase they occurred in was not successfully memorised as a whole (with the possible exception of inflections), whether they are nativelike or not.

Table 5.13: Example morphological deviation

Aiko-1-3	
MU	I’m familiar with classical music because I learned the piano as a child.
RP	I'm familiar with classical music because I learned **a piano** as a child.

For example, consider the expression *I learned the piano as a child* in Table 5.13. Aiko’s non-nativelike deviation to *I learned a piano*, suggest that the expression was not fully memorised whole (and this would still be the case even for a nativelike deviation such as *I learned piano*). On the other hand, it is still possible that she may have memorised (or already known) the frame *I learned X as a child*.

However, it is important to acknowledge that a lack of deviations within any expression does not necessarily mean it has been holistically stored. For example, in Aiko’s case, the expression *as a child* may have been reconstructed based on existing knowledge of the frame *as a(n) X* and then concatenated onto the *I learned X* construction.

5.4.3.2. Lexical deviations

The majority of lexical deviations were native-like and a large number involved substitutions by a synonym. In most cases, this was a key word in the phrase and showed clearly that the original version was not stored holistically. In some cases however (e.g. the first example in Table 5.14 below), it could be argued that the phrase was stored as a frame, with the synonym being substituted by an alternative completion.

Nativelike omission or insertion of a lexical content word was also common. Some insertions (e.g. Kaori-1-6 in Table 5.14) could be viewed as ‘qualifiers’ inserted into a holistically stored phrase. Similarly, some omissions occurred when the original (phrase within the) utterance included a qualifier that was viewed as non-essential (e.g. the word ‘*coach*’ in Sachiko-3-12 in Table 5.14). As noted in Section 5.4.2.2, this is particularly the case for adjuncts. Adjunct deviations may also signal points at which two holistically stored phrases are joined. For example, the missing *because* in Sachiko-3-12 may highlight that *It’s a special memory* is a separate component within the utterance for her.

Table 5.14: Examples of lexical deviations

Kaori-2-6	
MU	One of the **good** things about our event was that it helped to build a sense of community
RP	And one of the **best** things about our event was that helped to build a sense of community in those areas.
Sachiko-3-12	
MU	It’s a special memory for me **because** within our **coach** tour we managed to create a strong spirit of camaraderie
RP	It was a special memory for me - within our tour we managed to create very strong spirit of camaraderie.
Aiko-1-11	
MU	Compared with other world cities, Tokyo has a much more homogenous **population**.
PP	Compared with other countries, Tokyo has very much more homogenous **people**

The performance of the participants showed that if a lexical word was novel, it was generally recalled well. For example, in the final example in Table 5.14, Aiko recalls ‘homogenous’ while substituting other synonyms for other lexical words in the utterance. In general, this seemed to be because such words were highly salient to the participants and the necessary focus was applied to learn them. The same was true, to some extent, for novel phrases (such as *build a sense of community* and *strike up a conversation* in the examples above). However, even new phrases could be subject to change if elements within them were non-salient or not perceived as an essential part of the phrase (e.g. Sachiko recalls ‘a strong spirit of camaraderie’ but without the initial indefinite article).

5.4.3.3. Phrasal deviations

For phrasal deviations, about half were nativelike and half not. A possible reason is that, while new phrases may be salient (as suggested above), they may also be harder to learn than new words (because they are longer and there is more potential interference arising from familiar component words). In addition, some phrases may not be clearly marked as such within the utterance and therefore not treated as phrases (leading to non-nativelike changes). Others may be synonymous with more familiar phrases using similar words (leading to nativelike changes).

The majority of phrasal deviations were insertions or substitutions. Insertions tended to be more nativelike and this is likely to be because the participant has inserted an expression they already know. This is illustrated in Table 5.14 earlier, where Kaori adds in the phrase '*in those areas'*, and where adjunct phrases are added (or changed) for reasons of conversational flow (see Section 5.4.2.2). Substitutions and omissions on the other hand may show that the original phrase was recognised as being a holistic chunk, but the whole thing was not memorised sufficiently to be recalled. This may be the case in the first example in Table 5.15, where a similar (known) expression '*in particular'* replaces the original.

Table 5.15: Examples of phrasal deviations

Aiko-3-2	
MU	Fashion has a very wide meaning, so what **particular area** do you focus on?
RP	Fashion has a very wide meaning, so what **in particular** do you focus on?
Akira-1-8	
MU	Many people complain but **they don't do anything about it**
RP	So yeah the many people complain **about it** but **they don't do anything**

There were also a number of phrasal deviations involving word order changes. Around half of these were non-nativelike and signalled that the phrase was not memorised holistically. The nativelike word order changes typically involved moving shorter phrases within the longer utterance. For example, in Akira-1-8 in Table 5.16, a longer phrase (*they don't do anything about it*) is divided into two parts (*they don't do anything* and *about it*) and these are moved legitimately in the RP utterance. In examples like this, the smaller phrases appear to be holistic for the speaker because they are manipulated (successfully) as units.

5.4.3.4. Putting it together

For any utterance then, an analysis of the different types of deviation may give some clue as to how it was memorised and stored. Table 5.16 illustrates one way of doing this. Here a possible breakdown of the MU is given, based on evidence from the RP response. The potential holistic sub-components of the utterance (for Akira) are '*old mattress'* (without the definite article) and '*straight down the stairs'*. The key verb form '*threw*' is retained (as well as '*they*' for the protagonists), but '*just*' appears to be a non-salient element.

Table 5.16: Example for interpreting deviations

Akira-1-3	
MU	They | just | threw | the | old mattress | straight down the stairs.
RP	They took my old mattress and threw it straight down the stairs.

The analysis of deviations more generally seems to suggest that participants memorise and recall the utterances as a combination of parts or frames which are then pieced back together. Such parts may be recalled more easily because they are already known, or because they are new expressions that are particularly memorable or useful in themselves. However, some new phrases within the utterance seem vulnerable to changes (often non-nativelike). This may occur if not enough attention is paid either to the boundary points linking them with the rest of the utterance or to non-salient function words within the phrase. Familiar words or phrases were also subject to changes (often nativelike) via substitution, suggesting that the conceptual element of each part is linked to the utterance, but the specific way of expressing it may have received less focus.

5.4.4. *Attrition – what makes utterances stick?*

Attrition is where a new word, phrase or utterance has been learnt and used appropriately at some stage, but is later forgotten. The Delayed Performance (DP) stage took place 2-3 months after the main study and there was no instruction to continue memorising and practising the utterances in between. So, there was ample scope for attrition and the performance differences found suggested this was a common occurrence. In general, Propensity to attempt and Closeness were significantly lower at the DP stage compared to RP or PP, and the number of deviations was much higher. It is interesting therefore to focus on those utterances that were attempted successfully (i.e. resisted attrition), and to see what they suggest about long-term retention. Of the 139 utterances that were attempted, 22 had a closeness score above 0.7. Some points relating to these examples are given below.

5.4.4.1. Length and simplicity

The majority of utterances retained (18/22) were quite short (12 words or less) and tended to express useful things in a simple natural way. By definition, they had been new to the participant, but in most cases used familiar words or phrases. For example the following questions were all retained accurately:

- *What kind of things do you do in your free time?*
- *What have you been up to recently?*
- *What do you do when you're not working?*

Other short expressions included:

- *I met my friends and we discussed our latest news.*
- *Actually I have been to Edinburgh a couple of times before.*
- *Apart from me, no-one else speaks English in my office*

These may have been retained because they contained useful new expressions ('our latest news', 'a couple of times', 'apart from me') arranged in a familiar and relevant sentence.

5.4.4.2. Use and relevance

Of the longer utterances retained, some were likely to have been used regularly after the study because they were particularly useful. For example, Mari retained the very first utterance she learnt: *I am working in a foundation called XXXX which is responsible for administering the TOEIC test in Japan*.

5.4.4.3. Retention of parts

Looking at the DP responses is also useful in supporting the earlier conjectures in Section 5.4.3 about how utterances have been segmented in memory. Some examples are given in Table 5.17.

Table 5.17: Examples from DP responses

Hikari-2-6	
MU	It's got a very distinctive guitar riff as well as a more melodious part
DP	it has a very distinctive part - one is guitar riff and the other is a more melodious part
Kiyomi-1-6	
MU	Currently most software development is taking place in South-East Asian countries
DP	Recently (1) in south (.) east Asian countries (..) taking place (..) software development.

In Hikari-2-6, the novel expression *a more melodious part* is retained whole while the other new expression *distinctive guitar riff* is remembered in two parts. Kiyomi-1-6 is even clearer in this respect, as the expressions *South-east Asian countries*, *taking place* and *software development* are all remembered as part of the utterance, but not how to put them together as a whole. These examples also support the idea that unfamiliar, potentially difficult new words or phrases (*distinctive*, *guitar riff*, *melodious*, *South-east Asian countries*) may be retained because they are particularly salient and more time may have been focussed on getting them right.

5.4.5. ***Feedback from the participants***

The participants were very positive about the process and all said they found it very enjoyable and useful, but also quite challenging. Some participants described a mixture of nervousness and excitement at the opportunity to use the utterances with a new native speaker. Hikari: "I had good opportunities to speak English with unknown native speakers, so it was enjoyable and a bit thrilling." For most participants it was a 'unique' method and they found it useful and practical to learn natural ways of expressing the things they wanted to say "because I could actually use the phrases I memorised" (Mari). It gave them confidence that they were using appropriate English in a natural way. Some mentioned that it took a lot of time and effort and the utterances may be too specific to apply in other situations, and one participant (Yumi) felt she needed to be at a higher speaking level to take advantage of this method fully. To use the method as a way of teaching English, some suggestions were: to have normal and slower versions of the model sentences, to allow them to check against a written text, to give longer time between sessions, and to provide variable responses to give more flexibility in the conversations.

The instruction to not write anything down elicited mixed views with participants. Some mentioned that it helped them improve their listening and helped them focus on the spoken form. Some felt it affected their intonation and pronunciation in a positive way. However, others said it was difficult as they were so used to relying on the written form and it may have made the memorisation process take longer. Also, Kiyomi mentioned that she had to focus hard on grammatical words which were not easy to catch and, without the written form, she was less confident about her grammar. Kaori on the other hand said it made her really focus on the way that utterances were constructed out of shorter phrases rather than just hearing and understanding.

Most participants mentioned that the second cycle they did was easier than the first because they had got used to the method and they had already met the native speaker, making it easier to relax. In the second cycle, there was also more awareness about the type of utterance that could be easily brought into the conversation. For example, Yumi mentioned that, due to her experience from the first cycle, she was more able in the second iteration to plan how to lead the conversation so it would stay on topic. In the second cycle generally, there were a lot more questions, and participants realised the usefulness of specific conversation starters ("Do you mind if I talk about one of my hobbies which is reading") and references back to the first meeting ("I remember you said you've lived in Japan for a long time"). There were a couple of exceptions: Kiyomi felt that in her second cycle she became 'more of a listener' because she had already met the person once before. Hikari felt the first cycle was easiest because her utterances were easier. All participants that used Skype suggested that these sessions were more difficult.

Native speakers said that they found the second real performance with a participant more natural and easier. Overall, participants indicated that, going forward, the study had made them realise the importance of preparing good sentences to use and of being able to manage the conversation.

5.5. Conclusion

This replication of Wray and Fitzpatrick (2008) provided a useful method for introducing target utterances to L2 speakers and analysing the results of their later performance in real and practice conversations. The 14 participants here demonstrated that they could memorise long utterances which were new and relevant to them, and then re-use them successfully.

Overall, the findings suggest a general picture of participants (consciously or otherwise) tending to break down utterances into a combination of parts or frames and then piecing them back together. The processes involved in this segmentation process are highly relevant to the acquisition and storage of formulaic expressions. However, although the analysis of deviations conducted here can give clues to this process, deviations alone are insufficient to ascertain segmentation. In particular, the absence of deviations within any part of the utterance does not necessarily mean that part is holistically stored or processed. This will be addressed in Chapter 6 which will explore example utterances in greater detail, incorporating phonological features of the individual's performance at different stages.

CHAPTER 6: Extended analysis of Study S3

Fluency and deviations as a way to explore formulaicity in the memorised utterances

6.1. Introduction

Study S3 explored what happened when Japanese speakers of English were given novel, individually relevant target utterances to memorise and then use again in subsequent planned conversations. Evidence from the output and from participant feedback suggested that some degree of segmentation of the target utterances took place: either as a strategy during memorisation or as a way of reconstructing the utterance during the performance. It was conjectured that this segmentation may be connected to the presence of formulaic sub-sequences within the target utterances (or formulaicity of the whole utterance). Such sub-sequences may have already been formulaic for the individual or may have become so during the memorisation and reproduction process. In the approach adopted in this research, formulaicity is defined internally with respect to the individual and may be identified on the basis of phonological coherence in production and unitary form. One modification to the original study made in the Chapter 5 replication was to record fluency information. This allowed for further phonological analysis of the output, so that the possibility of internal formulaicity within segments could be explored. In addition, since memorisation of the original utterance was followed by a practice performance, a real performance, and later a delayed performance, there is potentially useful data on how segmentation occurs over time. This chapter describes some analyses of this data in order to explore how formulaic segments within a longer utterance may be created by the speaker and then used to reconstruct that utterance.

The exploration in this Chapter has two parts. In the first, quantitative data relating to fluency of reproduction during the real performance is examined and compared with the data on deviations discussed in Chapter 5. Based on the idea that fluency is a necessary condition for internal formulaicity, dysfluencies within the utterance may be taken to indicate points at which the utterance has been segmented into smaller parts. In addition, as suggested in Chapter 5, the location and type of deviations also give clues as to how the utterance has been memorised and stored internally by the speaker. The analysis therefore compares points and types of deviation with points of dysfluency in participant utterances. While there is no reason to expect deviation and dysfluency to coincide generally, exploring the extent and nature of any coincidence may help to determine what types of deviation could be indicators of segmentation. It also looks at individual differences in fluency of reproduction across the participants.

The second part takes a closer look at some specific examples. This is a qualitative analysis of particular utterances used by two participants. The aim is to explore in detail how particular utterances may have been segmented based on evidence of phonological coherence and deviations at each of the different stages. As well as dysfluencies, the data allows the opportunity to incorporate more detailed phonological analysis including temporal (associated with speed of delivery) and word boundary features. The analysis explores how formulaic segmentation might be occurring over stages and also provides some further insight into the identification of formulaic expressions. Additional data is provided as a result of having included a dictation task at the end of the Delayed Performance (DP) meeting. This was designed to provide further information about the way the utterance was stored in the mind of the individual by the end of the study.

6.2. Quantitative Analysis

As described in Chapter 5, 14 Japanese speakers of English each memorised two or three sets of model utterances (MUs) of varying lengths, each set to be used in a Real Performance (RP) conversation with a native speaker. Their attempt at using each MU was recorded and transcribed. Indicators of dysfluency were noted, including pauses > 02s, hesitation markers (e.g. *er*, *mm*); and false starts; and reformulations. A cleaned-up version of the utterance attempt was transcribed for analysis. It included a marker for each point of dysfluency ('/'). Table 6.1 illustrates this for one MU.

Table 6.1: Hikari-2-3

MU	I used to go out with someone who always played Eric Clapton songs on the guitar
RP (original)	I used to go (..) som- go with someone er who always- er always played Eric Clapton songs (.) on the guitar mm guitar
RP ('cleaned')	I used to go / with someone / who always / always played Eric Clapton songs / on the guitar.

As described in Section 5.2.3.3, each RP attempt was analysed in terms of its deviations from the original MU. Deviations were classified according to whether they were Morphological, Lexical or Phrasal, and into sub-categories within each. For example in Table 6.1, the missing word '*out*' (from '*I used to go out with ...*') was classified as a Lexical deviation of sub-type 'content word omission' (CW-omit).

Each deviation within an RP utterance was checked to see if it coincided with a dysfluency. Specifically, if the dysfluency occurred immediately before or after the omission, substitution or insertion of a word or phrase or an inflectional deviation (or within an inserted phrase), the deviation was marked as being associated with a dysfluency. For example, in Table 6.1 the deviation (omission of *'out'*) coincides with a dysfluency, and there are two more dysfluencies in the RP utterance which are not associated with any deviation. This analysis provided a means for quantifying the relationship between deviations and dysfluency in order to understand how deviations may be associated with the developing formulaicity of segments within the utterance.

6.2.1. ***Results***

Of the 404 MUs introduced in the study, 313 (77%) were attempted at RP. The mean length of utterances attempted was 13.3 words. Overall, there were 577 deviations with a deviation every 7.9 words on average, excluding adjuncts that were added at the beginning the utterance. There were 654 dysfluencies with a dysfluency occurring every 7.2 words on average. So, the number of deviations and dysfluencies overall was comparable. However, the number of occasions when a deviation and a dysfluency coincided was only 228. This meant that 65% of dysfluencies were not associated with deviations and 60% of deviations occurred away from any dysfluency. In other words, there was only a moderate degree of overlap and this confirms that, in general, deviation and dysfluency are separate phenomena.

However, it may be that some types of deviation do tend to coincide more with dysfluency and are therefore potentially related to any segmentation of the utterance that takes place during memorisation. To explore this possibility, Table 6.2 shows the number of deviations occurring for each category and sub-category of deviation and the number (and percentage) of dysfluencies that coincided with these deviations. Overall, phrasal and lexical adjunct (single word) deviations had the highest proportion of cases which matched dysfluencies, especially for insertions and substitutions. Morphological deviations and omissions of any words or phrases tended to have a lower overlap with dysfluency. The implications of this are discussed in Section 6.2.2.

Table 6.2: Deviations and coinciding dysfluencies by deviation category

Category		**deviations**	**coinciding dysfluencies**	**dev-dys overlap %**
MORPHOLOGICAL	INFL	35	8	23%
	FW-omit	33	13	39%
	FW-ins	18	7	39%
	FW-sub	13	7	54%
	ART-omit	44	13	30%
	ART-ins	17	5	29%
	ART-sub	25	7	28%
	PAR-omit	14	2	14%
	PAR-ins	0	0	-
	PAR-sub	13	3	23%
	VAR-sub	16	6	38%
	TOTAL	71	228	31%
LEXICAL	CW-omit	47	12	26%
	CW-ins	37	12	32%
	CW-sub	52	23	44%
	ADJ-omit	17	7	41%
	ADJ-ins	18	10	56%
	ADJ-sub	13	8	62%
	TOTAL	72	184	39%
PHRASAL	PHR-sub	70	43	61%
	WO	19	8	42%
	PHR-omit	38	10	26%
	PHR-ins	38	24	63%
	TOTAL	85	165	52%
Total		**228**	**577**	**40%**

Note: INFL=inflection; PAR=Particle; VAR=variant; CW=Content word; ADJ=single word adjunct; PHR=Phrase; WO=word order error; -omit=omission; -ins = insertion; -sub=substitution.

Deviation and dysfluency data were also calculated for each participant as shown in Table 6.3. The table gives the number of RP utterances attempted for each participant, their mean length, and the number of them that had no dysfluencies. The frequency of deviations and dysfluencies for each participant is shown in terms of how often they occur (average number of words for each occurrence). As can be seen, fluency of utterance delivery varies across participants with the most fluent (Kiyomi and Mari) having a dysfluency once every 10 or more words, and the least fluent (Takako) it is once every 4.2 words. As with the deviation data, this difference is not associated with proficiency. Further, there is no correlation between deviation and dysfluency rates across the participants, and the percentage of deviation-dysfluency overlap also varies considerably. Again this suggests that, overall, the tendency to deviate and the tendency to segment an utterance (as based on fluency) may be independent of each other.

Table 6.3: Mean deviation and dysfluency data for each individual participant

	RP utterances			**deviations v dysfluencies**		
	RP utts	**words / utt**	**fully fluent utts**	**words / deviation**	**words / dysfluency**	**dev-dys overlap %**
Satoshi	27 (100%)	11.1	9 (33%)	12.5	9.4	25%
Aiko	35 (85%)	13.3	2 (6%)	6.3	5.2	41%
Kaori	26 (100%)	13.5	4 (15%)	10.6	6.6	30%
Kazu	17 (80%)	13.6	1 (6%)	5	5.8	39%
Hikari	34 (87%)	13.6	3 (9%)	8.4	5.3	47%
Akira	15 (62%)	11.5	3 (20%)	7.2	8.2	38%
Yumi	26 (70%)	13.7	5 (19%)	7.7	8.3	37%
Kiyomi	16 (61%)	12.8	8 (50%)	12	12.8	18%
Yoshie	21 (91%)	14	1 (5%)	5.2	6.3	40%
Sachiko	30 (76%)	15	4 (13%)	6.6	5.9	38%
Hisako	23 (95%)	12.9	1 (4%)	10.6	5.4	39%
Takako	15 (55%)	13.3	0 (0%)	4.1	4.2	50%
Yoichi	11 (42%)	14.5	1 (9%)	7.3	6.7	55%
Mari	17 (70%)	14.1	8 (47%)	6.8	10.4	37%
All	**313 (77%)**	**13.3**	**50 (16%)**	**7.9**	**7.2**	**40%**

Participants listed in order of increasing proficiency (by YesNo score)

6.2.2. ***Discussion***

If fluency in spoken production is a necessary condition for internal formulaicity as suggested by Myles and Cordier (2017) and explored in Chapters 3 and 4, then the majority of target utterances produced in the real performance (RP) of the replication study were not formulaic: most (84%) contained at least one dysfluency at some point in the utterance. However, since the target utterances were generally quite long and multi-clausal, breaks within the utterances are not surprising and are consistent with the suggestion that longer utterances are broken down (segmented) into smaller segments for memorisation and reproduction. The dysfluency points within the utterance therefore may potentially provide an indication of where the utterance has been internally segmented (by that speaker).

However, since fluency is not the only condition for internal formulaicity, it is not necessarily the case that the fluent segments within an utterance are internally formulaic. The other necessary condition is that the segment should demonstrate at least one typical condition that shows a holistic dimension. The results suggest that while 40% of all deviations do coincide with a fluency-based segmentation point, more do not. There are two possibilities for these cases: either the deviation occurs within or as part of a formulaic segment, or the segment is not formulaic (and the deviation may be an indicator of this). The analysis for the deviation subcategories shows that insertions or substitutions of phrases and adjuncts tend to occur more often at segmentation points than other types of deviation. This is not surprising for adjuncts (whether words or phrases) as they typically mark change points between clauses and are associated with pauses in normal speech. Indeed, where adjuncts were included within target utterances, they were frequently occurred with pauses, even though there was no deviation.

In the case of phrasal insertions or substitutions, as the results in Chapter 5 showed, these were typically ‘nativelike’ deviations, suggesting that a known formulaic expression was being inserted or substituted. Morphological deviations, on the other hand, were generally non-nativelike and tended not to coincide with dysfluencies. In these cases, it suggests that the deviation is part of an inadequately memorised segment (due to the lack of saliency of these elements) which is nevertheless delivered fluently. In other words, the speaker does not memorise every detail of the expression, just an approximation in which the correct morphological element may be omitted or changed. It may be that a segment containing a morphological deviation represents an incompletely learned formulaic expression in that the overall structure (the lexical words linked together appropriately) is solid but the fine-tuning (some of the function words, particles, inflections etc.) is not. This chimes with Peters' view (e.g. Peters & Menn, 1993) that partial analysis and partial acquisition are features of language acquisition in general. Depending on subsequent practice and exposure, the partial acquisition may result in an expression which is formulaic for a speaker but remains inaccurate (fossilisation), or one which is subsequently fine-tuned to the canonical form over time.

The results also show that there are individual differences between participants in terms of the fluency of their output. These differences did not seem to be associated with their proficiency levels or with the length of their target utterances. For example, Kiyomi, Satoshi and Mari had greater fluency (more fully fluent utterances and higher mean words/dysfluency scores) than other participants. However, Satoshi was at a lower proficiency level, and Mari had relatively long utterances. As discussed in Chapter 5, the amount of time each participant spent memorising the expressions is an unknown factor, and it is likely that this would influence subsequent fluency and accuracy. However, there did not appear to be any association between number of dysfluencies and number of deviations across the participants, suggesting that any effect of memorisation time is not consistent for these two variables.

6.3. Qualitative analysis

6.3.1. *Overview*

The aim of the qualitative analysis was to select some particular target utterances and analyse how the speaker may have broken these down when memorising and reproducing them. In particular, the analysis attempts to identify which sub-sequences (segments) within the target may be (or have become) internally formulaic for the speaker.

The analysis involves first segmenting responses for the selected target at each stage of the study on the basis of fluency, then using deviation evidence to select potentially formulaic sub-sequences, and finally applying additional phonological checks. As outlined in Chapter 5, the stages of the replication study included an initial meeting with the researcher to determine the model utterances to be learnt (MU). This was followed a week later by a practice performance (PP) with the researcher, a real performance (RP) with a native speaker a further week later, and a delayed performance (DP) with the researcher two to three months after that. In addition, at the end of the study after the DP, the participant was given a dictation task (DT) designed to further highlight possible utterance segmentation. Thus, for many utterances, spoken examples and their transcriptions were available at various stages of its development, memorisation and use by the participant.

In the study so far, fluency has been used as the main indicator of phonological coherence, as a necessary condition for internal formulaicity. As discussed in Chapter 2, other phonological features, such as articulation rates (Siyanova-Chanturia & Van Lancker Sidtis, 2019) and prosody (Lin, 2019) may also be linked to formulaicity. These include the distinctive stress patterns of non-literal idioms compared with canonical forms (Ashby, 2006; Van Lancker et al., 1981), phonemic reduction of high-frequency sequences (Strik et al., 2010) and the idea that formulaic expressions are contained within intonation units (Lin, 2013; Lin & Adolphs, 2009). Based on these points, some additional phonological analysis was applied to the target utterances to provide support in identifying segments as internally formulaic

6.3.2. *Selection of case study participants and utterances*

Since detailed exploratory phonological analysis of spoken output is extremely time-consuming, a case study approach was adopted, focussing on two participants and a small selection of their target utterances. In the main study, there were 14 participants but, for reasons of logistics and availability, only 10 participants completed the delayed performance and the additional dictation task. Two participants (Aiko and Kazu) were selected on the basis that they had a good number of responses at the delayed performance stage, they had a relatively high number of deviations and dysfluencies (see Table 6.2) and they were available for follow-up if required. The specific utterances selected were ones which were attempted at all the stages.

6.3.3. *Data collection and analysis*

6.3.3.1. Dictation task

The dictation task (DT) was based on a procedure developed by Schmitt et al. (2004) and was described in Section 2.2.4 in the context of reviewing their study. The task is to present sentences orally to the participant who then repeats them back as exactly as possible. In the original study by Schmitt et al, potentially formulaic strings were embedded in longer sentences and these were the focus of the analysis. In the current version, the whole model utterance from study S3 was used as the cue, and dysfluencies in the resultant repetition were analysed to indicate potential segmentation points. The premise is that the utterance is stored as a combination of words and formulaic segments which are then reconstructed during reproduction. Formulaic segments within the longer sentence, if repeated back at all, will be fluent and intact, even if the overall reconstruction is not fully fluent or accurate. The process is made sufficiently challenging by giving a distraction task (counting down from 100 in threes for 10 seconds) between presentation of the stimulus and repetition to necessitate some degree of reconstruction.

6.3.3.2. Transcription and initial segmentation

Using the cleaned up data from the original study, the selected utterances were segmented using dysfluencies (/) for each stage as described in Section 6.2. Intonation Unit (IU) boundaries (as suggested by pitch re-set, tone unit, syllable lengthening and longer pauses) were also marked (//). To this was added the data from the dictation task output (DT), analysed and transcribed in a similar way. Table 6.4 illustrates the output for one utterance. Original recordings and full transcriptions were retained for further analysis of selected segments.

Table 6.4: Aiko-1-11 (Segmented on the basis of fluency)

MU	Compared with other world cities // Tokyo has a much more homogenous population.
PP	Compared with other countries // Tokyo has / very / much more / ho- / homogenous / single race / homogenous people /
RP	Compared with other countries // Tokyo has a / much more homogenous population
DP	Compared with other cities // Tokyo has a homogeneous population
DT	Compared with other world cities // Tokyo has much more / homogenous population

6.3.3.3. Selection of potentially formulaic segments

The next stage of the analysis was to identify the sequences that may be formulaic for the speaker at the time of speech. The practice performance (PP) represents a stage when the whole utterance is still being memorised and practised, and the real performance (RP) can be thought of as the culmination of this learning. The delayed performance (DP) represents what has been retained a few months after the RP, and the dictation task (DT) is a more controlled way of observing the utterance production (which, unlike the other examples of production, involves giving a direct oral cue of the utterance). The key stages for the analysis were therefore taken to be RP, DP and DT. Any fluent segment which was repeated fluently in all three of these stages was marked as being potentially formulaic. For example, in the case of Aiko-1-11, the segments '*compared with other*', '*homogeneous population'* and '*Tokyo has'* were considered potential formulaic segments on the basis that they consistently formed (or are part of) a phonologically coherent segment in the RP, DP and DT stages. The potentially formulaic segments were then analysed further to see how other phonological indicators may support the formulaicity conjecture. Two possible sources of evidence considered were word boundary behaviours and phrase articulation rates.

6.3.3.4. Word-boundary behaviour

A possible way of exploring phonological coherence is to consider features of the boundaries between words in the utterance and see how those within segments compare with those between segments. Features that might be expected within potentially formulaic segments are the presence of indicators of connected speech such as intrusion ("*I want to(w)eat*"), linking ('*It sno joke'*) and elision ('*It mus(t) be'*) and the absence of hesitancy indicators such as pauses and elongated syllables. Kuiper (1996) suggests that such hesitancy indicators are used to signal to hearers that there is a choice point coming and they should pay attention, as this is where the speech output is not so predictable. On the other hand, Erman (2007) suggests these could be associated with greater cognitive load in trying to construct the subsequent phrase. Hesitancy indicators would therefore not normally be expected to occur within a fixed formulaic expression, although they may occur before the variable slot in a formulaic frame.

Each utterance occurrence was therefore analysed to identify cases of connected speech (intrusion, linking and elision) and elongated syllables and to note where these occurred in relation to the potential formulaic segments. Note: Pauses are already taken into account by the segmentation process. Following previous research (Cordier, 2013; Riggenbach, 1991) a syllable was considered to be elongated if it is longer than 0.3s.

6.3.3.5. Phrase articulation rates

As detailed in Chapter 2, a variety of evidence suggests that formulaic expressions tend to be articulated faster than the same sequence would be if constructed in real time on a word-by-word basis. While it is not possible to do such a comparison on the potential formulaic segments here (since there are not two different versions to compare), it is possible to compare the segment's delivery rate with that of other sequences uttered by the same speaker (i.e. all the ones not repeated consistently at RP, DP and DT). An initial conjecture is that a formulaic n-gram (a segment of n contiguous words) would be delivered faster (in syllables/minute) than average n-grams in the speaker's speech (delivered in equivalent circumstances). One potential problem with this is that syllables have different lengths (e.g. diphthongs and monophthongs) and the speed of delivery of syllables within a word or phrase varies according to features such frequency and stress (Greenberg, Carvey, Hitchcock, & Shuangyu Chang, 2003) and position within the intonation unit (e.g. final syllable lengthening). This may be partially alleviated by ensuring that the words in the potentially formulaic sub-sequence to be checked are also included in the other n-grams used to create the average for that speaker.

Table 6.5: Articulation rates for 2-grams in Aiko-1-11

2-gram	**Sylls**	**AR (syll/s)**		
		RP	**DP**	**DT**
Compared with	3	2.727	3.093	3.788
with other	3	3.641	5.435	3.690
other countries	4	4.739	-	-
other cities	4	-	3.115	-
world cities	3	-	-	3.165
Tokyo has	3	4.190	3.676	4.630
has a	2	2.857	2.976	-
has much	2	-	-	3.436
much more	2	2.066	-	2.614
more homogenous	5	3.828	-	-
a homogeneous	5	-	4.864	5.092
homogeneous population	8	4.737	4.515	5.387

Note: Articulation rates are calculated by dividing the no. of syllables of the n-gram with the duration.

Therefore, for each utterance, the set of possible n-grams was analysed (at all three stages) to obtain an average articulation rate for each sequence length. The expectation was that formulaic expressions would be delivered faster in general than other n-grams (i.e. above the n-gram average). To illustrate this method, the articulation rates of all the 2-grams and 3-grams (excluding those that traverse pauses) for the target utterance Aiko-1-11 are given in Table 6.5.

6.3.4. ***Results and discussion***

6.3.4.1. Potential utterances and expressions

From her 35 model utterances (MUs), Aiko had five for which she was able to recall (a version of) the utterance at every stage. Kazu had 17 MUs of which six were recalled at every stage. Applying the segmentation process, the repeated multi-word sequences (at RP, DP and DT) for both participants are given in Table 6.6.

Table 6.6: Repeated segments (at RP, DP and DT) for Aiko and Kazu

Utterance	**Detail and number (n) of segments repeated**
Aiko-1-11	compared with other; Tokyo has; homogenous population (3)
Aiko-3-5	second time; Rosslyn Chapel (2)
Aiko-3-8	conclusion of the novel (1)
Aiko-3-11	(0)
Aiko-3-12	to see how old (1)
Kazu-1-2	I talk (1)
Kazu-1-4	from US (1)
Kazu-1-5	(0)
Kazu-1-9	flat structure
Kazu-1-10	(0)
Kazu-2-11	I play golf; for business (2)

Three of the utterances (Aiko-1-11, Aiko-3-8, and Kazu-2-11) were selected for further analysis

6.3.4.2. Aiko-1-11

Looking at the stages of the utterance reproduction (see Table 6.3), Aiko repeated the segment *compared with other* consistently and fluently each time (including at

PP), but used different ways to complete the expression, namely *countries*, *cities* and the target *world cities* (the latter only reproduced on the DT). There was evidence of elision between *compared* and *with* at DP and DT and also an extended syllable at the end of the segment at RP and DP, possibly indicating a point where she considered how to complete the expression. The segment *Tokyo has* was also repeated consistently with different completions. For the segment *homogeneous population* (which was not reproduced at PP), it is useful to consider how Aiko developed this though the stages. Based on discussion at the initial interview, the word '*homogenous*' was new to her while '*population*' was already known. As a combination, it was difficult for her to remember at first, and at PP she used it with a different noun (*people*). Subsequently, she was able to use the correct expression consistently at RP, DP and on the dictation. There was evidence of linking between the words at the DT stage, and also the '*a*' before *homogenous* was extended at DP and DT, possibly indicating a search for the required phrase coming next.

For all three segments therefore, there is some support for the contention that they were (or became) formulaic for the speaker during the memorisation process. While expressions such as *compared with other*__ or *Tokyo has* __ may not be typical formulaic frames in the language, they may be useful for this speaker. The former may have arisen because Aiko already knew the construction (but had more difficulty in remembering the variable element). For the latter, it may useful and salient for a Tokyo resident like Aiko when she wants to talk in English about her home city to an outsider. In the case of *homogenous population* we may be seeing the conversion of the single words into a formulaic segment. Establishing these segments as formulaic (frames or fixed expressions) could be part of the process of chunking up (or segmenting down) the utterance in order to help memorise and reconstruct it during on-line production.

Table 6.7 shows the articulation rates (AR) of the potentially formulaic segments compared with other n-grams at each stage. The segments *Tokyo has* and homogeneous population had consistently higher articulation rates than other 2-grams in the utterance, but this was not the case for *compared with other in* relation to other 3-grams.

Table 6.7: Articulation rates in Aiko-1-11

Segment	**AR (syllables/s)**		
	RP	**DP**	**DT**
Tokyo has	4.190	3.676	4.630
homogenous population	4.737	4.515	5.387
Mean (other 2-grams)	**3.310**	**3.897**	**3.631**
compared with other	3.369	3.817	3.663
Mean (other 3-grams)	**3.763**	**3.922**	**4.056**
Mean (all)	3.631	3.932	3.969

6.3.4.3. Aiko-3-5

The responses at each stage for the utterance Aiko-3-5 are given in table 6.7. The two potential formulaic segments were *second time* and *Rosslyn Chapel*.

Table 6.8: Aiko-3-5 (Segmented on the basis of fluency)

MU	The second time in particular / was really impressive // because I had the chance to visit Rosslyn Chapel.
PP	OK so // the second time / when I visited Edinburgh // I had the chance to visit / Rosslyn Chapel
RP	But the second time is very / in particular / impressive because / I had a chance to visit / Rosslyn Chapel
DP	but / second time is / really impressive because / I went to / Rosslyn Chapel
DT	Second time / is particular / impressive because / I had the chance to visit Rosslyn Chapel

Based on the model utterance (MU), it might be expected that Aiko would segment the initial part of the utterance as *the second time* plus *in particular* (or combine these into one). Her output over the stages suggest that, rather than memorise the whole segment *the second time in particular*, she attempts to reconstruct it unsuccessfully, with only *second time* being consistently and fluently repeated over the stages. There was elision between *second* and *time* in all stages analysed, suggesting these words were connected formulaically for Aiko. Her attempts at *in particular* suggest that, while she may have known this expression, she was not familiar with its use in the noun-phrase construction 'X in particular', leading to confusion during reconstruction of the target. This may be supported by the finding of syllable elongation at the end of *time* for DP and DT. Another likely segment *I had the chance* also seemed to rely on reconstruction. While there was consistent linking between *I* and *had* and relatively fast AR, inconsistency with the article (and omission at DP) suggest it was not fully established as a formulaic expression. It was also substituted by an expression (*I went to)* which was likely already formulaic for Aiko (as it was delivered fluently, articulated very quickly and the words were connected).

The segment *Rosslyn Chapel* is a good candidate for formulaicity because it is a proper noun phrase which was repeated by the speaker many times when discussing her visit there. This was supported by the segmentation analysis. However, the expression had a relatively low articulation rate as shown in Table 6.9. The table also shows that delivery of *second time* was particularly slow (compared to the usual trend) at DT.

Table 6.9: Articulation rates in Aiko-3-5

Segment	**Articulation rates AR (syllables/s)**		
	RP	**DP**	**DT**
second time	4.717	3.886	2.985
Rosslyn Chapel	3.311	2.770	3.865
Mean (other 2-grams)	**4.223**	**2.828**	**4.816**
Mean (all)	4.193	2.995	4.506

6.3.4.4. Kazu-2-11

The responses at each stage for the utterance Kazu-2-11 are given in Table 6.10. The two potential formulaic segments were *I play golf* and *for business*.

Table 6.10: Kazu-2-11 (Segmented on the basis of fluency)

MU	Sometimes I play golf // but that’s mainly for business reasons.
PP	Yeah // yeah I play golf but / that's mainly for business reasons.
RP	Yeah // sometimes I play golf // but that’s mainly for business
DP	I play golf sometimes // but // mostly / I play golf for business.
DT	Sometimes I play golf // mainly for business

The expression *I play golf* was repeated consistently (and twice within the utterance at RP) and was clearly a useful expression for Kazu. It seemed that he was able to link this with *sometimes*, but did not memorise the combination as a fixed expression. The second half of the utterance suggests that some elements were not salient (e.g. *that's mainly*, *reasons*) and were forgotten or replaced by simpler synonyms. The expression *for business* was however consistently repeated. This was already known to Kazu (based on his use of it in the initial interview) and there was evidence of faster articulation as shown in Table 6.11.

Table 6.11: Articulation rates in Kazu-2-11

Segment	**Articulation rates AR (syllables/s)**		
	RP	**DP**	**DT**
for business	4.630	2.994	5.548
Mean (other 2-grams)	**3.573**	**3.868**	**3.989**
I play golf	3.741	3.866 5.034	3.636
Mean (other 2-grams)	**3.419**	**3.029**	**4.489**
Mean (all n-grams)	3.624	3.682	4.265

6.3.5. ***Concluding thoughts***

6.3.5.1. Segmentation and formulaicity

The analysis in this Chapter was concerned with how the L2 learners in the replication study might have gone about memorising their target utterances in order to reproduce them as accurately (and fluently) as they could in the performances. The output over the stages suggests that they broke down the longer utterances to some extent, and points to some of the ways this might have been done. When faced with the targets, they tended to know most of the individual words, many of the constructions and some of the phrases and expressions contained within them. They therefore had options: remembering strings of words in chunks then piecing them together, remembering basic frames and adding in variables, or remembering key words or phrases and constructing the utterance around these based on existing knowledge of syntax and vocabulary, or any combination of these options.

One part of the memorisation process, then, may be in choosing chunks (sequences of words within the utterance) to learn as wholes. Possible chunks could include expressions that are already known to the speaker (e.g. *Rosslyn Chapel*) or ones which are constructed as part of the memorisation (e.g. *compared with other*). They could also be extensions of known expressions (e.g. *for business* + *reasons*). The learner may have to work hard to learn a particular segment, for example if it involves a new word (e.g. *homogenous population*), although the novelty of the word or expression may also help by making it more salient. The segments chosen may be fairly random and not necessarily indicate particularly useful or obvious candidates for formulaicity (e.g. *impressive because* was a repeated segment in Aiko-3-5 even though it traversed an IU boundary in the original MU).

