The Relationship between Formulaic Sequences and EFL Reading Comprehension

THESIS SUBMITTED

IN ACCORDANCE WITH THE REQUIREMENTS OF THE UNIVERSITY OF LIVERPOOL

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

MOLOUD BABAEE KHOU

MAY 2008

PAGE NUMBERS CLOSE TO THE EDGE OF THE PAGE SOME ARE CUT OFF

TEXT BOUND CLOSE TO THE SPINE IN THE ORIGINAL THESIS

DECLARATION

This work is original and has not been submitted previously in support of a degree, qualification or other course.

.....Babaee.....

Moloud Babaee Khou

TO MY FAMILY

ACKNOWLEDGMENT

I would like to express my deepest gratitude to all the people who assisted me in this study. I am greatly indebted to my first supervisor, Dr Mike Scott, for his invaluable guidance, helpful comments and constructive criticism throughout the writing of this thesis. I would also like to thank Mr. Geoff Thompson, my second supervisor, for his support and encouragement.

My special thanks go to my family, especially my husband Hossein and my daughter Asal for their love, support and patience through this journey.

ABSTRACT

The Relationship between Formulaic Sequences and EFL Reading Comprehension

Moloud Babaee Khou

The main focus of this study was to investigate the relationship between recognition and understanding of formulaic sequences (FSs) and L2 reading comprehension. It also sought to find out whether the interaction between recognition and understanding of these sequences affects reading comprehension.

In addition, six secondary hypotheses were examined. They investigated the relationships between recognition and understanding of formulaic sequences and the frequency of occurrence of the target sequences in the BNC and the subjects' overall English proficiency and their sex.

Data were gathered both quantitatively and qualitatively. On the quantitative side of the continuum two designs were employed, an ex post facto design (correlation) and a factorial design (two-way ANOVA). 53 Iranian university students with an upper-intermediate level of English proficiency participated in this part of the study and four tests (TOEFL, a reading comprehension, a test of recognition of FSs and a test of understanding FSs) were used to collect data.

On the qualitative side of the continuum, a combination of interview, recall and introspective methods were used to gather data. 6 Iranian students studying in the University of Liverpool at the graduate level voluntarily participated in this part of the study.

The results of the statistical analyses showed a moderate positive relationship between recognition of FSs and reading comprehension, however, a relatively low relationship was found between understanding these sequences and reading comprehension. The results also suggested a lack of interaction effect between recognition and understanding as independent variables.

Regarding the secondary hypotheses, no relationship was found between frequency of occurrence of target FSs and the subjects' recognition and understanding of these sequences. In addition, a positive moderate relationship was found between their overall proficiency and FSs. It was also found that males and females performed similarly on the recognition and understanding tests of FSs.

The results of the qualitative analysis showed that whether or not understanding of FSs contributes to better reading performance depends on several factors including the nature of the sequences and their functions in a given text. It was also found that readers use a number of linguistic and non-linguistic cues and knowledge sources to overcome difficulties arising from unknown sequences in a text.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter I
Introduction

1.1 Introduction to the Chapter

Over the last twenty years or so much has been done in the field of reading and vocabulary and the relationship between the two in the context of acquisition of English as a second or foreign language. Vocabulary is no longer considered "a neglected aspect of L2 learning process" (Meara 1987, p. 143) as mentioned by Zimmerman (1997), who then maintained "the teaching and learning of vocabulary have been undervalued in the field of second language acquisition (SLA) throughout its varying stages and to the present day" (p.5). This neglect of vocabulary was due to the dominance of syntactically-oriented approaches to language teaching including the Audio-lingual Approach. Even the emergence of approaches which emphasized communication as their pivotal point did not put vocabulary at the centre of attention.

Unlike vocabulary, reading has always been considered essential in developing syllabi for general English and English for specific purposes apart from the heyday of the Audio-lingual Approach which considered oral skills as having priority over written ones. However, research on reading has undergone dramatic changes. The approaches that favoured bottom-up processing of text which was once regarded as the only way to derive meaning out of text gave way to approaches that deemed top-down processing as important, which, in turn, were superseded by approaches bringing both processing models together. It is therefore argued by the proponents of the interactive approaches to reading that both bottom-up and

top-down processes are at work when the reader attempts to comprehend a text.

One of the elements of a text that influences its successful bottom-up processing is vocabulary. Different aspects of vocabulary knowledge, including vocabulary size, knowledge of vocabulary meaning, word recognition have also proved to be influential in reading. However, one of the most important aspects of vocabulary, that is, the relations between words (syntagmatic and paradigmatic), has not received sufficient attention in the field of L2 reading comprehension.

According to Sinclair (2004) in describing a language "the two fundamental axes of language patterning, the paradigmatic and the syntagmatic" (p. 168) should be included. He asserts that according to corpus linguists both axes are responsible for creating meaning. The syntagmatic dimension specifies the relationships between words and unlike the paradigmatic dimension which provides the possible choices at particular positions on the syntagmatic axis, the relationship on the syntagmatic axis is linear. That is, paradigmatic choices are made on the basis of their position in the syntagmatic axis and their relationships with other words. For example, in the sentence

Susan has a keen | **ear* / **head* / *eye* | *for details.*

it is the paradigmatic (vertical) axis that provides the possible choices at the particular position on the syntagmatic (horizontal) dimension, however, the possible choice which is 'eye' is selected on the basis of the syntagmatic axis which "provides a framework for the interpretation of any choice to be made

on the paradigmatic axis" (ibid, p. 170). The relations between collocations and other multi-word sequences are of this type as they function as rules for combining smaller units of any level of a language into bigger ones. As Sinclair puts it

> *instead of expecting to understand a segment of text by accumulating the meaning of each successive meaningful unit, here is the reverse, where a number of units taken together create a meaning, and this meaning takes precedence over the 'dictionary meanings' of whatever words are chosen (*ibid. *p.134).*

In this study, the relationship between the knowledge of syntagmatic relations between words and reading comprehension is investigated.