Deviations as a result of choosing chunks may arise from two sources. If a chunk is initially derived by extending a known expression, deviations may arise if the extension is not well established (e.g. Kazu reverting to *for business*, leading to the omission of *reasons*). Secondly, an element of a chunk may be omitted or changed because it is not recognised as an important part of the expression (e.g. Aiko dropping the determiner *the* from *the second time,* and changing *had the chance* to *had a chance*). This type of case is consistent with the finding in the quantitative analysis that morphological deviations tended to occur within fluent segments. Deviations within fluent chunks may be examples of the process described by Bardovi-Harlig (2019) where a formulaic expression becomes more target-like over time by a process of correcting or filling-in or fine-tuning the deviant (often non-salient) elements within the expressions as the speaker gets more exposure and practice of the expression.

As well as memorising particular chunks holistically, the participant also needs to remember how to piece them together. One aspect of this may involve remembering a (known or novel) expression along with the fact that it requires a variable or amplifier (e.g. *compared with other* + *cities/countries*; *much more* + *homogenous population*; *mainly/mostly* + *for business*). Deviations can occur if the variable or amplifier is not remembered and a synonym or approximation is put in its place. Memorisation of (parts of) the utterance may also involve focussing on key words or expressions (and their meanings) and trying to build that utterance (part) during reproduction, based on existing knowledge of syntax and vocabulary. For example, in memorising *The second time in particular was really impressive*, Aiko consistently remembers *second time*, *particular*, and *impressive* but pieces them together erroneously (see Section 6.3.4.3). In practice, as this example illustrates, speakers may use any of the different possible ways of memorising (choosing chunks, remembering key words, adding in variables and intensifiers, piecing all the parts together based on existing knowledge) in combination, depending on the particular utterance they are memorising.

6.3.5.2. Identifying formulaic segments

While the quantitative data suggested that dysfluencies and deviations in delivery of the targets do not consistently coincide, the analysis of the three examples demonstrates how these features may be used together to identify segments of the utterance which may have been formulaic for the speaker. The use of additional prosodic features was helpful in supporting this identification in some cases. For example, where evidence of connected speech between words (elision, intrusion, linking) occurred, it tended to be within the segments identified as potentially formulaic (i.e. those repeated in the same form at RP, DP and DT). In addition, the occurrence of elongated syllables was never within these segments, and was often immediately before. Since, like pauses, such syllable elongation at the end of a word may indicate some degree of construction or selection of what comes next, it seems appropriate to include syllable elongation as a segmentation point (along with pauses and other dysfluencies).

The information from articulation rate analysis was much less consistent in matching potential formulaicity. While some of the potentially formulaic expressions (such as *Tokyo has*, *homogenous population*, *for business*) did have relatively fast AR compared to the average, this was not so for other clear candidates for formulaicity such as *Rosslyn Chapel* and *second time*. As the examples demonstrate, AR may be affected by a number of other factors as well as the formulaicity of the expression. These may include: the degree to which the expression has been encountered or spoken in the same session (Kazu's second use of *I play golf* in the same utterance at DP was much faster than the first); lengthening of final syllables due to uncertainty of what comes next (e.g. *compared to other*) or being at the end of an IU; or differences in the ease with which two words can be connected in speech (e.g. *had the* may be more difficult to connect than *had a* due to the need to rearticulate with the tip of the tongue). AR may also be complicated by features of constituent words overshadowing any articulation benefits of the whole. Such features could include average word length (longer words may generally have faster AR) or words that are particularly easy (*Tokyo*) or difficult (e.g. *Rosslyn*) for a typical Japanese speaker to pronounce. The use of AR to identify individual cases of formulaicity therefore may not be particularly helpful. However, AR may be useful in illustrating the more general point that delivery of a given expression tends to vary depending on the stage of learning (delivery at RP tended to be faster than at DP) and on the task type (delivery at DT tended to be faster than at DP).

6.3.5.3. Conclusion

This analysis looked at how formulaic segments may be created as part of the memorisation of longer target utterances. It suggests that this may be done by chunking individual words that are known to go together (within a longer expression) or by segmenting the longer utterance into manageable pieces. The memorisation task here is not necessarily the same as the process by which an expression becomes formulaic over time. However, it does raise the question of whether the way in which a longer utterance is memorised can affect which sequences become chunked and therefore which expressions within the utterance become formulaic for the speaker. It also highlights that the process by which any chunking happens depends on the particular individual and on their experience with the words and expressions within the target utterance. Building on this, an interesting question is whether memorising a short target sequence within a longer utterance facilitates or hampers its journey to becoming formulaic, and whether drawing attention to the way in which an utterance is segmented has an influence too. These questions are explored in Chapter 7.

CHAPTER 7: Exploring targeted memorisation (Study S4)

Segmentation and embedding in the memorisation of spoken text by L2 learners of English

7.1. Introduction

While the approach to introducing novel expressions described in Chapter 5 was not specifically focussed on acquiring formulaicity, it provides a good starting point for exploring acquisition further. An important aspect of the acquisition process is how the approach to introducing the targets influences how they (or segments within them) become internally formulaic. The model utterances in study S3 were relatively long (and multi-clausal) and each one was specific to the needs of the individual participant learning it. The advantage of this design was that the novel targets were genuinely useful for real conversations, ensuring that participants were highly motivated to spend time on memorising and practising them. However, it also meant that the target utterances varied considerably in terms of length, internal structure, novelty of component words, and other features thought to influence the learning of formulaic expressions (such as idiomaticity and imageability). The method used in S3 therefore made it difficult to control the many potential factors that might influence how any expression becomes formulaic. The demands of memorising long, multi-clausal utterances may further complicate the process. An alternative method suitable for addressing these difficulties would be one in which a pre-selected set of single clause target sequences (of similar length and form) is used, and the method by which they are learnt is tested in a controlled way.

Based on the experience of study S3, a number of questions and considerations arise for planning the next step. For example, will participants be sufficiently motivated to learn target sequences that they did not select themselves? To what extent is it possible to control the input process, bearing in mind that individual participants may have different preferred learning styles and speeds? Alongside these methodological considerations, Chapter 6 raised some interesting empirical questions related to the segmentation of input for memorisation and the embedding of formulaic chunks within longer expressions. This chapter reports on a small exploratory study which is designed to test out a methodology and address some of the questions raised above. The study involves developing a set of novel target sequences and then manipulating how they are introduced to L2 participants, in order to check the effect of two possible factors (the segmentation and embedding of input) which may impact on the memorisation process.

7.2. Background

While, it is widely acknowledged that implicit and unconscious learning play a major part in the learning of a new language (e.g. Ellis, 2015; Ellis et al., 2008), for the deliberate memorisation of target expressions, it is important to look at explicit learning processes that may be involved, and the role of working memory. If a new sequence is to be memorised for subsequent oral reproduction, it needs to be rehearsed and repeated in working memory, with the aim of transferring it into long-term memory. As discussed in Section 2.5.2, the storage limit in working memory (WM) is thought to be about four items of information (Cowan, 2010). In the view of McCauley and Christiansen (2015), this limit means that chunking is a necessary part of production fluency, and it may also be relevant to the internalisation of new expressions. From the learner's perspective, a long novel expression to be learnt may consist of known or unknown words and phrases. In memorisation, those various items need to be put together, along with the meaning of the whole (and possibly meanings or labels representing sub-parts) and pragmatic information about where the full expression fits into the conversation or given context. In most cases, there will initially be too many items to hold and manipulate in WM, so some intermediate strategies must be in play. Various strategies are possible in theory. For example, learners could memorise some key words and expressions along with the overall meaning, and then attempt to piece the utterance back together during reproduction. Alternatively, they could divide the utterance into manageable chunks, memorise each chunk (one chunk at a time) and then put them together. Analysing examples from Chapters 5 and 6 suggests that a combination of such approaches was used depending on the participant and the particular utterance to be learnt.

7.2.1. *Segmentation of input*

One factor that might influence the strategy used is the way that learners are presented with the novel expression in the first place. In study S3, participants were given an audio recording of the target utterance, delivered naturally as a whole. They could then memorise it as they wished (using various possible strategies as suggested above). One common approach (according to feedback from participants) was to listen to the target and to repeat it (or parts of it) out loud. As suggested in Chapter 2, there is a range of research showing that repetition of a new expression facilitates memory for recall (e.g. Nelson, 1977) and for oral production fluency (e.g. Au & Entwhistle, 1999; Yoshimura & MacWhinney, 2007) provided that sufficient repetitions are undertaken and the expression is linked to a meaning or context. Baddeley (2003) suggests that such repetition utilises the phonological loop in working memory to transfer the phonological form of the repeated part into long-term memory.

However, research highlighting the benefits of repetition has generally focussed on short expressions. The benefits of repetition will not necessarily extend to long utterances presented in one go (unless they are easily segmentable by the learner). Results from study S3 seem to bear this out. In that study, the very shortest target utterances were all reproduced fluently and accurately. However, there were frequent cases in the dictation task (Chapter 6) where, despite hearing a known long utterance immediately beforehand, the participants were unable to accurately reproduce it. This is in keeping with the idea that, in general, it is hard to hold a long utterance in WM in order to repeat it.

It seems likely, therefore, that the way in which a learner segments a long target sequence will influence the effectiveness of the memorisation. A useful way to explore this would be to guide the participants into using different segmentation strategies through the way that the target utterances are presented. Based on the four item capacity limit for WM suggested by Cowan (2010), a good approach might be to present participants with the long target already divided in a natural way into a small number of manageable segments, each to be memorised whole via repetition. This presentation might guide them to a strategy optimised for reproducing the target utterance in fast and fluent segments. Other approaches, such as presenting long utterances whole or, at the other extreme, on a word-by-word basis, might be expected to be less effective.

7.2.2. *Embedding of input*

As discussed above, memorising shorter expressions is likely to be easier in general than memorising longer ones. The results from study S3 tend to bear this out, at least in terms of the rate of deviation and dysfluency in participant output. While repeating shorter expressions may help them to be learnt as a single unit, it is important for subsequent recall and appropriate usage that they are well linked to meaning and the context of their use. Including context in the memorising of vocabulary has been shown to be beneficial (e.g. Prince, 2012; Rodríguez & Sadoski, 2000). In study S3, participants were frequently able to reproduce particular shorter expressions within the longer utterance, and do this fluently, accurately and consistently, even though they could not reproduce the longer utterance as effectively.

One approach to memorisation that provides a specific example of use in context would be to embed a short sequence in a longer utterance and have the participant memorise the whole thing. Whether this is more effective than simply memorising the shorter sequence by itself (non-embedded) is an open question. On the surface, it seems that there would be less to learn in the latter, non-embedded case (i.e. fewer words). However, there are some potential benefits of embedded learning which could out-weigh this. Firstly, the longer exemplar provides a concrete example of the meaning and usage of the short sequence, making it easier to learn and link to a given context. Secondly, the process of learning the longer exemplar could help the learner view the short sequence as a single chunked item, particularly if it is repeatedly manipulated in WM as a chunk during the memorisation. This may be especially relevant in a situation where 'incomplete' short sequences (such as verb phrases) are to be learnt.

7.3. Overview and hypotheses

7.3.1. *Overview*

The aim of this study (S4) is to explore how the presentation of a target sequence affects the way it is learnt by an L2 speaker. The study focuses on two different manipulations of the presentation: the degree to which a sequence is segmented; and whether or not a sequence to be learnt is embedded (i.e. memorised as part of a longer expression). To do this, a series of target sequences was presented to six Japanese speakers of English to memorise, and their ability to use them later was assessed in terms of recall (whether participants could retrieve the targets given different cues) and the accuracy and fluency of the targets when reproduced.

Overall there were 12 pairs of target sequences. Each pair consisted of a short target (ST) sequence (e.g. *runs the risk of*), of four or five words, and a corresponding long target (LT) sequence of 13-17 words which contained the short target (e.g. *"If the company doesn't change, it runs the risk of losing staff to its competitors*"). Each participant was required to memorise all 12 short sequences, but while six of them were presented in their ST form, the other six were presented within their LT, with the instruction to memorise the entire LT form. The expressions to be learnt were presented within three context stories, each containing four sequences (two STs and two LTs). There were three different segmentation conditions for learning the expressions: as a whole unit (W), in two or three segments (S) or broken into individual words (I). A different segmentation condition was assigned to each story (i.e. to each set of four target sequences). In this way, each participant learnt two LTs and two STs via condition W, two LTs and two STs via condition S and two LTs and two STs via condition I.

Figure 7.1: Basic distribution of targets and conditions to stories

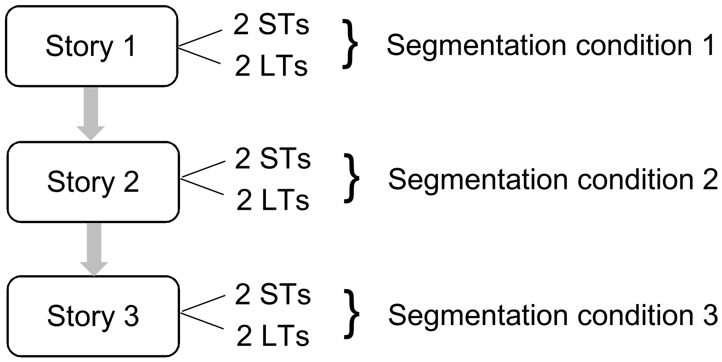


Figure 7.1 summarises the basic distribution of targets and conditions to stories. However, it should be noted that the combination of LT and ST sequences within stories, the assignment of conditions to stories, and the order that each was presented was varied for each participant to ensure a balanced distribution of target types and conditions. Further details of this process and of how the LT and ST sequences were presented are given in the next sections.

A key feature of the design of this experiment is that it allowed one set of data to be analysed in two different ways. As well as comparing the effect of the segmentation conditions (W, S, I) on how well the long (LT) and short (ST) targets were learnt, it was also possible to use the same data to explore the effect of embedding (as described in Section 7.2.2). This was possible because each LT (e.g. *the manager breathed a sigh of relief that he was safe*) contained within it a short expression equivalent to the corresponding ST (e.g. *breathed a sigh of relief*), and this short expression was therefore learnt incidentally as part of learning the LT. For the embedding analysis therefore, the learning of an LT can be re-cast as the incidental learning of an 'embedded ST' (E-ST). Data for the performance of these E-STs was therefore also collected, allowing for the comparison of an embedded condition (E) with a non-embedded (NE) condition. Although the design was complicated (see comments in 7.9.3), it did offer a controlled means of exploring two different aspects of the presentation of target sequences simultaneously.

7.3.2. ***Hypotheses***

Hypotheses are presented for three different cases: the effect of segmentation (W, S or I) on long targets (LTs); the effect of segmentation (W, S or I) on short targets (ST); and the effect of embedding (E or NE) on short expressions (i.e. according to whether they were memorised as short targets or within long targets).

7.3.2.1. Effect of segmentation (LTs)

Figure 7.2: Illustration of segmentation conditions

WHOLE (W):

[If the company doesn't change, it runs the risk of losing staff to its competitors]

SEGMENTED (S):

[If the company doesn't change] + [it runs the risk of] +

[losing staff to its competitors]

INDIVIDUAL (I):

[If] + [the] + [company] + [doesn't] + [change] + [it] + [runs] +

[the] + [risk] + [of] + [losing] + [staff] + [to] + [its] + [competitors]

Figure 7.2 illustrates how the LTs were broken up for presentation in the three segmentation conditions. The process by which the parts were presented and repeated by the participants is given in more detail in the procedure (Section 7.6).

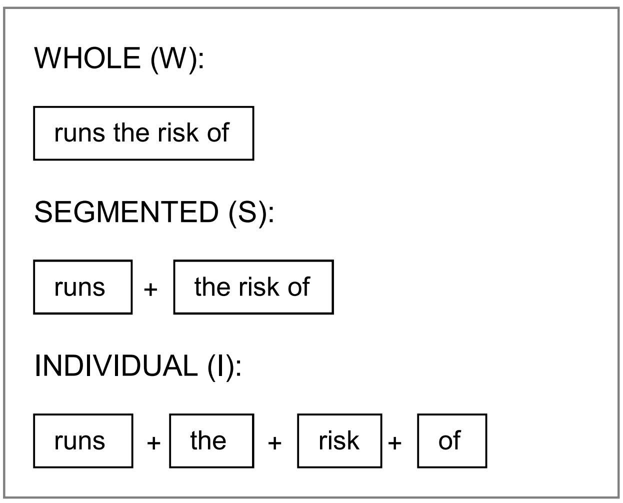
Following the discussion in Section 7.2.1, some conjectures about the effects of the segmentation conditions can be made. If the LT is presented as a whole (W), it constitutes one unit, but this unit is likely to be too large for memorisation purposes. Therefore, the participant will need to adopt a strategy for dealing with this. This may be to focus on particular salient parts, or to focus on key lexis and gist or to segment it on-line. As repetition of the whole is required during learning, adopting these strategies on-line may take additional resources in working memory and impact adversely on the memorisation process. Presenting the LT in a segmented form (S), on the other hand, focusses the participant on memorising each segment separately through repetition. Each segment is short enough to be chunked into a single unit (in working memory) and therefore easily repeated. Subsequently, the process of ordering the three segments and linking the form to the overall meaning of the LT is also manageable in working memory. In the case of presentation and repetition of the individual words (I), processing each word is manageable, but the challenge comes in putting them together. Here, too many items are involved to be processed effectively in one go. So, either some form of additional strategy will need to be adopted on-line during memorisation, or the learner will need to rely on retrieving the individual words and reconstructing the LT during reproduction. It therefore seems that segmentation (S) represents the sweet spot for learning LTs through repetition, while the other approaches are likely to result in less efficient use of working memory and chaotic segmentation or reconstruction strategies.

The hypothesis for LTs, then, was that segmented input (S) would optimise the memorisation process and lead to better recall and accuracy of output than the whole (W) or individual word (I) approaches.

7.3.2.2. Effect of segmentation (STs)

Figure 7.3 illustrates how the STs were broken down for presentation in each of the three segmentation conditions. In this case, the segmented case (S) involves splitting the short expression into two parts.

Figure 7.3: Illustration of segmentation conditions for STs



Following the reasoning set out on in the section above, it is expected that the short target sequences (STs) are already optimised (in terms of length and form) for direct acquisition. So, presentation of the whole expression (W) should be the optimum approach because it facilitates repetition of the full ST without any need for intermediate stages (i.e. piecing together sub-segments). While the other segmentation conditions (S and I) are also likely to be manageable in WM, they require an additional stage of reconstruction. This may be disadvantageous in terms of modelling the phonological form of the expression as a whole and in perceiving it as a single unit.

The hypothesis for STs, then, was that whole input (W) would lead to better recall, accuracy and fluency than the other segmenting conditions.

7.3.2.3. Effect of embedding or non-embedding

The embedding analysis is considered independently of the segmentation conditions and focuses on all 12 short sequences. For every participant, six of the short sequences were presented for memorisation on their own (as STs), while the other six were learnt incidentally as E-STs within LT sequences. This effectively led to two different embedding conditions for a short expression:

- NON-EMBEDDED (NE): the short sequence (ST) was learnt on its own.
- EMBEDDED (E): the short sequence (E-ST) was learnt (incidentally) as part of the corresponding longer sequence.

As described in Section 7.2, there are competing considerations when predicting the relative effect of these conditions. Presenting the expression in a longer expression means there is more to remember overall. So, the short expression (E-ST) is in competition with the rest of the LT and is not guaranteed to get sufficient memorisation time. However, if the embedded expression is sufficiently salient (due to its novelty and/or idiomaticity), it may become a focus or memorable feature of the longer expression. In this case, the context provided by focussing on the longer expression could help E-STs to be better recalled (in providing more links to context and meaning) and give a clearer model of how to use it in context.[25]

[25] The salience of embedded expressions could also be raised by the participant noticing how similar they are in length and nature to other expressions presented as STs.

A tentative hypothesis, then, was that short expressions in the embedded condition (E) would have higher recall, but lower accuracy and fluency compared with those in the non-embedded condition (NE).

7.4. Design

While the focus is on the two delivery approaches described, there are, as discussed in previous chapters, a number of other factors that that could influence how well the target sequences are memorised. These include individual differences between the participants in proficiency, approach and familiarity with the expressions or elements within them. In addition, features of the expressions themselves may influence memorability, such as form, level of transparency, pronounceability, salience of component parts and imageability. Finally, there is the possibility of order effects in the presentation of the stimuli.

To control for these possibilities, three key elements were incorporated in the design: careful selection of target sequences; a Replicated Greco-Latin Square Design; and a highly regulated directive for the memorisation task.

7.4.1. *Selection of target utterances*

The criteria for selection of the short sequences were that they should be

- Of similar form
- Distinct from each other (i.e. not share any lexical component words)
- Unfamiliar to the participants, but with familiar lexis
- Not fully transparent (i.e. not obviously ‘constructible’ from component words)
- Useful and easily understandable from context
- Of similar salience

To satisfy these criteria, it was decided to use verb phrases of four or five words, each with one familiar verb and one or two other familiar lexical words, and two function words. The following process was used to derive the examples:

1. Gather a list of about 100 high frequency 4 or 5 word verb phrases from the Phrases in English (PIE) on-line corpus (W. Fletcher, 2011).

2. With two other native speakers, sort and check these for consistency and familiarity, giving a list of 34 candidates.[26]

3. Create a test to check familiarity and administer to the participants.[27] This left 20 items that were unknown to the L2 learners.

4. Select 12 items (of different lexis) which could reasonably be included within three different created stories of about 150 words. These items are the short target sequences (STs).

5. Create the stories ensuring that the short sequences are embedded in sentences of 13-17 words long. These sentences become the long target sequences (LTs).

Table 7.1: Target utterances

Story	#	Short Target (ST)	#	Long Target (LT)
X	S1	get the hang of	L1	Once you get the hang of it, you can solve the puzzle quite quickly.
	S2	see the light of day	L2	Most of his designs didn't see the light of day, but the cube was an immediate success.
	S3	stood the test of time	L3	As a puzzle it has stood the test of time and is still very popular today.
	S4	paved the way for	L4	The success of the Cube paved the way for many new kinds of 3D puzzle.
Y	S5	runs the risk of	L5	If the company doesn't change, it runs the risk of losing staff to its competitors.
	S6	like the sound of	L6	Most employees like the sound of this idea and think it will improve their working lives.
	S7	sitting on the fence	L7	The Board are sitting on the fence and will not make a clear decision either way.
	S8	remains to be seen	L8	It remains to be seen whether the new technology will actually make any difference.
Z	S9	turned a blind eye to	L9	He turned a blind eye to the bad behaviour because Joe was such a great player.
	S10	came to a head	L10	The problems came to a head when he was late one day for an important match.
	S11	breathed a sigh of relief	L11	When he finally arrived, the manager breathed a sigh of relief that he was safe.
	S12	fell into the hands of	L12	The story fell into the hands of a local journalist who then wrote an article about it.

[26] This consultation with native speakers was also applied in the selection of any new target sequences used in the subsequent studies S5 and S6.
[27] Details of the test are given in Appendix 7.1.

The set of target utterances resulting from the process are given in Table 7.1. From the 24 possible short (ST) and long (LT) target sequences, two different sub-sets (A and B) were created, each containing an equal mix of ST and LT sequences. Set A consisted of S1, L2, S3, L4, L5, S6, L7, S8, S9, L10, S11, L12 (shaded in blue type in Table 7.1), while Set B contained the reverse combination (shaded in black type in Table 7.1). The sets were chosen so that ST and LT sequences alternated in the story and differences in short target length and 'transitivity' (whether the short sequence required additional words to complete the phrase or not) were spread evenly across the sets.

Figure 7.4: Story X for Set A

X – A: Rubik's Cube

The Rubik's Cube is a popular puzzle invented by Erno Rubic who was a designer from Hungary. The puzzle has six sides and each side has 9 squares of different colours. The aim is to make each side of the cube the same colour by twisting the cube. It is difficult to do at first, but with practice it becomes much easier. Once you **get the hang of** it, you can solve the puzzle quite quickly.

Rubik invented many different kinds of toys and puzzles. **Most of his designs didn't see the light of day, but the cube was an immediate success.** When it was released in the 1980s, millions were sold and it became the best-selling toy in the world. As a puzzle it has **stood the test of time** and is still very popular today. Over 350 million were sold in 2009. **The success of the Cube paved the way for many new kinds of 3D puzzle**. There is also a World Cube Association which organises competitions and keeps official world records.

For each subset (A or B), the relevant LT and ST targets were highlighted within the stories. For example, Figure 7.4 shows story X for set A. The stories for both sets were identical except for which version (ST or LT) of each target was highlighted. The full set of stories and accompanying pictures are shown in Appendix 7.2.

7.4.2. *Use of Greco-Latin Square design*

To mitigate potential memorisation differences across participants and targets a design was chosen that ensured that the experimental delivery conditions (segmentation and embedding) were spread evenly across participants and target sequences. Each participant (and each target) had an equal exposure to each segmenting condition and each embedding condition, and the order of these conditions was varied to control for possible order effects. To ensure that order and experimental conditions were applied in an even ‘orthogonal’ way, a Greco-Latin design was applied[28]. The design is summarised in Table 7.2.

Table 7.2: Greco-Latin Square Design

		Story / Target sequence		
		X-A (S1,L2,S3, L4)	**Y-A** (L5,S6,L7,S8)	**Z-A** (S9,L10,S11,L12)
Participant	**P1**	W – 1st	S – 2nd	I – 3rd
	P2	S – 3rd	I – 1st	W – 2nd
	P3	I – 2nd	W – 3rd	S – 3rd
		X-B (L1,S2,L3, S4)	**Y-B** (S5,L6,S7,L8)	**Z-B** (L9,S10,L11,S12)
	P4	W – 1st	S – 2nd	I – 3rd
	P5	S – 3rd	I – 1st	W – 2nd
	P6	I – 2nd	W – 3rd	S – 3rd

Note: X-A represents Story X, Set A with target sequences S1, L2, S3, L4, etc. P1 – P6 represent the different participants. W, S and I represent the different experimental presentation methods. The ordinal (1st, 2nd or 3rd) shows the order in which each set is delivered.

7.4.3. ***Memorisation task***

To ensure that, as far as possible, the only thing that varied in the memorisation task for each expression (whether LT or ST) was the experimental delivery condition (W, S or I), the following procedure was adopted. Firstly, as indicated above, the expressions were included in a story in order to provide context to support understanding of meaning and a cue for subsequent recall and production. Research (e.g. Wray, Bell, & Jones, 2016) suggests that context is important for understanding formulaic sequences, which is in turn a key factor in memory. Also, embedding target vocabulary in stories has been shown to support memorisation (Prince, 2012). Secondly, the main means of memorisation was oral repetition of the sequences, as this has been shown to be a highly controllable approach (e.g. Ellis & Sinclair, 1996). The number of repetitions of each word in the expression was the same across all conditions. This is important because the number of repetitions during memorisation

[28] In this design, the sub-sets A and B are simply a means for ensuring that the LT and ST in every pair is tested evenly across the whole set of participants. In the design, A and B are just labels for the two replicates of the Greco-Latin square and there is no variable distinguishing them. See Appendix 7.2 for more on how these replicates are treated in the statistical analysis.

has been shown to be a significant factor in recall and subsequent fluency of expressions (Yoshimura & MacWhinney, 2007). Further, because the long and short expressions were equally divided for each participant, the overall memory 'load' was equivalent.

7.5. Participants

For this small exploratory study, there were six Japanese speakers of English (JSE), all female with ages ranging from 36 to 58. Participants were chosen on the basis of availability, level of English and because they were interested in taking part. Informed consent was obtained and they were assured about the anonymity of their contributions. On volunteering they were given two simple vocabulary tests: V_YesNo (Meara & Miralpeix, 2016) and Lex30 (Fitzpatrick & Clenton, 2010; Meara & Fitzpatrick, 2000) both accessed from the Lognostics website (Meara, 2014). While there was some degree of variation in proficiency, the tests suggested that all were at an intermediate level of English.

Table 7.3: Participants (using pseudonyms)

#	Name	Sex	Age	YesNo	Lex30
P1	Yoshie	F	40+	6240	68
P2	Fumi	F	40+	5200	52
P3	Fukiko	F	50+	6340	55
P4	Mariko	F	30+	5800	60
P5	Junrei	F	30+	5333	59
P6	Kyoko	F	50+	6400	70

7.6. Procedure

7.6.1. *Initial meeting*

Each participant individually attended an initial meeting (of about 75 minutes). In this, they first memorised the targets in three sub-sessions, corresponding to the three stories/delivery conditions as prescribed in Table 7.2. For each story, the procedure was as follows:

- The researcher read the story out loud. The participant listened while observing the two representative pictures (but without looking at the transcript). The pictures were included to provide a visual symbol of the story as an aid to memory and as a cue for recall.
- The researcher read the story again. This time the participant could follow the transcript of the story. In the transcript, the target sequences (two short, two long) were in bold text (see Figure 7.4).
- The researcher checked that participant understood the story and offered to clarify any (non-target sequence) vocabulary. (In fact, this was never required).
- The researcher presented each target sequence in turn. The participant was first shown the target sequence on flash cards corresponding to the segmentation condition (whole, segmented, or individual words) as illustrated in Figures 7.1 and 7.2. At the same time, they were shown a card showing the Japanese translation (of the whole expression). They then repeated the expression out loud a number of times, as shown in Table 7.4.
- Finally, all four expressions were repeated (in whole, segmented or word-by-word form) twice.

Table 7.4: Repetition procedure for each condition

	Whole (W)	**Segmented (S)**	**Individual words (I)**
With script	whole expression x 4	each segment x 3 segmented expression x 1	each word x 3 word-by-word expression x 1
Without script	whole expression x 6	each segment x 3 segmented expression x 3	each word x 3 word-by-word expression x 3

'segmented expression' means the full expression was repeated on a segment-by-segment basis.

'word-by-word expression' means the full expression was repeated on a word-by-word basis.

7.6.1.1. Elicitation of memorised expressions

After each story input, there was a short break of 3-4 minutes where the researcher engaged the participant in simple small-talk (unconnected to the task).

The participant was then asked to recall the story, prompted by the pictures. They were encouraged to use any words and expressions they remembered (Initial Story recall). They were then asked to list the four sequences that they had memorised, first without any prompt (Initial Elicit), and then again when given a prompt consisting of the Japanese translations of the expression meanings (Initial Prompt).

After all three stories had been covered, there was another 3-4 minute break. Participants were then asked to recall all 12 target sequences (6xST, 6xLT) that they had memorised (Second Elicit). Finally, they were given prompts (Japanese translations) for all 12 short sequences and asked to recall and say these (Initial ST Prompt). That is, they were asked to produce the six STs that they had learnt, and the six E-STs that they had not learnt explicitly but which had been embedded within LTs that they had learnt. This was to provide data for a comparison between the non-embedded (NE) and embedded (E) conditions.

7.6.2. *Follow-up meeting*

A second, follow-up meeting (about 30 minutes) was undertaken a week later. In the intervening time, participants were instructed not to practise the sequences. After a brief warm-up, participants were asked to recall again the sequences on a story-by-story basis. For each story, they were shown the picture and asked to re-tell the story (Delayed Story). They were then asked to list the sequences, first without any other prompt (Delayed Elicit) and then with the Japanese translation (Delayed Prompt). After all stories had been covered, they were then given the prompts for all 12 short sequences (Delayed ST Prompt).

Participants were then interviewed about their experience, particularly with regard to how motivated they were to learn the target sequences and how they found the different input approaches.

7.7. Dependent Variables

All interactions with participants were recorded on a Sony IC recorder and the responses were transcribed and analysed. The variables that were calculated are listed in Table 7.5.

To compare the effect of the segmenting, results for the LTs and STs were analysed separately. For each participant, the relevant measures listed above were calculated for each of the targets they had memorised (six STs and six LTs). To compare the effects of embedding versus non-embedding, the focus was on the 12 short expressions, that is, the six STs they had learnt (Non-embedded) and six E-STs memorised incidentally as part of an LT (Embedded).

Table 7.5: Dependent variables

Whole v Segmented v Individual words		**Embedded v Non-embedded**
Target: LTs	**Target: STs**	**Target: STs & E-STs**
Initial Recall	Initial Recall	Initial Recall
Initial Accuracy	Initial Accuracy	Initial Accuracy
Delayed Recall	Initial Fluency	Initial Fluency
Delayed Accuracy	Delayed Recall	Delayed Recall
	Delayed Accuracy	Delayed Accuracy
	Delayed Fluency	Delayed Fluency

7.7.1. ***Accuracy***

Accuracy was calculated as follows. For each utterance elicitation (e.g. Initial Story, Initial Elicit, Initial Prompt, Second prompt),

$$\text{Accuracy} = \frac{\text{(no. matching words) – (no. of inserts) - (no. of WO errors)}}{\text{(total no. words in utterance)}}$$

Note: *Incorrect forms of correct lemmas count as 0.5*

Hesitations, reformulations/repetitions do not count as insertions

WO (word order) errors = minimum no. of phrase swaps/moves required to correct

This calculation measures the number of words that are correct as a percentage of a perfect presentation, but penalises if additional words or phrases are inserted or if the words are in the wrong order.

Table 7.6: Fumi (L10, Initial prompt)

TARGET	The problems came to a head when he was late one day for an important match
ELICITED	The problem (..) came to the head when <laugh> when he (3) when he- er he was too late at a- an (.) at an important match (1) one day.

Total words = 16; Exact matches = 13, lemma matches = 0.5 (problem); inserts = 1 (too); WO = 1 (one day)

Accuracy = (13.5 - 1 – 1) / 16 = 0.719

Table 7.6 gives an illustration of the calculation for an elicited long target. For each utterance/participant, Initial Accuracy is the average of accuracy scores over all initial elicitations (Initial Story, Initial Elicit, Initial Prompt, Second Prompt). Similarly, Delayed Accuracy is the average over all delayed elicitations.

7.7.2. ***Recall***

To measure recall, there needed to be some way to determine whether an attempt to retrieve a target sequence indicated they had recalled it to some extent (even if they were unsure of the exact form). For each elicitation of a target sequence, a recall score based on accuracy (as described in 7.7.1) was therefore calculated as follows:

$$\text{Recall} = \begin{cases} 0 & \text{for accuracy} < 0.3 \text{ (no recall)} \\ 0.5 & \text{for accuracy } 0.3 - 0.6 \text{ (partial recall)*} \\ 1 & \text{for accuracy} > 0.6 \text{ (recall)} \end{cases}$$

Note: for short utterances, if more than 1 lexical word is missing, recall = 0

For each participant-target item, Initial Recall was calculated as the average of recall scores over all initial elicitations (Initial Story, Initial Elicit, Initial Prompt, Second Prompt). Similarly, Delayed Recall was the average recall over all delayed elicitations.

7.7.3. ***Fluency***

Fluency (defined in terms of filled and unfilled pauses) is considered a necessary indicator of psycholinguistic formulaicity in this and other research (Myles & Cordier, 2017). Fluency was therefore measured for each elicitation of a short expression (ST or E-ST) as follows:

$$\text{Fluency} = \frac{(3 \text{ x no. of word gaps}) - \sum(\text{dysfluency values})}{(3 \text{ x no. of word gaps})}$$

This was derived according to the following steps:

1. Each word gap in the elicited expression was examined and given a ‘dysfluency value’ depending on the severity of any pause found there:

    0 = no dysfluency

    1 = minor dysfluency [pause > 0.2s; repetition dysfluency (false starts where repetitions precede a fluent continuation)]

    2 = medium dysfluency [pause > 0.5s; short filled pause; elongated previous syllable (>0.4s)]

3 = major dysfluency [pause > 1.0s; long filled pause; any other major pause phenomenon]

2. The sum of dysfluency values was calculated and subtracted from the maximum possible level of dysfluency for the expression (i.e. 3 x no. of word gaps)
3. The resulting value is expressed as a percentage

An example illustrating this measurement is given below in Table 7.7:

Table 7.7: Fumie (S11, Second elicit)

TARGET	breathed a sigh of relief
ELICITED	breathed- (1) breathed the (.) sigh er to relief
EFFECTIVE	breathed /// the / sigh / to relief

(n) = pause of n seconds; (.) = pause < 0.5s

Total word gaps = 4; Minor dysfluency (/) = 2; Medium (//) =0; Major (///) = 1;

*Fluency = (3*4 – 5) / (3*4) = 0.583*

For each participant-utterance, Initial Fluency is the average of fluency scores over all initial elicitations (Initial Story, Initial Elicit, Initial Prompt, Second Prompt), Similarly, Delayed Fluency is the average fluency over all delayed elicitations.

7.8. Results

7.8.1. *Degree of segmentation (Long targets)*

Recall and accuracy figures were calculated for the 12 LTs (each of which was memorised by three different participants). The resulting 36 participant-target items had been divided equally between the three conditions, resulting in 12 items for each condition. The mean recall and accuracy figures are given in Table 7.8.

The hypothesis was that segmented input (S) would lead to better recall and accuracy of output. However, as can be seen from the table, there was no clear picture consistent with the predicted effect.

Table 7.8: Mean recall and accuracy of LTs by segmentation condition

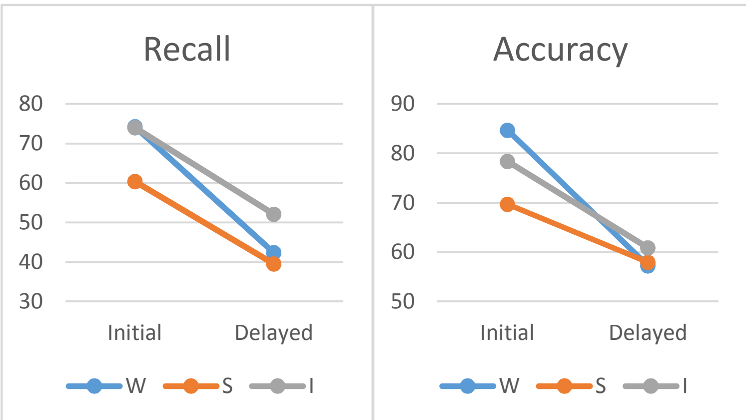
	Initial		Delayed	
	Recall %	**Accuracy %**	**Recall %**	**Accuracy %**
W	74.3 sd=0.328 n=12	84.7 sd=0.143 n=11	42.4 sd=0.294 n=12	57.3 sd=0.218 n=11
S	60.4 sd=0.266 n=12	69.7 sd=0.189 n=12	39.6 sd=0.339 n=12	58.0 sd=0.145 n=10
I	74.0 sd=0.244 n=12	78.4 sd=0.130 n=12	52.1 sd=0.302 n=12	60.9 sd=0.130 n=11

sd = standard deviation; n = number of items measured (max = 12).

Accuracy was not calculated for items that were not recalled.

The graphical representation of the data in Figure 7.5 shows that segmented input in fact has lower recall and accuracy than the other two conditions. The diagram also shows how recall and accuracy were lower at the delayed stage compared to the initial. An analysis of variance applied to Replicated Latin squares (see Appendix 7.2 for details) found no significant difference between the conditions at either stage.

Figure 7.5: Mean recall and accuracy of LTs by segmentation condition



7.8.2. *Degree of segmentation (Short targets)*

Recall, accuracy and fluency figures for the short target (ST) expressions are shown in Table 7.9.

Table 7.9: Mean recall, accuracy and fluency of STs by segmentation condition

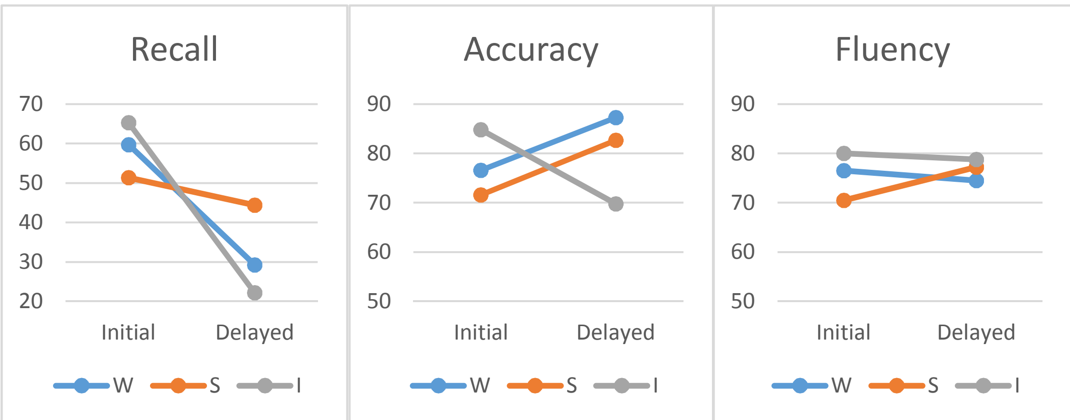
	Initial			Delayed		
	Recall %	**Accuracy %**	**Fluency %**	**Recall %**	**Accuracy %**	**Fluency %**
W	59.7 sd=0.271 n=12	77.0 sd=0.196 n=12	76.5 sd=0.213 n=12	29.2 sd=0.327 n=12	87.3 sd=0.193 n=7	74.5 sd=0.244 n=7
S	51.4 sd=0.398 n=12	71.6 sd=0.183 n=9	70.5 sd=0.177 n=10	44.4 sd=0.484 n=12	82.7 sd=0.125 n=7	77.2 sd=0.269 n=7
I	65.3 sd=0.361 n=12	84.8 sd=0.169 n=10	80.0 sd=0.122 n=10	22.2 sd=0.378 n=12	69.8 sd=0.291 n=7	78.8 sd=0.269 n=7

sd = standard deviation; n = number of items measured (max = 12).

Accuracy and Fluency were only calculated for those items that were recalled.

The hypothesis was that the whole input (W) would lead to better recall, accuracy and fluency. As the table shows, for all three measures, mean values for the W condition are not higher than those for the segmented (S) or individual word-by-word (I) approach. This can be observed more clearly in the graphical representation in Figure 7.6. Again, no significant difference in mean scores between the conditions was observed in either the initial or delayed case for any of the measures.

Figure 7.6: Mean recall, accuracy and fluency of STs by segmentation condition



While not the main focus of the study, it is interesting to compare results for STs with those for the LTs. The overall mean recall values for STs at initial and delayed stages (59.3% and 31.9) were lower than those of mean recall values for the LTs (70.0%

and 44.7%). Although the differences (analysed via paired t-tests[29]) were not found to be significant, these results, and the fact that overall mean accuracy values for ST and LTs were very similar, suggest that it was no more difficult to memorise long targets as short ones in this study.

7.8.3. ***Embedding of short sequences in longer sequences***

The recall, accuracy and fluency at initial and delayed stages were calculated for all the short expressions (STs and E-STs). As described in Section 7.3.1, the E-STs were learnt as part of the longer expressions (embedded) and the STs were learnt on their own (non-embedded). Results for initial and delayed performance (recall, accuracy and fluency) were calculated based in the Initial ST Prompt and the Delayed ST prompt, respectively. For each of the six participants there were scores for six embedded (E) and six non-embedded (NE) sequences, giving a total of 36 items for each condition.

Results were initially calculated separately for each of the three different delivery method styles (W, S and I). However, since as observed above, there was little difference between these, the results were combined (i.e. delivery method was ignored). These results are given in Table 7.10 and illustrated in Fig 7.7. In all instances, the embedded sequences scored higher than the non-embedded ones, though not by much and an analysis of variance on the Replicated Latin Squares Design (see Appendix 7.2) indicated that differences between the means for these results were not significant.

Table 7.10: Mean recall, accuracy and fluency by embedding condition

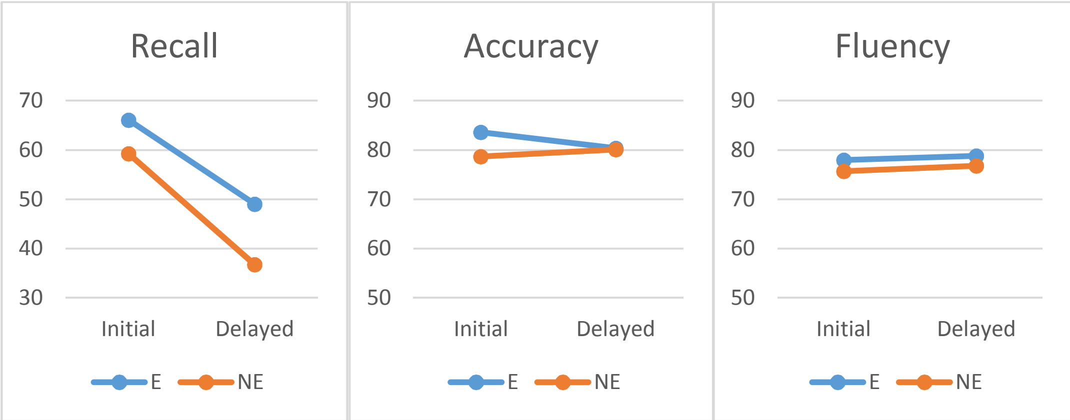
	Initial			**Delayed**		
	Recall %	**Accuracy %**	**Fluency %**	**Recall %**	**Accuracy %**	**Fluency %**
E	66.1 sd=0.300 n=36	83.6 sd=0.188 n=34	78.0 sd=0.207 n=34	49.0 sd=0.418 n=36	80.4 sd=0.294 n=26	78.8 sd=0.210 n=26
NE	59.3 sd=0.321 n=36	78.7 sd=0.198 n=32	75.7 sd=0.176 n=32	36.8 sd=0.406 n=36	80.1 sd=0.243 n=21	76.8 sd=0.248 n=21

sd = standard deviation; n = number of items measured (max = 36).

Accuracy and Fluency were only calculated for those items that were recalled.

[29] In the paired t-tests, corresponding STs and E-STs were paired on the basis of the segmentation condition by which they were learnt.

Figure 7.7: Mean recall, accuracy and fluency by embedding condition



7.8.4. ***Feedback from participants***

Feedback comments and observation during the study indicated that the participants were highly motivated to do as well as possible, and that they enjoyed the study and found it useful. The stories and the target sequences were well understood by all participants who were each able to reproduce the stories in detail, even if they could not remember the target sequences. None of the target sequences was known to the participants before the study, but all were well understood as a result of the input.

Most participants said they preferred the 'Whole sequence' delivery:

> JUNREI: "I think the last way to remember, the whole thing is a little bit easier because I understand the meaning of the sentence. So the most difficult way is the one-by-one."

Some participants also expressed a preference for the embedded sequences over the non-embedded:

> YOSHIE: "with the longer expressions it's easier to find a hook for my brain's drawers"

> FUMI: "I like to remember the whole sentence because the meaning is a reminder of the phrase"

The overall level of performance varied across the participants but, for all, recall was lower on the delayed performance than on the initial. However, mean accuracy and fluency (of the targets that were recalled) was fairly consistent between performances.

7.8.5. *Variation in target memorability*

To help appraise the appropriacy of the method, it is useful to check the performance of the long and short target utterances across the participants. For long utterances, there seemed to be a wide variation in how well they were memorised. Based on the recall and accuracy data, the best performances were for the following:

- *The Board are sitting on the fence and will not make a clear decision either way.*
- *If the company does not change, it runs the risk of losing staff to its competitors.*
- *The story fell into the hands of a local journalist who then wrote an article about it.*

The most difficult seemed to be:

- *Most of his designs didn't see the light of day, but the cube was an immediate success*

For short sequences there was less of a performance spread, but overall '*sitting on the fence'* was best memorised and '*see the light of day'* worst. Neither length (whether 4 or 5 words) nor 'transitivity' seemed to be factors that affected performance. From the perspective of the delivery context, story X (about the Rubik's Cube) was overall slightly less well remembered than the other two.

7.9. Discussion

7.9.1. *Effect of input conditions*

Due to the small number of participants in this exploratory study, it is not possible to draw any definitive conclusions about the hypotheses. However, the results observed do not lend support to the idea that segmenting sequences into manageable chunks for repetition aids memorisation or fluency. Indeed, for the long targets (LTs), the segmented approach seemed to be the least successful one. The initial conjecture was that the different degrees of segmentation during repetition would change the way that the sequence was perceived in the mind of the participant (as a whole, as three or four expressions linked together, or as a set of individual words joined according to the syntax of the language). However, since in all three conditions the participants also heard and saw the whole utterance in context in the story, this factor may have been more salient to the participant than the way the input was segmented for repetition. It may be that, in order to influence the way that an expression is taken on board, clear differences in the processing approach (e.g. elaborative or analytic approaches versus repetition) are required. This will be explored further in Chapter 9.

There was some indication that embedding (i.e. learning a short target sequence by placing it in a longer sentence and memorising the whole thing) could be helpful. For example, recall of embedded targets (E-STs) was slightly better, and some participants indicated that embedding helped them to understand and remember the expression better. However, any differences in recall as a result of embedding in this way were not statistically significant, and there did not appear to be any clear effect on fluency or accuracy either. It could be conjectured that any possible advantage of embedding (e.g. through the provision of context or the facilitative effect of the cotext) would be off-set by the increased difficulty and effort in having to memorise a longer expression. However, since recall and accuracy values for LTs and STs were in fact similar, results from this study do not support this conjecture. Overall, there does seem to be some value in providing some context for the expressions to be memorised by including them in a story and hearing them repeated in an example sentence. This was provided for all short target sequences in the study. However, the crucial difference between embedded (E-ST) and non-embedded (ST) expressions was whether the example sentence was memorised or not, and the advantages or disadvantages of this feature remain to be established.

7.9.2. *Reflections on experimental methodology*

Overall, the design of the experiment ensured that all participants, sequences and order of delivery were shared evenly across the experimental conditions. This was important because it ensured that the results were balanced across inherent variation in these elements. As it turns out, although the results suggested no effect of order on the overall memorisation, the performance of participants and the different sequences did vary. Even though the participants selected were of similar background and level of proficiency in English, the results suggested a fair degree of variation in their ability to memorise the sequences using this method. Similarly, despite the fact that the expressions were carefully selected to be similar in form, length, familiarity, salience and constructability, only the short expressions were memorised consistently (with a couple of exceptions), while there was quite a lot of variation across the longer ones.

Importantly, the learning of pre-selected sequences did not seem to lessen the participants' motivation to memorise and reproduce these expressions as well as they possibly could. While the circumstances of their reproduction were not as authentic as a real conversation, asking the participants to re-tell the story first (and use the memorised sequences in this context) does seem to have been a positive experience for them, and provided a natural, if highly focussed, way to elicit the sequences in the first instance. In terms of the input approaches, the participants commented that they preferred the presentation of the whole utterance most (and individual words the least). Interestingly, although they often initially struggled to repeat whole long utterances (without the script), this seemed to have the effect of making them focus more carefully, guiding them to parts where they needed to pay more attention in the next repetition. So, the improvement over repetitions was more marked than in the other cases. It may be that a degree of initial struggle makes something ultimately more memorable because there is a greater depth of processing (see Ellis, 1995; Ellis, 2006).

7.9.3. *Conclusion*

This exploratory study showed that introducing a set of novel target sequences can be an effective means of testing aspects of acquisition, provided that sufficient context is given and story re-telling is part of the assessment. This method therefore provides further scope for investigating factors that may influence how target sequences become formulaic for L2 speakers over time. Overall, the choice of the short target sequences seems appropriate and the memorisation of these kinds of multi-word verb phrases appears to be a realistic and useful goal for intermediate / advanced Japanese speakers of English.

While the design of the study was effective in controlling for the many variables that could influence the process, exploring multiple conditions made it quite complex. One challenge arising is that, even with a larger number of participants, there will be small numbers of items in individual cell conditions, reducing the power of the study to see anticipated effects. Therefore, it would be better in future studies to focus on only one variable condition in the input. The results from the study also suggest that embedding (in the form done here) may not influence the memorisation process much, at least not in terms of formulaicity. Based on these points, the approach for the next empirical study is to utilise the basic method used here, with a similar set of short sequences and a focus on one input condition.

An advantage of using shorter target sequences is that they can potentially be learnt in a holistic way without necessarily needing to be reconstructed. However, as this and previous studies show, this does not necessarily happen even when participants are presented with targets as whole expressions. The next part of the research, then, will focus more closely on the kinds of condition that might influence the way that a learner takes the target on board, and the effect on the subsequent formulaicity of the expression in reproduction. As this study showed, simply varying the way a target is segmented in repetition may not be sufficient to affect the way it is stored and processed. The next chapter will therefore return to the literature to better understand how formulaic expressions are represented in the mind and how the acquisition might be modelled.

CHAPTER 8: Models of acquisition and storage

Review of research for modelling the acquisition and storage of formulaic expressions

8.1. Introduction

The studies in Chapters 3-7 have explored the acquisition and use of formulaic expressions in L2 speech, with formulaicity defined with respect to the individual speaker's storage and processing of a given expression as a single holistic unit (Myles & Cordier, 2017; Wray, 2002a). A key methodological approach has been to present novel target sequences to L2 speakers to memorise and then to assess how they are used in subsequent spoken production. Since the unitary nature of the expressions is pre-determined, formulaicity of the expression (for the individual speaker) is then determined principally in terms of the phonological coherence of the expression, as represented by the fluency and accuracy of its delivery and the ease of recall.

Overall, the end-point of formulaic acquisition is taken to be when an expression becomes internalised in some way with a unitary semantic or procedural representation in the lexicon. Findings from the empirical studies (S1-S4) suggest that, in many cases, the targeted expression does not become formulaic immediately. There often appears to be an intermediate stage marked by dysfluency, inaccuracy or low recall. Shorter sub-sequences within longer targets may become formulaic first (as part of the process of the longer sequence becoming formulaic). In other words, while some expressions may become internalised as formulaic immediately, others may go through a period where they are reconstructed, before (with practice and repetition) the components become fused into a single unit.

The next phase of the research is to explore and model these different possible routes to formulaic acquisition, and to consider factors that might affect which route is chosen. In order to do this, it is useful at this stage to consider how formulaic expressions might be represented as holistic units in the mental lexicon of the speaker. This chapter therefore goes back to the literature to review ideas about holistic storage and processing, and to consider existing models of lexical representation and acquisition that are relevant. There are three main sections to the chapter. The first relates to the processing and storage of formulaic expressions and looks at the empirical evidence for holistic storage and automaticity in processing.

The second part considers some different models of lexical access and storage which may be relevant to formulaic expressions. The final part consider how these models may relate to the acquisition process.

8.2. Are formulaic expressions stored as holistic units in the mental lexicon?

The underlying view in much of the literature is that formulaic expressions are stored and processed as whole units without the need to construct them on-line (Dahlmann & Adolphs, 2007; Schmitt & Carter, 2004; Wray, 2002a). It is generally argued that such storage and processing provides the processing efficiencies necessary to enable the fluent connected multi-clause discourse of native speakers (Pawley & Syder, 1983; Tremblay & Baayen, 2010). Support for the holistic storage and processing view is also implicated in the idea that production of formulaic expressions involves automatic processing. Schmidt (2001) associates such processing with direct access to long term memory, and contrasts this with controlled processing which is associated with construction of strings from sub-components that are individually held in long term memory, but assembled in working memory. Further, as highlighted in Chapter 2, formulaic sequences have been specifically associated with holisticity in the mental lexicon (Wray, 2002a) or as morpheme equivalent units (Wray, 2008a) as described below.

8.2.1. *The heteromorphemic lexicon*

A particular conceptualisation of holistic storage is provided by Wray (2002a, 2008a) in the form of the Heteromorphemic Lexicon (2008a p.12). This is the idea that the individual's mental lexicon contains not only morphemes and words but also multi-word strings and lexicalised frames that are treated as if they were single entities. Such a lexicon is not stream-lined (as it contains both sequences and their constituent parts) but it allows for more efficient processing. The elements of the lexicon are termed Morpheme Equivalent Units (MEUs) and effectively make up the basic building blocks of the language. The implication is that if a given text were broken up into the constituents that a given speaker used to construct it, some of them would be small and simple (morphemes, words) and some would be longer and more complex. Wray equates the MEUs with formulaic sequences[30] and proposes

[30] Strictly speaking, Wray (2008) describes the MEUs as being the (theoretical) units that are stored in the lexicon while 'formulaic sequences' are sequences used by individuals that "appear to be" MEUs for them.

that, for each individual, the set of these is dynamic and constantly changing as well as being unique to that individual. This is consistent with the research (explored in Chapter 2) that suggests that individual L2 learners vary considerably in their use of formulaic expressions, both compared to other learners and also compared with themselves over time.

While the Heteromorphemic Lexicon implies holistic storage of formulaic expressions, the nature of that storage is not made explicit. In some respects, this may be considered an advantage because it allows the model to be compatible with other approaches that do not specify a particular area of storage for lexis in the brain. For example, connectionist approaches (Dell, Chang, & Griffin, 2010) identify memory in terms of strength of connections between neurons in a network. So, in this sense, 'holistic storage' may be considered a metaphorical representation of the situation.

Another perspective might be to consider formulaicity in terms of how an expression is indexed by semantic or procedural nodes. For example, a fixed expression like *See you later* may be indexed for the single meaning (e.g. 'a casual way to say goodbye'), while a similar, constructed expression, such as *I expect to see you at 7 tonight* would require several (e.g. an idea of personal expectation, an idea of meeting, and a time). A formulaic frame such as *not only X but also Y* may also by indexed by one idea (with the open-class items X and Y requiring two further nodes). However, even in this sense, the implication is that the access and processing of a formulaic expression should be similar to that of a single word, with a unitary representation in the mind. While this has strong intuitive appeal, it is useful to consider the extent to which it is supported by empirical research.

8.2.2. *Challenges to the idea of holistic storage*

Siyanova-Chanturia (2015) draws on a variety of research to question the extent to which formulaic sequences can be considered to be stored and processed as a single unit in the mental lexicon. She argues that although many studies (e.g. Jiang & Nekrasova, 2007; Kim & Kim, 2012; Underwood et al., 2004) show a processing advantage for certain formulaic expressions (see Chapter 2), fast processing alone does not necessarily imply holisticity of processing (or storage). For example, processing speeds could be related to word-to-word transitional probabilities between words in the expressions. It is known that speakers are sensitive to the statistical information regarding word combinations as a result of their experience of the language (Ellis, 2012). So, more familiar expressions may have higher transition probabilities leading to faster construction of the sequence during processing. To demonstrate holistic storage, Siyanova-Chanturia (2015) argues, experiments would have to be specifically designed to examine ‘the activation or prominence of the individual components within a larger chunk’ (p.291). There have been some studies that do take the individual components into account. For example, Kapatsinski and Radicke (2009) used a process whereby participants detected the particle *up* in two-word verbs (e.g. *turn up*). After controlling for differences in particle duration, they found that detection was slower for phrases with very high frequency than for those with medium frequency (in the language), and concluded that this was because the high frequency phrases were processed as a unit. However, such a result still does not necessarily imply holistic processing or storage because the processing effect observed for more frequent two-word phrases could, as before, be the result of transitional probabilities creating stronger links between the individual words in sequence. In other words, while their finding does not preclude holistic storage, we cannot use it to arbitrate between the two possible explanations. Therefore, other criteria or measures beyond fast processing (or fluency) may be required.

Some studies have also found that measures of processing related to formulaicity point towards a smooth change rather than indicating a clear distinction between formulaic and non-formulaic expressions. This is taken as evidence against holistic storage, which is thought to be a binary condition. The studies showing this are based on multiword sequences, where high frequency of occurrence within a corpus is taken as the main criterion for formulaicity. For example, in the psycholinguistic study by Arnon and Cohen Priva (2014) described in Chapter 2, they found that processing speeds for the middle word of trigrams increased smoothly with increasing frequency, with no clear cut-off point for formulaicity. In a neurolinguistic study, Tremblay et al. (2016) found that brain activity associated with online construction of n-grams appeared to flatten off with increasing n-gram frequency, but that this occurred for n-grams at a much lower frequency than would be expected if high n-grams frequency represented formulaicity. However, as discussed in Chapter 2, high frequency ‘in the language’ is not a reliable indicator of formulaicity, and is not considered a necessary condition in the approach taken in the current research. Further, any binary finding (e.g. a clear distinction between formulaic and non-formulaic processing) within an individual item can be masked (appear continuous) when considering averages over grouped data. In any case, there have also been neurolinguistic studies that suggest there are distinct mechanisms underlying the processing of highly predictable versus novel sequences. For example a study by Siyanova-Chanturia et al. (2017) using an EEG approach (as described in Chapter 2) demonstrated that processing for familiar binomial phrases was qualitatively different from that of similar novel phrases (as measured by differences in N400 and P300 ERPs), and this difference is down to the phrase rather than associations between the words. The research overall is therefore inconclusive about whether a clear distinction between formulaic and non-formulaic expressions can be observed using measures associated with processing (such as delivery speed and brain activity).

8.2.3. *The role of constituent parts of formulaic expressions*

Siyanova-Chanturia (2015) presents evidence showing that some decomposition of formulaic sequences takes place during processing. For example, she cites syntactic priming experiments (Konopka & Bock, 2009; Snider & Arnon, 2012) which found that idiomatic expressions primed their syntactic constructions as much as literal expressions, suggesting that these idioms have an internal structure that is accessed during processing. There has also been a study by Sprenger, Levelt, & Kempen (2006) showing that individual constituents of an idiom are accessed separately during planning and that literal word meanings of constituents become active during production. [Note: this study will be described in more detail in Section 8.3.3]. It has also been shown in a number of studies looking at idiom comprehension that components (words and syntax) within idioms do play a role and that even figurative or semantically opaque idioms may have some degree of compositionality. For example, Cacciari & Tabossi (1988) showed that the literal meanings of words within idioms are available and activated even when the context implies the idiomatic meaning. In such cases, they may not play any functional role in final comprehension, but their activity can be measured. It is also recognised that many idioms are 'semantically decomposable' (Nunberg, Sag, & Wasow, 1994), meaning that different parts may be mapped metaphorically to different aspects of the overall meaning (e.g. '*ice*' in '*break the ice'* is linked to a 'cold social atmosphere').

Regarding access to the syntax within idioms, a number of studies have shown evidence of syntactic priming in idiomatic expressions (Peterson, Burgess, Dell, & Eberhard, 2001) even for those that are syntactically non-flexible. It is worth noting, however, that in many experimental studies, the process of experimentation elicits greater analysis than usual, so access to the component word or syntax may be less prevalent in normal processing[31]. However, more generally, it is certainly the case that many formulaic sequences are capable of being modified in a way that requires some access to the syntax and/or individual words. For example, it is necessary to isolate the verb in a multi-word verb phrase in order to successfully inflect it.

The finding that constituent parts of a formulaic expression may be accessed during processing does not necessarily mean that the expression could not be stored holistically. For example, it is instructive to consider the case of compounds (e.g. *blackboard, handstand*). Whether or not these are defined as formulaic expressions, they do share two features which are highly relevant: they are in the form of a larger unit which is comprised of smaller meaningful ones; and they have meanings which are not necessarily derived in a consistent way from component words (e.g. *blackboard* and *blackout*, *handstand* and *grandstand*). Some research (Badecker, 2001; Libben & Titone, 2008) suggests that compounds are decomposed during recognition, and can be accessed via component lexemes. Other research (Mondini, Luzzatti, Saletta, Allamano, & Semenza, 2005; Pollatsek, Hyönä, & Bertram, 2000) has found that a parallel access approach is adopted, whereby compounds are accessed via both individual component words and holistically. In either case, since compounds are word-like and clear candidates for holisticity and have their own entry in the mental lexicon, the finding that their on-line processing may access the component words may be taken as an argument against defining holisticity in terms of whether the individual components are also accessed or not.

Access to component parts of a holistic unit also features in other theories of lexical processing. For example, the cohort model of speech recognition (Gaskell & Marslen-Wilson, 2002; Marslen-Wilson, 1987) suggests that when we hear an individual word, (initial) phonemes within that word activate a cohort of other words that share those phonemes. This parallel activation of multiple forms has been reliably attested through priming experiments and suggests that constituent elements of even a mono-morphemic word are accessed during processing of that word.

8.2.4. ***Automaticity and formulaic expressions***

The meaning of 'faster processing' in the context of holistic storage, raised in Section 8.2.2, is important because it relates to how formulaicity is measured and also to the

[31] Wray (1992) argues that experimental methods force processing into the left hemisphere, which favours breaking things down, while in non-experimental conditions the processing is done differently.

issue of what is meant by processing automaticity. In the empirical studies (S1-S4) undertaken so far, fluency in the delivery of an expression has been taken as a necessary condition for formulaicity. However, based on the experience of applying the fluency criterion (particularly in Chapters 4 and 6) and the issues raised in the chapter so far, it is useful to consider in more detail the extent to which fluency in a unitary expression does indeed indicate a processing advantage, whether due to automisation or holistic storage. In particular, there is a question of whether there is something qualitatively different about the speed and fluency of a formulaic expression that marks it out from expressions that are constructed on-line but nevertheless delivered fast and fluently

So far, little has been said about automaticity, though formulaicity has long been linked with it (Pawley & Syder, 1983; Wray, 2002a) and it features in the definition by Myles and Cordier (2017) that has been adopted here. Indeed, the folk understanding of an automatic skill as something that is done "without conscious or deliberate effort" could be taken as a hallmark of formulaic processing. The nature of automaticity in relation to formulaicity however has in general not been clearly specified theoretically, nor has there been much empirical research exploring automaticity in formulaic sequences. Some have approached formulaicity from a perspective of Andersons ACT* model (see below) and linked the use of formulaic sequences to proceduralisation (Towell et al., 1996) or to spreading activation (Erman, 2007), as discussed in Chapter 2. To understand further how features of automaticity may be usefully linked to formulaic sequences, it is helpful to look at how automaticity has been defined in general.

A wide variety of possible criteria or characteristics for automatic processes have been suggested in the literature. For example, DeKeyser (2001) describes automatic processes as being fast, parallel, effortless, capacity-free, unintentional (hard to suppress/control), the result of consistent practice, unconscious (no attention or monitoring), and associated with low error rate. From a theoretical perspective, two main strands have been proposed: rule-based and memory-based. The rule-based perspective (Anderson, 1992) equates skilled processing with the use of production rules (proceduralisation), and describes automaticity as coming from ever more efficient use of those rules. The memory-based view (Logan, 1992) equates processing with the single step retrieval of exemplars from long-term memory. Automaticity in this case is thought to come through the accumulation of more exemplars so that any one becomes easier to retrieve.

An amalgamation of these approaches is found in the exemplar-based random walk model (Palmeri, 1997; Rickard, 1997) in which multiple exemplars are retrieved sequentially, based on their similarity to the required process, each incrementally adding information and adjusting what is required from the next exemplar. According to DeKeyser (2001), these theories agree that the observable characteristics of automaticity are the result of highly specific repeated practice that leads to faster memory retrieval, and that improvement follows a power law (in particular a levelling off of performance as processes become more automatic). In addition, automatic processes exhibit load independence (i.e. adding other processing does not impact on performance). These approaches also imply that automaticity represents a gradual quantitative change.

An alternative view of automaticity is given by Segalowitz (2003, 2010). While he agrees that automatic processing is effortless, load-independent and unconscious, he maintains that automaticity represents a qualitative change in processing rather than a gradual increase in efficiency. He suggests that although automatic processing is fast, it is difficult to quantify how much speed would be required to claim automaticity. So, just looking at speed of processing alone is not enough to gauge if something is automatic or not (Segalowitz, Trofimovich, Gatbonton, & Sokolovskaya, 2008). Segalowitz makes the point that multiple activities and sub-processes are involved in a complex skill and the speeding up of some of these (e.g. through practice) may alter which sub-processes are ‘rate determiners’ and the order in which they take place (i.e. create structural change within the overall process. Holistic storage or processing (compared to on-line construction) could represent the kind of qualitative structural change required for automaticity.

If this is indeed the case, then observing fluency and other indicators associated with formulaicity such as faster delivery may not be sufficient to indicate a fully automatic process. It may therefore be the case that further indicators of automaticity are useful for establishing the full psycholinguistic formulaicity of a sequence. Segalowitz (2010) highlights some experimental methods for exploring automaticity in the speech of L2 speakers. These have been applied to explore the effectiveness of vocabulary teaching approaches for the development of speed and automaticity in lexical access (Akamatsu, 2008; Pellicer-Sánchez, 2015). However, as far as I am aware, these or other measures of formulaicity have not been directly applied to the learning or use of formulaic expressions.

8.3. Modelling formulaic storage and processing

As well as considering the empirical evidence associated with the idea that formulaic expressions are stored and processed as single lexical units, it is useful to consider specific ways in which such units could be modelled in speech production and comprehension. An influential model of speech production is that of Levelt (1993). This model describes lexical access and the production process, and gives a detailed description of the mental lexicon and how it might be structured. A number of researchers have discussed the processing of formulaic sequences with respect to the model (Cutting & Bock, 1997; Kuiper, van Egmond, Kempen, & Sprenger, 2007; Nattinger & DeCarrico, 1992; Sprenger et al., 2006). This sub-section will therefore summarise Levelt's model and explore how formulaicity can be integrated within it. It will also look at hybrid models of formulaicity (Cutting & Bock, 1997; Kuiper et al., 2007; Nattinger & DeCarrico, 1992; Sprenger et al., 2006) which can account for the influence of component words in formulaic expressions as well as for their unitary aspects.

8.3.1. *Levelt's blueprint for speech production*

There are different refinements of Levelt's blueprint (Levelt, 1993), but essentially it consists of three main processes: conceptualising, formulating, and articulating speech. The Conceptualiser draws on encyclopaedic knowledge, situational knowledge and discourse knowledge in order to plan and produce a pre-verbal message. The Formulator is the grammatical component which produces the appropriate message in the language required. First there is grammatical encoding which involves a search for appropriate basic word (or sequence) forms stored in the lexicon. Then phonological encoding takes place, whereby information about the morpho-phonological form is accessed from the lexicon and a phonetic plan (internal speech) is developed. Finally, the spoken output is actually produced via the Articulator.

Levelt gives a detailed description of how lexical entries in the mental lexicon are structured and accessed during speech production (1993, pp.181-234). A key idea is that all grammatical and phonological coding is derived from features of the lexical items activated to convey the message, rather than from the message itself. Levelt suggests that each item in the lexicon has at least four key features: meaning, syntax, morphology and phonology. He takes a lexical entry to mean the set of all inflections of the item, and defines a 'lemma' as being the part of the lexical entry consisting of meaning and syntactic information since these are relevant for grammatical encoding. The semantic information in a lemma specifies what conceptual conditions in a message have to be fulfilled for the lemma to be activated. The syntactic features include its category (e.g. verb), the grammatical functions it requires and a specification of which conceptual category is to be mapped onto each grammatical function. In addition, there is a 'pointer' to the 'address' of the form information (these terms are used metaphorically) and a list of relevant variable settings (related to tense, aspect, person, pitch accent, etc.) which ensure the correct form is selected.

In his approach, Levelt (1993, p. 187) suggests that idiomatic collocations are entries in the mental lexicon, but he does not give any specific details about how these or other formulaic expressions are integrated into the model. The implication is that a formulaic expression is considered a lexical item in its own right, and the lexical entry would contain any inflected versions (e.g. *come to a head*, *came to a head*). Similarly to word lemmas, a 'sequential lemma' would be the part of the entry consisting of meaning and syntax.

According to Levelt, the conceptual specification within the lemma defines elements that encapsulate the meaning of the item and the type of arguments required for its grammatical deployment. This could, for formulaic expressions, presumably include defining the conceptual categories of variables that fit into the open slots in frames (e.g. 'insert a noun-phrase that is animate'). There would also be a limited set of inflectional variations within the lexical entry, the choice of which to use being defined by an appropriate variable. Levelt suggests that single word items have associative connections with other words on the basis of shared meaning, morphology or phonology. For formulaic expressions, such associative connections may exist for the component words as well as for the whole expression. For example, the expression *came to a head* may have connections with *leader* as well as to *problems* or *situation*. There would be other differences between words and sequences as well. For example, the pitch-accent variable mentioned by Levelt (1993, p. 191) as part of the lemma might need to be extended to specify the areas in the sequence that would take accent should the sequence be the point of focus. As discussed in Section 2.4.2, formulaic expressions often have a distinctive phonology different from a simple collection of the word phonology, so this would need to be represented within the lexeme too.[32]

Within the Levelt model, the existence of such a 'sequential lemma' would provide a processing advantage over a similar sequence that was composed on-line. In the case of a compositional novel sequence (where an expression is fully constructed on a word-by-word basis), its processing in the Formulator can be described by a series of steps (Levelt, 1993). It starts with the conceptual meaning of a key component word, the initial fragment of the message to be delivered. A search then finds the equivalent lemma to match its conceptual specifications. This lemma will then call an appropriate procedure to construct sub-phrases and the process continues going back and forth between the lemma for each word selected and the conceptual needs of the message. The lemma of each component word is accessed and processed in the formulator. On the other hand, if there was a single lemma for the expression, only one iteration of access and processing would be needed, resulting in faster processing.

8.3.2. *Models of idiom processing*

As highlighted in Section 8.2, the research evidence seems to suggest that formulaic expressions have both a unitary and a componential nature. From the perspective of idiom comprehension, Cacciari and Tabossi (1988) explain this in terms of their Configuration Hypothesis. This proposes that an idiomatic phrase activates all the lexical items (individual lemmas and associated syntax) that would be activated in the literal compositional version of the idiom. The idiom is processed word-by-word until enough information is accumulated to recognise the sequence as a memorised idiom. The point at which this occurs is known as the idiom key, and access to the idiomatic meaning is only triggered when that point is reached. The idiom key can be in different positions in different idioms, determining how early the string is recognised as an idiom. The configuration hypothesis emphasises the compositional nature of idioms and can account for features such as syntactic priming and the targeted nature of internal modification. Libben and Titone (2008) expand this hybrid view of idiom comprehension by suggesting that activation of the figurative meaning is subject to the satisfaction of certain 'constraints' during the ongoing compositional analysis of the idiom. Such constraints may relate to the familiarity of the idiom, its

[32] Indeed, an entire formulaic expression can fall into the intonation slot of the single word with which it is in parallel, e.g. *I saw an excellent film* and *I saw a never-to-be-forgotten film*.

degree of compositionality, its literal plausibility, and component word frequencies (Libben & Titone, 2008, p. 1117).

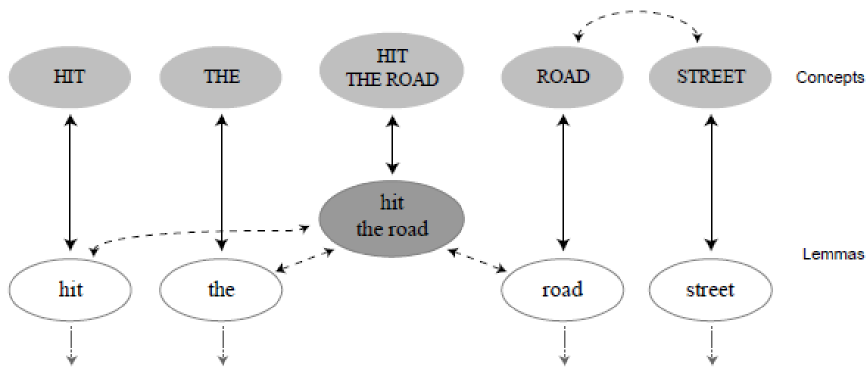
In the case of idiom production, a model was proposed by Cutting and Bock (1997). They based this on a series of experiments they undertook in which speech errors involving the blending of pairs of idioms were induced. Participants read pairs of idioms, and then were cued to say one out loud. Responses were measured in terms of production latency and the degree of blending. Cutting and Bock compared pairs, which varied in terms of whether their syntax or figurative meaning matched or not, and pairs that varied in whether they were decomposable or not. The results suggested that idioms are syntactically analysed during production and literal word meanings are also activated. Further, they suggested that representations of decomposable and non-composable idioms are the same as they enter the production process. Based on these conclusions, Cutting and Bock (1997) proposed a hybrid account of idiom production linked to existing models of speech production incorporating the multi-level representations of the lexicon (Levelt, 1993) and spreading activation between nodes at the different levels (Dell, 1986). In this, idioms have a unitary entry at the conceptual level which will activate individual lemmas of constituent words at the lexical-syntactic (lemma) level. The unitary conceptual entry is also linked to special syntactic information in the form of a prefabricated phrasal frame. Thus, in this model, while idioms are stored as a whole at the conceptual level, they are not word-like as such because they have analysable internal structure represented by the constituent lemmas and the phrasal frame, and do not have separate entries as lemmas themselves.

To some extent, the Configuration Hypothesis for idiom comprehension and the hybrid model for production show a consistent theoretical way of viewing the processing of idioms. In particular, they both involve processing at the lexico-syntactic and conceptual levels, with production broadly representing a top down approach (from concepts to lemmas) and comprehension the other way round. However, a potential disadvantage of both models is that the unitary nature of the idiom is only represented at the conceptual level. The only difference between the idiom and a similar literal form is either a special configuration key (in the configuration hypothesis) or a link to a prefabricated phrase form (in the hybrid model). In particular, the idiom is not represented as a lemma at the lexico-syntactic level in either model. In addition, the models introduce different features of the lexicon for comprehension and production. This runs counter to the view of some researchers (e.g. Kempen & Harbusch, 2002) that the same network of concepts and linguistic representations is used for both comprehension and production.[33] An adapted version of Cutting and Bock's hybrid model has therefore been proposed by Sprenger et al. (2006).

8.3.3. *The 'superlemma' model: Sprenger et al. (2006)*

Sprenger et al. argue for a unitary lemma representing the whole idiom (Sprenger et al., 2006). This 'superlemma' as they call it, "is a representation of the syntactical properties of the idiom that is connected to its building blocks, the simple lemmas" (p.176). This solution retains the hybrid nature of the model and also complements the blueprint of Levelt (1993). In this way, the selection and processing of an idiom is similar to the processing of a single word in terms of lexical competition and co-activation. At the same time, it retains the idea that idioms have a syntactic structure related to the individual constituents at the lexico-syntactic level. Thus the 'superlemma' replaces the syntactic prefabricated frame posited in the original hybrid model and represents the specific syntactic constraints of the idiom. Sprenger et al. illustrate the model for the idiom '*hit the road'* (see Figure 8.1).

Figure 8.1: Illustration of the 'superlemma' model (Sprenger et al, 2006, p.176)



meaning relationship

'element of' relationship / co-activation

All connections between processing levels denote the same relationship

The idiom is represented at both the conceptual level and also the lemma level. The superlemma is connected to the simple lemmas of the constituent words by 'element

[33] There are other views. For example, some neurological studies suggest that comprehension and production processes do not necessarily mirror each other. These have shown double dissociation, whereby someone loses the capacity to understand but not produce, and vice versa, which might suggest independent routes.

of/contains' relationships, while lemmas are connected to their conceptual meanings by 'meaning' relationships. This addresses a challenge in the original hybrid model which mixes these relationship types across levels.

Sprenger et al. (2006) derived their model following a series of experiments they ran to explore the original hybrid model. In contrast to Cutting and Bock's studies which were based on speech errors, Sprenger et al. used an error-free approach that involved priming and response times for producing idioms. They were interested in exploring the role of individual words in the production of idiomatic and non-idiomatic fixed expressions.

8.3.3.1. Sprenger et al. (2006): Experiment 1

Their first experiment was based on the use of an 'identity prime' – that is, priming an expression (e.g. *hit the road*) by using a word which is the same as one which appears in the expression (e.g. *road*). This was compared with the effect of using a phonologically and semantically unrelated control prime. In the experiment, they presented 16 pairs of expressions (in Dutch), all finite verb phrases with the same basic syntactic structure, to 16 university undergraduates (with L1 Dutch). In each pair, was an idiomatic expression (e.g. *viel buiten de boot* = fall outside the boat, a Dutch idiomatic phrase meaning 'be excluded') and a linked literal expression (e.g. *ging met de boot* = go with the boat, literally 'take the boat'). The shared noun (e.g. *boot* = boat) served as the identity prime in each case. The participants memorised one expression from each pair (eight idiomatic and eight literal) along with a prompt word for each. The prompt for any item pair was a Dutch name e.g. *Jan*) which formed a complete sentence when paired with the expression.

The memorising process was divided into two experimental blocks (i.e. 8 expressions at a time). In each block, once the expressions were fully memorised, the test began. The participant was presented with the prompt word (written) along with a spoken prime (either the identity prime or the control) and asked to repeat the appropriate memorised phrase as quickly and accurately as possible. For each block, there were 128 trials covering 8 phrases (each with 8 identity and 8 control primes) and production latency (speech onset) times were recorded.

Sprenger et al. found no difference in speed and accuracy of response between the idiomatic and non-idiomatic phrases, but there was an interaction between the idiomaticity and prime type. In particular, the effect of prime type was significantly stronger for the idiomatic phrases. The authors concluded that this supported the hybrid model in two ways. Firstly, the identity prime did indeed prime idiomatic expressions along with the literal phrases, suggesting that the individual word was activated in both cases. Secondly, they reasoned that the identity prime activated the whole phrase, thereby in turn activating all the other simple lemmas associated with it. On the other hand, for the non-idiomatic phrases, the identity prime helped speed up production of the corresponding word in the expression, but not the whole phrase or the other constituents. Since the literal expression would have no entry in the lexicon, there would be nothing to bind the identity noun with the other constituent words. Hence the priming effect is weaker.

It may be argued that, since participants had to memorise phrases, there was a lot of scope for variation between the participants and across phrases in how well they were memorised. However, the researchers did run checks to ensure all the phrases were learnt so that they could be recalled fluently in advance, and the mixed orthogonal design of the experiment and the counter-balancing order of the blocks across participants ensured that the effects of individual differences and order were well controlled for. A further point could be that the idioms, being well-known expressions were more easily recalled than the literal expressions, representing a potential bias. However, the results indicated that reaction times were actually slightly faster for the literal expressions (but not significantly so).

8.3.3.2. Sprenger et al. (2006): Experiment 2

Their second experiment explored the extent to which literal word meaning becomes active during idiom production. It was an idiom completion task on 16 well-known syntactically similar idioms. The participants were initially given a written cloze test in which they had to supply the missing noun, indicating that the expressions were indeed known. The experiment then presented the expressions in written form with the noun again missing. Participants were instructed to say the missing noun (i.e. complete the expression) as quickly as possible. They also received an acoustic prime which was either semantically related to the missing noun, phonologically related, unrelated, or else they received no prime. In addition, the prime was given with stimulus onset asynchrony (SOA) of -150ms, 0, 100ms, or 200ms. There were 71 participants each undertaking 512 trials (4 SOAs x 16 items x 4 prime types). They found that response types were speeded up for the semantic and phonologically related primes and slowed down for the unrelated prime compared to the no prime case. The phonological priming was strongest for SOA=0 and the semantic for SOA=-150ms. They concluded that idioms are primed via constituent words both phonologically, supporting previous research suggesting the phonological priming for idioms, and semantically, supporting the idea that individual word meanings are accessed, as suggested by the hybrid model.