1.2 Definition of Terms

The key words used in the title of the study are defined in this section. These words are:

Formulaic Sequences: In this study a formulaic sequence refers to:

> *A combination of two or more words which co-occur more frequently than expected by chance and which carries a single meaning.*

For the study described in chapters 3-5, a minimum threshold of 5 occurrences in the BNC was used. Collocations like '*growth industry*', idioms like '*to cover a multitude of sins*', phrasal verbs like '*to turn out*' and poly-words like '*in the making*' are different types of formulaic sequences used in this study.

Reading comprehension: The process of deriving meaning from text.

Understanding Formulaic Sequences: Some formulaic sequences are not semantically transparent, that is, their meaning is not the sum of the individual component meanings. Idioms are included in this category. In this study, the knowledge of the meaning of some opaque formulaic sequences like idioms is tested.

Recognition (awareness) of Formulaic Sequences: This phrase refers to the readers' ability to recognise sequences of words which are formulaic. Recognition is defined by Harley (2001) as "identifying an item as familiar" (p. 141).

1.3 Significance of the Study

In the context of learning English as a foreign language, reading plays an important role. In most EFL situations, especially in higher education, English is the medium of instruction and the resources and academic materials are in English. Moreover, students in EFL contexts do not usually have access to native speakers and authentic materials and texts are the only reliable source of target language for them. As a consequence, reading for these students is not only a source of information, but it is also a means through which they may learn English or improve it. In this vein, Carrell and Grabe (2002, p. 234) point out, "L2 readers in academic settings most often need to develop 'reading for understanding' and 'reading to learn'".

On the one hand, in the literature on second language reading comprehension, various accounts have been given of the factors which either promote or impede effective reading. One of these factors which has attracted the attention of researchers is vocabulary. It is widely acknowledged by scholars

that vocabulary knowledge is the key to successful apprehension of text (Anderson 2000; Read 2000; Qian 2002). In this vein, Laufer (1997, p. 20) maintains, "no text comprehension is possible, either in one's native language or in a foreign language, without understanding the text's vocabulary". In the past decades or so, many studies have been done, many articles have been written and many books have been published to address the significant role vocabulary plays in understanding text.

Three different views regarding the relationship between vocabulary and reading have been introduced (Nation 1993; Anderson & Freebody 1981). According to the Instrumental view, vocabulary knowledge is a prerequisite to reading comprehension. The Aptitude view considers that a large vocabulary and good reading comprehension are both the result of high mental ability and no causal relationship between the two is expected to be seen. According to the Knowledge view, vocabulary knowledge equals world knowledge which, in turn, increases the possibility of comprehension. With respect to the relationship between vocabulary knowledge and reading some practitioners have proposed that "the threshold for reading comprehension is, to a large extent, lexical" (Haynes & Baker 1993, p.21).

On the other hand, in recent years, due to improvements in computer analyses and the appearance of corpus linguistics, scholars in the field of language acquisition have shifted their attention from single words to multi-word sequences as meaning bearing units of language. In this respect, Sinclair (1996, p. 82) concludes that "units of meaning are expected to be largely phrasal" suggesting that two principles are in action when producing or comprehending a piece of language: the open-choice principle and the idiom principle. According to the open-choice principle constructing meaning is "the result of a large number of complex choices. At each point where a unit

is completed (a word or a phrase or a clause), a large choice *opens up* and the only restraint is grammaticalness" (Sinclair 1991, p. 109, emphasis in the original). Sinclair does not see the open-choice principle as the only explanation of language, and maintains that "in many instances 'semi-preconstructed phrases' are used" (Jones and Haywood 2004, p. 269). According to the idiom principle, these semi-preconstructed phrases act as single choices. Therefore, when using language, once the initial choice of meaning is decided upon, it is often the selection of preconstructed phrases rather than a series of single words that follows.

As a result of this reorientation in language description, the units of semantic analysis are not considered to be individual words but word combinations. Individual words are not regarded as basic units for description of meaning any longer as it is argued that it is the environment of occurrence, that is the other items surrounding a particular word, that determine its meaning. Accordingly, it is argued that students need much more than single words and their dictionary meanings. They need something to be stored in memory as a whole which can be retrieved and processed easily, without the need for every single word to be processed separately. These ready-made chunks are referred to differently by different scholars, for instance 'formulaic sequences' (Wray 2000, 2002a), 'fixed expressions' (Moon 1998), 'lexical phrases' (Nattinger & DeCarrico 1992), 'lexicalized sentence stems' (Pawley and Syder, 1983), 'routine formulae' (Coulmas 1979) to name a few. Different classifications have also been developed based on a variety of criteria including fixedness, transparency and length.

There is a consensus among scholars that a considerable proportion of our daily language production is comprised of formulaic sequences and studies addressing these sequences have proved their significance in language

acquisition, both in L1 and L2. However, almost all of these studies are dedicated to investigating the role of formulaic sequences in acquisition of language in general and the connection between the knowledge of these sequences and speech fluency and writing (Schmitt 2004, Wray 2004, Wood 2006). As a consequence, the connections between formulaic sequences and language comprehension in general and reading comprehension in particular have not come to the attention of researchers. This is to some extent due to the lack of a clear-cut, established method of estimating learners' knowledge of formulaic sequences to be used to investigate their role in comprehension. In this thesis, it is argued that if formulaic sequences are helpful in language acquisition and particularly language production, then it becomes important to determine if they would improve language comprehension, particularly reading. If so, then the results will be useful in designing courses and writing materials to facilitate reading comprehension task for EFL learners. In other words, this study is significant, as its results would contribute to the literature on foreign language learning in general and L2 reading comprehension and formulaic sequences in particular.

1.4 Statement of the Purpose

Having taught English in Iran for about twelve years, I noticed that although Iranian students have a better ability in reading in comparison to other language skills, they still lag behind native speakers in this area of language comprehension. My students often complained about their inability to understand a text with ease despite being able to get its gist. In order to be able to answer the comprehension questions which ask for details, they have to grasp the meaning of the whole passage as well as the meaning of individual sentences and phrases.

This is rooted in the methodologies adopted and textbooks designed to teach English at both high school and university levels. In Iran, students start studying English at the age of 13 as one of the subjects to be covered at guidance school, the level between primary and high school. They continue for four more years at high school. However, English is given priority. Each year students just pass 64 hours in English classes which are mostly spent on grammar taught in their mother tongue. Students and teachers communicate in Farsi and speaking in English is limited to drills and exercises. Students usually read the reading texts aloud taking turns and do the comprehension exercises at the end of each text. Most of the time teachers translate parts of the texts which are problematic. In the textbooks, new vocabularies are introduced in sentence long contexts along with their definitions in English which are translated into Farsi either by the teacher or by the students using bilingual dictionaries. Students are encouraged to memorize the new words and their meanings and are rarely asked to use these words in new contexts. It should be mentioned that recently the Ministry of Education appointed a group of scholars to design new textbooks. The new textbooks are intended to increase students' communicative competence as new sections are added to the textbooks to familiarize students with different language functions. However, old habits die hard and teachers still resort to old teaching methods and focus on grammar as there is a myth among EFL learners in Iran that knowing English equals knowing grammar. In addition, most students plan to continue their education at university and as it requires them passing the university entrance examination (Konkour), they prefer much of the class time to be spent on techniques for the multiple choice tests which are mostly grammar based.

At universities, students majoring in English pass the first two years studying basic courses such as reading, writing, grammar, listening and speaking.

Teachers who teach these courses are not native speakers and the graduates from English speaking countries do not usually teach basic courses and prefer to teach courses on language theories and teaching methodologies. The reading textbooks are usually designed by Farsi speakers in an attempt to reduce the cultural impact of original textbooks. Teacher–student interaction has to be in English, however in the first terms, this is limited to students who have studied English in private language schools. Whenever communication between the teacher and students breaks down they resort to Farsi. In reading classes, teachers mostly focus on new words and new vocabularies are introduced in the texts and their definitions are given in English. When necessary, teachers give the Farsi equivalents of new words. Although teachers try to encourage students to use monolingual dictionaries, students prefer to use bilingual dictionaries and write the Farsi equivalents of new words above the words in their textbooks to help them get the meaning of the text while reading. Reading classes are mostly devoted to reading passages students have already read at home and extensive reading is limited to a couple of short stories or simplified story books, which teachers do not usually find the time to check, to see whether students have got problems understanding them. Strategies for better reading such as guessing the meanings of new words, skimming and scanning are also introduced in reading classes, however little practice is done in this regard. As there is no separate course devoted to learning vocabulary, a reading class is a vocabulary class as well and 50% of questions in the final exam are devoted to new vocabularies.

ESP (English for Specific purposes) classes are reading based. The main goal in ESP classes is set to be improving the reading ability of the students in order to enable them to understand texts from the original sources in their fields. Despite the priority that reading has received, not much has been done

to improve the teaching of reading comprehension in ESP classes. In a typical ESP class, students have to read passages which are mostly technical and answer comprehension questions and do some grammatical exercises. The passages are translated into Farsi by the teacher as students do not usually have a good command of English. Farsi is the medium of instruction and both teachers and students speak Farsi. These classes are teacher-centered and students are made to memorize new vocabularies and their Farsi meanings and learn grammatical points, and strategies for better reading are not usually introduced or practiced in these classes.

The similarity between classes for English majors and ESP students is that students are not normally familiarized with the notion of collocations and lexical phrases and the focus is on single words as units of meaning. Even new vocabularies in the textbooks are not introduced along with their collocates. Formulaic sequences in general and idioms in particular are not the focus of class. In both English as a specific purposes and general English courses, I had students who struggled to get the meaning of phrases although they knew the meaning of the individual words comprising the sequence. In addition to this problem, most of the students are not able to identify meaningful chunks of language while reading. As a consequence, they put too much effort and time in processing each word separately, and the process of putting the meanings of single words together to extract the meaning of the phrase does not always lead to satisfactory apprehension of the sequence, which in turn hampers comprehension of the whole sentence.

Scholars investigating L1 and L2 reading comprehension have also pointed out that a good reader "chunks a text into **sense groups**, units of meaning each consisting of several words" (Nuttall 2000, p. 55, emphasis in the original). It is argued that it is quicker and takes less effort to fit together "the

sense of two or three chunks than to do the same with all the separate words that compose them" (ibid, p. 55).

Having this in mind, I came across a book which included a chapter entitled 'The nature of Lexis' from Michael Lewis's *The Lexical Approach.* The chapter provided me with inspiration to implement some of the principles of the Lexical Approach in my classes, and it had remarkable results. Two intact reading classes were selected and using the TOEFL test, the homogeneity of the classes and the level of their reading ability were checked. The same teacher and teaching methods were used in both classes with the exception that the experimental group received instruction on the nature of formulaic sequences and their functions. Within weeks, the students could identify lexical phrases in the passages and could use them while attempting to explain the ideas in the texts. They not only outperformed the control group in the reading comprehension section of the TOEFL test at the end of the term, they also got better results on their achievement test which was based on their textbook. So, was it the knowledge of formulaic sequences that helped them improve their reading comprehension? This question was the incentive for this thesis.

The purpose of this thesis is therefore to investigate the relationship between the knowledge of formulaic sequences and reading comprehension. It differs from the studies investigating vocabulary in that in vocabulary studies the researchers are mostly concerned with the relationships between depth and breath of vocabulary knowledge and reading comprehension. In the case of formulaic sequences, it is much harder to measure one's knowledge of formulaic sequences as no standardized test like Nation's Vocabulary Levels Tests (Nation, 1993) has been developed. In addition, as formulaic sequences refer to a complex notion and since knowledge of them can be assessed

considering different dimensions including their form, function (syntactic, semantic, discourse or pragmatic), meaning and even intonation, the scope of this study has been restricted to two aspects of these sequences, that is, recognition (awareness) and understanding of their meaning and the relationships of these variables with reading comprehension. In doing so, two separate studies, a main and a follow-up, were conducted. In the main study which was quantitative, gathering data through tests, the following research questions were posed:

1. Is there any relationship between Iranian EFL learners' recognition of formulaic sequences and their performance on a test of reading comprehension?

2. Is there any relationship between Iranian EFL learners' understanding of formulaic sequences and their performance on a test of reading comprehension?

3. Does the interaction of understanding and recognition (awareness) of formulaic sequences affect Iranian EFL learners' performance on a test of reading comprehension?