8.3.3.3. Sprenger et al. (2006): Experiment 3

A third experiment reversed the priming effect in order to show that not only are idioms primed via constituent words, but also production of the idiom itself primes associated words (i.e. the effect is bi-directional). In this experiment, participants were presented with idiomatic expressions with the final word missing. They were then instructed to say the missing word on presentation of a question mark. They were also informed that some trials would be different (in fact half the trials were) in that a word would be presented instead of the question mark and they had to say this word as quickly as possible instead. The latter cases were the key ones. The surprise word was either phonologically related, semantically related or unrelated to the missing word. Results confirmed significant effects on response times for phonologically and semantically related words compared with those that were unrelated to the missing idiom word. This effect was predicted by the hybrid account which proposes that lexical concept nodes of constituent lemmas are activated as parts of the idiom representation.

Overall, Sprenger et al. concluded that the results support their amended version of the hybrid model of idiom production. They argue that this ‘superlemma’ model offers an elegant and intuitive account in that it does not posit separate types of storage for the comprehension and production in the processing of idioms. Further, since it does not propose special processing for the idiom, it is flexible enough to accommodate fixed formulaic expressions more generally. Further research on this model has been undertaken by Kuiper et al. (2007) who looked at naturally occurring speech errors in multiword formulaic expressions in English (from the Tuggy corpus of over 1000 naturally observed slips). They found that activation patterns for idiom blending predicted by the ‘superlemma’ theory matched what they found in the data. The results were corroborated by reference to a smaller corpus of formulaic expression slips in Dutch (the Kempen corpus).

While, as the researchers acknowledge, further development and corroboration of the model is required, it can provide a useful theoretical basis for exploring the processing and storage of formulaic expressions.

8.4. Modelling the acquisition of formulaic expressions

Levelt's blueprint and the related models of Cutting and Bock (1997) and Sprenger et al. (2006) discussed above do not describe the acquisition of new lexical entries directly. However, some models of L2 vocabulary acquisition which draw on the basic framework of Levelt have been proposed. Two of these (De Bot, Paribakht, & Wesche, 1997; Jiang, 2000) are described below and they have some implications for the acquisition of formulaic expressions.

8.4.1. *De Bot, Paribakht, & Wesche (1997)*

De Bot, Paribakht, & Wesche (1997) use a version of Levelt's model of speech processing (1993) to provide a structure for describing and explaining aspects of L2 word acquisition. De Bot et al. highlight the idea that the lemma has distinct elements of knowledge including syntactic and semantic components which are, in turn, separate from the morphological and phonological components of the lexemes to which the lemma is linked. They suggest that when a learner encounters a new word, an 'empty' lemma structure is created. The learner then uses semantic and syntactic information from context (and morphological information from the lexeme depending on their experience of the language) to fill in this structure.

De Bot et al. present empirical findings from an introspective study of lexical inferencing to provide support for their use of the model. In their study, ten participants at an intermediate level of English were each given two different kinds of reading task. In the Summary Task, they read a text and summarised each paragraph as they completed it. In the Question Task, they answered reading comprehension questions about the whole text. The data for the study was based on an 'immediate retrospective' think aloud protocol, whereby participants gave reports on how they dealt with unknown words immediately after they completed a task. Participants reported using a variety of different inferencing strategies to deal with the unknown words, including the use of sentence level grammatical knowledge, word morphology, world knowledge and homonymy. The results were interpreted as being compatible with the model in that the categories could be mapped onto different components of the processing model. However, the study did not specifically address the stages for separating the semantic, syntactic and phonology components of a word during processing or how the 'empty lemma' is filled.

8.4.2. ***The lemma mediation model***

Jiang (2000) extended the idea of the 'empty lemma' to create his lemma mediation model of L2 vocabulary acquisition. This recognises that the learning of new L2 vocabulary usually proceeds with only limited exposure to the full contextual information about a word in the L2. At the same time, the L2 speaker already has a full set of lexical and semantic information related to the word in the L1. Based on evidence from studies of the bilingual lexicon and of L2 production, Jiang suggests three stages of lexical development. In the first (formal) stage, the written or phonological form of the word is learnt and a lexical entry created. However, this entry does not contain semantic, syntactic or morphological information. The meaning (and grammatical information) is provided via associated links to the L1 translation or definition. The second (lemma mediation) stage occurs as experience with the word increases and stronger bonds with the lemma of the L1 translation develop through simultaneous activation of L2 form and L1 lexical information. At this stage, the L2 lemma is filled by L1 lemma information, and links to the conceptual representation are mediated via the L1. The final (integration) stage occurs if and when sufficient exposure and use enable L2 semantic, syntactic and morphological information to be integrated into the L2 lemma structure and a strong direct link with the conceptual meaning is established.

Jiang (2000, p. 54) suggests that L2 lexical development often fossilises at the second stage and this has consequences for processing. While the mediation stage allows for greater automaticity of processing than the formal stage, it still relies on L1 semantic and syntactic information. This can lead to lexical errors (since semantic and syntactic information between L2 words and L1 translations rarely overlaps fully) and grammatical errors (due to a lack of morphological specification within the L2 lemma). The lemma mediation stage also predicts relatively weak connections between the L2 words and their conceptual representations, since these are mediated via the L1 semantic information. Jiang (2000, pp. 60-61) cites several masked cross-language priming experiments showing that L1 translated words tend to prime L2 targets, but not the other way round. He argues that, since cross-language priming is mediated via the conceptual level, this asymmetry demonstrates weak connections between L2 words and concepts. Whether such weak connections are only a feature of early stage learners, however, is not something he directly addresses.

8.4.3. *Application to formulaic expressions*

While there does not appear to be a conceptual framework describing the acquisition of formulaic expressions explicitly, the various models described above provide some useful pointers regarding what such a framework might require. For example, to accommodate the idea that any theory of lemma access must involve both parallel processing and convergence to a single choice of lemma, Levelt (1993, p. 212) introduces the notion of "core" meanings which are unique and specific for each lemma. In the case of a compositional expression (one that is constructed on-line during production), each constituent word is a lemma retrieved during the process, with a different set of core concepts that define it along with its own conceptual categories, functions and sets of variables. On the other hand, if the phrase is formulaic, it may instead have a single 'superlemma' as suggested by Sprenger et al. (2006, p. 176) and discussed in Section 8.3.3. This would have associated links with its component words. The lemma must have its own conceptual meanings, categories and specifiers, and a unique set of core concepts (which differentiate it from any other lemmas, including those of its component words). Acquisition of the phrase as formulaic (whether as a whole new expression or through a process of fusion) would require a new sequence lemma to be created; one which matched the conceptual meaning of the whole expression.

Jiang's lemma mediation model (Jiang, 2000) suggests one way that this may occur. However, the multiword nature of the formulaic expression means each stage is more complex, potentially involving L1 translations of both the whole phrase and of component words (or sub-sequences).This may be illustrated in a study by Yamashita and Jiang (2010) which applied the lemma mediation model to the acquisition of collocations by Japanese EFL and ESL speakers. Yamashita and Jiang used a phrase-acceptability judgement task to check knowledge and processing of congruent and incongruent collocations. Congruent collocations were those where the English and Japanese translations matched word-for-word (e.g., *make lunch*). Incongruent collocations (e.g. *kill time* and *slow learner*) could only be translated using non-matching words (e.g. *crush/break time* for *kill time*) or reformulations (e.g. *someone who learns slowly* for *slow learner*). The results suggested that L1 congruency does affect acquisition of collocations and that incongruent collocations are generally difficult to acquire. However, they also suggested that, once established in memory, L2 collocations are processed independently of the L1. In the context of the model, the assumption was that collocations are holistic units with their own entry in the mental lexicon and follow the model of word acquisition. An implication of this, hypothesised by Yamashita and Jiang (2010, p. 663) is that congruent collocations would initially require less frequent exposure than incongruent collocations because they could rely on individual word associations to provide the lemma information. However, congruent collocations may require more exposure subsequently to establish L2 lemma information and connections to the conceptual meaning, due to the influence of the L1. Conklin and Carrol (2019) also highlight the relevance of L1 congruence in models of the processing of formulaic expressions. They cite a number of studies showing that full or partial L1 congruence influences the processing of idioms in L2 speakers at the earlier stages of learning, and suggest this is due to “fast, automatic activation of L1 translation equivalents while processing in the L2” (p.73). They suggest that, as with the Jiang model, at early stages of learning, links from L2 formulaic representations (e.g. super-lemmas) to conceptual representation are mediated by links to the L1 formulaic representations.

8.5. Conclusion

8.5.1. *Implications*

The research discussed above has some important implications for the idea of holistic storage and the processing of formulaic expressions. Overall, the evidence does not preclude holistic storage but it does highlight the challenge in being able to collect evidence for or against its existence. For example, the case has been made that faster processing seen in formulaic expressions does not necessarily mean that they must be stored holistically (Siyanova-Chanturia, 2015) or processed automatically (Segalowitz, 2010). Additional ways to determine when an expression is fully formulaic may therefore be required. There is also evidence to suggest that, even if expressions are stored holistically, component words and structures may also be accessed during processing (Arnon & Cohen Priva, 2014; Konopka & Bock, 2009; Siyanova-Chanturia, 2015; Sprenger et al., 2006). This implies that any model of holistic processing and storage should be able to cater for this.

As outlined above, one model that does do that is the amended hybrid model of Sprenger et al. (2006) which posits a ‘superlemma’ for the formulaic expression with associated links to the component word lemmas. This and related models such as those Levelt (1993) and De Bot et al. (1997) also highlight a meaning link between the lemma and a unique core concept. While none of these models specifically covers the process of acquisition for formulaic expressions, they do imply certain aspects that may be part of that process. In particular, formulaic acquisition must include developing a core concept for the expression and creating the expression lemma (superlemma). The core concept may already exist (e.g. if there is a unitary expression for it in the L1) or it may develop as a unitary concept as the expression becomes formulaic. The expression lemma may be created via an 'empty lemma' gaining components and structure (De Bot et al., 1997) or through the components of a constructed sequence becoming fused together in some way. In either case, links must remain to the component words, for example by 'idiomatising' the elements (Levelt) or through same-level links (Sprenger et al., 2006). According to the lemma mediation model (Jiang, 2000), the situation may be further complicated by the potential influence of the L1 translation of the phrase or of components words, particularly in the early stages of acquisition. This will also be affected by the level of congruence of the new L2 formulaic expression with its L1 translation.

8.5.2. ***Questions for further exploration***

The research and empirical studies covered so far have suggested that there may be different ways for a novel unitary expression to become formulaic (in the context of an L2 speaker being given such an expression to memorise and reproduce). In particular, the expression may be memorised and stored whole at the time of learning, or it might acquire formulaicity over time. Based on the review in this Chapter, two key questions are:

1. How can a model of formulaicity that includes holistic storage accommodate the gradual acquisition of formulaicity over time?
2. Is (consistent) fluency of delivery sufficient to indicate formulaicity as modelled here? What other measures might be used?

In addressing these questions, it useful to note that gradual or staged acquisition of formulaicity may be compatible with holistic storage in a number of ways. One option is that expressions learnt by an individual may instantaneously flip into formulaicity after a certain amount of practice and repetition. (If this point is different for different expressions, the proportion of formulaic expressions in a sample could increase gradually over time giving the appearance of gradually increasing formulaicity). Alternatively, an individual expression could go through a period where it has a holistic representation, but one which is not yet easily accessed (e.g. it is weakly connected to other networks in the mind and not linked to many relevant cues). If a testing cue does not trigger the holistic form, the speaker may rely on reconstruction or not recall the expression at all. A third possibility is that the expression may become increasingly fused together with practice, as indicated by a corresponding development of fluency, accuracy and consistency. However, these indicators, while necessary, may be insufficient to guarantee that one has identified holistic storage or automated processing. If so, a further measure may be required to observe the switch to full formulaicity.

To investigate these options and the acquisition process in more detail, two related empirical studies were carried out. These are described in Chapters 9 and 10.

CHAPTER 9: Comparing routes to formulaicity (Study S5)

Dramatic repetition and semantic-formal elaboration in the L2 memorisation of target sequences

9.1. Introduction

The research and studies up to this point suggest that there may be different types of route to acquiring formulaic expressions. The two main ones can broadly be defined as ‘holistic acquisition’, whereby a common sequence is learnt and processed as a single holistic unit immediately, and ‘fusion’ whereby an expression, initially constructed in some way, becomes formulaic through subsequent practice and usage, joining the components into a single whole and fine-tuning accuracy of form.

This chapter describes an empirical study (S5) designed to explore these different possible processes for internalising the sequences. The study builds on the previous study (S4), using a similar set of target sequences applied to a new (larger) group of Japanese participants. The method is adapted to compare two different strategies for memorising the targets, intended to promote either holistic acquisition or fusion. The overall premise of the study is that applying resources to establishing a holistic form of the sequence directly during initial memorisation may be more effective in establishing formulaicity long-term than starting with knowledge-based understanding, since the latter may encourage reconstructive practice and production.

Choosing how to control the input method is an important consideration, since it is likely to have a significant effect on the learning outcome. For example, as discussed in Chapter 2, the principle of Transfer Appropriate Processing (Roediger et al., 2002) proposes that any processing strategy is linked to a particular outcome. In particular, the initial learning processes determine the qualitative nature of the trait encoded (Craik, 2002). For the intentional learning of formulaic expressions, different forms of semantic or formal elaboration have been suggested. These include: drawing attention to L1 congruence (Conklin & Carrol, 2019); analysing the structure and component words through matching or cloze style activities (Boers, Demecheleer, et al., 2014); linking metaphorical meanings of non-compositional idioms (Boers et al., 2007); and utilising imageability (Steinel et al., 2007). These may lead to learning benefits in terms of long-term recall and accuracy, but their effect on fluency is not clear.

Insofar as internal formulaicity is defined in terms of holisticity and identified by delivery features such as fluency, approaches to memorisation that are geared towards this outcome may be more effective in promoting ‘holistic acquisition’. As discussed in Chapters 2 and 7, oral repetition as a means for achieving fluency when memorising target sequences is enhanced if it is accompanied by a clear link to meaning (Au & Entwhistle, 1999) and close attention to the imitation process (Ding, 2007). A particular example of the latter is the strategy, used by actors, of ‘actively experiencing’ a target utterance as it is repeated during practice (Noice & Noice, 2006). As mentioned in Section 2.5.3.2, this strategy has been taught to non-actors and shown to be highly effective for accurate, fluent recall and reproduction of learnt target sequences. It seems, therefore, that an approach to repetition which highlight context and the mimicking of appropriate delivery may be usefully applied to the memorisation of target formulaic expressions. This kind of approach should help promote accurate acquisition of the complete phonological form and a strong automatic link to overall meaning.

Based on these considerations, two different input strategies were developed for the study. The first, Dramatic Repetition (DR), focuses on accurate and fluent reproduction of the sequences, while the second, Semantic-Formal Elaboration (SFE), is a deeper, more analytic approach focussing on meaning and form. The effect of these initial processing strategies on formulaicity is assessed over time in terms of the fluency and accuracy with which the expressions are recalled. The aim is to create conditions which facilitate holistic acquisition in some sequences and fusion in the others, so that these two routes to formulaicity can be compared and explored. In particular, it is expected that DR will promote holistic acquisition to a greater extent than SFE.

9.2. Method

9.2.1. ***Participants***

Ten Japanese speakers of English (JSE) at an intermediate/advanced level of English were recruited. All were working adults living in Tokyo, chosen on the basis of availability, level and because they were interested to take part. They completed a Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) and answered a background questionnaire about their English usage and study experience. A list of the participants along with their most recent TOEIC scores (ETS, 2019) is given in Table 9.1.

Table 9.1: List of participants (using pseudonyms)

Grp	Name	Sex	Age	TOEIC	VLT
P1	Ayane	F	30+	700	97
P4	Mari	F	30+	845	123
P2	Naomi	F	30+	850	109
P3	Sachiko	F	40+	890	123
P4	Keiko	F	30+	900	127
P1	Akemi	F	40+	920	127
P3	Tetsuko	F	40+	925	122
P2	Kaori	F	20+	930	112
P2	Kentaro	M	30+	930	126
P1	Shizuko	F	20+	940	128

Grp = experimental group (see Section 9.2.2)

VLT = sum of scores on 2K, 3K, 5k, Academic, & 10k levels of the Vocabulary Levels Test

Five of the participants (Sachiko, Tetsuko, Kaori, Kentaro and Shizuko) had experience of living or studying abroad in an English-speaking country. Two (Akemi and Shizuko) currently used spoken English regularly in their work, and three (Mari, Kaori and Shizuko) used it occasionally. All had studied English to some extent after school, but only two (Ayane and Naomi) were currently doing so.

9.2.2. ***Design***

The target sequences were selected according to the process outlined in S4 (Section 7.4.1) and are listed in Table 9.2. All were verb phrases of four or five words selected from the Phrases in English (PIE) on-line corpus (Fletcher, 2011). Each had high frequency lexical words (with no repetition across the sequences), was non-congruent with the L1 Japanese and was confirmed to be unknown to the participants via a pre-study check. The sequences were embedded in four stories[34] (each of about 150 words) and the stories were paired to form two sets (AB and CD) containing six sequences each. Sequences were balanced across the sets for length (words and syllables) and ‘transitivity’. Each story was assigned a pair of suitable pictures as a visual cue. The stories are given in Appendix 9.1 and a list of the targets along with their given Japanese translations (see Section 9.2.3) is given in Appendix 9.2.

[34] Three of the stories (A, B and C) were adapted from the ones used in study S4

Table 9.2: List of target sequences

Set 1 (AB)	A1	turned a blind eye to	
	A2	came to a head	
	A3	breathed a sigh of relief	
	B1	run the risk of	
	B2	go a long way towards	
	B3	like the sound of	
Set 2 (CD)	C1	set his sights on	
	C2	stood the test of time	
	C3	get the hang of	
	D1	knew better than to	
	D2	toyed with the idea of	
	D3	remains to be seen	

To provide a realistic comparison between conditions, the time given for the memorisation of targets was precisely measured to be the same for both conditions (18 minutes for 6 sequences), and the subsequent assessments of sequence recall and performance were identical.

Table 9.3: Ordering of sequences and conditions by participant group

	1st	**2nd**
P1	AB (DR)	CD (SFE)
P2	AB (SFE)	CD (DR)
P3	CD (DR)	AB (SFE)
P4	CD (SFE)	AB (DR)

In order to mitigate against the possible confounding effect of differences between participants or sequence memorability, a cross-over design was used whereby participants, sequences and order of learning were balanced across the two conditions. To facilitate this, participants were randomly assigned to one of four groups, as shown in Table 9.3. For example, participants in group P1 learnt the sequences in A and B via the DR approach, then the sequences in C and D via the SFE approach.

9.2.3. ***Procedure***

In the first session, participants learnt the 12 sequences (six by DR and six by SFE) in the order stipulated for their participant group. Each participant listened to a story (A or C) without any script, but while looking at the pictures (to provide a cue for later). The three sequences in that story were introduced for learning using either the DR or the SFE input approach (described below). The process was repeated for the second story (B or D), using the same input approach. To consolidate the learning, the standard assessment process (see Section 9.2.4 below) was then applied to all six target sequences just learnt. Next, the procedure was repeated for the other two stories, this time with the sequences learnt using the other input approach. After all sets had been learnt and assessed, the participants listened to each story once more. Following a ten minute break, there was a further assessment to establish performance at the end of the learning session (W0).

After one week and three weeks, participants were given further assessments (W1 and W3 respectively). At the end of the third week session, there was also a Usage Test (see Section 9.2.5 below). Participants were instructed not to review or practice the sequences between assessments.

The input sessions varied according to the condition as follows:

9.2.3.1. DR input

The dramatic repetition (DR) input approach focussed on consistent repetition of the expression with an emphasis on accurate imitation of prosody, intonation and rhythm, and 'active experiencing' of the sequences. The basic meaning of the expressions was provided by the story and the Japanese translations, but was not further elaborated on. For each sequence, participants listened to the full sentence containing it and read a translation of the sequence to check meaning. They then did a series of repetitions of the sequence following the exact intonation and rhythm of the model provided. (Where necessary this was slowed down to ensure accuracy). They interspersed this with repeating the whole sentence and also practised responding quickly to the Japanese translation of the sequence (as a cue card). Participants were encouraged to mimic the exact prosody and intonation of the delivery whenever they repeated each expression and "to imagine they were performing in a radio play". All engaged willingly with the process and appeared to enjoy doing it.

9.2.3.2. SFE input

After listening to the story, participants were given a gap fill exercise based on the story script in order to try to generate the sequences. After finishing, they corrected this using the answer script. They then did exercises looking at the structure of each sequence (count the verbs and nouns), and compared the sequence with its Japanese translation by rating its ‘closeness’ (in terms of words used). They were also asked to consider what might help them remember each sequence (e.g. particular words or images) and wrote example sentences for each which were then corrected if necessary by the researcher.

9.2.4. ***Assessments and measures***

Exactly the same assessment process was applied at all stages: immediately after the initial input, at the end of the learning session (W0), after one week (W1) and after three weeks (W3). Each assessment covered all 12 target sequences and consisted of the following:

1. Context recall: Given the pictures and title, participants retell the story trying to use the target sequences.

2. Cued recall: Cue cards (featuring the L1 translation of each sequence) are presented in random order and participants recall the appropriate sequence out loud. If they cannot do so, the researcher says the first word as a further cue.

3. Written recall: Participant writes down the expressions given the L1 translation.

4. Read out loud (ROL): Targets are presented on a computer screen in random order and the participant reads them out loud. This was included as a controlled way for participants to review and repeat the targets at each assessment stage, but also to explore whether changes in voice onset time and articulation rates might be observed across conditions and stages.

The assessments were recorded, transcribed and analysed to calculate a variety of measures for each participant-sequence.

9.2.4.1. Recall

Recall is a way to measure whether a participant can recall the target sequence (even if not fully accurately) given a context, meaning or first word cue. In order to ensure that a recall attempt was actually targeting the sequence in question, recall was deemed to have occurred if a sequence was recalled with over 50% accuracy on either of the recall tasks (context or cued). This is equivalent to the condition that at least three words were correctly remembered in the correct order (allowing for errors in grammatical inflections e.g. *turn* for *turned*; *eyes* for *eye*).

9.2.4.2. Accuracy

Accuracy is a measure of how close the recalled item is to the original target. The sequence was considered ‘fully accurate’ if it exactly matched the target on either of the two recall tasks (context and cued). In addition, a graded value for accuracy was calculated for each response using the method for measuring accuracy introduced in study S4 (see Section 7.7.1). The accuracy score (A-score) for an item was the maximum accuracy value over context and cued responses.

Equivalent measures for written responses were calculated on a similar basis, with no penalties for incorrect spelling.

9.2.4.3. Fluency

For each recall attempt, the output was analysed for dysfluencies. These were defined as any of the following:

- Unfilled pauses > 0.2 seconds
- Filled pauses (e.g. *er, umm, ah*)
- Syllable lengthening > 0.4 seconds
- Repetition or repair/retracing

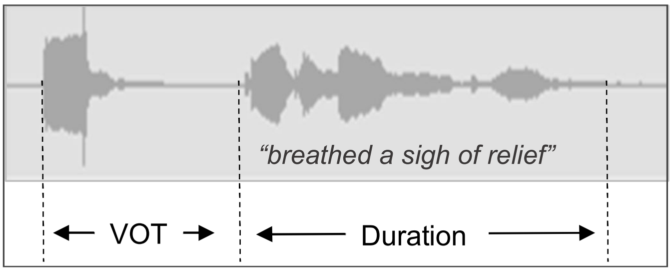
A recalled sequence was considered ‘consistently fluent’ if it was delivered with no dysfluencies and with consistent form across the oral recall attempts. In addition, a value for the fluency of each recalled sequence was calculated using the method for measuring fluency of responses introduced in study S4 (see Section 7.7.3). The fluency score (F-score) for an item was taken as the fluency value of the most accurate of the context or cued responses.

Since consistent fluency is a potential indicator of internal formulaicity in this research, it was hypothesised that, at W0, there would be more ‘consistently fluent’ sequences and a higher mean F-score resulting from the DR approach compared with SFE.

9.2.4.4. *ROL voice onset time (VOT) and articulation rate (AR)*

In the read out loud (ROL), each sequence appeared on screen simultaneously with a 'beep' sound. The participant repeated the sequences out loud and the process was recorded. From the recorded audio, voice onset time (VOT) and speech duration were obtained from the recorded audio file as indicated in Figure 9.1.

Figure 9.1: Audio analysis of read-out-loud responses



Articulation rates (AR) were calculated by dividing the duration (in seconds) by the number of syllables in the response. The aim was to compare average VOT and AR between the conditions at each stage. It was hypothesised that AR would tend to be faster for the DR condition at W0 because more targets had been learnt holistically at the beginning than for SFE.

9.2.5. ***Usage assessment***

At the end of the study (in week 3), a usage test was given. This gave participants the opportunity to recall the targets sequences in a different context (and using a different form of the verb from the original presentation). In the test, participants heard a brief description of a situation (audio cue and were then given a few written words that would cue them to restate the audio cue using one of the target sequences. Their task was to complete the sentence verbally using an appropriate target sequence. An example item is given below:

Table 9.4: Example Usage Test item

AUDIO CUE:	The police in Japan seem to ignore people cycling on the pavement. They never stop them doing it.
WRITTEN CUE:	The police in Japan …
MODEL RESPONSE:	"The police in Japan turn a blind eye to people cycling on the pavement"

Responses were recorded and transcribed and scored on the basis of: selection of correct target; accuracy of sequence; appropriacy of sequence adaptation; and, accuracy of whole response. A copy of the full test is given in Appendix 9.3.

9.3. Results

9.3.1. *Recall, accuracy and fluency*

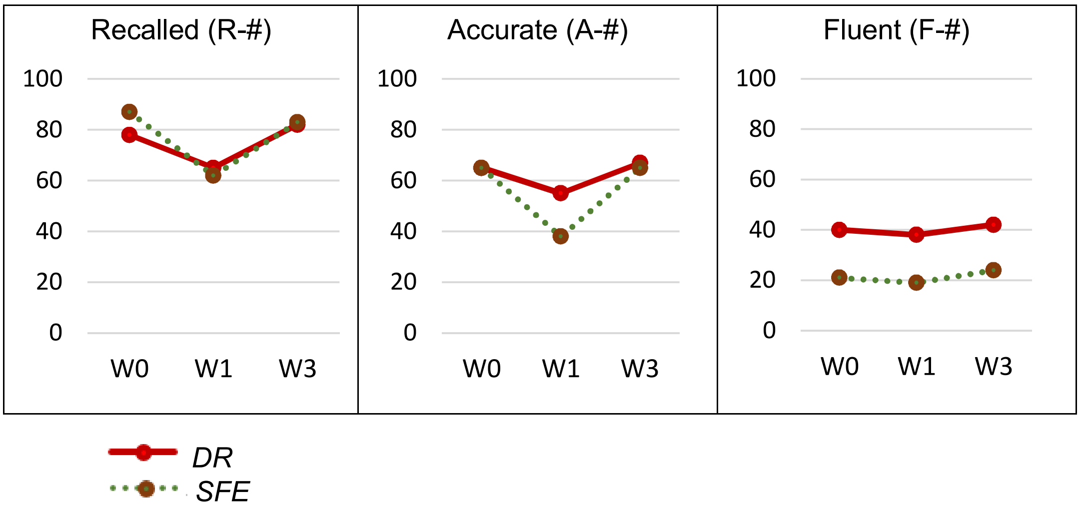
Overall, the ten participants, each learning six sequences via DR and six via SFE, provided 120 participant-sequence combinations (60 for each condition). The numbers of sequences that were recalled (R-#), fully accurate (A-#) and consistently fluent (F-#), along with mean accuracy scores (A-mean) and mean fluency scores (F-mean) are given in Table 9.5. Results are given across the two conditions for each of the assessment stages: i.e. immediately after learning (W0), after new week (W1) and after three weeks (W3).

Table 9.5: Recall, accuracy and fluency by condition and assessment phase

		Recall	Accuracy		Fluency	
Phase	**Cond**	**R-#**	**A-mean**	**A-#**	**F-mean**	**F-#**
W0	DR	47 (78%)	0.819	39 (65%)	0.864	19/47 (40%)
	SFE	52 (87%)	0.866	39 (65%)	0.813	11/52 (21%)
W1	DR	39 (65%)	0.710	33 (55%)	0.879	15/39 (38%)
	SFE	37 (62%)	0.641	23 (38%)	0.776	7/37 (19%)
W3	DR	49 (82%)	0.819	40 (67%)	0.892	22/49 (44%)
	SFE	50 (83%)	0.850	39 (65%)	0.835	12/50 (24%)

The initial effect of the two input methods can be seen in the results immediately after learning (W0). These show that recall was slightly better for sequences learnt via SFE, while accuracy is similar across the two conditions. For fluency, the proportion of recalled sequences that were consistently fluent (F-#) is higher in the DR condition (40%) than in SFE (21%) and this difference was significant according to a chi-square test of independence, χ^2 (1, N=99) = 4.36, p=0.0367. For the subsequent assessments, the general pattern of results for recall, accuracy and fluency is a dip from W0 to W1 followed by a return to earlier levels at week 3 (W3). This can be seen graphically in Figure 9.2.

Figure 9.2: Percentages of recalled, accurate and fluent sequences by condition



The mean accuracy and fluency scores show a similar pattern to the counts. In general, written accuracy was similar across conditions and assessment stages to spoken accuracy.

9.3.2. ***Read-out-loud analysis***

To provide a further comparison between the output of targets learnt by DR and by SFE, the voice onset time (VOT) and articulation rate (AR) was measured for each sequence spoken in the read-out-loud task.

Table 9.6: Read out loud (ROL) analysis summary

Phase	Cond	Mean VOT(s)	Mean AR (syll/s)
W0	DR	1.082	4.106
	SFE	1.071	3.837
W1	DR	1.056	3.989
	SFE	1.123	3.847
W3	DR	1.079	4.135
	SFE	1.040	3.913

Table 9.6 gives the mean scores for VOT and AR at each stage. As hypothesised, AR was slightly faster on average for DR compared with SFE all stages. Based on independent samples t-tests (one-tailed), these differences were significant at W0, $t(53)=1.750$, $p=0.415$, and at W3, $t(59)=1.861$, $p=0.0326$, although the effect sizes

were not large (Cohens d = 0.291and 0.340 respectively). There was no clear difference between the VOT scores across conditions or phases.

9.3.3. *Usage*

The usage test administered at the end of the week 3 checked participants' ability to use the learnt sequences in a different context. The test involved them selecting the appropriate sequence, adapting its form to the new context and fitting it into the new context appropriately. For example, adaptation of form could mean changing *set his sights on* to *set her sights on*, or *stood the test of time* to *stand the test of time*. Table 9.7 shows the performance of items according to whether they were learnt via DR or SFE.

Table 9.7: Usage test results

	Selection		**Usage**			
	Selected	**score**	**Accurate**	**Adapted**	**Context**	**U-score**
DR	51/60	85%	43	30	38	73%
SFE	48/60	80%	38	29	35	71%

Selected = no. of correct target sequences selected

Accurate = no. of original sequences remembered accurately

Adapted = no. of sequence forms adapted appropriately

Context = no. of sequences fitted appropriately into wider sentence (e.g. "she set her sights on winning a medal")

The overall usage score (U-score) was calculated for each sequence that was correctly selected by averaging percentage scores for the three usage components. The mean across each condition is given in the table. From the results, there did not seem to be any difference in the ability to adapt sequences for usage, at least on the basis of this test.

9.4. Discussion

The aim of the current study was to provide an input method based on oral repetition (DR) that would encourage the acquisition of target sequences directly as single holistic forms. The hypothesis was that expressions acquired in this way would be more likely to be spoken fluently and consistently in subsequent recall than those acquired initially through construction of the sequence (SFE). This discussion therefore focusses on the extent to which that effect was found and other differences between DR and SFE. Overall performance over the stages of assessment, and across participants and targets is also considered.

9.4.1. *Evidence of holistic input and fusion*

Following Myles and Cordier (2017), the approach taken in the research so far has viewed the fluent, consistent delivery of a target (unitary) sequence as a good indicator of its internal formulaicity for that speaker. F-# provides a count of such sequences. At the end of the initial learning session (W0), F-# was significantly higher for sequences learnt via DR than for those learnt by SFE and this difference was maintained through the assessments. While there was some fall in the value for F-# in W1 (for both conditions), this was largely due to the lower recall at that point (i.e. expressions could not be retrieved whether or not they were formulaic). Otherwise, targets tended to remain fluent across assessments. This suggest that the DR input did support holistic acquisition more than SFE and may have resulted in more expressions becoming formulaic for the speaker straight away. This is supported by the finding that articulation rates were faster in the DR condition at W0. However, while around 80% of the sequences were recalled at W0, even in the DR condition only about 40% of these were fully fluent. This may indicate limits on the numbers of sequences that can be memorised holistically in the given time period.

On the other hand, the fact that the actual number of fully fluent sequences did not change much between W0 and W3 indicates that few additional sequences became formulaic for the speakers over the three weeks. For most targets, a reconstructive approach continued to be applied during recall. A typical pair of responses is given in Table 9.8, where there is increased fluency and accuracy at W3, but without yet being sufficient for the target to be considered internally formulaic for the speaker.

Table 9.8: Context recall responses by Kentaro for 'breathed a sigh of relief'

W0	he breathe on / breathe a / this one he breathe something / breathe / sigh / of relief / that …
W3	he breathed / a / bre- breathed a sigh of relief / because …

Note: / indicates point of dysfluency

It may be that the assessment tasks at W0 and W1 (effectively the only way the speakers had to 'practise' the expressions) were not sufficient to move more sequences into formulaicity, but further practice could have done so.

9.4.2. ***Differential effects of DR and SFE***

As shown above, the DR approach seemed to facilitate holistic acquisition to a greater extent than SFE (as measured by initial and continuing fluency). At the same time, this approach did not appear to have a detrimental effect on the recall or accuracy of the learnt expressions, which were similar for both conditions across the stages.

The DR approach involved only repetition (in context) and did not entail any detailed analysis of the component words or how they might link to the overall meaning. However, this did not seem to have any effect on the ability of participants to apply and adapt the sequence to a different context, at least as far as the Usage test measures this capacity. Results from the test at W3 were very similar for sequences learnt by DR and SFE and in both cases a high proportion of the sequences that were recalled were applied appropriately.

9.4.3. ***Overall recall performance over time***

The overall pattern of performance in the recall and accuracy of the target sequences was for a reduction in week 1 followed by an improvement in week 3. Since all participants confirmed that (as instructed) they had not reviewed the targets between tests, the overall reduction in performance at week 1 may be part of a natural decay in memory (Baddeley, 1997). However, the increased recall, accuracy and fluency at week 3 (W3) is more notable. Since the only additional learning or review of the sequences following the initial input session was the week 1 (W1) assessment check, the week 3 results suggest that this had a positive effect on the longer-term learning. This interpretation may be linked to work on spaced retrieval (Kornell et al., 2015) which suggests that the recall of learnt items (e.g. words learnt via flash cards) is enhanced by each attempt to retrieve them, and this effect occurs whether or not that attempt is successful, provided the correct answer is subsequently given. Although their work was not specifically on the learning of sequences, the retrieval conditions in the assessments used here were comparable. So, the repeated assessments may have had a cumulative effect in supporting the enhanced performance at week 3 as this was the fourth time the sequences were retrieved.

9.4.4. ***Differences across participants and targets***

Despite the specific nature of the participants (Japanese working adults at intermediate/advanced levels of English), there were considerable differences in their performance. For example, Naomi delivered a target response with consistent fluency only 6% of the time (over all assessments) while Akemi did so 36% of the time. The variety in participant performance was only partly due to differences in proficiency level (as measured by the TOEIC scores and pre-vocabulary tests). Another key factor seemed to be the extent to which they currently used English in their daily (working) life.

There were also differences across the target sequences in terms of how well participants remembered and used them. One aspect observed frequently in the participant feedback was the degree to which participants felt the meaning expressed by the Japanese translation (which was used on the cue card for the cued recall test and for the written check) matched the English sequence. This was quantified to some extent as, during SFE input, participants were asked to rate each sequence in terms of how 'close' they thought it was to the translation (on a word by word basis) on a scale of 1 (not close at all) to 10 (very close). Interestingly, across the targets, there was a correlation between mean closeness and overall recall, $r(10)=0.604$, $p=0.037$. However, there was no correlation with other measures such as accuracy or fluency. Indeed, the highest rating on closeness was for *breathe a sigh of relief* which was also the sequence least often delivered fluently. For sequences, the word length seemed to be a more important factor, particularly for the measure of fluency, with the four-word targets tending to be delivered more fluently than the five-word ones.

Overall, the study was carefully controlled to ensure that differences in participant and sequence performance did not have a systematic effect on results. However, wide variation across the data resulting from differences in the relative performance of participants and sequences would serve to reduce the power of the study to demonstrate any effects present.

9.1. Conclusion

This study has provided some evidence that an initial focus on repetition (without looking at the written form, and delivered with sufficient intonation and feeling) might help establish the holistic storage of target sequences early on, and that this may help maintain fluency and accuracy of output over time. Further, such a focus on repetition, provided it is sufficiently linked to a particular context and meaning, does not appear to impact negatively on recall or the ability to adapt the sequence to another context. It is not suggested, however, that the DR procedure on its own be recommended as a means of acquiring formulaic sequences. Indeed, a combination of DR and SFE type approaches would likely be most beneficial. However, research on processing specificity (e.g. Barcroft, 2002) suggests that, in a situation of limited time and processing resources, there are trade-offs in output. So, if the goal is formulaicity, sufficient time on repetition at an initial stage would appear to be important.

CHAPTER 10: Measuring sequence automaticity (Study S6)

Comparing fluency and automaticity in the routes to acquisition of formulaic expressions

10.1. Introduction

So far, fluency has been used as a key indicator of formulaicity, and this follows a precedent set by previous research. However, Chapter 8 raised the possibility that consistent fluency alone may not always be sufficient to indicate that a given unitary sequence has holistic storage or automatic processing. A possible implication of this is that fluency is a staging post towards formulaicity rather than evidence of having reached the destination. In order to explore this further, this chapter describes a study which utilises two further checks on formulaicity (based on automatic processing) and applies them to the same participants and target sequences from study S5. Since the target sequences were learnt and carefully monitored in that study, this final empirical study (S6) provides a good way of comparing fluency with measures of automaticity and also a further opportunity to monitor the acquisition of formulaicity in the targets learnt.

10.1.1. *Automaticity and formulaic expressions*

The processing advantage and fluency benefits of formulaic expressions are often described in terms of ‘automatic’ processing (Myles & Cordier, 2017; Pawley & Syder, 1983). However, few studies have explored this aspect of formulaicity explicitly. In the case of individual words, automaticity is most commonly associated with lexical access (DeKeyser, 2001). For formulaic expressions, however, automaticity may also be associated with the absence of a conscious or effortful on-line construction of the sequence during speech production. While a variety of approaches have been applied to automaticity in speech processing generally (as discussed in detail in Chapter 8), one that seems most usefully applicable to formulaicity is that of Segalowitz (2010). He considers automaticity an element in the cognitive processing of L2 speakers that leads to ‘utterance fluency’. A key view of Segalowitz is that automaticity is more than a simple speeding up of cognitive processes; it involves a qualitative change in the way a process is organised or structured. The holistic storage of formulaic sequences may represent such a qualitative difference when compared with simply constructing the sequence more and more fluently through repetition. Segalowitz (2010, pp. 80-90) offers two ways of describing automaticity linked with qualitative restructuring. These are automaticity defined as ballistic processing and automaticity defined as processing stability.

Automaticity defined as ballistic processing ('ballistic automaticity') captures the idea that automatic processes are unstoppable once they start and involve involuntary operations. It was illustrated by Favreau and Segalowitz (1983) in a priming task where a prime word describing a category influenced the reaction times to a lexical decision task when the prime was related to the target, even when the participants were explicitly instructed to expect a target unrelated to the prime. They proposed that it was ballistic automaticity that underlay participants not being able to help processing lexical information prompted by the prime even though it was not expected to be relevant.

Automaticity defined as processing stability ('stable automaticity') is the idea that automatic processes will be subject to less variation compared to non-automatic processes. This was measured by Segalowitz and Segalowitz (1993) in terms of intra-individual reaction time (RT) variability in a set of lexical processing task items. They used a measure called the coefficient of variation (CV) which is the ratio of the mean RT to its standard deviation (SD). Usually, for an RT measuring some kind of processing, the SD varies linearly with mean RT. So, when a system of processing speeds up without change to the underlying process, the CV remains constant. However, if the CV reduces (e.g. after a period of training or practice), it suggests that some of the less efficient (slow / variable) parts of the process have been eliminated, leaving more stable components of the underlying mechanism. CV as a measure of automaticity has been used by Akamatsu (2008) and Pellicer-Sánchez (2015) to test the effect of targeted learning on fluency and automaticity in L2 learners.