There were also some secondary questions which were not the main focus of the study but were investigated. They are as follows:

1. Is there any relationship between frequency of formulaic sequences and recognition of these sequences?

2. Is there any relationship between frequency of formulaic sequences and understanding these sequences?

3. Is there any relationship between Iranian EFL learners' recognition of formulaic sequences and their overall proficiency?

4. Is there any relationship between Iranian EFL learners' understanding of formulaic sequences and their overall proficiency?

5. Is there any difference between male and female subjects' performance on the recognition and understanding tests?

In the follow up study, which was qualitative, data were collected using recall and interview methods. It was designed to find out more about the role of formulaic sequences in reading comprehension using a design different from the main study. In addition it tried to answer the following question:

What kinds of knowledge sources do readers employs to infer the meanings of unknown formulaic sequences?

1.5 The Structure of the Thesis

This thesis is composed of five chapters. The first chapter is an introductory chapter in which some background information regarding the topic of the thesis has been provided. It has addressed the statement of purpose including research questions and the significance of the study as well. A section has also been devoted to the definition of the terms used in the title and the questions. Chapter two is devoted to an account of the relevant literature on formulaic sequences, reading comprehension and the connection between the two. In the third chapter the methodology used in pursuing the research is presented. This includes information about the design of the study, the participants, the instruments and the procedure employed in conducting the research. Chapter four addresses empirical results of the main study as well as

the results of the follow-up study. In the final chapter, discussion of the findings and conclusions drawn on the basis of them are presented, with some suggestions for future research, pedagogical implications of the findings, and limitations of this study.

Chapter II
Review of the Literature

2.1 Introduction

This chapter presents a review of the literature related to formulaic sequences, reading comprehension and the relationship between the two. The first part of this chapter provides information about the historical background of formulaic sequences, what they are, how they are defined and classified the criteria used to distinguish them. It also addresses broader issues such as the role of these sequences in first and second language acquisition and language processing and the processing of these sequences in language comprehension.

In the second half of this chapter issues related to reading comprehension are addressed. Background information is presented on reading process as well as different models and approaches suggested to describe reading process. There is also a discussion of the relationship between vocabulary knowledge and reading comprehension, the relationship between word recognition and reading ability followed by empirical studies supporting these views.

As this study concerns the connection between reading comprehension and the knowledge of formulaic sequences, that is, understanding and recognising them, the last part of this chapter is devoted to a discussion of the relationship between these variables.

2.2 Formulaic Sequences

2.2.1 Historical Background of Formulaic Sequences

Formulaic sequences play a pervasive role in human language. Nattinger and DeCarrico (1992) maintain that "just as we are creatures of habit in other aspects

of our behavior, so apparently are we in the ways we come to use language" (p.1). They attribute a ubiquitous role to ritualization in language behavior and argue that

> *routinized formulas and other sorts of prefabricated language chunks, which are products of this ritualization, seem to play a large part in both acquiring and performing language (ibid, p. 1).*

Coulmas (1981) also highlights the existence of 'routinized speech' and concludes that

> *much of what is actually said in everyday interaction is by no means unique. Rather, a great deal of communicative activity consists of enacting routines making use of prefabricated linguistic units in a well-known and generally accepted manner (p. 1).*

Wray (2000) refers to the widespread presence of formulaic sequences in normal languages and maintains that "any quantity of our language could be formulaic" (p. 466). She makes reference to the findings of corpus research with respect to the proportion of everyday discourse that can be accounted for by formulaic sequences. For example, Erman and Warren (2000) in their analysis found that 68.6% spoken and 52.3% of written English discourse consist of formulaic sequences. Altenberg (1990) concludes that 70% of language used by adults may be formulaic. In a similar vein, Pawley and Syder (1983, p. 213) assert that "the number of sentence-length expressions familiar to the ordinary, mature English speaker probably amounts, at least, to several hundreds of thousands".

There are, however, a number of negative reactions towards the status and function of ritualization in language acquisition and performance. Krashen and Scarcella (1978), for instance, having examined a number of studies and explanations on formulaic speech conclude that the learners have recourse to routines whenever conversational demands are present and whenever they have insufficient knowledge of the language. They cite Wong Fillmore's (1976) study which reveals her subjects learning English employ previously heard patterns to interact with others and assert that

> *clearly, the sort of early output demands Fillmore's subjects had imposed on them, and the routinized [i.e., formulaic] predictable input are not present in most language acquisition situations (1978, p. 295).*

Although Krashen and Scarcella (1978) acknowledge that formulaic sequences are "useful in establishing and maintaining relations" (p. 295), they do not assign a primary role to them in language acquisition and performance.

Krashen and Scarcella are not the only scholars who attribute little significance to the use and functions of formulaic speech. In fact, it was not until the late 1980s that formulaic speech was given its due attention in the context of language acquisition. This delay was due to the dominance of the structural approaches to language learning since the 1940s with their emphasis on syntax and phonology at the expense of lexis. Charles Fries, the founder of audio-lingual method, attributed most problems encountered by L2 learners to the "conflict of different structural systems" (Zimmerman 1997, p. 10). Fries was against the idea of presenting large amounts of vocabulary to students, asserting that the key to successful language learning is the mastery of basic sentence patterns and that "it is these basic patterns that constitute the learner's task. They require drill, drill, and more drill, and only enough vocabulary to make

such drills possible" (Hockett 1959,cited in Richards & Rodgers 2001, p. 52). In Fries's view, words are of four different kinds: function words, substitute words, words of negative/affirmative distribution and content words. The first three of these which are in fact grammatical have to be completely learned, but regarding the last group just a small number would do (Fries 1945, p. 39, cited in Carter & McCarthy 1988, p. 40). This state of negligence of the importance of vocabulary continued for decades.

The recent findings of research in language learning revealing that syntax, though essential in its own part, does not guarantee learners' success in effective communication (Meara 1980, 1983, Carter & McCarthy 1988, Carter 1998) and the investigations on the mental lexicon in psycholinguistics (Levelt et al. 1991, Singleton 1999) led to the emergence of a new wave of interest in lexis in EFL/ESL learning. As Richards (2000, p. xi) puts it,

> *vocabulary and lexical units are at the heart of learning and communication. No amount of grammatical or other type of linguistic knowledge can be employed in communication of discourse without the mediation of vocabulary.*

As a corollary of this reorientation in the role of lexis, the bridge between vocabulary and grammar collapsed and lexis started to be viewed "as having a syntagmatic 'horizontal' parameter alongside its paradigmatic 'vertical' parameter" (Farghal & Obiedat 1995, p. 317). As Sinclair (1987, p. 9) notes:

> *It thus appears that a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary one.*

Sinclair considers lexis and grammar "two interpreting ways of looking at form" (1996, p. 114). Language is therefore viewed as consisting of "grammaticalized lexis, not lexicalized grammar" (Lewis 1993, p. 89).

The recent wave of interest in lexis brought with itself a renewed impetus for the study of formulaic language. The study of vocabulary is no longer limited to the acquisition and retention of single words and their role in language learning and communication. Rather, it is argued that the native speaker's mental lexicon does not exclusively list "the syntactically atomic elements" (Webelhuth 1995, p. 32), an idea once supported by the proponents of Generative Grammar. On the contrary, it is composed of anything from a single word to a string of words or "big words" as Ellis (1996, p. 111) calls it which are syntactically irregular and semantically opaque.

With the advent of the computer and the emergence of corpus linguistics, research on formulaic language has grown considerably and attempts have been made to find its true status in language learning and use. Formulaic sequences have been the focus of research in language acquisition, sociolinguistics, psycholinguistics, phraseology, lexicography, corpus linguistics, and so on. In the context of language acquisition, formulaic sequences have been examined to determine their role in child first and second language acquisition and adult second language learning (Peters 1977, 1983; Nelson 1981; Bates et al. 1988; Wong Fillmore 1976; Ellis 1994; Weinert 1995; Granger 1998; Howarth 1998, Schmitt et al. 2004), aphasics (Van Lancker 1987; Van Lancker and Kempler 1987) language teaching (Nattinger and DeCarrico 1992; Lewis, 1993, 2000), and syllabus design (Willis 1990).

2.2.2 Definition and Characterization of Formulaic Sequences

Although several attempts have been made to date to define multi-word expressions in a comprehensive way (Becker 1975, Bolinger 1976, Coulmas 1979, 1994, Yorio 1980, Pawley and Syder 1988, Moon 1992, 1998, Nattinger and DeCarrico 1992; Howarth 1998, Hudson 1998), none of the definitions is inclusive enough to embrace all the characteristics and functions of these expressions. Wray (2002a) presents about fifty labels which have been used in research studies to refer to "a larger or smaller part of the set of related phenomena" (p.8). Unlike Weinert (1995) who concludes that "while labels vary, it seems that researchers have very much the same phenomenon in mind" (p.182), Wray (2002a) argues that "some of the terms shared across different fields do not mean entirely the same thing in all instances" (p. 8).

In effect, as multi-word units of language are multi-faceted in nature, scholars in different fields have looked at them from different perspectives and have focused on one or some of their salient features while trying to define them. Some definitions focus on aspects of form including length, stress, fixity and or grammaticality. As early as 1924, Jespersen makes a distinction between what he calls "formula or formular units and free expressions" (1924, p. 18). He proposes that "somethings in language–in any language-are of the formula character; that is to say, no one can change anything in them" (ibid, p. 18). He then gives '*How do you do*?' as an example of formulas and proposes that in this phrase "everything is fixed: you cannot even change the stress, saying '*How do you do?*' or make a pause between the words" (p. 18). Similarly, Pawley and Syder (1983) include features of form like variability and length of the strings in their definition of what they refer to as 'lexicalized sentence stems':

> *A lexicalized sentence stem is a unit of clause length or longer whose grammatical form and lexical content is wholly or largely fixed; its fixed elements form a standard label for a culturally recognized concept, a term in the language (pp. 191-2).*

This definition also includes the notion of lexicalization. Lexicalized units are stored and processed unanalyzed as if they were a simple lexical item. This concept is used by many other scholars in their definitions of multi-word units.

In some definitions of multi-word units, their function in communication is reflected, for example:

> *[Routine formulas] are highly conventionalized prepatterned expressions whose occurrence is tied to more or less standardized communication situations (Coulmas 1981, p. 2-3).*

Coulmas (1979) presents various arguments against considering the descriptive meaning of 'routine formulas' as the only criterion in their definition and continues:

> *an adequate analysis of the meanings of R[outine] F[ormula]s depends heavily on a proper description of their respective situational context. Only knowledge of the relevant dimension of social situations (and their relative weight) guarantees an understanding of the meaning of formulae which are tied to them (1979, p.242).*

Frequency of occurrence, as a distinguishing feature of multi-word combinations, comes from the field of corpus linguistics. It is argued that "if a sequence is frequent in a corpus, this indicates that it is conventionalized by the speech community, at least to some extent" (Schmitt 2004, p. 2). Nattinger and Decarrico (1992), refer to multi-word strings as 'lexical phrases' and give the

following definition which include, among other things, the concept of frequency:

> *They are multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time (p.10).*

However, Hickey (1993) maintains that "we must not rule out the possibility that an utterance which does not occur repeatedly is a formula" (p.33). In a similar vein, Wray and Perkins (2000) argue that there are some formulaic sequences which are culturally familiar to a speech community in spite of "their comparative rarity in a real text (e.g ...Time for bed, said Zebedee; ...)" (p. 7).

In her definition of what she refers to as 'formulaic sequence', Wray (2002a), points to the notion of storage and retrieval of these sequences as single units. Her definition includes a wide range of multi-word units and draws upon criteria from both linguistics and psycholinguistics. She defines a formulaic sequence as:

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

This definition encompasses a wide range of strings from "tightly idiomatic and immutable strings, such as *by and large*, which are both semantically opaque and syntactically irregular" (Wray & Perkins 2000, p. 1), to sequences which are

transparent and flexible, like the syntactic frame *a ... ago* that different lexical items fit into to make *a year ago*, *a day ago* or *a short while ago*.