10.2. Developing tests for automaticity

In this study, the ideas of ballistic and stable automaticity were recruited to develop two further ways of checking potential formulaicity in target sequences (in addition to fluency). The main check was one for 'holistic automaticity' (HA), a feature that describes a type of ballistic automaticity associated specifically with holistic storage. In HA, the idea is that, when the start of an internally formulaic expression is cued, the rest of that expression is invoked automatically (involuntarily called to mind) due to its holistically stored form. A second method, based on stable automaticity, was also developed, expanding on the definition of CV given by Segalowitz and Segalowitz (1993) and the read-out-loud method introduced in study S5.

10.2.1. ***Holistic automaticity (HA) test***

In holistic automaticity (HA), when the first word of a target sequence is activated (by hearing the word as an auditory prime), the speaker cannot help but process the whole sequence for potential speech production. In particular, subsequent words in the sequence will be activated and, given a suitable cue, preferentially selected over other candidate words in a word response test. The reasoning for this draws on the amended hybrid model of speech processing of Sprenger et al. (2006) introduced in Chapter 8.

Figure 10.1: Adapted version of 'superlemma' model

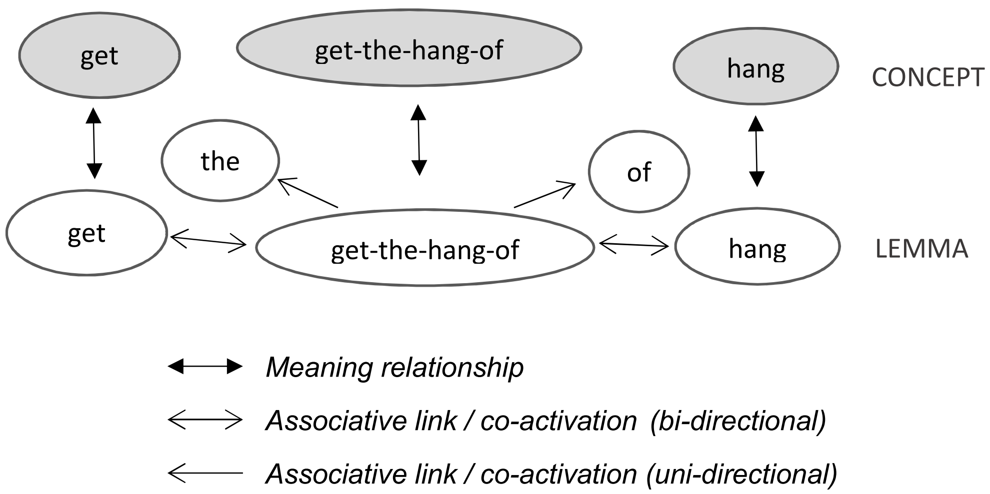


Figure 10.1 shows a simplified version of the model (2006, p. 1760) as applied to a target sequence from the current study. If the sequence is formulaic, the contention is that a superlemma (*get-the-hang-of)* exists and it is linked both to its conceptual meaning directly and to the lemmas of its constituent words via associative links. When the identity prime ('get') is heard, the lemma for *get* is activated which then activates the lemma for *get-the-hang-of*. This in turn activates the other constituent word lemmas, including the lemma for *hang*. When the letter cue ('h_') is then seen, it triggers a search for words beginning with 'h'. Since '*hang'* is already active, selection of this word is facilitated above other candidates.

This idea is also linked to research on final word completion in idioms and high frequency multiword phrases. For example, in an eye-tracking study, Underwood et al. (2004) showed that final words in idioms are processed more quickly than those in matched compositional expressions. In a neurolinguistic study, Connolly, Phillips, and Forbes (1995) found that event-related potentials (ERPs) associated with semantic and phonological processing are absent when participants encounter high cloze probability sentence endings. Segalowitz (2010) and DeKeyser (2001) suggest that the absence of expected brain activity during a task is associated with automatic processing. This is consistent with the proposal that the fast, preferential production of the target word in the test (given the identity prime and letter cue) is indicative of automatic processing of the sequence.

10.2.2. *Testing for stability of processing*

Segalowitz (2010) suggests that automaticity can be measured by looking at the stability of processing over repeated performance. In particular, automaticity for a processing act is indicated when the Coefficient of Variation (CV) for a given measure of that act (e.g. reaction times on a lexical task, articulation rates or durations) is reduced to a minimum level. In this study, for each participant, a repeated read out loud (ROL) task was used to calculate stability of response for each target sequence. Two measures were used: voice onset time (VOT), that is, the time from presentation of the written sequence to the beginning of speech; and articulation rate (AR) the speed of articulation in syllables per minute. Reduced mean CV scores across particular groups of target sequences compared with others was taken as an indication of stable automaticity for that group.

10.3. Method

10.3.1. *Participants*

The 10 Japanese speakers of English (JSE) recruited for the original study described in Chapter 9 all agreed to take part in this follow-up. As described in Section 9.2.1 all were working adults living in Japan and were at an intermediate/advanced level of English. They had completed a Vocabulary Levels check (Schmitt et al., 2004), and answered a background questionnaire about their English usage and study experience. Informed consent was obtained and they were assured about the anonymity of their contributions.

10.3.2. ***Target sequences***

Along with the 12 sequences previously learnt by the participants in study S5, six new control sequences were introduced. These were selected using the same principles as the originals and confirmed to be unknown to the participants. In particular, they contained familiar lexis, they were formulaic ‘in the language’ and they did not share any lexical words.

For the HA testing, the initial verb of the sequence was taken as the prime and one of the key lexical words in the remainder of the sequence was the target word. For example, for the sequence *get the hang of*, *get* was the prime word and *hang* the target. Each sequence was to be presented twice: once with a cue letter corresponding to the target word (T-cue), once with a cue letter unconnected to the sequence (NT-cue). The list of sequences, primes and cue letters is given in Table 10.1.

Table 10.1: List of target sequences

	Sequence	**Prime**	**T-cue**	**Target**	**NT-cue**
A1	turned a blind eye to	turned	b	blind	f
A2	came to a head	came	h	head	b
A3	breathed a sigh of relief	breathed	r	relief	t
B1	run the risk of	run	r	risk	l
B2	go a long way towards	go	l	long	s
B3	like the sound of	like	s	sound	t
C1	set his sights on	set	s	sights	t
C2	stood the test of time	stood	t	test	i
C3	get the hang of	get	h	hang	r
D1	knew better than to	knew	b	better	r
D2	toyed with the idea of	toyed	i	idea	l
D3	remains to be seen	remains	s	seen	l
E1	look on the bright side	look	b	bright	f
E2	rolls off the tongue	rolls	t	tongue	h
E3	scared the life out of	scared	l	life	h
F1	walk on thin ice	walk	i	ice	s
F2	reserve the right to	reserve	r	right	i
F3	lie at the heart of	lie	h	heart	f

10.3.3. ***Procedure***

Each participant undertook the following procedure individually, between two and three months after their final attendance in study S5:

- Fluency assessment of targets
- Brief review of sequences
- ROL stability check
- Introduction of control words
- HA test

10.3.3.1. Fluency assessment

Although participants had attempted to learn each of the 12 sequences in the original study, there was variety in the level of recall, accuracy and fluency across the participants and sequences even after the final session. In addition, a period of around two months had elapsed between the final session and the current study. Therefore, the current fluency of the learnt sequences was assessed for each participant.

To do this, the two recall tasks given in the original study were repeated for the current study, and the participant's output was assessed in exactly the same way as before. On the basis of this, participant-sequences were categorised into one of the following:

a. No recall (NR): The sequence was not recalled with sufficient accuracy in either task

b. Major dysfluency (D-major): Major or multiple dysfluencies in either task

c. Minor dysfluency (D-minor): Only one minor dysfluency in one or both tests

d. Fluent - low recall (F-low): Recalled on one test and fully fluent in that one

e. Fluent - high recall (F-high): Recalled on both tests and fully fluent and consistent in both

This categorisation was chosen to separate out those sequences that were judged formulaic for that speaker at that time (d, e) from those that were not (a, b, c). In addition, it enabled exploration of the extent to which ease of recall (of the whole sequence) and the 'degree' of fluency of a sequence may be relevant to automaticity. A minor dysfluency was defined as a single short pause (between 0.2s and 0.5s) occurring in one or both of the tests.

10.3.3.2. Brief review of sequences

Following the assessment, the six new control sequences were read out to the participant and then shown on a written list with a Japanese translation. The participant read each one out loud once to ensure it could be said smoothly with no pronunciation difficulties. This process was designed to familiarise the participants with the new sequences. However, it was assumed that this limited exposure and the subsequent read out loud would not be sufficient to render these sequences formulaic for the participants.

10.3.3.3. Read out loud (ROL) stability check

After a short break, the 18 target sequences were presented to the participant one at a time in written form on a computer screen. The participants were given instructions to say the sequences out loud “as quickly and naturally as possible”. Three sets of the 18 sequences, each in a random order, were presented, resulting in 54 items for each participant. For each item, VOT and duration were measured from the audio recording as described in Chapter 9. Articulation rates (AR) in syllables/second were calculated by dividing the no. of syllables in the sequence by the time taken to articulate it. The aim was to compare average VOT and AR between the conditions, and also to calculate the CV (coefficient of variation) for each:

$$\text{VOT-CV} = \frac{\text{standard deviation of VOT score}}{\text{mean of VOT score}}$$

$$\text{AR-CV} = \frac{\text{standard deviation of AR score}}{\text{mean of AR score}}$$

10.3.3.4. Introduction of response word controls

To provide some degree of control over the possible responses in the HA test, a set of 40 words was introduced immediately before the test. The presentation of these words was designed to control any possible effect that prior exposure to words might have on the test. In particular, it ensured that the target words were not preferentially in mind before the test (e.g. as a result of exposure to the target sequences earlier in the session). The 40 words contained 8 different starting letters which matched the range of cue letters of the test. All 18 target words were included along with 22 dummy words[35] of similar form, resulting in 5 words for each initial letter. The set of words is given in Table 10.2.

Table 10.2: List of control response words

better	fast	hang	ice	learn	rang	seen	test
blind	fight	head	idea	letter	relief	sights	tight
bread	find	heart	interest	life	run	song	tongue
brown	fate	hike	item	long	right	sound	teeth
bright	fish	here	icon	lion	risk	space	tree

Participants were presented with the words one-by-one on cards in random order. After repeating each one out loud (and confirming it was known), they performed a simple grouping exercise based on initial letter and repeated them again. After a break and immediately prior to the HA test, a brief check was done in which the participants were presented with each cue letter and asked to say out loud any word they could think of. The presentation was prefaced by a beep and they had 4s for each letter. The purpose of this was to ascertain whether target words were preferentially in mind before the test.

10.3.3.5. HA test and analysis

The computer-based HA test consisted of 36 items (two for each target sequence). For each item, there was a fixation point on the screen accompanied by a beep. After 2.5s an auditory prime of the cue word (the first word of a sequence) was played and a further 750ms later, the cue letter appeared. Each auditory prime lasted between 500-600ms, leaving a short gap (150-250ms) before the letter cue was shown. The 36 items were presented in pseudo-random order to ensure that: (a) the two occurrences of each sequence were well separated, (b) the same cue letter was not repeated sequentially, and (c) cue letters did not follow presentation of a prime word with the same beginning letter. This was to minimise cross-item interference. Participants were given the following instruction:

> “You will hear a word. You will then see a letter. Say a word beginning with that letter as quickly as you can. NOTE: You may like to use one of the words

[35] All 22 new words were ranked 4000 or less (in lemma frequency) on the iWeb online corpus (Davies, 2018) and were considered likely to be well-known to the participants.

introduced earlier but you don't have to. The aim is to respond as quickly as possible."

The aim was to encourage participants to choose words from the list but without compelling them to think too consciously about it. Each test was recorded and the participant response and response time (RT) noted for each item. To determine whether the target word had been activated and spoken quickly and in preference to other possibilities, a set of criteria was applied for each target sequence:

- The expected target word must be chosen in response to the T-cue.
- The RT for this word should be faster than that for the NT-cue word for the same prime.
- If there are other occasions when the same target word is given (i.e. as an NT-cue response to a different prime), all of these should also have slower RTs.

If all criteria were satisfied for a sequence for the participant, it was marked as a 'holistic hit'. To illustrate, Table 10.3 gives a typical example of a possible set of participant responses involving the prime 'get' and the response 'hang' (for testing the sequence *get-the-hang-of*). In this example, the appropriate target response is given, and its RT is faster than any other response involving the prime 'get' or the response 'hang'. So, it would be marked as a holistic hit.

Table 10.3: Example HA test responses

Prime	Cue	Response	RT (s)
get	h_	hang	1.125s
get	b_	boy	1.491s
lie	h_	hang	1.662s
rolls	h_	hang	2.010s

Assuming that participants were choosing from the given set of response words, the chances of the automaticity criteria being passed by chance is less than 10%. To illustrate this, consider the null hypothesis case where *hang* has no relationship with *get*. On presentation of the prime 'get', the chance of a participant randomly choosing *hang* from the 5 possible response words beginning with 'h_' is 20%. In only half of those cases (i.e. 10%) will this choice be faster than whatever the participant chooses for the prime *get* with cue 'b_'. This percentage is further reduced by the possibility that *hang* will be randomly chosen for other 'h_' cues and delivered faster than when it was chosen for *get*.

10.4. Results

Across the 10 participants, a total of 49 of the original sequences were deemed to be formulaic (23 low recall and 26 high recall), while 56 were non-formulaic (38 with major dysfluencies, 18 with a minor dysfluency) and 15 were not recalled at all. This information was used to divide the results into categories for subsequent analysis.

10.4.1. *Holistic automaticity*

In the word check test, 61% of responses were from the list of 40 control words given at the start of the session. Of these, 34% were target words from the original sequences and 16% were target words from the control sequences. These figures are close to the percentages expected if the words were chosen at random (12/40=30% and 6/40=15%, respectively). This was the anticipated result and confirmed that the target words were not preferentially activated before the test compared to other possible choices of words.

Table 10.4 gives the numbers and proportions of holistic hits (i.e. cases where all the automaticity criteria were satisfied for a participant-sequence) across the main sequence categories. As the table shows, the memorised sequences deemed formulaic by the criteria had a much higher percentage of holistic hits compared with non-formulaic learned sequences. The control sequence results are similar to those of the original sequences which were not recalled. Excluding the No Recall group, a chi-square analysis comparing Control, Non-F and Formulaic groups shows that the differences are significant, χ^2 (2,N=165)=*25.257, p<0.00001*. The effect size (given by Cramer's V) was 0.28, representing a medium to large effect (Cohen, 1988).

Table 10.4: Proportion of holistic hits over main categories

Sequence type	N	Holistic Hits	Hit rate
Control	60	9	15.0 %
Not recalled	15	2	13.3 %
Dysfluent	56	15	26.8 %
Fluent	49	31	63.3 %

Looking at the more detailed categories, Tables 10.5 and 10.6 present the effects of 'degrees of fluency' and recall, respectively. Table 10.5 shows that the proportion of holistic hits rises steadily from major dysfluency to minor dysfluency to fluent. Table 10.6 shows that, within sequences categorised as fully fluent, it rises from low recall to high recall. The results are suggestive that the likelihood of a sequence being holistically automatic increases the more formulaic it appears to be (as measured by fluency) and the more easily it is recalled.

Table 10.5: Holistic hits by 'degrees of fluency''

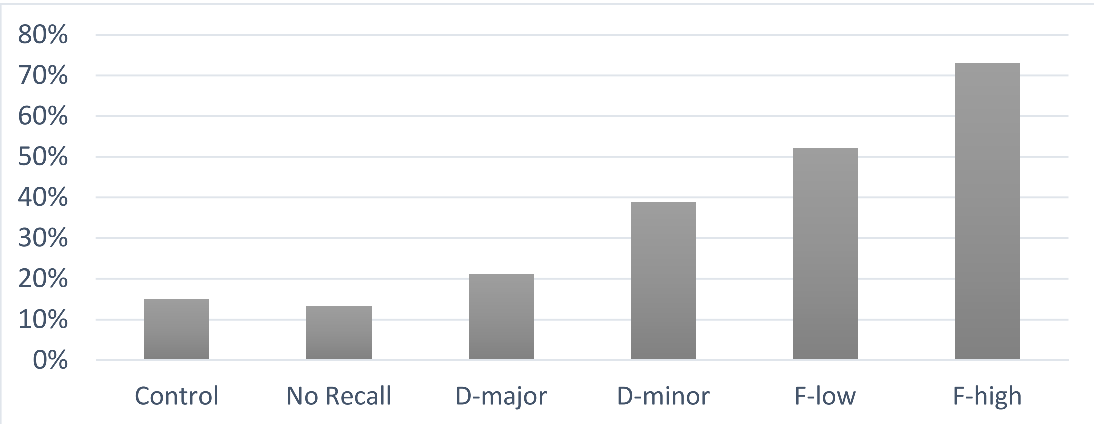
Sequence type	N	Holistic Hits	Hit rate
Dysfluent - major	38	8	21.1 %
Dysfluent - minor	18	7	38.9 %
Fluent	49	31	63.3 %

Table 10.6: Holistic hits by recall (over consistently fluent sequences)

Sequence type	N	Holistic Hits	Hit rate
Fluent - low recall	23	12	52.2 %
Fluent - high recall	26	19	73.1 %

Figure 10.1 summarises the results, showing the continuous rise in holistic hits (representing holistic automaticity) through the categories (representing increasing degrees of fluency).

Figure 10.1: Percentage of 'Holistic hits' per sequence type



10.4.2. ***Processing stability check***

For each participant, the means and standard deviation for voice onset time (VOT) and articulation rate (AR) were calculated for each sequence over the three repetitions. CV (coefficient of variation) values were then calculated and the mean values for these over all participants and category groups are given in Tables 10.7 and 10.8.

10.4.2.1. Voice onset times

Table 10.7 shows that the mean VOT and the mean VOT-CV score are lower for the F-high (fluent with high recall) category. However, the combined fluent group (including both high and low recall fluent expressions) has a very similar CV score to the non-fluent categories. Thus, while it was expected that fluent expression would show stable automaticity, the results suggest that it was only high recall fluent expressions that show this, at least according to the VOT-CV indicator.

Table 10.7: Summary of VOT results

<table>
<tr><th>Sequence Type</th><th colspan="2">Mean VOT (s)</th><th colspan="2">Mean VOT-CV</th></tr>
<tr><td>Control</td><td>0.807</td><td></td><td>0.127</td><td></td></tr>
<tr><td>Not recalled</td><td>0.791</td><td></td><td>0.134</td><td></td></tr>
<tr><td>Dysfluent (Major)</td><td>0.790</td><td rowspan="2">0.779</td><td>0.119</td><td rowspan="2">0.116</td></tr>
<tr><td>Dysfluent (Minor)</td><td>0.754</td><td>0.111</td></tr>
<tr><td>Fluent (Low recall)</td><td>0.805</td><td rowspan="2">0.762</td><td>0.172</td><td rowspan="2">0.119</td></tr>
<tr><td>Fluent (High recall)</td><td>0.725</td><td>0.072</td></tr>
</table>

A separate post hoc analysis was therefore done to check the influence of recall more generally on VOT-CV scores. To do this, the original sequences were categorised in terms only of their ease of recall (Table 10.8) and all high recall sequences in term of their fluency (Table 10.9).

Table 10.8: VOT results by recall (over all original sequences)

Recall type	N	Mean VOT (s)	Mean VOT-CV
Low recall	66	0.796	0.139
High recall	54	0.746	0.096

Table 10.9: VOT results by fluency (over high recall sequences)

Recall type	N	Mean VOT (s)	Mean VOT-CV
Dysfluent	28	0.766	0.1184
Fluent	26	0.725	0.0718

While the mean VOT was similar across categories, high recall sequences had a much lower VOT-CV than low recall sequences, and an independent samples t-test showed this difference was significant, $t(118)=2.717$, $p=0.0075$. In addition, for high recall sequences only, those that were fluent had a significantly lower mean VOT-CV than those that were non-formulaic, t(40)=2.496, p=0.0084. This suggests that both fluency and ease of recall may be relevant to stable automaticity (as measured by minimising VOT-CV).

10.4.2.2. Articulation rates

The articulation rate (AR) results are given in Table 10.10. As can be seen, the sequences categorised as fluent were delivered with a higher articulation rate and also had a lower CV than the dysfluent sequences and controls. The mean CV for the No Recall category was low however. Excluding this category, there is a difference in mean CV between the main Control, Non-formulaic groups and Formulaic groups, but this is not statistically significant.

Table 10.10: Summary of AR results

Sequence type	syllables	Mean AR (syll/s)		Mean AR-CV	
Control	4.67	3.579		0.0928	
No Recall	4.93	4.035		0.0661	
Dysfluent - major	4.97	4.080	4.0296	0.0858	0.0879
Dysfluent - minor	4.78	3.923		0.0925	
Fluent- low recall	4.35	4.313	4.2851	0.0664	0.0662
Fluent - high recall	4.65	4.260		0.0661	

Note: syllables = average no. of syllables for expressions in that group

10.5. Discussion

10.5.1. *Fluency development over time*

While the main focus of the study was on comparing the fluency and automaticity measures and how they relate to the acquisition of formulaic expressions, it also provided a good opportunity to track the target sequences originally learnt in study S5. An important finding from the initial assessment of the 12 original targets was that 105 (88%) of the participant-sequences were recalled and 49 (47%) of these were classified as consistently fluent following the context and cued recall tasks. This shows that the overall numbers for recall and fluency rose in the two months from the final assessment (W3) in study S5 to the assessment to the start of study S6. In the case of fluency, this increase was considerable, rising from 34% (34 out of 99) in W3 to 47% (49/105). While it is possible that some participants came across the sequences during the two months, this increase may be further evidence of a spaced retrieval effect as described in Section 9.4.3. Further, of the 49 fluent sequences, 33 were originally learnt by DR and 16 by SFE. For the 31 fluent sequences that also had holistic hits, that ratio was consistent (21 to 10). This suggests that the long-term benefit (in terms of fluency and formulaicity) of the DR input is maintained.

10.5.2. *Fluency and holistic automaticity*

The HA test was a way of seeing whether the initial word of the sequence primed the whole sequence sufficiently that a subsequent target word within the sequence was selected by the participant in preference to other candidates (and with a faster RT compared to other responses not involving the target sequence). As would be expected if fluency is a necessary indicator of formulaicity, the fluent sequences had a significantly higher proportion of holistic hits than the dysfluent and control sequences. The proportion of hits rose steadily through the categories, suggesting that holistic automaticity may be related to the relative fluency of the sequences and the ability to recall them. To some extent, the results also support the idea that automaticity is a 'stronger' condition than fluency on the road to formulaicity, with some fluent sequences yet to have reached the holistic automaticity stage

In the approach taken here, holistic automaticity is taken to indicate that the expression is a holistic unit in the mind of the speaker (e.g. as a 'superlemma') which is processed automatically as a single unit. There is a question therefore of why some dysfluent participant-sequences had holistic 'hits'. These are those cases where a participant-sequence passed all the test criteria for HA even though it was earlier found to have a dysfluency in the context and cued recall tasks given at the beginning of the session. One reason is that the HA test is to some extent probabilistic. Based on random but appropriate choices (i.e. they give a word starting with the given cue letter) from the 40 control words and the criteria for a holistic 'hit', the predicted false positive rate would be around 10% as discussed in Section 10.3.3.5. There may be other factors associated with the test which could also cause false positives. For example, it may be that some primes and targets are linked associatively (e.g. because they have been heard together before) even though the overall sequence is not formulaic. Another type of possibility, raised in Chapter 4, is that some internally formulaic sequences (for a speaker) are occasionally be delivered with a dysfluency. This will be discussed further in Chapter 11.

Of course, the probabilistic nature of the HA test works both ways, and it is possible that a genuinely formulaic expression could fail to show a holistic hit on occasions. In particular, there was the possibility of interference from other potential associates of the prime. For example, for the prime ‘run’, one participant chose a similar sounding word *rung* rather than *risk*, even though *run the risk of* was consistently fluent for them. Also, some target words were chosen correctly (e.g. *idea*, for *toyed-with-the-idea-of*), but then the same word was used as a later response more quickly, thus failing the adopted criteria for holistic automaticity in this test, and this could have been due to its previous priming. Also, since the lower recall fluent sequences had fewer holistic hits than the higher recall, it seems likely that the ability to recall a formulaically stored expression also affects performance on the HA test.

Overall, however, the results do suggest a staged increase in holistic hits for increased levels of fluency, complementing the idea that formulaicity in a sequence may be acquired over time and with practice. The findings also seem to suggest that there could be increasing ‘degrees of formulaicity’. However, if formulaicity is also conceptualised in terms of holistic storage and a ‘superlemma’ model, it does raise questions about what it could mean to be partially or ‘nearly’ formulaic. These questions are addressed in more detail in the general discussion in Chapter 11.

10.5.3. ***Measuring automaticity due to stability***

The idea that automaticity may be indicated by stability of processing was explored in the study by analysing the variability of two measures associated with delivery of the target sequences in the read-out-loud test. In particular, processing was measured according to how quickly participants responded to and read out the target sequences over repeated trials, using voice onset times (VOT) and articulation rates (AR) to measure this. Stable automaticity is observed when the Coefficient of Variation (CV) associated with the variables is minimised.

For the VOT, both the onset times themselves and the CV were lowest for the F-high category which means that the variation may have been minimised for fluent, high recall sequences compared with the others. At the same time, means and CV values for the F-low (fluent, low recall) group were high and comparable with the non-fluent sequences. Comparing high and low recall (over all learnt sequences), the significantly lower mean VOT-CV for the high recall group suggests that the ability to recall an expression is an important factor affecting the variability of voice onset time over repeated measures. This is possibly because such easily recalled sequences are recognised more quickly and this speeds up preparations to speak. However, even within the high recall sequences, those categorised as fluent had a significantly lower CV, suggesting that fluency still influences stable automaticity the VOT-CV even when the effects of recall are removed.

For articulation rates, mean AR was higher for the learnt sequences than the controls and for fluent sequences compared with dysfluent. Recall that fluency was measured in terms of pauses during production of the sequence in the assessment tasks, while AR was measured independently in the read-out-loud task. The AR-CV scores show a similar pattern although the differences were not significant. It should be noted that, while the average number of syllables across the controls and learnt sequences was comparable, sequences which were not recalled (No Recall) or with major dysfluencies (Dysfluent-major) on the assessment task did have higher average numbers of syllables, suggesting that the longer sequences were more difficult to learn and recall fluently.

Overall, the CV results are in line with the hypothesis that fluent sequences have greater stable automaticity, but the lack of significance highlights some potential challenges with the method. Firstly, the processing measured by the test involved reading out loud, which may not be sufficiently taxing in processing terms to show significant differences, particularly for articulation rates. Also, repetition of the sequence only three times may not capture the subtle differences in variability required by the method. A more pressing concern however is in establishing at what point the CV can be said to be minimised. In the methodology applied by Segalowitz and Segalowitz (1993) and Pellicer-Sánchez (2015), the comparisons in CV were intra-participant between time periods (before and after a period of study) and calculated over all items attempted. In the current application of the method, the CV is calculated for individual participant-sequences making it potentially susceptible to outliers (although there did not appear to be any in the data) and the comparison was done across categories of sequence fluency type. So, while the significantly lower VOT-CV for the fluent high-recall group suggests stable automaticity, it does not guarantee it (since there is scope for the value to be yet lower).

10.6. Conclusion

This study built on the previous study S5 in order to explore automaticity in the processing of formulaic sequences by L2 speakers. An advantage of linking the two studies was that the provenance of the target sequences was well-known for each participant, making it easy to identify potentially formulaic sequences via fluency and recall, using the same assessment process. Using the two psycholinguistic tests of automaticity adapted from previous research, the study showed that consistently fluent sequences were more likely to be automatic than dysfluent sequences and controls, but that not every fluent sequence was automatic as tested here.

While results from the stable automaticity check (particularly for VOT) showed expected associations with fluency, some questions remain about its theoretical grounding, particularly in respect of how best to adapt the methodology to test the automaticity of individual sequences in a speaker. The holistic automaticity (HA) test does, however, appear promising as a tool for exploring automaticity in the processing of formulaic sequence as holistic units. Some refinements to the method may be useful to further control interference from prime word associates.

Overall, the results support the idea that, for many learnt sequences, formulaicity develops over time, and that consistent fluency is not necessarily the end point in the journey to becoming a formulaic expression. The study also provides further input on the potential routes involved, how these link with a holistic storage view of formulaic sequences, and the possible interplay of recall, automaticity and fluency in observing acquisition. These points will be discussed in detail in the discussion in Chapter 11.

CHAPTER 11: General discussion

Identifying and modelling the routes to formulaicity in the acquisition of target formulaic expressions by L2 speakers

11.1. Introduction

This final chapter is an opportunity to draw together the empirical research and theoretical insights covered across the previous chapters. In doing this, we return to the key research questions asked at the beginning:

- What are the psycholinguistic processes by which targeted expressions become internally formulaic for L2 speakers?
- How do different approaches to memorisation (operational/learning process/sequence of actions) influence the acquisition process?

In tackling these questions, the research has progressed through a number of stages, building on findings and questions arising from the literature review in Chapter 2. The first two empirical studies (S1 & S2) explored ways of identifying formulaic expressions in the speech of L2 speakers. This helped to build a picture of formulaic usage and the extent to which expressions can be observed going through the process of becoming formulaic. The next study (S3) replicated a method of introducing target utterances to L2 speakers and monitoring their subsequent usage. This was extended to analyse how the targets (or segments within them) may become formulaic for the speakers. Building on that, an approach was developed in a small exploratory study (S4) whereby specific target sequences were introduced to a small group of L2 speakers. Following further exploration of the literature regarding how formulaicity may be represented psycholinguistically in the mind of a speaker, the approach was adapted and applied in a further set of studies (S5 and S6). The aim of study S5 was to compare two different ways of memorising the targets, specifically designed to induce either 'holistic acquisition' (where a speaker takes on the expression as single whole unit) or 'fusion' (where the expression is initially reconstructed but later becomes unitary). The follow-up study (S6) used an additional psycholinguistic method to check the formulaicity of the learnt expressions a few months later.

In drawing together these findings, this discussion first looks at two key threads through the studies. The first covers the challenge of identifying internal formulaicity on the basis of spoken output. The second looks at what the studies say about the acquisition of internal formulaicity over time (given an initial target to learn). Building on these points, and on the results for the final studies and theoretical research on lexical acquisition, a simple model for two different paths to acquisition is then presented and discussed. Although empirical work can only go so far in finding out about psycholinguistic processes, the model provides a way of summing up the findings and a platform for further exploration.

11.2. Identifying internal formulaic expressions

A key part of this research is establishing how the individual creates and uses formulaic expressions, irrespective of whether or not they are considered formulaic in the language as a whole. The main approach has been modelled on the hierarchical criteria proposed by Myles and Cordier (2017). In this, the first necessary condition for internal formulaicity is phonological coherence, and the second (applied after the first has been satisfied) requires the sequence to show signs of unity. In practice, fluency is used as a measure of phonological coherence and the first condition is applied as a way of segmenting text, with points of dysfluency marking the boundaries between runs of fluent text. The unitary condition is then applied to pick out the potentially formulaic expressions from the fluent runs.

This approach was applied in the first empirical study (S1) which explored the use of internal formulaic expressions in a group of intermediate / advanced level Japanese speakers of English. As in the Cordier (2013) application of the approach to advanced L2 learners of French, the findings suggested that the use of formulaic expressions by L2 speakers was higher than that indicated by previous research using other identification methods such diagnostic criteria (e.g. Wood, 2009) or comparison with lists of external formulaic expressions in learner corpora (Granger, 2019). At the same time, it remained lower than most estimates for native speakers, and this was borne out in S1 when the same sampling method and hierarchical approach was applied to two native English speakers for comparative purposes. The mean percentage of text that was formulaic was 34.6% (range 29.6% - 40.3%) for the Japanese speakers of English (JSE) and 47.3% for the native speakers. The study also supported previous research suggesting that L2 speakers, even of similar proficiency, vary considerably in the way they use formulaic expressions. Speakers with experience of living or studying in an English-speaking country or who used spoken English at work tended to use more formulaic expressions than those without that experience. This suggests that exposure to appropriate expressions is important.

The sequences used were mainly referential (verb phrases, noun phrases, time / place complements) and few examples of grammatically or functionally irregular sequences were found. There was evidence that learners chose or created expressions geared towards their needs and experience, such as those related to work (e.g. '*procedures for foreigners*', '*put the cheque in*', '*test administration*', '*month end*').

Study S1 demonstrated that the hierarchical approach of Myles and Cordier can be applied effectively to speech samples in a reasonably consistent way (although a considerable amount of time is required to carefully transcribe multiple samples of spoken text). The study also provided the opportunity to examine the fluency and unitary form criteria in more detail. A number of aspects of their use, both practical and theoretical, were explored and then developed further over the course of the research. These relate to the identification of unitary form, the significance of 'inconsistent fluency' (i.e. when an expression is fluent on some occasions but not on others), the relationship of fluency to processing advantage, and the role of accuracy in the analysis of formulaic expressions. These are discussed in the following sections.

11.2.1. ***Challenges with identifying 'unitary form'***

A challenging aspect of the hierarchical approach is determining which part of a fluent run (following segmentation via the first criterion) is an expression with unitary form. The possible criteria for 'unitary form' suggested by Myles and Cordier (2017) are that the expression has semantic/functional unity, or that the sequence was learnt holistically. Semantic/functional unity means that the expression has a holistic form-meaning or form-function mapping. It may have a single meaning (e.g. an expression for place or time) or single function (e.g. as a filler like *I don't know*, or to introduce a view, *in my opinion*). Often semantic or grammatical irregularity will be a good indicator (since in these cases the whole cannot be derived from the constituent parts). Many formulaic expressions are represented by a single grammatical constituent such as a nominal or prepositional phrase. However, this does not have to be the case, since expressions like sentence starters (e.g. *I think…*) and formulaic frames (e.g. *not only X but also Y*) have a function but are not complete as grammatical units until the sentence ending or variables are entered. The other potential criterion that Myles and Cordier give is that the expression was learnt holistically, and this relies on knowledge of or assumptions about the speaker's previous learning experience.

Overall then, the unitary form criteria are diagnostic and (as outlined in Chapter 2) there can be challenges in applying them consistently. Some features may be reasonably clear (e.g. grammatical unity or semantic opacity), but others may be less so. One particular challenge is in deciding whether a fluent expression represents a fixed formulaic expression or a frame with open slots. For example, in S1, Yayoi delivers a fully fluent utterance *it's partially the subcontractor's job to train proctors*. It is hard to say which parts of this are formulaic for her (or if the whole utterance is) on the basis of the criteria. It may be that *it's the X's job* to and *train proctors* are formulaic and have been combined with other variable items (*partially* and *subcontractor*) on this occasion. This is the phenomenon of 'nesting' described in Chapter 3. Deciding whether an expression has been learnt holistically requires knowledge of the individual and their learning experience. The working adults featuring in the studies had experienced a wide variety of learning contexts (school/university classes, adult conversation classes, everyday use of English abroad or at work). So, it would be difficult to say what they might have come across. In particular, each individual speaker will have their own experience and idiolect, and certain word combinations may have taken on a unitary meaning or function for them even if they are not thought to have one in the wider language.

The case study (S2) used a different approach for identification because its aim was to identify potentially formulaic items, deemed 'repeated unitary expressions' (RUEs) and then check their fluency. The initial segmentation of the text involved an algorithmic approach based on Brooke et al. (2014). This utilised the frequency of recurring n-grams (for n>1) in the individual learner corpus. The algorithm was designed to resolve issues of overlap across repeated expressions or the nesting of one in another. The use of frequency in this process was possible because there was a much larger sample of the individual's speech and due to the design of the speaking task, repetition of expressions was much more prevalent. In S2, there were many expressions that obeyed the frequency conditions and were also consistently fluent across all instances of their use. Most of these could also be identified as having unitary form. However, there were a number of consistently fluent repeated expressions such as *and the*, *in this*, *for that* which were not unitary, and others like *busy this week* and *make the vessel*, for which it would be hard to make such a case.

This highlights a point of debate in the wider research about the extent to which a unitary form condition is necessary or appropriate. A case in point is the treatment of 'lexical bundles', defined by Granger (2019) as contiguous corpus-driven sequences extracted “disregarding any pre-defined linguistic categories” (p.236). While she refers to these as a type of formulaic sequence, other researchers (e.g. Kesckes, 2019) exclude them due to their lack of psychological salience. Wray (2019) suggests re-casting lexical bundles as a separate type of formulaic expression, one which provides the ‘mortar’ of text into which the ‘bricks’ (typical formulaic sequences that can be ascribed a dictionary meaning) are contained. She argues that this type of formulaic expression is different in kind to more unitary forms, but is nevertheless vital for fluent speech. While such strings were not a focus of the current research, it is interesting to speculate on the extent to which they provide a processing advantage through automation even though they are not obviously holistic in form.

In the end, the choice of whether to include or reject such expressions as formulaic depends on the theoretical stance of the researcher and the aims of the particular research study. In the current research, a unitary condition is considered important but with a recognition that, for each individual speaker, what is unitary (in terms of the sequence being linked to a single semantic or procedural node) could be very broad indeed. In study S2, a unitary criterion was applied to filter the repeated expressions that resulted from the algorithmic segmentation (not part of the original Brooke et al. method). Due to the challenges highlighted above, a very broad interpretation of the unitary criterion was adopted. In particular, the only sequences that were rejected as being non-unitary were:

- sequences starting with conjunctions/discourse markers (e.g. *but they, so I, yeah so*)
- sequences consisting only of functional/closed class words (e.g. *and the, on this, for that, us to*)

There were 23 sequences removed according to the above two criteria (14%). Most strings rejected on the first criterion above were, in fact, dysfluent anyway, but most from the second were fluent. Adopting a criterion such as the second one above may therefore be useful in filtering out non-unitary forms. The later studies circumvented the unitary condition issues by focussing on targeted expressions, which were unitary by definition.

11.2.2. *The meaning of ‘inconsistent fluency’*

In study S1, there were some expressions (deemed unitary) which were repeated by a speaker across their samples of speech. In some cases, these were consistently fluent and so the identification of these as formulaic was unproblematic (at least insofar as fluency indicates formulaicity). Indeed, the repetition added weight to that identification, in line with the third (graded) criterion of frequency. However, while not many, there were some repeated expressions for which fluency was inconsistent across repetitions, and this was a common occurrence in subsequent studies (which were more geared to observing the phenomenon).

If fluency is supposed to indicate formulaicity, what could it signify when an expression is sometimes delivered dysfluently by a speaker? Several issues need to be considered in addressing that question. Firstly, the analytic approach taken in the first study excluded consideration of dysfluent versions of an expression, because the criteria for identification filtered them out. This could be argued to skew our perception of the data. The expression might alternatively have been deemed *non*-formulaic, on the basis that the speaker was not able to produce it fluently every time. Either classification, however, would overlook what is perhaps the most interesting aspect of the inconsistency, namely that (in keeping with the dynamic nature of the individual lexicon) it could be indicative of the expression becoming formulaic through a process of gradual fusion over time.

The longitudinal nature of the case study (S2) allowed observation of changing fluency in expressions repeated (spontaneously) at different time points. It showed that fluency tended to increase over time (i.e. most expressions with inconsistent fluency tended to go from dysfluent to fluent), supporting the conjecture that inconsistent fluency may indicate fusion in action. However, there were some expressions for which this was not the case. These could be interpreted as being temporarily formulaic (see discussion in Section 11.3.4 below) or at an early stage of acquiring formulaicity (which may or may not be realised later). This highlights that in a one-off sample of text, some expressions may appear fluent at that point but would not be if they were repeated.

Another issue concerning the use of fluency is the possibility (discussed in Chapters 2, 8 and 9) that formulaic expressions are not always delivered fluently. In natural discourse, such pausing or hesitation could be for planning speech while holding one's turn (Wray, 2019) or for socio-pragmatic reasons, such as appearing sincere (Bardovi-Harlig, 2019). While these particular reasons for pausing are unlikely in the current context (neither turn-taking nor sincerity was really an issue here), they do highlight a possibility to be aware of in the design of studies. Overall, fluency and dysfluency in a single occurrence of an expression in a one-off sample need to be interpreted with caution. However, in study S2, most repeated unitary expressions were in fact consistently fluent. So, in a sample of authentic speech, observing fluency in an expression (even if it occurs only once) may be a reasonable indicator that the expression would be delivered fluently more generally and therefore is formulaic for the speaker. In the more artificial experimental situation of studies S4 and S5, inconsistent fluency was much more prevalent in the target sequences, suggesting that, in those cases, the expressions had not yet reached formulaicity.