In the above examples of different definitions of formulaic sequences, aspects of frequency, lexicalization, variability, length, memorization and function were introduced as the defining features of these sequences. There are some other features which are used by some scholars as the criteria for determining if a given string is formulaic. One criterion is non-compositionality or internal structure of strings, that is, the string does not have to be grammatically regular or semantically logical. Institutionalization, that is "the process by which a string or formulation becomes recognized and accepted as a lexical item of the language" (Bauer 1983: 48 and *passim*, cited in Moon 1998, p. 7) is another condition for a string to be classified as a formulaic sequence. And the phonological criterion which refers to whether a sequence is "produced with an intact intonation contour (suggesting the sequence is stored as a whole)" (Schmitt & Carter 2004, p. 2) is also included by some scholars.

It should be mentioned that not all formulaic sequences include all of these features and that having just one or two of these characteristics does not necessarily distinguish a word string as formulaic.

2.2.3 Categorization of Formulaic Sequences

The literature on formulaic sequences is full of suggested categories, (Alexander (1978, 1984); Bolinger (1976); Coulmas (1979, 1994); Howarth (1998); Lewis (1993, 1997); Nattinger & DeCarrico (1992); Moon (1998) and Yorio (1980)) including some or all of the types that follow: phrasal verbs, polywords, idioms, irreversible binominals, collocations, verbatim texts, situational utterances, proverbs, quotations, clichés, catchphrases, discourse structuring devices,

formulas, similes, sayings, meta-messages and metaphors. There are some categories not included in any of the classifications made and some which have been used in most taxonomies but under different labels. Simple collocations, for instance, are "variously included and excluded from taxonomies" (Wray 2002a, p. 46). In this regard, Hudson (1998) asserts that "an inherent shortcoming of these typologies is that the categories are neither discrete nor comprehensive" (p.13).

The inclusion or exclusion of categories in taxonomies may be on the grounds of their formal features, their function or their meaning. However this does not mean that these criteria are mutually exclusive. Some classifications may use a combination of form, function and meaning criteria. An example of each type of classification will be given below. In the next part, two types of formulaic sequences that have been included in almost all classifications of these sequences, and have caused disagreement as how they should be defined, that is, collocations and idioms, are dealt with in more detail.

Carter (1998, pp. 68-69) has made his classification of multi-word units using two form-based criteria, that is, collocational restriction and syntactic structure, plus semantic opacity which has to do with meaning. Wray (2000) concludes,

> *Cross-associations such as these between form and function, and also between form, meaning and provenance, or between any other subsets, are probably nearer the truth than single-parameter categorizations (p. 48).*

Among function-based classifications of formulaic sequences Yorio's (1980) taxonomy is commonly cited in most articles addressing formulaic language. His taxonomy has two main types: idioms and routine formulas with the latter having five sub-types of formulaic sequences. It includes situational formulas,

which fit certain conversational parameters, e.g., *this hurts me more than it hurts you*; stylistic formulas, which are specific to particular registers of language, e.g., *ladies and gentlemen*; ceremonial formulas, which are used in ritualistic interactions, e.g, *Dear Mrs.*; gambits, which organize interactions or activities, e.g., *Pardon me* or *it's your turn;* and euphemisms which are "avoidance formulas used to deal with situations that require direction" (Yorio 1980, p. 238), e.g., *Mr. Smith is no longer with this company.*

Meaning is another criterion which is used by some scholars. Cowie (1988), for instance, maintains, "word combinations can be divided into major groups, which differ according to the kinds of meaning which their members convey and to the structural level at which they operate" (p.133). Pure idioms, that is, idioms in the strict sense, such as *blow the gaff* and figurative idioms which are slightly less opaque, such as, *catch fire* are among the categories that he used in his classifications of formulaic sequences. Biber *et al* (1999) apply "opacity of meaning" as one of their criteria for categorizing multi-word strings. Yorio (1980) classifies conventionalized language into five categories from a semantic point of view. These are transparent, semi transparent, opaque, situationally ambiguous and ambiguous with respect to the intention of the speaker idioms or routine formulas.

A summary of the classifications suggested by some scholars is presented in Table 2.1. It shows that some scholars have got inspiration from the work of their pioneers in their categorization of multi-word expressions. For example, Nattinger's (1980) classification of lexical phrases is to a large extent similar to that of Becker (1975). He selects five categories from Becker's classification and substitutes 'meta messages' for 'deictic locutions'. However, these labels refer to the same lexical item. Lewis's (1993) categorization of multi-word items also

seems to be based on Nattinger and DeCarrico's (1992) classification of lexical phrases.

Scholar	Classification
Becker (1975) **lexical phrases**	1. **polywords:** *for good* 2. **phrasal constraints:** *by sheer coincidence* 3. **meta-messages:** *for that matter ...* 4. **sentence builders:** *(A) gave (B) a (long) song and dance about (a topic)* 5. **situational utterances:** *How can I ever repay you?* 6. **verbatim texts:** *Better late than never*
Nattinger (1980) **lexical phrases**	1. **polywords:** *put up with* 2. **phrasal constraints:** *a year ago* 3. **deictic locutions:** *as far as I know* 4. **sentence builders:** *not only X but Y* 5. **situational utterances:** 6. **verbatim texts:** *the public seldom forgives twice*
Yorio (1980) **conventionalized language**	1. **idioms:** *red-herring* 2. **routine formulas:** a. **situation formulas:** *You had to be there* b. **stylistic formulas:** *to whom it may concern* c. **ceremonial formulas:** *Dear Miss* d. **gambits:** *Let's call it a day* e. **euphemisms:** *he passed on/away*
Alexander (1984) **fixed expressions**	1. **idioms:** a. **phrasal verbs:** *to give away* b. **'tournures':** *to kick a bucket* c. **irreversible binomials:** *cash and carry* 2. **discourse-structuring devices:** a. **greetings/formulae:** *Long time no see!* b. **connectives, 'gambits':** *For a kick off ...* 3. **proverbs:** *You scratch my back and I scratch yours.* 4. **catchphrases:** *What's up doc?* a. **clichés:** *live happily ever after* b. **slogans:** *Your country needs you.* 5. **quotations/allusions:** *You've never had it so good.*
Cowie (1988) **multi-word units**	1. **pragmatically specialized:** *good morning* 2. **semantically specialized/ idiomatic:** *kick one's heels*
Nattinger & DeCarrico (1992) **lexical phrases**	1. **polywords:** *for the most part* 2: **institutionalized phrases:** *a watched pot never boils* 3. **phrasal constraints:** *a day ago* 4. **sentence builders:** *I think (that) X(assertion)*

Lewis (1993) **multi-word items**	1. **polywords:** *of course* 2. **collocations:** *prices fell* 3. **institutionalized expressions:** *Just a moment, please*
Howarth (1998) **word combinations**	1. **functional expressions** a. **idiomatic:** *for a kick off* b. **non-idiomatic:** *What's up doc?* 2. **composite units** a. **grammatical:** (preposition+noun) *under the table* b. **lexical:** (verb+noun) *blow a trumpet*
Moon (1998) **fixed expressions and idioms (FEIs)**	1. **anomalous collocations:** a. **(syntactically) ill-formed collocations:** *at all* b. **cranberry collocations:** *kith and kin* c. **defective collocations:** *in effect* d. **phraseological collocations:** *in/into/out of action* 2. **formulae:** a. **simple formulae:** *alive and well* b. **sayings:** *an eye for an eye ...* c. **proverbs:** (1). **metaphorical:** *every cloud has a silver lining* (2). **Non-metaphorical:** *enough is enough* d. **similes:** *as good as gold* 3. **metaphors:** a. **transparent metaphors:** *behind someone's back* b. **semi-transparent metaphors:** *pecking order* c. **opaque metaphors:** *kick the bucket*

Table 2.1 Summary of classifications of formulaic sequences (adapted from Nelson 2000)

It should be noted that most of the classifications of formulaic sequences have to rest on the idea that formulaic sequences like certain other aspects of language fall along continua, "since it is sometimes difficult to distinguish sharp boundaries between categories" (Nattinger & DeCarrico 1992, p.38). As Nattinger and DeCarrico (1992, p. 34) point out,

> *it is more likely that what constitutes a pattern and what does not is relative, a matter of degree instead of kind, for one usually finds a continuum in the amount of variation involved,*

from more invariable and frozen forms (such as idioms and clichés) to less invariable (non-canonical) forms.

As Howarth (1998) asserts , "it is essential to see the categories as forming a continuum from the most free combinations to the most fixed idioms, rather than discrete classes" (p. 35). In general, formulaic sequences can be classified as falling along different continua such as variability vs fixity or fixedness, continuity vs discontinuity, and compositionality vs non-compositionality or transparency vs opacity involving. Variability vs fixity refers to the degree of flexibility a formulaic sequence allows, both syntactic and semantic. Using this continuum, collocations are placed at one end of the line and idioms at other. Continuity vs discontinuity makes reference to whether a multi-word expression consists of an unbroken sequence of words like the poly-word *in a nutshell*, or whether it is interrupted by variable lexical fillers like '*My point in all this is that...*' which is broken by the lexical filler *in all this*. Compositionality vs non-compositionality relates to meaning. Non-compositional sequences do not take their meaning from their component parts. Examples include: *out of the blue, raises its ugly head, talk turkey*, and *cut the mustard.* Free collocations and pure idioms fall at the two ends of this continuum. A schematic representation of these continua is shown in Figure 2.1.

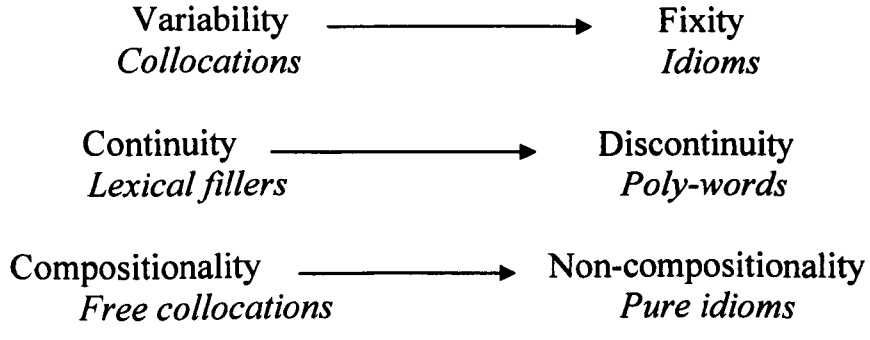


Figure 2.1 Schematic representation of continua used to classify FSs

To sum up, different classifications of formulaic sequences have been proposed by scholars using a variety of criteria. Some types of word combinations are used in most classifications under different labels. When formulaic sequences are placed on different continua, a better classification of these sequences can be achieved, as it is not always easy to draw sharp boundaries between categories.

As mentioned at the beginning of this section, two types of formulaic sequences have always caused controversies among scholars as to which class they belong, how they should be defined and what their main features are. These are collocations and idioms which are dealt with in the following subsections.

2.2.3.1 Collocations

Collocations as a type of formulaic sequence have attracted huge attention from scholars and have been defined and treated differently. Some like Moon (1998) and Wray (2002a) have classified collocations as one type of formulaic sequence and some like Palmer (1933) and Nation (2001) use collocation as an umbrella term covering different types of word combinations. That is why Bahns (1993, p. 57) asserts, "regrettably, collocation is a term which is used and understood in many different ways". Whatever the approach, there is one feature in common in all definitions: co-occurrence of words.

Collocation was first used by Palmer in his monograph '*Second Interim Report on English collocations*'. Palmer (1933) concludes that "each [collocation] ... must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts" (p.4, cited in Nation 2001, p. 317). Hence, in Palmer's view collocations refer to lexical items that must be memorized and retrieved as wholes.