11.2.3. ***Does fluency alone indicate a processing advantage?***

In the studies where L2 speakers learnt new target sequences, the principal indicator of formulaicity was taken to be (consistent) fluency of delivery (as the targets were of unitary form by definition). A question was raised during the research regarding the extent to which fluency of delivery in a unitary expression, even if it is consistent, is sufficient to mean the expression is formulaic. This question centres on the extent to which consistent fast and fluent delivery necessarily means that the expression has a processing advantage due to holistic storage or automisation.

One aspect of this may be to ask whether fluency is the only or best way of establishing phonological coherence of the expression. Chapter 6 explored this question by looking at some other possible indicators. In the examples selected, it was clear that looking at speed of delivery (using articulation rates, AR) was of limited usefulness due to the variety of factors that influence AR and the difficulty of comparing formulaically delivered expressions with how the same expression would be delivered by the same speaker if it was not formulaic. Looking at the connection between words may be useful because features such as intrusion, assimilation and elision seem to be good indicators of phonological coherence (at least between the two words that are connected). It may also be reasonable to say that such phonological features represent a kind of fusion of two parts into a unitary form. However, since such a connection does not tend to occur for every word boundary within an expression and may not occur at all, it has value only as a supporting feature. Indeed most of the potentially formulaic expressions across the various studies did not feature such word boundary connection features. The exploration in Chapter 6 however, did suggest that the definitions of dysfluency might also usefully consider phenomena like elongated syllables, as proposed by some researchers (e.g. Erman, 2007).

Perhaps the key challenge with using fluency concerns its sufficiency as an indicator of formulaicity. In particular, as discussed in Chapter 9, Siyanova-Chanturia (2015) argues that fast and fluent delivery is not sufficient to show holistic storage, and Segalowitz (2010) suggests that automaticity must involve some qualitative structural change in the way that expressions are processed. Study S6 sought to operationalise these psycholinguistic features (holistic storage and automatic processing) by developing a check for 'holistic automaticity'. This is the idea that the elements of the expression are linked together (e.g. in holistic storage) in such a way that once processing of a formulaic expression starts, it cannot stop. The test for holistic automaticity developed for the study was based on this principle, drawing on the theoretical model of Sprenger et al. (2006) which posits a single lemma (the superlemma) as the basis of how a formulaic expressions is represented in the mind.

Insofar as the test measures this holistic automaticity (HA), the results suggested that, while increasing degrees of fluency are associated with greater likelihood of HA generally, the two measures do not fully coincide. In particular, there were target sequences that appeared to be fully fluent but were not HA. While the HA test was experimental and was developed for this study (i.e. it has not been independently tested itself), it does support the idea that consistent fluency of a (target unitary) expression may not always indicate that the expression is internally formulaic (i.e. it is stored holistically or processed automatically). Since psycholinguistic features such as holistic storage or automatic processing cannot be measured directly, a variety of tools can be useful to triangulate findings when exploring whether expressions are fully formulaic for a speaker. With suitable refinements (e.g. to further control interference from prime word associates and reduce the rate of false positives), a method such as the HA test appears promising as a tool that could be used alongside measures such as fluency and accuracy.

11.2.4. ***Role of accuracy***

While fluency was the principal means of identifying formulaicity, accurate reproduction is clearly important when learning formulaic expressions. To be useful for communication, a formulaic expression needs to be understood by the hearer and this can obviously be compromised by errors in storage or delivery of the form. In some cases (e.g. in S3, when Yumi says "*I prefer to teach free-lance compare with working for company"*) the meaning can still be made out, and often errors with function words such as articles or particles may not be fatal. However, when expressions are idiomatic or at a higher level of sophistication (grammatically or semantically) than the speaker's constructed speech, attention to the precise form may be essential. Such cases were not uncommon in the replication study (S3). For example, when Kiyomi says “*Putting on a kimono is a surprisingly time-consuming process*” in the real performance, the native speaker is audibly surprised (and, as he confirms later, very impressed) at this unexpected but entirely apposite use of English. On the other hand, when Sachiko says “*Japanese learning arts usually starts appreciating the style”*, the morphological omissions (*the*, *of* and *from*) rendered the utterance unintelligible to the native speaker. On a related note, Bell and Skalicky (2019) make the observation that when L2 speakers deliberately change a known idiom (e.g. for humorous effect), they are frequently misunderstood as having made an error.

While accuracy is important for unambiguous communication, the extent to which it can also help to indicate formulaicity is less clear. In studies S1 and S2 there are examples of expressions (such as *he and me*, *hasn’t got much of idea*) which seem to have become formulaic for the speaker (on account of being fluent and having a unitary dimension) despite being inaccurate or non-standard. There are two possibilities: either the speaker constructed the expression incorrectly or they heard the standard expression being used but failed to remember it accurately. This may arise because elements are not particular salient or contingent to meaning (Ellis, 2006) and/or because the learner has not paid it enough attention, as in the case of the Japanese speaker of English described by Schmidt (2010) who spoke fluently but with many errors in the expressions he used (see Section 2.5.3.6). Further, where the meaning is reasonably clear, the L2 speaker may not get corrected (which is common in non-classroom usage). Thus initial errors within acquired expressions may lead to fossilisation (if speakers are unlikely to hear a correct form of the expression or if they are particularly inattentive), or the speaker may start to amend or fill in the details over time if they have sufficient exposure. The latter case is suggested by Bardovi-Harlig (2019, p. 110) who describes the pragmatic acquisition of ‘conventional expressions’ by L2 speakers as following a series of stages.

> non-target-like response → target-like response but non-target-like lexical resources → target-like lexical core → full conventional expression

Whether inaccurate or non-standard expressions used formulaically by L2 speakers are considered as formulaic expressions is down to the definition of formulaicity adopted by the researcher. If the expression is being fine-tuned in the manner described by Bardovi-Harlig above, then there may be a case for saying that inaccurate versions have not yet reached full formulaicity. On the other hand, the choice could be made to include inaccurate or non-standard expressions. For example, one of the diagnostic criteria of Wray (2008a, p. 120), as described in Chapter 2, could be interpreted to deal with such cases:[36]

> Although the word string is novel, it is a clear derivation, deliberate or otherwise, of something that can be demonstrated to be formulaic in its own right.

The approach in the current research has been to treat inaccurate or non-standard expressions as formulaic for the speaker provided they satisfy the unitary form and processing criteria. In the case where a speaker's speech is sampled, then accuracy is not a guide to formulaicity unless the provenance of the expression and its development are known. In study S3, where specific target utterances were memorised and reproduced in speech, the participants were known to have experienced the correct versions. The focus regarding inaccuracies (deviations) here was on understanding how participants may have gone about memorising the targets and their approach to the task. Unlike in the original (Fitzpatrick & Wray, 2006), participants in most cases did not deliberately change expressions, so deviations were considered to have arisen from misremembering component parts or errors in reconstruction of the expressions during speech.

The analysis in Chapter 6 suggested that deviations and dysfluencies did not tend to coincide significantly, and deviations, particularly morphological ones, often occurred in fluent segments. In the subsequent studies (S4 and S5), which focussed on the learning of given target sequences, accuracy was measured (to give a broader understanding of the effect that different memorisation tasks had). Measures of accuracy also contributed to assessing formulaicity in that only targets over 70% accurate and delivered with consistent form (for the context and cued recall assessments at that stage) could be candidates for formulaicity at any of the stages of assessment. This allowed cases where the target was delivered with just one word incorrect (e.g. *like a sound of* for *like the sound of*) to be considered for formulaicity (provided they satisfied the fluency or holistic automaticity conditions). In fact, there were not many cases where a target was reproduced fluently and yet in the wrong form.

---

[36] While this criterion is quite broad and includes deliberate derivations, it could be adapted if a researcher wanted a more explicit way to include erroneous versions.

Overall then, provided the expression was recognisable, minor errors were not considered a bar to formulaicity in the research studies. Looking at deviations in the studies was however useful in helping to understand how expressions may have been memorised. More generally, acquiring formulaic expressions has wider benefits than just fluent communication. For example, they also function as standard means of communicating, signs of being part of community, and ways of conveying nuanced ideas quickly and effectively. So, accurate reproduction is important in these respects as well as for unambiguous communication.

11.3. L2 acquisition of targeted expressions over time

A main aim of this thesis has been to explore how new expressions given to an L2 speaker to learn are internalised. In particular, the research is concerned with determining if and when the expressions become formulaic for the speaker, and the factors that influence this process. Two main areas of focus can be considered: the initial learning (or memorisation) of the target sequences, and the subsequent changes in how the expressions might be stored and processed by the speaker over time. This section looks at ideas and evidence from the studies regarding these two aspects, then considers the patterns of acquisition that the findings suggest.

11.3.1. *The initial memorisation process*

It might be expected that, given sufficient time and motivation, an L2 speaker would be able to memorise any expression presented as a fluent semantically unitary entity and reproduce it exactly as given. However, as this and other research (e.g. Fitzpatrick & Wray, 2006) has shown, there are a variety of factors that will affect the extent to which this is possible. As the replication (S3) showed, individual differences in aptitude and approach towards the task seem to account for differences in outcome (as measured by recall, accuracy and fluency of reproduction). In addition, features associated with the target, such as length, complexity and the learner's existing knowledge of component parts also seemed to play their part. In exploring these latter factors, it is instructive to consider the process by which both longer, multi-item expressions and shorter, single clause expressions might be memorised under the constraints of working memory.

11.3.1.1. Memorising long utterances

It has been shown that it is possible to memorise long texts such as the Koran (e.g. Gent & Muhammad, 2019) or oral poetry (e.g. Finnegan, 2017) verbatim, without necessarily understanding the detail. Such techniques broadly work by automising sentences and linking them together via procedural cues. In order to understand the possible psycholinguistic processes required for memorising the long utterances of S3, it is useful to consider how this kind of technique works at the sentence level.

Essentially, the process seems to involve a number of stages in each of which items are chunked together into a larger unit and linked to a semantic or procedural cue. These stages are repeated until the sentence is a single chunk with a single cue. The number of stages depends on the number of items in the target and is constrained by the limitations of working memory (Cowan, 2010) to around four items at a time. To illustrate, consider the process applied to the sentence *She expected that, as a young Japanese woman, I would know how to put a kimono on properly* (a target for Kiyomi in S3). This sentence may contain several items, each with its own semantic or procedural anchor, determined on the basis of the learner's initial knowledge of the component parts. For example, for Kiyomi these could be:

- *She expected that* = idea 1
- *as a X* = idea 2
- *young Japanese woman* = idea 3
- *I would know* = idea 4
- *how to X* = idea 5
- *put X on* = idea 6
- *a kimono* = idea 7
- *VERB X properly* = idea 8
- A, B, C (order of clauses) = idea 9

If formulaicity involves a single form-meaning or form-function link then the task should be to pull the parts together into a holistic phonological form and link it to single semantic or procedural cue. Such a cue could be something like 'what I say to describe my host mother's expectations' or 'that sentence I needed to learn starting with "she expected" in it'. Since the number of items is too many to manage in one go, some intermediary stages would be required. For example, this could involve first memorising each of the clauses as units and then putting them together: e.g. idea 1 + (ideas 2 & 3) + (ideas 5, 6, 7 and 8) + idea 9.

Kiyomi's responses at the practice performance (PP) stage can give an indication of how she internalised the target.

> *She expected // as Japanese- / young Japanese woman / I would know how to put a kimono on properly.*

This segmentation by major (//) and minor (/) dysfluencies may indicate that there were indeed three potentially unitary segments corresponding to the main clauses of the target.[37] Deviations in the first and second segments (the absence of the function words *that* and *a*) illustrate how potentially formulaic parts may nevertheless be incomplete. The reformulation (*as Japanese-)* could indicate that the second segment was known as a whole unit, but not easily recalled as such (see Section 11.3.4 for further discussion on such examples).

11.3.1.2. Memorising short utterances

While longer utterances necessarily require a strategic approach to memorisation, there is the possibility that shorter (semantically unitary) utterances can be learnt in one fell swoop as there are fewer items involved. Study S5 suggested that, while this did occur, it was by no means the norm. For example, based on fluency and accuracy measures after the initial learning (Week 0) in S5, only 30% of targets were recalled with consistent fluency.

There may be a number of reasons why many of the short targets were not immediately acquired holistically. In particular, some targets, despite being four or five words long, might still have had more items than could be handled in a single pass. For example, a target like *breathe a sigh of relief* might (for a particular speaker) need to be broken down into two or three short term memory units for the words, plus another one for their order, plus a particular focus on some tricky aspects of pronunciation, plus a fixing with the overall meaning. This is more than can be managed in one go in short term memory, so there would need to be some sort of nesting. If not enough time is available during the input phase, it may not be possible to go through all the stages to fuse these items into one. In cases where the memorisation is incomplete, recall and reproduction of the expression will necessarily

---

[37] When reproducing longer utterances, it will often be appropriate for there to be pauses (for example, between clauses or intonation units). A case could therefore be made that long utterances with appropriate pauses can be formulaic (with the pauses as part of the memorised phonology of the whole unit). However, this is not the approach taken in the current research which focusses on the formulaicity of single-clause units within longer targets or as targets themselves.

require a degree of reconstruction. For example, Mari's response during the learning phase was *he / breathed a relief / sigh*. Here, she may have established that *breathed a X*, *sigh* and *relief* were part of the expression, but not gone beyond that. As well as ordering issues and incomplete items, reconstruction may also increase the possibility of interference, such as when a semantic cue for one item (or for the whole thing) leads to a synonym being used (e.g. *get the idea of* for *get the hang of*).

A further complicating factor for short targets was that, as verb phrases, they required a subject to make them meaningful and, for the transitive type expressions, an object too e.g. *they run the risk of losing staff to their competitors.* Studies S4 and S5 were designed so that the short targets were presented in context within an example sentence related to a story. So, participants may have included aspects of the longer sentence in their memorising process, in order to better link the target to a specific meaning and usage. This could explain why the exploratory study (S4) found that memorising a longer sentence with a target embedded was not significantly different from learning just the short target.

It was conjectured that an important factor in how the short targets would be memorised was the way in which they were presented. In studies S4 and S5 the presentation of the targets was varied in order to guide participants into different ways of approaching the memorisation. Results from S4 were inconclusive regarding the effects of segmenting the targets into pre-defined chunks or embedding them in longer sentences. Study S5 compared an approach which focussed on the target as a single phonological form (Dramatic Repetition, DR) with one that was more analytic (Semantic-Formal Elaboration, SFE). Results suggested that the DR approach did seem to encourage more holistic acquisition as predicted. SFE, on the other hand, may have promoted a more reconstructive approach, leading to errors and dysfluency, as suggested above. However, even for the DR input, many targets were not learnt holistically. The DR approach focussed on getting participants to repeat the whole expression as a single unit linked directly to the meaning. Even though the emphasis was on the whole string of sounds rather than on the constituent words and their meanings, learners still needed to remember the 4-7 syllables, so some implicit segmenting into groups of familiar sounds was still likely to be necessary in order to repeat back the expression during memorisation. In cases where the limited repetition during memorisation was insufficient to establish the full phonological form as a single unit, the phonological parts would need to be recalled and put together during reproduction, leading to potential errors and dysfluency.

11.3.2. ***The effect of further practice***

As well as considering acquisition at the time of initial presentation, the studies also monitored use of the expressions over time. In the model utterance study (S3), practice on target utterances continued after the practice performance (PP) and participants incorporated feedback from this into the later real performance (RP) with a native speaker. In most cases, there appeared to be further development in the memorising of the target, leading to improved performance at RP despite it being a more pressurised situation. For example, in the case of Kiyomi described above, with additional practice, she was able to deliver the target almost perfectly at RP (except for the missing article), with pauses only between the main clauses.

For studies S5 and S6, additional practice was not part of the study except as provided by the regular assessments. Each assessment involved a variety of retrieval attempts, with correct responses being provided, and the opportunity to hear the expressions again (in the story) and oral presentation and repetition of each expression. In S5, assessments took place as part of the initial learning (IL), at the end of the learning session (W0), after one week (W1) and after three weeks (W3). There was also an assessment at the beginning of S6, 2-3 months later. The effect of these assessments on the recall, fluency and accuracy of the output was interesting. For all measures, mean values dropped at W1, and this may understood as resulting from a natural decay in memory (e.g. Baddeley, 1997, pp. 169-173) that would occur in the absence of any further practice during the week after the learning session. However, the measures rose again at W3, then rose further at S6. In the absence of any other practice, this increase would seem to result from a cumulative effect of the assessment. At time W1, there had been two previous assessments (IL and W0), at W3 there had been three (IL, W0 and W1) and at S6 four (IL, W0, W1, W1). Further, the two retrieval attempts preceding W1 (IL and W0) were only spaced by 20-30 minutes in the initial session, those preceding W3 and S6 were one week apart (W0-W1) and two weeks apart (W1-W3) respectively. The cumulative effect of the assessments over time supports research which highlights the importance of retrieval (Karpicke, 2012) and the beneficial effects of increasing the spacing between retrievals (Kornell & Vaughn, 2016). In particular, this spaced retrieval, feedback and repetition seemed to result in more targets becoming formulaic for speakers over time.

More generally, outside of the experimental situation, learners may have had the opportunity to fine-tune their use of learnt expressions, enabling them to become formulaic through exposure and use. However, this would very much have depended on having opportunities to hear or use the expressions, and these could be quite limited for many learners. Other factors such as salience (both of the expression as a unit and of elements of its form) and relevance to their needs are also likely to have been important. In studies S5 and S6, there were yet other factors that appeared to influence which participant-targets became formulaic in the long-term. As with study S3, there were individual differences across the participants, although these were not obviously associated with proficiency level. There were also differences between target sequences regarding how easily they appeared to become formulaic. The most successful were *came to a head* and *run the risk of*, and the least successful *breathe a sigh of relief*, *go a long way towards* and *set his sights on*. Generally, while length may have had a slight effect, transitivity and other features (such as alliterative lexical words, opacity of meaning) did not seem to be a factor. The targets were chosen to be similar to each other in form and non-congruent with the Japanese. However, it may be the case that some were more familiar or understandable to Japanese learners than other. (This is discussed further in Section 11.4.4 in relation to links with conceptual meanings). Regarding fluency and processing advantage, another relevant factor may be how easy an expression is to pronounce. For example, some participants suggested in feedback that *breathe* and *relief* were a challenge in this respect.

11.3.3. ***Patterns of acquisition***

In the case of the acquisition of formulaic expressions by L2 speakers through the memorisation of specific target sequences, results suggest that two broad routes are possible. These involve the immediate 'holistic acquisition' of the whole sequence as a single unit, and the fusion of component parts into a single unit over time. Studies S5 and S6 provided some evidence that this is the case, using consistent fluent reproduction of the target sequences after the initial learning (stage W0) as an indicator of holistic acquisition, combined with a specific test of 'holistic automaticity' as a sign of retained or acquired formulaicity at two months after the learning. In the two studies, the target sequences learnt by the Japanese participants were assessed at various points in time to determine the extent to which the expressions had become internally formulaic for the speakers. Overall, there were 31 participant-sequence combinations (26%) that were fluent and also demonstrated holistic automaticity at the time of S6.

Assuming that these are cases where internal formulaicity has indeed been attained, a closer look at these instances illustrates how the different routes are manifested in the output. A number of the expressions appeared to become formulaic straight away, particularly when learnt via the DR strategy. Throughout the assessments, these sequences remained more or less fluent and accurate, but varied in how consistently they were recalled. The example in Table 11.1 is typical of such cases. The target is delivered fluently and accurately in context and cued tasks at W0. However, at W1 the speaker requires a first word cue to deliver the expression, and in W3, she may be repeating the cue herself before delivering the sequence fluently.

Table 11.1: Kaori – 'run the risk of'

	Context recall	Cued recall
W0	*now /// run the risk of / losing staff*	*run the risk of*
W1	<no recall>	RUN => *run the risk of*
W3	*company /// run? /// run run /// run the risk of // losing staff*	*run the risk of*
S6	<no recall>	*run the risk of*

Note: RUN => indicates that the researcher needed to give the first word 'Run' as a cue on that occasion

Other sequences were not recalled fluently initially but became formulaic over time. This appeared to be facilitated by the practice and retrieval afforded by the regular assessments and suggests that some kind of fusion was taking place. Illustrative cases are given in Tables 11.2 and 11.3. In each of the examples, there is a mixture of fluent and dysfluent production (with the cued responses tending to be more fluent) and there is also evidence of reconstruction at the earlier stages.

Table 11.2: Tetsuko – 'toyed with the idea of'

	Context recall	Cued recall
W0	*he // he toyed // the idea of buying a new one*	*toyed with / with // the idea of / toyed with the idea of*
W1	*<no recall>*	*TOYED => // the idea of*
W3	*he // toyed / toyed with the idea of*	*toyed with the idea of*
S6	*he // toyed with the idea of buying a new one*	*toyed with the idea of*

Table 11.2 shows how some words ('*toyed*') and sub-sequences ('*the idea of*') may be known and linked as part of the expression. In joining these together during

reconstruction, non-lexical words ('*with*') may get missed out. Other examples from the studies include '*turned blind eye*' and '*breathed sigh of relief*'.

Table 11.3: Sachiko – 'set his sights on'

	Context recall	**Cued recall**
W0	*he set his / sight on / inventing ...*	SET=> *set his / sights on*
W1	*he set his /// he set his mind / of / creating a new game*	*set his /// set his / mind // set his /// target / it's not target*
W3	*then he set his / sights on / in- inventing new / games*	*set his sights on*
S6	*he / he set his sights on / inventing a new game*	*set his sights on*

Table 11.3 illustrates how existing knowledge, such as lexical associates of the component words (e.g. '*mind'*) or lemmas associated with the meaning (e.g. '*target'*), may interfere with reconstruction process. In these examples, the retrieval and corrective feedback of the assessments facilitated accurate fluent reproduction of the forms eventually. However, repetition without feedback could potentially lead to the fossilisation of non-target formulaic forms.

11.3.4. *Fluency, recall and 'degrees of formulaicity'*

While the general trend was towards increased formulaicity over time, there was some inconsistency. For example, in study S5, there were some sequences that were phonologically coherent at earlier stages but not later ones (i.e. it was not the case that fluency was always maintained from one stage to the next). In S6, although the results showed that the more consistently fluent an expression was the more likely it was to also show holistic automaticity, there were still some dysfluent expressions which appeared to have holistic automaticity.

There may be a number of reasons for this. As discussed in Chapter 10, the holistic automaticity (HA) is designed to establish if the initial verb of the target sequence automatically activates access to the whole expression, by checking if the other main lexical word in the expression is activated. The test is to some extent probabilistic (in that a cue 'h_' and prime *came* could produce the target *head* even if *came to a head* was not formulaic for the speaker) and therefore a minimal number of false positives is to be expected. A positive result could also be observed in the case that the expression is partially learnt, provided the two key words feature in the incomplete holistic form. For example, Mari produced the target *like the sound / of* with a dysfluency, but scored a positive response on the HA test for the same expression. In this case, we could say that *like the sound* was formulaic (but not quite the full target). Another situation, commonly observed in the studies, is where the apparent dysfluency occurs because the speaker is trying to self-cue their recall of the whole sequence. Possible examples are shown in the W3 context task responses of Examples 11.1 and 11.2 above. The same might apply in cases such as "*blind eye / a blind eye // he turned a blind eye to her behaviour*" in which the self-cue is a sequence ('*blind eye'*) within the expression. Such responses were marked as dysfluent due to the reformulations (which indicate breaks in the sequence). However, it could also be that the sequence is holistically stored but not easily recalled on this occasion. That would parallel the tip of the tongue (TOT) phenomenon (Ecke & Hall, 2013) where aspects of a word can be recalled (e.g. the first letter) but not the whole word (even though the word is presumably holistically stored). The self-cue word or phrase may act as a label for accessing the full expression, perhaps a phonological one (as in the case of TOT for words where the contributing part is a letter or phoneme) though other mnemonics might also apply.

This conjecture may be supported by the finding in S6 that a positive result on the HA test was far more likely when a sequence was easy to recall. An explanation is that, for low recall formulaic expressions, a holistic form exists but is not (yet) well-established in the lexicon (i.e. its connections with associated concepts and lemmas are still relatively few and weak). This could result in a lower level of activation in the HA test, making it more susceptible to interference from other more activated candidates. This reasoning could be extended to 'partially' formulaic expressions where a weakly established lemma exists, but there still remains the possibility of a speaker reconstructing the sequence in situations where the whole lemma cannot be accessed from the given cue. Thus, the identification of a sequence as 'formulaic' (via the fluency or holistic automaticity criterion) may depend not only on the existence of a holistic lemma but also on the strength and type of connections that that lemma has. This idea can explain the variation in holistic automaticity across categories, and also provides a way of understanding apparent 'degrees' of formulaicity within a holistic storage model such as that of Sprenger et al. (2006).

A further variation might also be the notion of 'temporary formulaicity'. In the first study (S1), there were cases where participants repeated back part or all of the interviewer's prompt question (e.g. *the most challenging part* [of my job] *is* …). In the case study (S2), Kensuke used specific expressions like *September and October* in a formulaic way to indicate a particular time period that was relevant at the time of the conversation, but not at other times. These illustrate that expressions may become formulaic for a speaker for short periods through imitation or creation on-the-hoof, linking to a particular pragmatic use at the time. The expression may not be strongly embedded in the mind however (e.g. the link to the cue is not particularly strong), and it could be lost (as a formulaic expression) if not further practised or used. This could presumably happen to target sequences learnt in S5 and S6, and could explain why a few expressions appeared formulaic at some point, but later were not so.

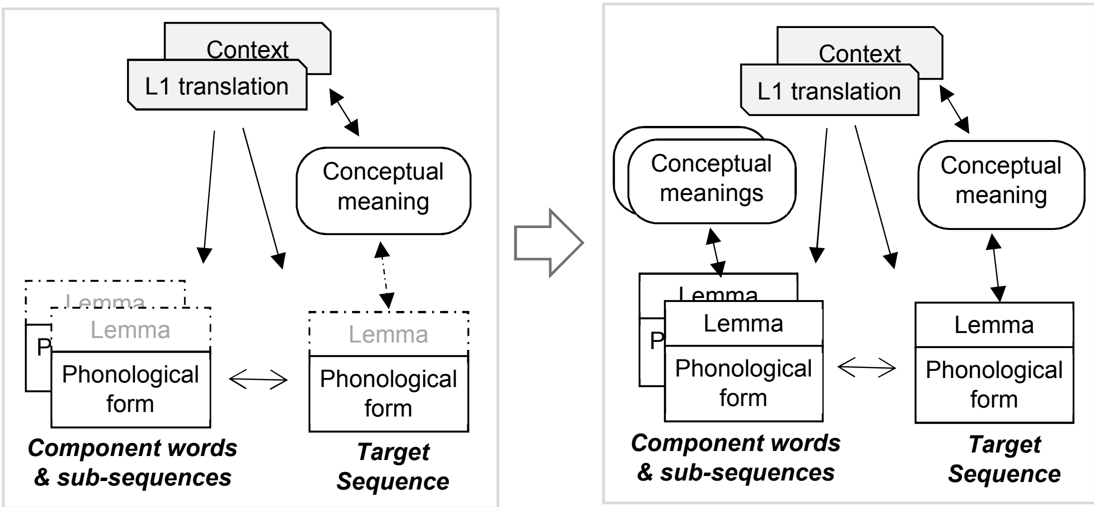
11.4. Modelling the routes to formulaicity

As discussed in Chapter 8, the 'superlemma' model of Sprenger et al. (2006) provides a useful way of showing how formulaic expressions may be represented in the mind, but it does not specifically address acquisition. There are, however, some more general models of vocabulary acquisition that may be adapted. For example, De Bot et al. (1997) provide a structure for describing and explaining aspects of L2 word acquisition based on Levelt's model of speech processing (1993). To recap, Levelt highlights the idea that knowing a lemma means having distinct elements of information including syntactic and semantic components which are, in turn, separate from the morphological and phonological components of the lexemes to which the lemma is linked. De Bot et al. suggest that when a learner encounters a new word, an 'empty' lemma structure is created. The learner then uses semantic and syntactic information from context (and morphological information from the lexeme depending on their experience of the language) to fill in this structure. Complementing these, Jiang (2000) in his lemma mediation model of L2 vocabulary acquisition suggests that, in the initial stages of acquisition, the phonological (or written) form of the word is stored and a lexical entry created. The semantic and syntactic (and morphological) information is then provided via associated links to the L1 translation or definition.

11.4.1. *Modelling holistic acquisition*

In terms of the models of De Bot et al. and Jiang, an outline hypothesis is that the DR approach in S5 helps to create a holistic phonological form of the target sequence in the mind of the speaker, facilitating the creation of an 'empty' lemma to which this lexeme is linked. The basic lemma structure is linked to the meaning (e.g. via the given L1 translation) and the context of the learning (e.g. via the story and the episodic memory of engaging with it). Holistic acquisition is achieved when there is sufficient targeted oral repetition of the sequence to create the (holistic) phonological form in memory and automate its retrieval given the appropriate elicitation cue. Accurate memorisation of the sequence in a fixed holistic form may then serve as a stable building block for further learning, to integrate semantic, syntactic and morphological aspects of the expression lemma.

Figure 11.1: Simple model of holistic acquisition for L2 speakers



Meaning relationship *Associative link / co-activation*

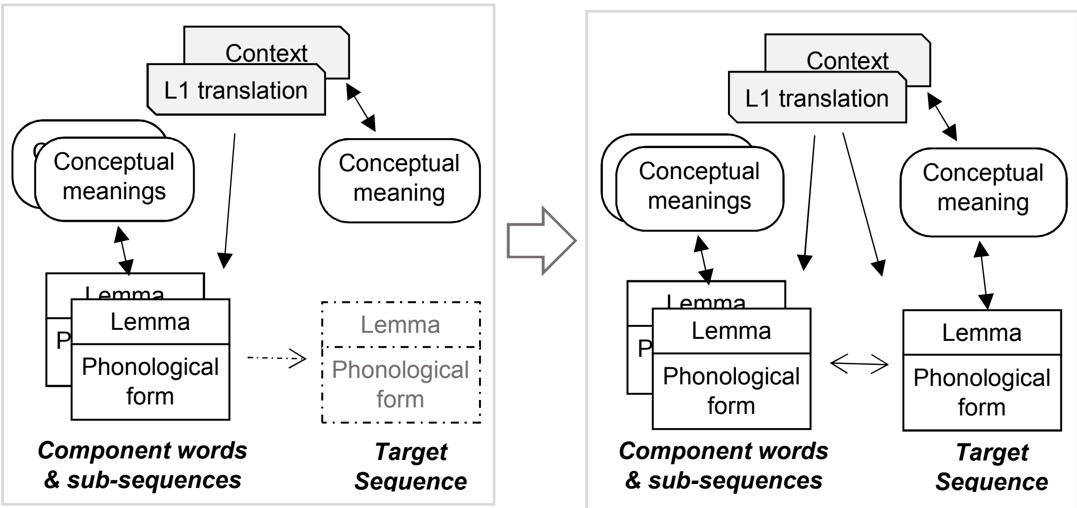
Dotted arrows indicate weaker links / dotted boxes indicate empty elements

Figure 11.1 presents a highly simplified model of this process, showing possible initial and final stages in the holistic acquisition of a formulaic expression. Initially, hearing the expression in context and seeing the L1 translation helps to set up the conceptual meaning. The holistic phonological form is established through the DR process and linked to the concept (and the strength of this may vary, as shown by the dotted arrow). The phonological form may also be linked associatively with phonological forms of words and sub-sequences, but direct links to their meanings are discouraged. As the target is retrieved and repeated over time, the link between the concept and the lemma is strengthened along with associative links to the lemmas of the component words and sub-sequences. This consolidates the holistic sequence lemma in memory and helps make it easier to recall.

11.4.2. ***Modelling fusion***

In the case of 'fusion', where a sequence is initially reconstructed to some extent before later becoming formulaic, components and sub-sequences (e.g. '*breathed*' and '*sigh of relief*'; '*turned*' and '*blind eye*') appear to be combined on-line, with dysfluencies marking their joins. In some cases, errors occur at the joins too (e.g. *breathed his sigh of relief*; *turned blind eye*) usually involving less salient function words (e.g. *a*, *to*), or occasionally the wrong choice of lexical word (e.g. *set his mind on*). In S5 and S6, there were also examples of morphological changes to the key lexical words (‘*breathe*’; *‘turn’*) compared to the given target. The morphological and lexical changes suggest that the meanings of the component words were being accessed during the reconstruction. Fusion therefore seems to involve a combination of the chunking together of known components and the correcting of erroneous or missing words. To some extent this process mirrors the latter stages of the acquisition sequence postulated by Bardovi-Harlig (2019) and discussed in Section 11.2.4. In particular, the targeted learning of given expressions can move learners quickly to the ‘target-like lexical core’ stage of Bardovi-Harlig’s proposed learning sequence, but further development is required to become fully formulaic. A possible model for this fusion process is given in Figure 11.2.

Figure 11.2: Simple model of fusion for L2 speakers



Meaning relationship *Associative link / co-activation*

Dotted arrows indicate weaker links; dotted boxes indicate empty elements

In the initial learning stage, while a conceptual meaning for the target sequence may be established, it is not linked to a holistic lemma or single phonological form. In order to recreate the expression therefore, it is necessary to access the lemmas of the component words and sub-sequences linked to the context and L1 translation. This link may involve the concepts of the component words if they are relevant to the concept for the whole (i.e. where the formulas are sufficiently transparent to enable a match of parts to whole at a semantic level). In cases where construction involves components which are not semantically linked to the whole, the link from parts may be more indirect, involving episodic memory or a created mnemonic. For example, the word *run* may be linked to the concept for *run the risk of* through an image involving running, an episodic memory of hearing or reading the story, or via an alliterative link to *risk* (or through a combination of these). In either case, while an 'empty' expression lemma may be created, it takes further retrieval and repetition to facilitate the chunking up and correcting required to develop a fused phonological form.

11.4.3. ***The unitary concept***

As suggested in Section 8.4.3, a feature of the models of Levelt (1993) and Sprenger et al. (2006) as applied to the representation of formulaic expressions is that each lemma is linked to a unique concept. The idea is that when this concept is required (during speech production), it automatically triggers the appropriate lemma. To ensure that this triggering process converges to a single choice of lemma, Levelt introduces the notion of 'core meanings' for a lemma and principles of uniqueness and specificity associated with them (1993, p. 212). Core meanings are associated with conceptual conditions that must be satisfied for a particular lemma to be selected. The Uniqueness Condition says that each lemma has a unique set of core meanings. The Specificity Condition says that of all the items whose core conditions are satisfied during a search, the one with the most specific conditions is chosen.

The acquisition of a formulaic expression (as modelled in 11.4.1 and 11.4.2 above) implies that a new concept associated with the 'superlemma' is also created (with a unique set of core meanings). There may be a number of ways in which this could happen. For example, if the new formulaic expression is congruent with a similar expression in the L1 (i.e. has the same lexical words and meaning/function), it could simply co-opt the concept associated with the L1 expression (with an additional core meaning indicating that it was an L2 expression). In the studies, the English targets were not congruent with L1 Japanese expressions, but for some there were unitary expressions in Japanese with the same meaning. For example, *turned a blind eye to* has the formulaic equivalent in Japanese 見て見ぬふりをした (literally: 'see but pretend not to see') and *get the hang of* has コツを掴む (literally: 'grasp the trick/art'). Target sequences like these could utilise the same concept as for the Japanese version. Conklin and Carrol (2019) suggest that, in general, where there are formulaic expressions in both languages for a particular concept, the more overlap between the L1 and L2 forms there is, the greater the processing advantage for the L2 speaker. Their reasoning is that the concept automatically triggers the L1 words for the expression, and they in turn trigger the L2 translations of those words, helping to cue the desired L2 expression. Even if they do not share component words, similar ideas or images generated from the L1 (e.g. the idea of looking but not seeing) may help cue the L2 formulaic expression in a similar way. However, the effect of this in terms of the acquisition of an L2 expression as internally formulaic may be more complex. While shared form may help recall, it may also promote a focus on L2 component words rather than the whole unit and it may create interference via the L1 words that are not congruent. The target sequences in S5 and S6 which shared conceptual meanings with Japanese formulaic expressions were not among the expressions which became formulaic most frequently. So, the studies are inconclusive about the extent to which sharing a concept supports or hinders formulaic acquisition. However, this would be a useful area to explore further in order to develop the acquisitions models proposed in 11.4.1 and 11.4.2.

In the models, the concept for the newly acquired formulaic expression is linked to the L1 translation or to the context from which the meaning was given. In cases where there is a clear word or expression equivalent in the L1, Jiang (2000) suggests that, as well as the meaning, other (lemma) information may be derived from the L1 version, including the subject and objects (if they exist) and the type of situations. However, these may only be helpful if there is a close alignment between L1 and L2 forms. For the target sequences in studies S4-S6 this type of information is more likely to come from the example sentences that the targets were presented in. For many of the targets, the L1 translation was more descriptive, without an obvious unitary concept. So, in this case the unitary concept may need to be developed over time, as part of the L2 expression becoming formulaic for the speaker. In general, even if there is a clear unitary concept to begin with (e.g. from the L1), the nuance of meaning and usage will not be entirely equivalent. So, acquiring a formulaic expression over time will involve refining the conceptual meaning at the same time as developing the fluent phonological form.

11.4.4. ***Developing the concept for formulaic acquisition***

The discussion around establishing a clear conceptual representation for the formulaic expression has so far been applied to both routes to acquisition. However, the type of acquisition may affect how it is achieved, particularly with regard to the component words. For example, the dynamic repetition (DR) approach worked on the principle of establishing the fluent form first with a link to a basic meaning, then potentially refining the concept over time. On the other hand, semantic-formal elaboration (SFE) focussed on understanding the meaning and the components parts first before developing the fluency. To illustrate how each route might affect conceptual formation, let us consider the learning of the target *toyed with the idea of* as a new formulaic expression

The L1 translation and context provide the conceptual meaning as something along the lines of 'considered somehow'. The learner may therefore adopt this (or simply 'consider' or 'think about') as the conceptual meaning. However, in order to satisfy Levelt's requirement for convergence (1993, p. 200), at some point, that meaning will need to be refined (e.g. with additional core meanings) in order for the expression to be formulaic (have a holistic entry in the lexicon). In the experimental situation, that refinement could be something like 'the kind of considering a person might do in the situation of the story' or even 'the expression I needed to learn which has "toyed" in it').

11.4.4.1. Conceptual development in holistic acquisition

In the holistic acquisition situation, the presupposition is that the whole phonological form is learnt in one go. So, a simple meaning may be adopted as above, and elements of the lemma information may be taken from the lemma for *consider*, or gleaned from its usage in the story. Thus, *toyed-with-the-idea-of* may be treated as a single lexical entry and, like *consider,* would have the syntactic category of V, conceptual arguments for actor and theme, grammatical functions of Subject and Object and a pointer to the appropriate grammatical inflection for past tense (*toyed with the idea of)*. The processing for production of the message (*He toyed with the idea of telling his landlord*) in the re-telling of the story, would therefore be similar to how it would be if a single verb like *consider* was involved, with the concept triggering the lemma *toy-with-the-idea-of* and the past tense form retrieved for phonological encoding. The ease with which the speaker could do this would depend on how strong the link was between concept and lemma. If it was fairly weak, even if the holistic phonological form was established, the speaker might fail to recall it or might show tip-of-the-tongue behaviour (see Section 11.3.4).

11.4.4.2. Conceptual development in fusion

In the case of fusion, the expression is not initially memorised as a single unit. There may be a number of components, either each word separately, or a combination of words and short phrases. For example, a combination observed in the study data (based on fluency and accuracy of responses) was a breakdown into *toyed (with)* and *the idea of* (with some participants omitting *with*). In this case, reconstructing the expression involves simply concatenating the parts. At the initial stage, the production may involve lemmas for (some of) the component parts (e.g. *toyed* and *the idea of*), the meanings of which support their memorisation. The trigger idea (concept) for the overall expression needs to link to these somehow. For example, the concept may include (or be linked to) an image which shows someone thinking hard or playing with an object representing an idea. A word (e.g. 'toyed') may be created as a label for the whole expression and the two constituent parts linked to it in some way. However it is done, at the point of formulaic acquisition, a new lemma must be created and linked to the concept. This will result in a situation similar to the holistic acquisition described above. However, the nature of the acquisition may have an effect on the final representation of the formulaic expression. For example, it seems likely that links to the associative component lemmas (and their meanings) would continue to contribute to the cognitive representation, and the unique concept and core meanings for the superlemma could reflect the transition stage (e.g. feature particular labels or images associated with the components), making it slightly different from how it would be if learnt holistically immediately.

In practice it will be difficult to establish whether an expression has been holistically acquired or simply acquired swiftly and effectively via fusion. In fact, there was evidence from study S5, that some expressions deemed formulaic at stage W0 (and therefore considered cases of 'holistic acquisition' in the study), didn't meet the threshold for formulaicity during the subsequent stages. It is also worth recalling that speakers could learn expressions formulaically (e.g. by holistic acquisition) but in an incorrect form. For example, there were consistently fluent uses of *toyed the idea of*, *turned blind eye* and *breathed sigh of relief.* These could be considered formulaic as they were delivered fluently and still represented (to the speaker, at least) the concept of the target. In some cases, subsequent uses helped the speakers refine and correct the form (while retaining fluency), mirroring the process described by Bardovi-Harlig (2019). It is feasible that, even though the expression is formulaic, the process of correcting or fine-tuning its form could induce occasions of dysfluency if the correction is being attended to consciously.

11.4.5. ***Application to general acquisition***

11.4.5.1. Constructing formulaic expressions naturally

In the experimental situation of studies S4-S6, particular target sequences were specifically memorised and, although additional retrieval and correction took place, the emphasis was not on activities to move the expressions to formulaicity after initial learning. In a more general situation, L2 speakers may acquire formulaic expressions through fusion as a result of their continued use of a particularly useful initially constructed expression. For example, an expression used by Kensuke in the case study was *our rival companies* which may have been initially constructed. At the construction stage of learning, each constituent word is a lemma retrieved during the construction, with its own set of defining core concepts, along with its own conceptual categories and functions. For this particular phrase to have been chosen, its main semantic components (i.e. 'our', 'rival', 'companies') would need to have been a part of the message. When later the expression becomes internally formulaic for Kensuke, a new lemma is created with its own set of conceptual meanings, categories and specifiers, and a new and unique set of core concepts (to differentiate it from any other lemmas). For example, the lemma for *our-rival-companies* would have syntactic category NP with specifiers indicating its plurality (Levelt, 1993, p. 196). The overall meaning may borrow heavily from the head of the syntactic phrase it evolved from (e.g. 'our rival companies' is a particular set of companies related to Kensuke's own company) and the core concepts would need to include the additional specification (e.g. they are rivals in the same industry). For that evolution to come about might require a threshold number of co-occurrences or it might be driven by re-conceptualisation (e.g. some example companies now forming a more abstract general set).

According to Levelt's model (Levelt, 1993), the new lemma would need to match the conceptual meaning of the message fragment in order to be selected for production. An interesting question, then, is whether a message such as *it was really hard to borrow cargo from our rival companies* would differ subtly to accommodate the more nuanced or abstract meaning of the formulaic expression compared to the constructed version (e.g. by de-emphasising the particular company features and emphasising the 'rival' aspect).

11.4.5.2. Developing nuanced conceptual meanings

The research by Sprenger et al. (2006) discussed in detail in Chapter 8 suggests that formulaic expressions retain links to their component parts and their meanings. This is accommodated in the models proposed earlier, particularly in the case of fusion. The process of becoming formulaic can be thought to represent an increasing prominence of the whole relative to the constituent parts. While this prominence is clear in the case of idioms and non-canonical sequences, it can also be demonstrated in formulaic expressions that are compositional. For example, Nordquist (2004) presented participants in a study with single words and asked them to create three sentences. In the case of the word 'I', participants tended not to create common sentence starters such as *I think*, *I guess* … and *I suppose* even though corpus data indicates that these are the highest frequency uses of I + verb type expressions. This demonstrates how the familiarity of the expressions as whole units overshadows their association with their component words (even though these are obviously accessible if required).

This prominence of the whole can also be seen in the case of multi-morphemic words where the constituent parts only contribute partially (e.g. *singer, baker*) or little (*really*) to the overall meaning. For example, Taylor (2012, p. 131) notes that *singer* is not just one who sings but one who sings well; *baker* is not just one who bakes, but a retailer who sells baked products. Taylor suggests that such adapted meanings are often the result of drift in a language, whereby the meaning of the compounded word moves away from its morphemic composition over time.

The idea of drift can also be applied to formulaic expressions more generally (Wray, 2009; Wray & Fitzpatrick, 2009) where the whole expression gains a more nuanced meaning or a clear functional purpose. For example, *I think/I guess/I suppose* are standard ways of introducing an opinion or thought (with varying degrees of confidence). Other examples from Taylor (2012) are *I don't know,* which may be used as a hedge, and *Would you believe it?* which usually serves as an expression of surprise rather than its literal meaning as a yes-no question. While the idea of drift is associated with changes in the language as a whole over time, there may be an analogous process occurring in the minds of individual learners as their use and experience with an expression becomes more familiar.

CHAPTER 12: Concluding thoughts

Insights, limitations and implications

This research in this thesis has sought to explore how L2 speakers acquire new internal formulaic expressions, with the aim of better understanding the stages that may be involved and the psycholinguistic processes that might underlie those stages. Drawing on the findings from the research literature and the empirical studies undertaken, a variety of suggestions and ideas regarding the acquisition process was discussed in Chapter 11. In particular, two broad routes to acquiring target sequences and the possible processes involved for each were modelled and discussed. This final chapter provides the opportunity to reflect on the insights and limitations of the approach, and to consider pedagogic and other implications of the results and acquisition models suggested. Ideas for future research are also highlighted as they arise through the chapter.

12.1. Methodological insights

The six studies that comprised the empirical part of this thesis used a variety of methods while offering a coherent development of the thinking through the research. They provided the opportunity both to replicate and assess the methodology of previous studies and to develop some new approaches. For example, study S1 applied the hierarchical approach for identifying formulaic expressions proposed by Myles and Cordier (2017) and applied it to sample text gathered in a similar way (by structured interviews and story-telling based on pictures). This demonstrated that the approach can be applied effectively for different L1 and L2 languages, and supported the original finding that formulaic material measured in this way is more prevalent in L2 speakers than found in previous research using different criteria. It also highlighted where enhancements could be made (e.g. in dealing with inconsistent fluency). Study S3, the replication of Fitzpatrick and Wray (2006), adopted the same method as the original for introducing and monitoring target utterances by L2 speakers. However, certain aspects needed to be changed in order to cater for the different circumstances in which the study took place (i.e. the ‘real performance’ task needed to be a conversation with a native speaker rather than an everyday transaction). Broadly similar results were achieved, confirming the importance of individual differences and the types of deviation typical of such targeted memorisation of longer utterances. The replication was also extended to look at the fluency behaviour of the participants, giving further insight into the segmentation of longer utterances and the identification of formulaic expressions in speech. Adaptation was also evident in the case study (S2) which, as well as applying a repeated interview approach to gather longitudinal data, used a method for identifying potential formulaic expressions which combined elements of Myles and Cordier's approach with the corpus-based algorithmic approach of Brooke et al. (2014). This adaptation was necessary for the specific needs of the study (checking the fluency behaviour of repeated expressions over time). The study also developed an analytic approach for testing the direction of fluency change, based on a 'weighted pivot' metaphor.

For the later studies, the main experimental approach was to present novel target sequences to L2 speakers in different ways and observe subsequent production of those expressions using a variety of assessment tasks. A number of methodological elements were tested in study S4 and then adapted and refined for the subsequent studies. These included the selection of a set of targets and the embedding of them into stories. This created a context for the participants to understand the targets and a natural means for prompting their production at the assessments (i.e. retelling the story).

Studies S5 and S6 also introduced some innovative ways of presenting targets for memorisation, and for assessing the holistic and automatic aspects of formulaicity. The dramatic repetition (DR) approach to introducing target sequences for memorisation was designed on the basis of previous research concerning the effect of repetition and other learning factors that encourage fluency. It provided a good opportunity to test this direct phonological approach against the more traditional semantic-formal elaboration (SFE) in a controlled way. Study S6 was an experimental follow-up to study S5, providing a way of assessing the formulaicity of the same targets which had been learnt a few months earlier. This final study also introduced a further means for assessing formulaicity, the holistic automaticity (HA) test, which adopted a specific priming paradigm based on the theoretical approaches of Sprenger et al. (2006) on holisticity and Segalowitz (2010) on automaticity. The study demonstrates how the use of phonological coherence (fluency) in the identification of internal formulaicity in speech output can be augmented by psycholinguistic methods.

12.2. Limitations

While there were a number of innovative approaches there were also some important limitations in studying the acquisition of L2 formulaic expressions in this way. Firstly, as acknowledged in Chapter 1, the particular scope of the research was limited in terms of the participants and the type of expressions studied. While this was a feature of the research rather than a bug, it does inevitably restrict how much the results can be generalised. In particular, all participants were Japanese, at an intermediate or advanced level of English, and based in Japan. In the model utterance study (S3), there were findings for this group that differed from the original which had featured participants (Japanese and Chinese) who were based in the UK (Wray & Fitzpatrick, 2008). For example, unlike in the original, the S3 participants tended not to deliberately change the targets in their performances and they performed better in the real performance than the practice. There may, therefore, be aspects of the findings in the other studies that are specific to the participant group. For instance, participants in S5 all appeared to embrace the DR approach enthusiastically, perhaps because of not usually having the chance to repeat in this way. For other types of participant, however, there may be a different effect.

More generally, the relationship between the targeted expressions and learner understanding is likely to vary across different L1s due differences in congruence or metaphorical usage. There also may be general traits in terms of pronunciation ease or difficulty, in attitudes towards accuracy or fluency, or with exposure to idiomatic formulaic expressions via native speakers. A similar point may be made regarding the set of targets used in studies S4-S6, which were of a very particular form (verb phrases). It may well be that certain other types of formulaic expression, such as exclamations (*I don't believe it*), fixed conventional expressions (e.g. *How's it going?*) or collocations (e.g. *absolute necessity*) are easier (or harder) to render formulaic via holistic acquisition. This would be a useful area for further study.

Another possible limitation relates to the wide variety of factors that could potentially influence the way a particular expression became formulaic for an individual. As suggested by the research reviewed in Chapter 2 and the first empirical studies undertaken here, these factors are likely to include differences between individuals (for example in their ability or inclination to acquire formulaic expressions), differences between the targets regarding how easy they were to acquire, and differences in the way in which the targets were presented for memorisation. The main focus of the later studies was on the exploring different types of presentation while controlling for the other factors. Control was provided through the choice of targets (similar length, form and degree of novelty) and the design of the procedure (using an orthogonal design that ensures that condition and order are shared equally across participants and targets).

Despite the controlled choice of participants and targets in studies S5 and S6, however, there was still considerable variation in performance across both. Rather than any systematic trends for particular features of the targets (e.g. length) or participant (e.g. proficiency level), it seems likely that the variation was due to a complex interaction between such features (and, no doubt, other factors), related in part to the speaker's particular experience with the words and sub-sequences within each expression. The studies were not designed to investigate these interactions, even though they are likely to influence the effect of the input conditions (e.g. DR and SFE) and the routes to acquisition. So, further research which manipulates known or unknown component words and sub-sequences within target sequences when being learnt by L2 speakers would be necessary to develop a more detailed picture of acquisition. It would also be important to investigate how other variables (such as length, prosodic features, imageability and L1 congruence) affect these routes.

The orthogonal design of the studies (S4-S6) did appear to be effective in balancing any potential confounding effects across conditions. However, this kind of design also brings some disadvantages, particularly when the numbers of participants and target sequences is relatively small. The main limitation was in reducing the power of the studies to show effects. For example, the Greco-Latin square used in study S4 (which had two sets of conditions) resulted in a small number of items in each cell. On top of this, the general variation across participants and targets, while balanced, nevertheless resulted in elevated standard deviations within cells. A single condition was therefore adopted for studies S5 and S6, along with more participants. While this raised the numbers of items per condition, the numbers of targets satisfying the conditions for formulaicity was quite low, and there remained considerable variation across participants and expressions. Because of this, the possibility of observing statistical significance was still reduced to some extent, even though trends appeared clear.

Collecting, transcribing and analysing speech data is a time-consuming process for a single researcher and this, along with other issues such as recruiting and scheduling volunteers for multiple sessions, placed practical restrictions on the numbers. Although the criteria for transcribing and analysing the data were clearly defined and did not rely on intuitive diagnostic assessments (except for assessing the unitary condition in the first study), and the data was checked and reviewed many times, there would still have been value in having an additional person involved in the process both for cross-validation and potentially to increase the numbers of participants involved. That said, the limited numbers and the design of the data collection allowed for every participant-target item to be examined in detail at each stage of each study, resulting in a rich set of qualitative data.

A final potential limitation of the research in general lies in the challenges associated with trying to determine psycholinguistic features occurring in the mind of the speaker (such as holistic storage and automatic processing) by using observable features of speech. As acknowledged in the discussion in Chapter 11, fluency in the delivery of a unitary expression may not be enough to guarantee that an expression is formulaic for the speaker. This possibility was raised in the literature (Chapter 8) and supported in the final study, S6, where the holistic automaticity check was used as an additional indicator of formulaicity. The research in this thesis may therefore have strayed into the territory of making claims about formulaicity in L2 speakers on the basis of an identification method which is itself still being explored and developed. In response, it is important to note that conclusions drawn from study S6 and the models suggested in Chapter 11 are based on formulaicity identified via the more stringent measure (i.e. fluency and HA combined). This marked a development from the first of the studies, in which formulaicity was identified purely in relation to fluency.

It should be noted that results for the percentage of formulaic syllables (FS%) in S1 (and in Cordier, 2013), and for the percentages of formulaic expressions acquired in S5, were developed in relation to (consistent) fluency. Percentages in these studies may therefore be inflated compared to what a more stringent measure for formulaicity would have shown. While the nature of the results in study S5 does not suggest that the relative difference between conditions such as DR and SFE would change, it would nevertheless be useful for future studies to verify the findings using the enhanced measures. Before doing this, it would also be important to undertake further studies to evaluate and the refine the HA test itself. This kind of study could involve using individual native speakers for whom it may be easier to compare 'known' formulaic expressions with novel constructed ones.

There may also be other ways to measure the automatic element of formulaicity. For example, automaticity may also apply to how quickly the full expression is recalled. Although this was touched on in the research (and discussed in relation to phenomena such as self-cueing), it was not measured in a systematic way. However, such an analysis could be usefully incorporated in future studies. The research also did not focus in detail on the appropriacy of the usage of target sequences (except in the context of the original story). Indeed it is recognised that the fluent, automatic delivery of a target sequence by a speaker on a series of specific tasks does not necessarily represent the full picture of acquisition, since further restructuring of the sequence is likely to occur as experience of its usage is gained and new knowledge is integrated within the current mental system (McLaughlin, 1990). However, the type of formulaicity explored in the studies may be thought of as a useful 'building block' in the development of deeper knowledge of the formulaic expression. The usage test applied in S5 (which suggested that acquired target sequences were easily adapted for the new situations presented in the test) could also be further developed as a useful means of exploring this continuing acquisition process.

12.3. General implications and future research

12.3.1. *Pedagogic implications*

As Myles (2004, p. 156) notes, the use of formulaic expressions by L2 speakers at an early stage of development is effective (and popular) because it allows them to produce more sophisticated language than they normally would. The current research shows that L2 speakers at an intermediate or advanced level can also be motivated to learn formulaic expressions and use them fluently in speech. While the studies presented were not designed with a specific focus on pedagogy, they do provide insights which may enhance the acquisition process. For example, the results suggest that the strategic memorising of given expressions can be a legitimate and effective way for L2 speakers at this level to acquire formulaic expressions. Study S3 showed that given clear auditory models, the Japanese L2 speakers could learn and use relevant and useful expressions effectively in real conversations with native speakers. This learning equipped them for expressing things they wanted to say in a clear and generally nativelike fashion. They also expressed satisfaction about doing so and a good sense of achievement. The process revealed that it is not always necessary to analyse expressions in detail or understand why something works the way it does in order to use it effectively. It may be necessary to segment expressions for the purpose of memorising them but this can be done on a pragmatic basis.

The study also highlighted the importance of selecting relevant expressions, providing supportive practice and feedback and, crucially, giving an opportunity to use the expressions and reflect on their use. As well as the pragmatic and motivational benefits of an approach like this, there is a potential generative effect. Although not explored in the research, the idea of acquiring and embedding specific exemplars first and later abstracting them (e.g. changing parts of the exemplar to apply to a different situation) has been suggested as a potential learning method by different researchers (e.g. Lewis, 1997; Woolard, 2013).

The DR approach (in S5) demonstrated that, even with a relatively short period of learning (3 minutes per expression), an initial focus on repeating the phonological form (as if in a radio play narrating a story) can help establish holistic storage of sequences early on and support fluency and formulaicity of output over time. In the DR approach, meaning was given by the context of the story and a simple Japanese translation. This seemed sufficient to establish an initial understanding of the meaning and usage: all participants fed back that they were confident they understood the expressions. Performance on the usage test in study S5 was good, and comparable with that for expressions learnt by SFE. In addition, the focus on repetition did not appear to impact negatively on recall or accuracy. Results also suggested that initial errors can be fine-tuned with practice and corrective feedback (as given by the regular assessments). This contradicts the suggestion from some research (e.g. Stengers & Boers, 2015) that initial errors in learning formulaic expressions are difficult to change. The difference may be due to the greater focus in the current studies on the form of the whole and on reproducing it exactly as given. Although the DR approach was not designed to assess a teaching method, the idea of targeting the complete phonological form of a new expression through meaningful repetition at the initial stage seems a simple and useful means for encouraging formulaicity, so long as context and an example usage are provided.

The models for formulaic acquisition proposed in Chapter 11 highlight the importance of establishing a unitary concept for the expression (with a unique set of core meanings). In practice, the full nuance of the meaning and usage may only develop through exposure to multiple examples and via opportunities to practise using it. However, the idea that formulaic expressions provide a particular meaning or function may be important for learners to be aware of. While research into the effect of raising awareness of formulaicity expressions in general (e.g. Boers et al., 2006; Lewis, 1997; Stengers, Boers, Housen, & Eyckmans, 2010) is mixed, for a given target sequence, specifically highlighting how it has its own unique meaning or function could be valuable. For example, it could help learners decide whether (or when) it is worth learning an expression such as *toyed with the idea of* instead of just using *thought about*.

The studies also support the idea of regular (spaced) retrieval and simple corrective feedback as a way of consolidating recall and formulaicity of the learnt sequences. As well as improving recall of the expressions, as predicted by the work of Kornell (Kornell et al., 2015; Kornell & Vaughn, 2016), retrieval also helped develop the fluency and accuracy of the expressions and therefore helped to promote or retain formulaicity in them. As shown in the studies, this retrieval and corrective feedback does not need to be very elaborate to be effective. In the on-going assessments, participants just retrieved the expressions by re-telling the story and via the L1 translation cue, then listened to and repeated a correct version. Encouraging such review on regular (but increasingly spaced) occasions would seem to be important for long-term acquisition. While such an approach is undoubtedly a feature of many beginner level classrooms already, the research here suggests it applies equally to intermediate and advanced L2 speakers trying to acquire new formulaic expressions.

12.3.2. ***Other implications***

The processes described for holistic acquisition and fusion in Section 11.4 show possible ways that formulaic acquisition may occur. A particular implication is that, in the case of fusion, the meanings of component words and sequences are accessed in the development stage in order to construct the expression, while this is not necessary for holistic acquisition. Fusion is therefore likely to be more susceptible to interference based on the speaker's existing knowledge of the component words or sub-sequences. Examples of this from the studies include cases where a component word is already collocated with another word (e.g. '*take the risk of'* for '*run the risk of'*) or when synonyms replace component words (e.g. '*set his target on'* for '*set his sights on'*). Holistic acquisition, on the other hand, de-emphasises the meanings of the component words making such interference less likely. However, although holistic acquisition has advantages, the research suggests that fusion is a common means for L2 speakers to acquire formulaic expressions. This can arise because a constructed phrase is useful to the speaker (e.g. *I am very surprised*; *our rival companies*) or because an expression could not (due to memory constraints) or did not (e.g. due to the learning method) get learnt holistically initially. There were frequent examples of this in the studies, even for expressions that were not easily constructed on the basis of component word meanings (e.g. *get the hang of*). In these cases, it would be interesting to investigate the extent to which the speaker incorporates meanings for the parts (e.g. *get the* and *hang of*) with the overall meaning of the whole expression.

The research also provides some indication how measures of accuracy and ease of recall are relevant to the acquisition of formulaic expressions. Ease of recall, as an indicator of the accessibility of the whole sequence given a cue, was shown to be an important element in the tracking of formulaic acquisition. It appears to work alongside fluency in contributing to the holistic or stable automaticity of an expression. Analysing the accuracy of production can provide clues to how an expression has been memorised, and to how the formulaicity of an expression may be developing. For example, certain inaccuracies (predominantly lexical or phrasal) were taken as indicators of reconstruction during fusion, while others (frequently morphological) were thought to suggest as yet uncorrected holistic segments. Thus, when monitoring the acquisition of formulaic expressions, recall and accuracy along with fluency and other indicators such as holistic or stable automaticity can be used together to give a picture of how the expression is being stored and processed by the speaker.

12.4. To sum up…

Based on the premise that helping L2 speakers increase their stock of usable formulaic expressions is a worthwhile endeavour, this thesis has sought to better understand how such expressions can be acquired using different types of memorisation and retrieval. The research has developed insights into the possible acquisition processes for L2 speakers and explored a variety of experimental methods along the way. While not the main focus, the studies have also demonstrated some potentially useful approaches for learning expressions and promoting fluency and formulaicity in L2 speech.

This latter aspect is particularly important because part of my motivation for researching formulaicity derives from the experiences of being both a learner of L2 Japanese and a teacher of English to adult learners in Japan. From this perspective, a notable element of the research was the evident satisfaction that participants took in undertaking the studies. Part of the motivation for that response seems to have been the sanctioning of a learning strategy whereby whole phonological forms could be memorised without the need to attend to how or why component words and structure contribute to the whole meaning. Insofar as my participants were representative of L2 learners more generally, there is a useful message here. For learners schooled in (or inclined towards) an analytical approach, holistic memorisation can be a liberating experience. Notwithstanding the necessary complexity of the research itself, a simple observation can be made: just repeating an expression out loud without analysis (but allied to a given context) can be an effective and legitimate strategy for learning formulaic expressions.

References

Aijmer, K. (1996). *Conversational routines in spoken discourse*. London, UK: Longman.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics, 29*(2), 175-193. doi:10.1017/S0142716408080089

Anderson, J. R. (1992). Automaticity and the ACT-asterisk theory. *American Journal of Psychology, 105*(2), 165-180. doi:10.2307/1423026

Anderson, J. R. (1996). ACT - A simple theory of complex cognition. *American Psychologist, 51*(4), 355-365. doi:10.1037/0003-066x.51.4.355

Anthony, L. (2018). AntConc (Version 3.5.7). Tokyo, Japan: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon, 9*(3), 377-400.

Ashby, M. (2006). Prosody and idioms in English. *Journal of Pragmatics, 38*(10), 1580-1597. doi:10.1016/j.pragma.2005.03.018

Au, C., & Entwhistle, N. (1999). *'Memorisation with understanding' in approaches to studying: cultural variant or response to assessment demands?* Paper presented at the European Association for Research on Learning and Instruction Conference, Gothenburg, Sweden.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. East Sussex: Psychology Press.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208. doi:10.1016/S0021-9924(03)00019-4

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4

Badecker, W. (2001). Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes, 16*(4), 337-366. doi:10.1080/01690960042000120

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*(3), 241-248. doi:10.1111/j.1467-9280.2008.02075.x

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*(2), 323-363.

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research, 22*(4), 487-497. doi:10.1191/0267658306sr276oa

Bardovi-Harlig, K. (2019). Formulaic language in second language pragmatics research. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 97-114). Oxford, UK: Routledge.

Barth-Weingarten, D. (2011). The fuzziness of intonation units: Some theoretical considerations and a practical solution. *Interaction and Linguistic Structures, 51*.

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., . . . Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*, 1-26.

Beebe, L. M. (1983). Risk-taking and the language learner. In H. W. Seliger & M. L. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 39-66). Rowley, Mass: Newbury House.
Bell, N., & Skalicky, S. (2019). Humor and formulaic language in seond language acquisition. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 115-131). Oxford, UK: Routledge.
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Grammar of spoken and written English. In *Edinburgh: Pearson Education Limited*.
Birjandi, P., & Ahangari, S. (2008). Effects of task repetition on the fluency, complexity and accuracy of Iranian EFL learners' oral discourse. *The Asian EFL Journal, 10*(3), 28-52.
Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*(1), 54-74.
Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261. doi:10.1191/1362168806lr195oa
Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics? *Language Teaching Research, 11*(1), 43-62.
Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System, 33*(2), 225-238. doi:10.1016/j.system.2004.12.007
Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83-110. doi:10.1017/s0267190512000050
Boers, F., Lindstromberg, S., & Eyckmans, J. (2012). Are alliterative word combinations comparatively easy to remember for adult learners? *RELC Journal, 43*(1), 127-135. doi:10.1177/0033688212439997
Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vial-Vigo International Journal of Applied Linguistics, 11*, 41-62.
Brazil, D. (1997). *The communicative value of intonation in English*: Cambridge University Press.
Brooke, J., Hammond, A., Jacob, D., Tsang, V., Hirst, G., & Shein, F. (2015). Building a lexicon of formulaic language for language learners. *Proceedings of NAACL-HLT, 2015*, 96-–104.
Brooke, J., Tsang, V., Hirst, G., & Shein, F. (2014). Unsupervised multiword segmentation of large corpora using prediction-driven decomposition of n-grams. *Proceedings of The 25th International Conference on Computational Linguistics, COLING 2014*.
Bygate, M. (2001). Effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks*. London: Routledge.
Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*(6), 668-683.
Chafe, W. (1987). Cognitive constraints on information flow. *Coherence and grounding in discourse, 11*, 21-51.
Chen, X. (2009). *The role of formulaic sequence in speech fluency development in ESL*. Paper presented at the Computational Intelligence and Software Engineering International Conference, Wuhan, China.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of experimental psychology. Learning, memory, and cognition, 31*(6), 1235-1249. doi:10.1037/0278-7393.31.6.1235

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.

Conklin, K., & Carrol, G. (2019). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic languages: A second language acquisition perspective*. New York: Routledge.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers. *Applied Linguistics, 29*(1), 72-89. doi:10.1093/applin/amm022

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32,* 45-61.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 94*(4), 276-287. doi:10.1016/0013-4694(95)98479-R

Cordier, C. (2013). *The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study (Unpublished doctoral thesis).* (PhD). Newcastle University, Newcastle, UK.

Cornwell, S., Simon-Maeda, A., & Churchill, E. (2007). Selected research on second-language teaching and acquisition published in Japan in the years 2000–2006. *Language Teaching, 40*, 119–134.

Corrigan, R. L., Moravcsik, E. A., Ouali, H., & Wheatley, K. M. (2009). *Formulaic language. Typological studies in language*. Amsterdam: John Benjamins.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: CUP

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science, 19*(1), 51-57.

Craik, F. I. M. (2002). Levels of processing: Past, present and future? *Memory, 10*(5/6), 305-318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.

Cutler, S. F. (2013). *Is formulaicity a factor in fluency? An investigation into the relationship between fluency and use of formulaic sequences in adult Japanese Speakers of English (Unpublished masters dissertation)*. (MA). School of Languages and Area Studies. University of Portsmouth. Portsmouth, UK.

Cutler, S. F. (2017). The use of psycholinguistic formulaic language in the speech of higher level Japanese speakers of English. *Vocabulary Learning and Instruction, 6*(1), 1-13. doi:10.7820/vli.v06.1.Cutler

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & cognition, 25*(1), 57-71.

Cyr, A.-A., & Andersen, N. D. (2014). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 841-850.

Dabrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics, 16*(3), 437-474.

Dahlin, B., & Watkins, D. (2000). The role of repetition in the processes of memorising and understanding: A comparison of the views of German and

Chinese secondary school students in Hong Kong. *Educational Psychology, 70*(1), 65-84. doi:10.1348/000709900157976
Dahlmann, I. (2009). *Towards a multi-word unit inventory of spoken discourse (Unpublished doctoral thesis).* (PhD). University of Nottingham, Nottingham, UK.
Dahlmann, I., & Adolphs, S. (2007). *Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)?* Paper presented at the Workshop on a broader perspective on multiword expressions, Prague.
Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge, UK: Cambridge University Press.
Davies, M. (2018). iWeb: The 14 billion word web corpus. In (May 2018 ed.).
De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL Reading. *Studies in Second Language Acquisition, 19*(3), 309-329. doi:undefined
DeKeyser, R. (1994). How implicit can adult second language learning be. *Consciousness & Second Language Learning: Perspectives on Form-Focused Instruction. AILA Review, 11*, 83-96.
DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-151). Cambridge: Cambridge University Press.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321. doi:10.1037/0033-295X.93.3.283
Dell, G. S., Chang, F., & Griffin, Z. M. (2010). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23*(4), 517-542. doi:10.1207/s15516709cog2304_6
DGA. (2014). Transcriber AG (Version 2.0.0). Retrieved from http://transag.sourceforge.net/
Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System, 35*(2), 271-280. doi:10.1016/j.system.2006.12.005
Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163-188. doi:10.1177/0267658309349431
Ecke, P., & Hall, C. J. (2013). Tracking tip-of-the-tongue states in a multilingual speaker: Evidence of attrition or instability in lexical systems? *International Journal of Bilingualism, 17*(6), 734-751.
Eggins, S. (2004). *Introduction to systemic functional linguistics*. London, UK: Continuum.
Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency*. Ann Arbor, MI: University of Michigan Press.
Elgort, I., Candry, S., Boutorwick, T. J., Eyckmans, J., & Brysbaert, M. (2016). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics, 37*, 1-23. doi:10.1093/applin/amw029
Ellis, N. C. (1995). Vocabulary acquisition: Psychological perspectives and pedagogical implications. *The Language Teacher, 19*, 12-18.
Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*(1), 91-126. doi:10.1017/S0272263100014698
Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194. doi:10.1093/applin/aml015

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17-44.
Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat & p. 3-23). (Eds.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins.
Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396. doi:10.2307/40264474
Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: putting language in good order. *Quarterly Journal of Experimental Psychology, 49A*(1), 234-250.
Ely, C. M. (1986). An analysis of discomfort, risk-taking, sociability, and motivation in the L2 classroom. *Language Learning, 36*(1-25).
Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics, 12*(1), 25–53
Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.
ETS. (2019). TOEIC Listening and Reading Test. Retrieved from https://www.etsglobal.org/fr/en/test-type-family/toeic-listening-and-reading-test
Eyckmans, J., Boers, F., & Lindstromberg, S. (2016). The impact of imposing processing strategies on L2 learners deliberate study of lexical phrases. *System, 56*, 127-139.
Eyckmans, J., & Lindstromberg, S. (2017). The power of sound in L2 vocabulary learning. *Language Teaching Research, 21*(3), 341–361.
Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & cognition, 11*(6), 565-574. doi:10.3758/BF03198281
Finnegan, R. (2017). *Oral poetry: Its nature, significance and social contex*. Eugene, OR: Wipf and Stock.
Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing, 27*(4), 537-554.
Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *VIAL, 1*, 55-74.
Fitzpatrick, T., & Wray, A. (2006). Breaking up is not so hard to do: individual differences in L2 memorisation. *Canadian Modern Language Review, 63*(1), 35–35.
Fletcher, M., & Munns, R. (2005). *Storyboard: 24 Stories through pictures - A photocopiable resource for busy teachers*. London, UK: Scholastic.
Fletcher, W. (2011). Phrases in English website. Retrieved from http://phrasesinenglish.org/ on 14 January 2016.
Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition, 6*(1), 35-53.
Forsberg, F. (2009). Formulaic sequences: A distinctive feature at the advanced / very advanced levels of second language acquisition. In E. In Labeau & F. Myles (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 M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, Testing*. London, NY: Longman.
Garner, J., & Crossley, S. (2018). A latent curve model approach to studying L2 n-gram development. *The Modern Language Journal, 102*(3), 494-511. doi:10.1111/modl.12494

Gaskell, G. M., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45*(2), 220-266. doi:10.1016/S0010-0285(02)00003-8

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411-458. doi:10.1016/0010-0285(83)90014-2

Gent, B., & Muhammad, A. (2019). Memorising and reciting a text without understanding its meaning: A multi-faceted consideration of this practice with particular reference to the Qur'an. *Religions, 10*(7), 425-439.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalisation in language*. Oxford: OUP.

Goldman-Eisler, F. (1964). Hesitation, information, and levels of speech production. *Disorders of language*, 96-111.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15*(2), 103-113. doi:10.1177/002383097201500201

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. *Phraseology: Theory, analysis, and applications, 145*, 160.

Granger, S. (2019). Formulaic sequences in learner corpora: Collocations and lexical bundles. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), (pp. 228-247). Oxford, UK: Routledge.

Granger, S., & Meunier, F. (2008). *Phraseology: An interdisciplinary perspective*: John Benjamins Publishing.

Grant, L. (2005). *Frequency of `core idioms' in the British National Corpus (BNC) International Journal of Corpus Linguistics*, 10(4), 429-451. doi:10.1075/ijcl.10.4.03gra

Greenberg, S., Carvey, H., Hitchcock, L., & Shuangyu Chang, S. (2003). Temporal properties of spontaneous speech: a syllable-centric perspective. *Journal of Phonetics, 31*, 465-485.

Gries, S. T., & Mukherjee, J. (2010). Lexical gravity across varieties of English: An ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics, 15*(4), 520-548.

Gruber, M. C. (2009). Accepting responsibility at defendants' sentence hearings: No formulas for success. . In R. Corrigan, E. A. Moravcsik, & K. M. Wheatley (Eds.), *Formulaic language: Acquisition, loss, psychological reality, and functional explanations* (Vol. 2, pp. 545–566). Amsterdam: John Benjamins.

Hamilton, M., & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language, 44*(1), 96-117. doi:10.1006/jmla.2000.2749

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language, 20*, 27-41.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484.

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London: Routledge.

Horwitz, E. K. (1987). Surveying student beliefs about language learning. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice Hall.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473. doi:10.1093/applin/amp048

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44. doi:10.1093/applin/19.1.24

Hsu, J. Y. (2010). The effects of collocation instruction on the reading comprehension and vocabulary learning of Taiwanese college English majors. *Asian EFL journal, 12*(1), 47-87.
Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97-112.
Isbilen, E. S., & Christiansen, M. H. (2018). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*, 1–14. doi:10.1111/tops.12376
Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E.-J. Van der Linden, A. Schenk, & S. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133-165). Hillsdale, NJ: Erlbaum.
Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77. doi:10.1093/applin/21.1.47
Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal, 91*(3), 433-445. doi:10.1111/j.1540-4781.2007.00589.x
Kanazawa, Y. (2016). Micro-level emotion as a factor of L2 vocabulary memory: The effect of lexical emotional valence on incidental recall performance. *Language education & technology*(53), 23-52.
Kapatsinski, V., & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. *Formulaic language, 2*, 499-520.
Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current directions in psychological science, 21*(3), 157-163. doi:10.1177/0963721412443552
Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151-162.
Kasahara, K. (2011). The effect of known-and-unknown word combinations on intentional vocabulary learning. *System, 39*(4), 491-499. doi:10.1016/j.system.2011.10.001
Kempen, G., & Harbusch, K. (2002, 2002). *Performance grammar: A declarative definition*.
Kersten, S. (2010). *The mental lexicon and vocabulary learning: Implications for the foreign language classroom*. Tubingen, Germany: Narr Verlag.
Kesckes, I. (2019). Formulaic lanuiage and its place in intercultural pragmatics. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 132-150). Oxford, UK: Routledge.
Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly, 46*(4), 831-841. doi:10.1002/tesq.66
Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology, 58*(1), 68-101. doi:10.1016/j.cogpsych.2008.05.002
Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32(2)*(145–164.).
Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming,. *Applied Cognitive Psychology, 23*, 1297-1317.
Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 283–294.
Kornell, N., & Vaughn, K. E. (2016). How retrieval attempts affect learning: A review and synthesis. *Psychology of Learning and Motivation, 65*, 1-32.
Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kuiper, K., van Egmond, M.-E., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon, 2*(3), 313-357. doi:10.1075/ml.2.3.03kui
Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757-786. doi:10.1002/tesq.194
Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716. doi:10.1093/applin/amn018
Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.
Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.
Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency*. Ann Arbor, MI: The University of Michigan Press.
Levelt, W. J. M. (1993). *Speaking: From intention to articulation* (Vol. 1): MIT press.
Lewis, M. (1993). *The lexical approach*. Hove, UK: Language teaching publications.
Lewis, M. (1997). *Implementing the lexical approach: Putting theory in practice.* Andover, UK: Heinle.
Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & cognition, 36*(6), 1103-1121. doi:10.3758/MC.36.6.1103
Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language, 30*, 333-370.
Lin, P. M. S. (2010). The phonology of formulaic sequences: A review. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication*. London, UK: Continuum.
Lin, P. M. S. (2013). The prosody of formulaic expressions in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics, 18*(4), 561-588.
Lin, P. M. S. (2019). Formulaic language and speech prosody. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 78-94). Oxford, UK: Routledge.
Lin, P. M. S., & Adolphs, S. (2009). Sound evidence: Phraseological units in spoken corpora. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 34-48). Basingstoke: Palgrave Macmillan.
Lindstromberg, S., & Boers, F. (2008a). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics, 29*(2), 200-222. doi:10.1093/applin/amn007
Lindstromberg, S., & Boers, F. (2008b). Phonemic repetition and the learning of lexical chunks: The power of assonance, *System,* 36, 423-436.
Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 883.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102. doi:10.1016/0010-0277(87)90005-9
McCauley, S. M., & Christiansen, M. H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D. C. Noelle & R. Dale (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11*(2), 113-128. doi:10.1093/applin/11.2.113

Meara, P. (2005a). Designing vocabulary tests for English, Spanish and other languages. In C. Butler, M. Gómez-González, & D. S. S. (Eds.), *The dynamics of language use*. Amsterdam: Benjamins.
Meara, P. (2005b). *LLAMA Language aptitude tests: The manual*. Swansea: University of Wales.
Meara, P. (2014). Lognostics website. Retrieved from http://www.lognostics.co.uk/
Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System, 28*(1), 19–30.
Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied Linguistics in Society*. London: CILT.
Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Bristol, Buffalo: Multilingual Matters.
Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics, 32*(111-129).
Mondini, S., Luzzatti, C., Saletta, P., Allamano, N., & Semenza, C. (2005). Mental representation of prepositional compounds: Evidence from Italian agrammatic patients. *Brain and Language, 94*(2), 178-187. doi:10.1016/j.bandl.2004.12.005
Moon, R. (1992). Textual aspects of fixed expressions in learners' dictionaries. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics*. London: Palgrace Macmillan.
Myles, F. (2004). From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society, 102*(2), 139-168.
Myles, F., & Cordier, C. (2017). Formulaic sequences (FS) cannot be an umbrella term in SLA. *Studies in Second Language Acquisition, 39*(1), 3-28.
Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48*(3), 323-364. doi:10.1111/0023-8333.00045
Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A Basis for Creative Construction? *Studies in Second Language Acquisition, 21*(1), 49-80.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Clarendon Press Formulaic language.
Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16*(151-171).
Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242. doi:10.1093/applin/24.2.223
Noice, H., & Noice, T. (2006). What studies of actors and acting can tell us about memory and cognitive functioning. *Current directions in psychological science, 15*(1), 14-18.
Nordquist, D. (2004). Comparing elicited data and corpora. In M. Achard & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 211-224). Stanford, CA: CSLI Publications
Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language, 70*(3), 491-538.
O'Donnell, M. B. (2011). The adjusted frequency list: A method to produce cluster-sensitive frequency lists. *ICAME Journal, 35*, 135-169.
Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(2), 324-354.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, 130–149.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. *Language and Communication*, 191-226.
Pellicer-Sánchez, A. (2015). Developing automaticity and speed of lexical access: the effects of incidental and explicit teaching approaches. *Journal of Spanish Language Teaching, 2*(2), 126-139.
Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 153-173). Oxford, UK: Routledge.
Peters, A. M. (1983). *The units of language acquisition*. Cambridge, UK: Cambridge University Press.
Peters, A. M., & Menn, L. (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language,* 69, 742-777.
Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of experimental psychology. Learning, memory, and cognition, 27*(5), 1223-1237. doi:10.1037//0278-7393.27.5.1223
Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 820-833.
Prince, P. (2012). Towards an instructional programme for L2 vocabulary: Can a story help? *Language Learning & Technology, 16*(3), 103-120.
Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Möhle, & M. M. Raupach (Eds.), *Second language productions* (pp. 114-137). Tubingen, Germany: Gunter Narr Verlag.
Rayson, P. (2009). Wmatrix: a web-based corpus processing environment. Retrieved from http://ucrel.lancs.ac.uk/wmatrix/
Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126*(3), 288-311.
Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14*(4), 423-441. doi:10.1080/01638539109544795
Rodríguez, M., & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning, 50*(2), 385-412. doi:10.1111/0023-8333.00121
Roediger, H., Gallo, D., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels of processing framework. *Memory, 10*(5/6), 519-332.
Rubin, D. C. (1995). Stories about stories. In R. S. Wyer Jr (Ed.), *Knowledge and memory: The real story* (Vol. 8, pp. 153-164). London: Psychology Press.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge, UK: Cambridge University Pres.
Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.
Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). Amsterdam, Netherlands: John Benjamins.
Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-152). Amsterdam: John Benjamins.

Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pegagogy*. Cambridge, UK: Cambridge University Press.
Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.
Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*. Oxford, UK. : Blackwell Publishing Ltd, .
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Abingdon: Routledge.
Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*(3), 369-385. doi:10.1017/S0142716400010845
Segalowitz, N., Trofimovich, P., Gatbonton, E., & Sokolovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon, 3*(1), 47-71. doi:10.1075/ml.3.1.05seg
Seki, T. (2004). *Attitudes to and motivation for learning English in Japan with special reference to first-year university students.* (PhD). University of Stirling, Retrieved from http://hdl.handle.net/1893/60
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11*(2), 285-301. doi:10.1515/cllt-2014-0016
Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175*, 111-122. doi:10.1016/j.bandl.2017.10.004
Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251-272. doi:10.1177/0267658310382068
Siyanova-Chanturia, A., & Lin, P. M. S. (2017). Production of ambiguous idioms in English: A reading aloud study. *International Journal of Applied Linguistics, 28*(1), 58-70. doi:10.1111/ijal.12183
Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. Oxford, UK: Routledge.
Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 38-61). Oxford, UK: Routledge.
Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In D. Divjak & S. Gries (Eds.), *Frequency Effects in Language Representation* (pp. 127–164): DE GRUYTER.
Sommers, M., & Barcroft, J. (2013). Effects of referent token variability on L2 vocabulary learning. *Language Learning, 63*(2), 186-210.
Sprenger, S. A., Levelt, W. J. M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*(2), 161-184. doi:10.1016/j.jml.2005.11.001
Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition, 29*, 449-484.

Stengers, H., & Boers, F. (2015). Exercises on collocations: a comparison of trial-and-error and exemplar-guided procedures. *Journal of Spanish Language Teaching, 2*(2), 152-164.
Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does "chunking" foster chunk-uptake? In S. De Knop, F. Boers, & A. De Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 99–117). Berlin, Germany: Mouton de Gruyter.
Strik, H., Hulsbosch, M., & Cucchiarini, C. (2010). Analyzing and identifying multiword expressions in spoken language. *Language Resources and Evaluation, 44*(1-2), 41-58. doi:10.1007/s10579-009-9095-y
Strong, B. (2016). *The effects of error-free and trial-and-error techniques for learning phrasal verbs*. Paper presented at the Vocab@Tokyo, Tokyo, Japan.
Stubbs, M. (2007). Collocations and semantic profiles: On the cause of the trouble with quantitative studies In T. W. & R. Krishnamurthy (Eds.), *Corpus Linguistics: Critical Concepts in Linguistics*. London & New York: Routledge.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & S. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & cognition, 37*(4), 529-540. doi:10.3758/MC.37.4.529
Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & cognition, 32*(5), 742-751. doi:10.3758/BF03195864
Tavakoli, P. (2010). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*. doi:0.1093/elt/ccq020
Taylor, J., R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: OUP.
Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84-119.
Tremblay, A., Asp, E., Johnson, A., Migdal, M. Z., Bardouille, T., & Newman, A. J. (2016). What the networks tell us about serial and parallel processing: An MEG study of language networks and n-gram frequency effects in overt picture. *The Mental Lexicon, 11*(1), 115-160. doi:10.1075/ml.11.1.06tre
Tremblay, A., & Baayen, 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*.
Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon, 6*(2), 302-324.
Tulving, E. (2001). Episodic memory and common sense: How far apart? *Philosophical Transaction of Royal Society London B, 356*, 1505-1515.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373.
Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (Vol. 9, pp. 153-172): John Benjamins Publishing.
Van Lancker, D., Canter, G., & Terbeek, D. (1981). Disambiguation of ditropic sentences. *Journal of Speech and Hearing Research, 24*, 330-335.
Van Lancker, D., & Canter, G. J. (1981). Idiomatic versus literal interpretations of ditropically ambiguous sentences. *Journal of Speech and Hearing Research, 24*(1), 64-69.
Van Lancker Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English. *Applied Psycholinguistics, 24*(1), 45-57.

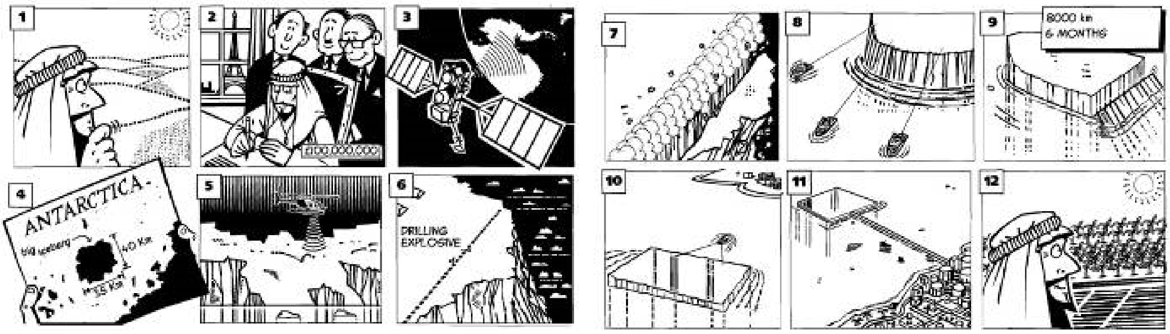
Van Lancker Sidtis, D. (2009). Formulaic and novel language in a dual process model of language competence: Evidence from surveys, speech samples, and schemata. In R. L. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language: Vol. 2. Acquisition, loss, psychological reality, and functional explanations* (pp. 445-470). Amsterdam: Benjamins Publishing Co.
Wahl, A., & Gries, S. (2018). Multi-word expressions: A novel computational approach to their bottom-up statistical extraction. In P. Cantos-Gómez & M. Almela-Sánchez (Eds.), *Lexical Collocation Analysis: Advances and Applications* (pp. 85-109). Springer International Publishing AG
Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25*(2), 261-274. doi:10.1016/S0346-251X(97)00013-4
Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics, 32*(3), 259-276. doi:10.1093/applin/amq051
Wichmann, A., House, J., & Rietveld, T. (2000). Discourse constraints on F0 peak timing in English. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 163-182). Dordrecht: Springer Netherlands.
Willis, D. (1990). *The lexical syllabus*. London: HarperCollins.
Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition, 35*(3), 451-482.
Wolter, B., & Yamashita, J. (2017). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing. *Studies in Second Language Acquisition*, 1-22. doi:10.1017/S0272263117000237
Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT Journal, 29*(2), 209-230.
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, 12*(1), 39-57.
Woolard, G. (2013). *Messaging: Beyond a lexical approach*. London: The Round.
Wray, A. (1992). *The focusing hypothesis: The theory of left hemisphere lateralised language re-examined*. Amsterdam: John Benjamins.
Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics, 21*(4), 463-489. doi:10.1093/applin/21.4.463
Wray, A. (2002a). *Formulaic language and the lexicon*. Cambridge: CUP Formulaic language.
Wray, A. (2002b). Formulaic language in computer-supported communication: theory meets reality. *Language Awareness, 11*(2), 114-131.
Wray, A. (2004). 'Here's one I prepared earlier': formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use (pp. 249-268)*. Amsterdam: John Benjamins.
Wray, A. (2008a). *Formulaic language: pushing the boundaries.* Oxford: OUP Formulaic language.
Wray, A. (2008b). The puzzle of language learning: From child's play to 'linguaphobia'. *Language Teaching, 41*(2), 253-271.
Wray, A. (2009). Identifying formulaic language. In R. L. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language*. Amsterdam: John Benjamins.
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*, 231-254.
Wray, A. (2019). Concluding question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & A.

Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language perspective* (pp. 248-269). Oxford, UK: Routledge.
Wray, A., Bell, H., & Jones, K. (2016). How native and non-native speakers of English interpret unfamiliar formulaic sequences. *European Journal of English Studies*, 20(1), 47-63. doi:10.1080/13825577.2015.1136163
Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorised language as a guide to nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123-148). Philadelphia, PA: John Benjamins.
Wray, A., & Fitzpatrick, T. (2009). Pushing learners to the extreme: the artificial use of prefabricated material in conversation. *Innovation in Language Learning and Teaching, 4*(1), 37-51. doi:10.1080/17501220802596413
Wray, A., & Namba, K. (2003). Use of formulaic language by a Japanese-English bilingual child: A practical approach to data analysis. *Japan Journal for Multilingualism and Multiculturalism, 9*(1), 24-51.
Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication, 20*(1), 1-28. doi:10.1016/s0271-5309(99)00015-4
Yamashita, J., & Jiang, N. (2010). L1 Influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*, 647-668. doi:10.2307/27896758
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2015). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing, 33*(4), 497-528. doi:10.1177/0265532215594643
Yang, S.-Y., Ahn, J. S., & Van Lancker Sidtis, D. (2015). The perceptual and acoustic characteristics of Korean idiomatic and literal sentences. *Speech, Language and Hearing, 18*(3), 166-178. doi:10.1179/2050572814Y.0000000061
Yoshimura, Y., & MacWhinney, B. (2007). The effect of oral repetition on L2 speech fluency: An experimental tool and language tutor*. Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*. International Speech Communication Association (ISCA) ISBN: 13: 978-0-615-16568-4

Appendices

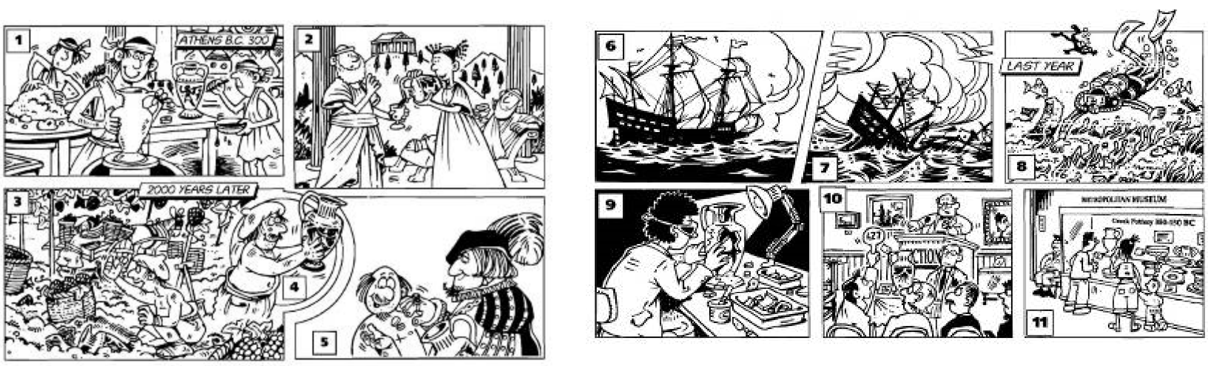
Appendix 3.1: Picture sequences for story narration task (S1)

1. Iceberg



2. New life

3. Treasure



Sequences adapted from M. Fletcher and Munns (2005)

Appendix 4.1: Algorithmic methods for identifying formulaic expressions

In order to select potential formulaic expressions in the sample (i.e. those that can reasonably be expected to be or to become formulaic), a non-phonological approach was required for study S2. However, methods based primarily on diagnostic criteria (see Section 2.4.1) were not considered appropriate for this purpose because of the difficulty in applying them consistently.

Alternative methods using frequency-based algorithms were therefore explored. These typically use measures of the co-occurrence of words based on frequency statistics from the corpus. For example, the collocation strength of bigrams can be estimated using lexical association measures such as Mutual Information (MI) or T-value (Stubbs, 2007), log-likelihood (Wahl & Gries, 2018), or variations of them such as MI3, which gives more weight to the raw frequency of the collocate (Oakes, 1998). For multiword units longer than two words, different methods again have been used. In these, the statistical measures are typically applied sequentially to give an overall measure for the coherence of the sequence as a unit. For example, MERGE (Wahl & Gries, 2018) successively combines bigrams using log-likelihood. In this approach, the log-likelihood of all bigrams that occur in the corpus is calculated (based on the component word frequencies). Then the bigram with the highest log-likelihood is given 'word' status – i.e. absorbed into the corpus as if a single word. All frequency measures are updated along with log-likelihood measure for the remaining bigrams. The process is then repeated until the 'winning' bigram reaches some minimum cut-off value. Since, during the process, bigrams can combine with other words or bigrams continuously, sequences can grow in length in an unrestrained way. Once the cut-off is reached, a list of individual lexemes remains, from single words to multiword units of any length.

Gries and Mukherjee (2010) have suggested a method that uses a lexical association measure called Lexical Gravity (LG). This is based on the sum of the forward and backward transitional probabilities of a two-way co-occurrence, each weighted by the type frequency (that is, the number of different word types) that can occupy its outcome slot, given its cue. In their extraction method, *n*-grams of various lengths are scored on the basis of the mean LG of their component bigrams. Those *n*-grams with mean LGs below a certain value are discarded. In the remaining list, any n-grams that are contained in an *(n+1)*-gram with a higher mean LG score, are also discarded. O'Donnell (2011) uses a similar approach but with just raw frequency counts as the metric of sequence strength. He first identifies all *n*-grams with frequency greater than some threshold (which he sets at 3). Then, in order of decreasing length, each n-gram is analysed in terms of both component *(n-1)*-grams. The count of these *(n-1)*-grams is decreased by the number of tokens of any *n*-grams in which it is a component and removed if it now falls below the threshold.

An approach from computational linguistics is that of Brooke and associates (Brooke (Brooke et al., 2015; Brooke et al., 2014). They developed a way of extracting lists of formulaic sequences from corpora of any size using an unsupervised algorithmic process (i.e. without the need for any subsequent hand curating). The approach explicitly references the lexical nature of formulaic sequences and the theoretical background of Wray (2002a, 2008a). Theirs is a top down method, starting with longer segments and then deciding whether to break them down. That decision is based on a metric that uses conditional probability to assess the predictability of words appearing within sequences. The overall algorithm works by analysing a corpus with regard to *n*-grams (within sentences) that are maximal with regard to a given frequency threshold (where 'maximal' = the largest *n*-gram occurring above the threshold frequency for which any *(n+1)*-gram which contains it does not occur above the threshold). They then segment all the sentences in the corpus according to these maximal n-grams. Overlaps are resolved by selecting breakpoints that maximise overall word predictability. This initial segmentation, by design, produces quite long sequences. A second segmentation process is then applied by analysing all the sequences that have been identified from the initial segmentation, working longest to shortest. A sequence is broken down into sub-sequences if it either a) is below the frequency threshold, or b) has a break point which would render the overall word predictability higher.

Appendix 4.2: Calculating sequence predictabilities (Study S2)

To calculate the predictability of each sequence (e.g. *our rival companies),* the predictability of each word occurring within that sequence (i.e. *our*, *rival* and *companies*) is first calculated. The sequence predictability is then taken as the product of the word predictabilities for all words in the sequence.

Word Predictabilities

Word predictabilities are calculated by looking at the conditional probabilities of that word appearing with the other words (or sub-sequences) in the target sequence, and taking the maximum value across these. Formally, for any word w_i within a sequence $w_{1,n}$ the predictability is defined as:

$$\text{pred}\,(w_i, w_{1,n}) = \max_{j,k} p(w_i \mid w_{j,k}) \quad (1 \le j \le i \le k \le n)$$

where $p(w_i \mid w_{j,k})$ is the conditional probability of the word w_i appearing in a sequence $w_j \ldots, w_{i-1}, w_{i+1}, \ldots w_k$. For i=j=k (i.e. when the 'sequence' is a single word), the predictability is just the marginal probability of the word appearing in the corpus. The idea is that we look at every sub-sequence (within the target sequence) that the word could appear in and take the highest value conditional probability of the word across these.

Example

To illustrate, consider the example of the sequence *our rival companies*. Word predictabilities must be calculated for *our*, *rival* and *companies*. To calculate the word predictability of *rival* within the sequence, each sub-sequence in which it appears (*our rival* or *rival companies*) is considered and the conditional probability of the word appearing within that sub-sequence calculated, using frequency data from the whole learner corpus:

p(*rival* | *our* __) = freq (*our rival*) /freq (*our)* = 0.068

p(rival | ___ companies) =freq(rival companies)/freq(companies) = 0.700

The word predictability for *rival* is then taken as the maximum of these (0.700). Similar calculations are done to give word predictabilities for the other two words in the sequence: *our* (predictability = 1.000) and *companies* (predictability = 1.000). The product of these then gave overall sequence predictability of 0.700.

Note

In Brooke et al. (2014) the log of the max word predictabilities is taken and the combination for each word summed. However, the above process is essentially the same as this, and is the one adopted in Brooke et al. (2015). However, in Brooke et al. (2015), they attempt to reduce the effect of syntax on word probabilities by dividing conditional word probabilities by the probability of that word given the syntactic combination it appears in (based on the word's part of speech). This was experimented with for the current study, but found not to reduce the influence of syntax on the predictability due to the small sample and idiosyncratic nature of the individual's own syntax. It seemed to favour combinations which involve fixed class words (e.g. '*to*', determiners, prepositions) where the number of different word types is limited. This resulted in similar probabilities for numerator and denominator, leading to higher values for conditional probability.

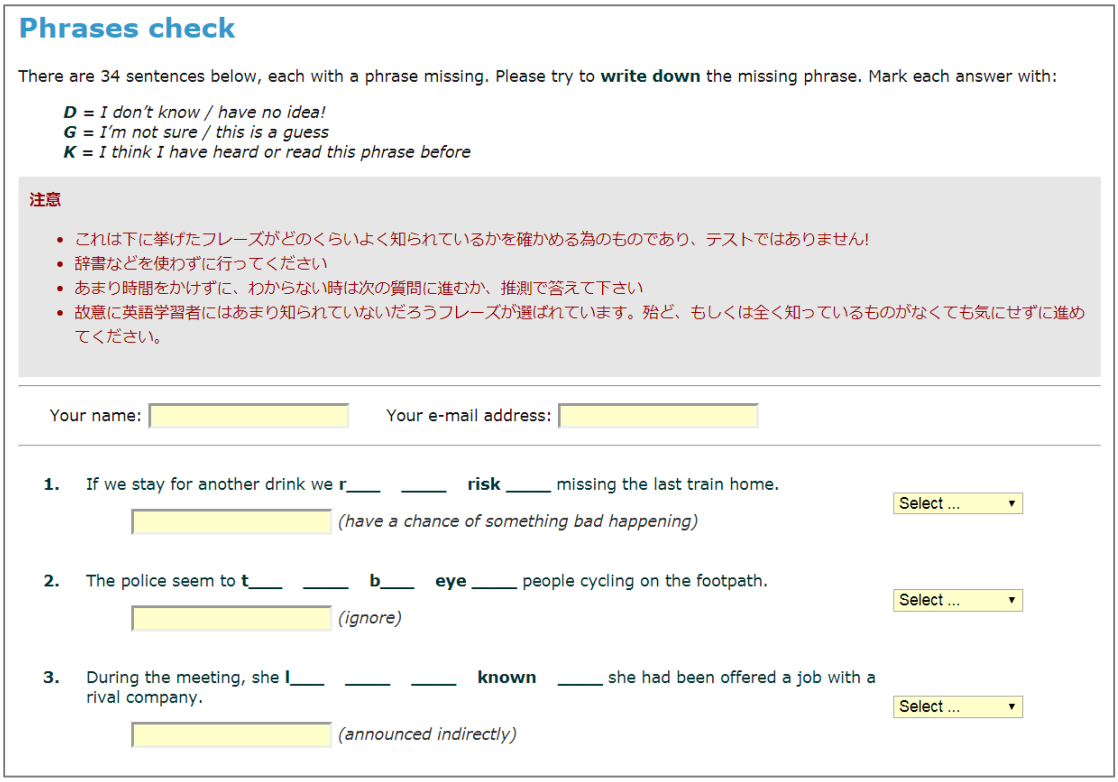
Appendix 7.1: Test to check familiarity of potential targets (Study S4)

A test was created to check which of the 34 potential target sequences were already known to the participants. This was available as an on-line test and results were collected automatically.

For each potential target, an example sentence containing the target was created. This was then presented with part of the target removed, and the participant was asked to write down the expression. A simple gloss to the meaning was also provided. For example, for the potential target *follow in the footsteps*, the following was presented:

> He wants to **f___ ___ ___ footsteps** of his father and become an actor.
> *(do the same thing as)*

The test-taker was also asked to indicate if it was known (K), a guess (G), or if they did not know at all (D). The test questions were checked with three native English speakers to ensure that the sentences prompted the desired responses when the targets were known. A screen-shot showing the instructions and part of the test is given below:


Phrases check

There are 34 sentences below, each with a phrase missing. Please try to **write down** the missing phrase. Mark each answer with:

D *= I don't know / have no idea!*
G *= I'm not sure / this is a guess*
K *= I think I have heard or read this phrase before*

注意

- これは下に挙げたフレーズがどのくらいよく知られているかを確かめる為のものであり、テストではありません!
- 辞書などを使わずに行ってください
- あまり時間をかけずに、わからない時は次の質問に進むか、推測で答えて下さい
- 故意に英語学習者にはあまり知られていないだろうフレーズが選ばれています。殆ど、もしくは全く知っているものがなくても気にせずに進めてください。

Your name: ______ Your e-mail address: ______

1. If we stay for another drink we **r___ ____ risk** ____ missing the last train home. *(have a chance of something bad happening)* Select ...
2. The police seem to **t___ ____ b___ eye** ____ people cycling on the footpath. *(ignore)* Select ...
3. During the meeting, she **l___ ____ ____ known** ____ she had been offered a job with a rival company. *(announced indirectly)* Select ...

Appendix 7.2: Context stories and picture cues (Study S4)

The 'A' set of stories:

The target sequences (STs and LTs) are highlighted in bold text.

<table>
<tr><td>
X – A: Rubik's Cube

The Rubik's Cube is a popular puzzle invented by Erno Rubic who was a designer from Hungary. The puzzle has six sides and each side has 9 squares of different colours. The aim is to make each side of the cube the same colour by twisting the cube. It is difficult to do at first, but with practice it becomes much easier. Once you get the hang of it, you can solve the puzzle quite quickly.

Rubik invented many different kinds of toys and puzzles. Most of his designs didn't see the light of day, but the cube was an immediate success. When it was released in the 1980s, millions were sold and it became the best-selling toy in the world. As a puzzle it has stood the test of time and is still very popular today. Over 350 million were sold in 2009. The success of the Cube paved the way for many new kinds of 3D puzzle. There is also a World Cube Association which organises competitions and keeps official world records.
</td><td>


</td></tr>
<tr><td>
Y-A: Working hours problem

ABC Company has a problem with their workforce. Currently staff have to work very long hours every day and often at the weekends too. Although they enjoy their work, the long hours are causing a lot of stress and sickness. If the company does not change, it runs the risk of losing staff to its competitors.

A project team has been researching the problem and have made a proposal for introducing a fixed 6.30 pm finishing time, 3 days per week. Most employees like the sound of this idea and think it will improve their working lives. However, some managers are worried that important work will not be completed on time if people finish work earlier. The Board are sitting on the fence and will not make a clear decision either way. They are planning to introduce new technology soon and hope that the situation will improve once this has been introduced. However, it remains to be seen whether the new technology will actually make any difference.
</td><td>


</td></tr>
<tr><td>
Z-A: The unreliable soccer player

Joe Mann was a very talented soccer player but he was unreliable and lazy. He was often late for training and argued with his team-mates. The team manager did not seem to worry about these problems though. He turned a blind eye to the bad behaviour because Joe was such a great player.

However, the problems came to a head when he was late one day for an important match. No-one knew where he was and there was no contact from him. The match started and he still had not arrived. Everyone was worried, thinking that Joe must have had an accident. When he finally arrived, the manager breathed a sigh of relief that he was safe. Actually, there had not been an accident, Joe had just over-slept again. However, the manager did not complain; he was just happy that Joe could play in the second half.

Some of Joe's team-mates were angry though and later talked about the situation with their friends. Soon, the story fell into the hands of a local journalist who then wrote an article about it. When the fans heard, they demanded that Joe be fired for showing disrespect to the team.
</td><td>


</td></tr>
</table>

The 'B' set of stories

These are identical to the A set, except that the reverse set of long and short targets (highlighted in bold) are used.

X – A: Rubik's Cube

The Rubik's Cube is a popular puzzle invented by Erno Rubic who was a designer from Hungary. The puzzle has six sides and each side has 9 squares of different colours. The aim is to make each side of the cube the same colour by twisting the cube. It is difficult to do at first, but with practice it becomes much easier. **Once you get the hang of it, you can solve the puzzle quite quickly**.

Rubik invented many different kinds of toys and puzzles. Most of his designs didn't **see the light of day**, but the cube was an immediate success**.** When it was released in the 1980s, millions were sold and it became the best-selling toy in the world. **As a puzzle it has stood the test of time and is still very popular today**. Over 350 million were sold in 2009. The success of the Cube **paved the way for** many new kinds of 3D puzzle. There is also a World Cube Association which organises competitions and keeps official world records.

Y-A: Working hours problem

ABC Company has a problem with their workforce. Currently staff have to work very long hours every day and often at the weekends too. Although they enjoy their work, the long hours are causing a lot of stress and sickness. If the company does not change, it **runs the risk of** losing staff to its competitors.

A project team has been researching the problem and have made a proposal for introducing a fixed 6.30 pm finishing time, 3 days per week. **Most employees like the sound of this idea and think it will improve their working lives.** However, some managers are worried that important work will not be completed on time if people finish work earlier. The Board are **sitting on the fence** and will not make a clear decision either way. They are planning to introduce new technology soon and hope that the situation will improve once this has been introduced. However, it **remains to be seen** whether the new technology will actually make any difference.

Z-A: The unreliable soccer player

Joe Mann was a very talented soccer player but he was unreliable and lazy. He was often late for training and argued with his team-mates. The team manager did not seem to worry about these problems though. **He turned a blind eye to the bad behaviour because Joe was such a great player**.

However, the problems **came to a head** when he was late one day for an important match. No-one knew where he was and there was no contact from him. The match started and he still had not arrived. Everyone was worried, thinking that Joe must have had an accident. When he finally arrived, **the manager breathed a sigh of relief that he was safe.** Actually, there had not been an accident, Joe had just over-slept again. However, the manager did not complain; he was just happy that Joe could play in the second half.

Some of Joe's team-mates were angry though and later talked about the situation with their friends. Soon, the story **fell into the hands of** a local journalist who then wrote an article about it. When the fans heard, they demanded that Joe be fired for showing disrespect to the team.

Appendix 7.3: Statistical analysis for the Replicated Latin Square Design

The following describes the statistical testing approach applied to the Replicated Latin Square (RLS) design for each of the three analyses:

1. Comparison of conditions W, S & I for the long target sequences (LTs)
2. Comparison of conditions W, S & I for the short target sequences (STs)
3. Comparison of conditions E and NE for the short sequences (E-STs and STs)

1. Comparison of segmentation conditions on LTs

The experimental design for applying the segmentation conditions (W, S and I) to the LTs is equivalent to the following pair of Replicated Latin Squares:

		Targets		
		Story X (L2, L4)	**Story Y** (L6, L8)	**Story Z** (L9, L11)
Ppts	**P1**	W	S	I
	P2	S	I	W
	P3	I	W	S

		Targets		
		Story X (L1, L3)	**Story Y** (L5, L7)	**Story Z** (L10, L12)
Ppts	**P4**	W	S	I
	P5	S	I	W
	P6	I	W	S

Here the items are blocked according to two factors: the participant (P1-P6) and the story (X, Y, Z), with two different LTs (e.g. L2, L4) in each story. The replicated Latin squares (equivalent to sub-sets A and B in experimental design) have different participants and targets. According to the literature (e.g. Open.Ed@PSU, 2016), RLS designs with different types of row and column factor are analysed as follows:

The statistical model (effects model) is: $Y_{hijk}=\mu+\delta_h+\rho_{i(h)}+\beta_{j(h)}+\tau_k+e_{hijk}$,

For each observation Y_{hijk} we have an overall mean (μ), plus

- an effect due to repetition squares: δ_h, h = 1,..n
- an effect due to participants: $\rho_{i(h)}$, i = 1, ..t (rows)
- an effect due to target pairs: $\beta_{j(h)}$, j = 1,..t (columns)
- an effect due to condition (τ_k), k=1,…t - with one value of k per (h,i,j) combination
- plus error (e_{hijk})

The analysis of variance table is as follows:

AOV	df	SS
replications = square (A or B)	$n - 1 = 1$	$\sum t^2(\bar{Y}_{h...} - \bar{Y}_{....})^2$
row = participants (1 to 3)	$n(t - 1) = 4$	$\sum t(\bar{Y}_{hi..} - \bar{Y}_{m...})^2$
column = stories (1 to 3)	$n(t - 1) = 4$	$\sum t(\bar{Y}_{h.j..} - \bar{Y}_{m...})^2$
treatment = condition (W, S or I)	$t - 1 = 2$	$\sum nt(\bar{Y}_{...k} - \bar{Y}_{....})^2$
error	$(t - 1)\,[n(t - 1) - 1] = 6$	(subtraction)
Total	$nt^2 - 1 = 17$	$\sum(\bar{Y}_{mijk} - \bar{Y}_{....})^2$

Where: df = degrees of freedom ; SS = sum of squares

n = no. of replicated squares = 2; t = no. of treatments (conditions) = 3

This analysis is then applied to the results from study S4. For example, inserting figures for the Initial Accuracy measure (with cell values equal to the mean of the two LTs in the story) gives:

Ppts	Targets		
	L2/L4	**L6/L8**	**L9/L11**
P1	0.726	0.943	0.813
P2	0.488	0.856	0.796
P3	0.705	0.943	0.862

Ppts	Targets		
	L1/L3	**L5/L7**	**L10/L12**
P4	0.344	0.460	0.866
P5	0.713	0.586	0.911
P6	0.879	0.939	0.717

	W	**S**	**I**
Mean	0.777	0.697	0.784

Comparing means using the ANOVA based on the above table gives:
$F(2,6) = 1.112$, $p=0.3883$

Similar results are obtained for the other dependent variables.

2. Comparison of segmentation conditions on STs

This is similar to the previous case: treatments applied to the STs are equivalent to the following repeated Latin Squares:

		Targets		
		Story X (S1,S3)	**Story Y** (S5,S7)	**Story Z** (S10,S12)
Ppts	**P1**	W	S	I
	P2	S	I	W
	P3	I	W	S

		Targets		
		Story X (S2,S4)	**Story Y** (S6,S8)	**Story Z** (S9,S11)
Ppts	**P4**	W	S	I
	P5	S	I	W
	P6	I	W	S

The analysis of variance would be the same as for the LT case.

3. Comparison of Embedded (E) v Non-embedded (NE) on STs

In this case we are interested in comparing the performance of the short sequences between the cases where they are E-STs embedded in the ST (condition E) and where they are STs delivered on their own (condition NE). We can consider this a set of repeated 2x2 Latin squares as follows:

W	**S2/S4**	**S1/S3**
P1	E	NE
P4	NE	E

W	**S5/S7**	**S6/S8**
P3	E	NE
P6	NE	E

W	**S10/S12**	**S9/S11**
P2	E	NE
P4	NE	E

S	**S5/S7**	**S6/S8**
P1	E	NE
P4	NE	E

S	**S10/S12**	**S9/S11**
P3	E	NE
P6	NE	E

S	**S2/S4**	**S1/S3**
P2	E	NE
P4	NE	E

I	**S10/12**	**S9/11**
P1	E	NE
P4	NE	E

I	**S2/S4**	**S1/S3**
P3	E	NE
P6	NE	E

I	**S5/S7**	**S6/S8**
P2	E	NE
P4	NE	E

Here the replicated squares represent different combinations of Story (X, Y, Z) and Segmentation conditions (W, S, I), the rows represent participants, and the columns represent the different targets (S1-S12). There are 2 targets for each column (e.g. S2/S4). This is more complicated than the previous example because the nine replications are mixed across targets and participants.

To simplify the analysis, we can pool all the embedded and non-embedded targets for each participant (i.e. ignore the W-S-I factor). This would be equivalent to:

	2/4/5/7 10/12	1/3/6 8/9/11
P1	E	NE
P4	NE	E

	2/4/5/7 10/12	1/3/6 8/9/11
B1	E	NE
B2	NE	E

	2/4/5/7 10/12	1/3/6 8/9/11
C1	E	NE
C2	NE	E

This is a RLS design with no. of replicates, n=3, and no. of treatments, t=2. In this case, the columns change but the rows are the same.

The statistical model (effects model) is again: $Y_{hijk}=\mu+\delta_h+\rho_{i(h)}+\beta_{j(h)}+\tau_k+e_{hijk}$,

The analysis of variance table is as follows:

AOV	**df**	**SS**
repetitions = square (1 to 3)	$n - 1 = 2$	$\sum t^2(\bar{Y}_{h...} - \bar{Y}_{....})^2$
row = participants (1 to 2)	$n(t - 1) = 3$	$\sum t(\bar{Y}_{hi..} - \bar{Y}_{h...})^2$
column = target set (1 to 2)	$t - 1 = 1$	$\sum nt(\bar{Y}_{..j..} - \bar{Y}_{....})^2$
treatment = condition (E or NE)	$t - 1 = 1$	$\sum nt(\bar{Y}_{...k} - \bar{Y}_{....})^2$
error	$(t - 1)[n(t - 1) - 1] = 4$	(subtraction)
Total	$nt^2 - 1 = 11$	$\sum(\bar{Y}_{hijk} - \bar{Y}_{....})^2$

Where: df = degrees of freedom ; SS = sum of squares

n = no. of replicated squares = 3; t = no. of treatments (conditions) = 2

Entering figures to compare the dependent variable for Initial Accuracy gives:

	2/4/5/7 /10/12	**1/3/6/ 8/9/11**
A1	0.840	0.740
A2	0.917	0.744

	2/4/5/7 /10/12	**1/3/6/ 8/9/11**
B1	0.770	0.811
B2	0.839	0.859

	2/4/5/7 /10/12	**1/3/6/ 8/9/11**
C1	0.854	0.679
C2	0.760	0.871

	E	**NE**
Mean	0.836	0.787

Comparing means using the ANOVA based on the above table gives:
$F(1, 4) = 1.187$, $p=0.3372$

Similar results are obtained for the other dependent variables.

Appendix 9.1: Context stories and picture cues (Study S5)

A: The unreliable volleyball player

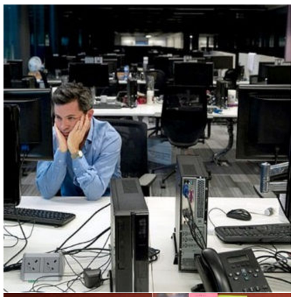
Jenny was a very talented volleyball player but she was unreliable and lazy. She was often late for training and argued with her team-mates. The team manager didn't seem to worry about these problems with Jenny though. He **turned a blind eye to** her behaviour because she was such a great player.

However, the problems **came to a head** one day when she was late for an important match. No-one knew where she was and there was no contact from her. The match started without her and the manager was really worried. When she finally arrived, the manager **breathed a sigh of relief** that she was safe. In fact, there was no accident, she had just over-slept again. So, finally the manager decided to take some action and dropped her from the team.

B: Working hours problem

In my company, staff have to work very long hours every day and often at the weekends too. Although most staff really enjoy their work, the long hours are causing a lot of stress and sickness. It's a problem for the company because, if nothing changes**,** we **run the risk of** losing staff to our competitors.

A project team has been researching the problem and have made a proposal for introducing a fixed 6.30 pm finishing time, 3 days per week. Although it's not a perfect solution, the team think it'll **go a long way towards** solving the problem. According to a staff survey, most employees **like the sound of** this idea and think it will improve their working lives. However, management are not so sure because they think it will result in less work getting done.



C: Rubik's Cube

The Rubik's Cube is a popular puzzle invented by a designer from Hungary called Erno Rubic. As a young man, Rubik wanted to be famous and successful. He **set his sights on** making a challenging puzzle that would be loved by people all across the world. When it was released in the 1980s, millions were sold and it became the best-selling toy ever. As a puzzle it has **stood the test of time** and is still very popular today. Over 350 million were sold last year.

The aim is to make each side of the cube all the same colour by twisting it around. It's very difficult to do at first but, once you **get the hang of** the puzzle, you can solve it quite quickly. Many people can now do it in less than 30 seconds!

D: My broken sofa

I live in a rented apartment and have a very scary landlord. He's a big angry man with a loud voice. Last weekend, I accidentally broke the sofa in the apartment and I wasn't sure what to do. I **knew better than to** call the landlord and tell him, because I was sure he'd make a terrible fuss - and probably make me leave the apartment. For a short while, I **toyed with the idea of** buying a new sofa, but then decided not to because they're very expensive.

In the end, I decided to fix it myself. So far, it seems to be fine and I think I did a good job. However, the landlord is coming round next week and it **remains to be seen** if it will be OK when he sits on it!

Note: Target sequences are in bold.

Appendix 9.2: List of targets and Japanese translations (Study S5)

A1	turned a blind eye to 見て見ぬふりをした
	He turned a blind eye to her behaviour 彼は彼女の素行を見て見ぬふりをした。
A2	came to a head 〔問題／危険が〕頂点に達した
	The problems came to a head one day ある日、問題は頂点に達した。
A3	breathed a sigh of relief ホッと一息ついた
	The manager breathed a sigh of relief that she was safe. 彼が無事だったことにマネージャーはホッと一息ついた。

B1	run the risk of ～の危険がある
	We run the risk of losing staff to our competitors ライバル会社に職員を取られてしまう危険がある。
B2	go a long way towards ～に大いに役立つ
	It'll go a long way towards solving the problem それは問題を解決するのに大いに役立つ。
B3	like the sound of 〔アイデアなどを〕いいと思う
	Most employees like the sound of this idea 殆どの従業員がこのアイデアをいいと思っている。

C1	set his sights on 狙いを定める
	He set his sights on making a challenging puzzle 彼はやりがいのあるパズルを作ることに狙いを定めた
C2	stood the test of time 長年に渡って賞賛された
	As a puzzle, it has stood the test of time それはパズルとして長年に渡って賞賛された。
C3	get the hang of コツを掴む
	Once you get the hang of the puzzle, you can solve it quite quickly. 一旦コツを掴んだら、かなりの速さでパズルを解くことができる。

D1	knew better than to ～するほど馬鹿でない
	I knew better than to call the landlord 大家に電話するほど馬鹿ではない。
D2	toyed with the idea of ～と、なんとなく考えた
	I toyed with the idea of buying a new sofa 新しいソファーを購入しようかと、なんとなく考えた。
D3	remains to be seen 現時点では不明
	it remains to be seen if it'll be OK when he sits on it そこに座っても大丈夫かどうか現時点では不明である。

Appendix 9.3: Usage test (Study S5)

In the test, participants hear a brief description of a situation (audio cue) and are then given a written sentence starter (written cue). Their task is to complete the sentence verbally using an appropriate target sequence.

1	Audio cue:	The police in Japan seem to ignore people cycling on the pavement. They never seem to stop them doing it.
	Written cue:	**The police in Japan …**
	Model response:	The police in Japan turn a blind eye to (people) cycling on the pavement.
2	Audio cue:	Last week Tom went out drinking late in the city centre with friends. It was nice but by staying out so late, there was a good chance that he would miss the last train home.
	Written cue:	**By staying out late, he …**
	Model response:	By staying out late, he ran the risk of missing the last train.
3	Audio cue:	I kept on getting very bad headaches. Then I took some new medicine. It didn't completely stop the headaches but it really helped me feel much better.
	Written cue:	**The new medicine …**
	Model response:	The new medicine has gone a long way towards stopping my headaches.
4	Audio cue:	I got a new job recently. At first the work was difficult but, after doing it for a short while, I became able to do it quite well and I now find it quite easy.
	Written cue:	**After a short while …**
	Model response:	After a short while, I got the hang of my new job
5	Audio cue:	We are buying a new car and trying to decide which type to get. We were briefly thinking about getting a fancy new sports car, but have now decided to get a normal family car.
	Written cue:	**For our new car, we were …**
	Model response:	For our new car, we were toying with the idea of getting a sports car.
6	Audio cue:	My friends went to a new restaurant last week and told me about their unusual desserts. From what they said, I knew I would enjoy the desserts too, so I went there last night.
	Written cue:	**I went to that restaurant because I …**

	Model response:	I went to that restaurant because I liked the sound of their desserts
7	Audio cue:	I am expecting John to come round to my house tonight. He said he would come, but I'm not sure if he actually will as he is quite unreliable.
	Written cue:	**He said he would come, but it …**
	Model response:	He said he would come, but it remains to be seen if he will
8	Audio cue:	The parents were anxious because their child was ill in hospital. When the doctor told them their child was going to be fine, I could see that they felt very relieved at that news.
	Written cue:	**I saw the worried parents …**
	Model response:	I saw the worried parents breathe a sigh of relief at the news.
9	Audio cue:	The company has been having many different financial difficulties recently. The problems have reached a very bad point now because they have just lost their best client.
	Written cue:	**Problems at the company …**
	Model response:	Problems at the company have come to a head because they lost their best client
10	Audio cue:	He is very sensitive about his appearance. You should not have talked about his diet when you were out with him as it made him angry.
	Written cue:	**When out with him, you should …**
	Model response:	When out with him, you should have known better than to talk about his diet
11	Audio cue:	The basic design of the bicycle has stayed the same for over 100 years and it is still very good. I think it will continue to stay popular and useful in the future too.
	Written cue:	**The design of the bicycle will continue to …**
	Model response:	The design of the bicycle will continue to stand the test of time
12	Audio cue:	That teenage girl is a very fast runner and she is training hard. She is aiming to win a medal at the 2020 Olympics
	Written cue:	**The young runner has set her sights on winning a medal at the Olympics …**
	Model response:	The young runner has set her sights on winning a medal at the Olympics