Although Palmer was the first to use the term, it is Firth who is known as the first linguist to use collocation as a technical term. He does not, in fact, give a definition for the term but maintains, "I propose to bring forward as a technical term, meaning by collocation, and apply the test of collocability" (Firth 1957, p. 194). He is in fact influenced by Greek scholars who believed that "words do not exist in isolation, and they may differ according to collocations in which they are used" (Robins 1967, p. 21). Collocations of a given word, according to Firth, "are statements of habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order" (Firth 1968, p. 181). He refers to meaning by collocation as a lexical meaning derived from the co-occurrence with other words in the same context. Therefore, "one of the meanings of night is its collocability with dark, and of dark, of course, collocation with night" (Firth 1957, p. 196). He uses syntagmatic and paradigmatic terms to refer to combinative and selective relations between lexical items. Syntagmatic or horizontal relations between words refer to their co-occurrence while paradigmatic or vertical relations concerns the words that are from the same class and can replace one another in the same context. For example, 'cold' collocates with 'night' as in 'dark night'; therefore they are syntagmatically related while 'cold' and 'dark' are in paradigmatic relations as both collocate with 'night' and can be substituted for one another and they are in the same class or set.

Neo-Firthians, among whom Halliday is one of the most well-known, followed Firth and developed his theory. Halliday (1966) refers to a collocation as "the linear co-occurrence relationship of lexical items, while a set is the grouping of members with like privilege of co-occurrence in collocation" (p. 153). Sinclair (1991) defines collocation as "the occurrence of two or more words within a

short space of each other in a text" (p.170). Partington (1998) refers to these definitions as the 'psychological' or 'associative' definitions and maintains that native speakers distinguish, by virtue of their communicative competence and lifelong exposure to the language, between what is usual and normal and what is unusual collocation.

Other scholars like Hoey (1991), Lewis (2000) and Williams (2002), put emphasis on the frequency of occurrence in their definitions of collocation. Hoey (1991) refers to the statistical definition of collocation as "the relationship a lexical item has with items that appear with greater than random probability in its (textual) context" (p. 6-7). In the same vein, Lewis (2000) refers to the statistical significance of collocations and asserts that collocations are "words which are statistically much more likely to appear than random chance suggests" (p. 29).

According to Gitsaki (1996), there are three schools of thought on collocation: the lexical composition approach, the semantic approach and the structural approach. The proponents of the lexical composition approach, with Firth as its pioneer, argue that "words receive their meaning from the words they co-occur with" (Gitsaki 1996, p. 10). Lexis is, therefore, considered as independent from grammar. However, it does not mean that the role of grammar is neglected. As mentioned before, Firth (1951) asserts that part of the meaning of a word is determined by the words with which it collocates and that the lexical relations should be perceived as syntagmatic rather than paradigmatic or structural. Halliday (1966) follows Firth in this respect and maintains that "it is part of the meaning of 'past' that it contrasts with 'present', and it is part of the meaning of *strong* that it collocates with 'tea'" (p.160). In the semantic approach, on the other hand, 'linguists attempted to investigate collocations on the basis of a semantic framework, also separate from grammar' (Gitsaki 1996, p.13; Cited in

Nelson 2000). The main concern was to discover why certain words collocate with each other. For instance, why 'blond' collocates with 'hair' but not with 'car'. The structural approach differs from the previous approaches in that it considers grammar as inseparable from lexis. It is argued that "collocation is influenced by structure, and collocations occur in patterns. Therefore ... the study of collocations should include grammar" (Gitsaki 1996, p. 17). In the same vein, Hunston et al. (1997) assert, "there are two main points about patterns to be made: firstly, that all words can be described in terms of patterns; secondly, that words which share patterns also share meanings" (p. 209).

Collocations are of two main types: lexical and grammatical. Carter (1998) refers to collocation as frequent co-occurrence of words in a language and maintains that

> *these patterns of co-occurrence can be grammatical in that they result primarily from syntactic dependencies or they can be lexical in that, although syntactic relationships are involved, the patterns result from the fact that in a given linguistic environment certain lexical items will co-occur (p.51).*

Examples of grammatical collocations include: *agree with, advantage over, adjacent to, by accident,* and *afraid of*... . "They consist of a noun, an adjective, or a verb, plus a preposition or grammatical structure such as an infinitive or clause" (Bahns 1993, p. 57). Lexical collocations consist of open words, i.e., nouns, verbs, adjectives and adverbs. For instance, the combination *'commit suicide'* is a lexical collocation consisting of two open class words occurring in the pattern 'verb + noun'. Other examples include: *naked eye*, *curry favour*, and *good luck*. A notion related to grammatical collocations is colligation suggested by Firth (1957) referring to "the grammatical associations a word may have" (Hoey 2003, p. 6), "the positions it prefers' and what a word '**typically** does

grammatically'" (Hoey 2000, p. 234 – emphasis in the original). In other words, "certain words are primed to occur in (or avoid) certain grammatical functions" (Hoey 2005, p. 13). For instance, *happen* is primed to avoid the passive voice.

There are inconsistencies as to what makes a collocation, that is, the notion of collocational span which refers to "the number of lexical items on each side of a node [an item whose collocations are studied] that we consider relevant to that node" (Sinclair 1966, p. 415). In the past, no optimal collocational span was set and in the absence of computerized lexical analysis software it was not easy to set one. As Sinclair (1966) puts it, "we reject, for the moment, the suggestion that degree of proximity within the chosen boundaries of collocation should be considered of primary importance, thus also rejecting a neat but dangerous way of fixing the boundaries of each collocation" (p. 414). However, Sinclair (1966) set a 3:3 span as he points out "we will take the span as 3 lexical items before and after each node". Later, in 1991, he changes the span to 4:4 when he maintains "little of collocational interest is found outside a span of 4:4" (Stubbs 1995, p. 4) and continues that he would do more research to find out what the optimal collocational span would be (Sinclair 1991, p. 106). Scott and Tribble (2006) tackle the problem concluding that "it is likely that co-occurrence of collocates beyond the span of four or five words from the node will ... cause useful information to drown in a sea of noise" (pp. 34-35). In Scott and Tribble's opinion, a suitable solution to the issue is to note that "the span determines what the researcher can in practice identify as collocates, not the true typical distance between node and collocate" (ibid. p. 36). Cheng et al. (2006) assert that when studying word associations, it should be noted that although 'node' and 'collocate' are "convenient terms to use, the term 'node' does not imply a hierarchy between it and its 'collocate', and that 'node' words that have

'collocates' are themselves collocates if the collocate is studied as the node" (p. 414).

Another issue which is related to the notion of collocational span is the way in which word combinations are identified and categorized. To find instances of multi-word units, n-gram, gapped n-grams or skipgram and concgram searches are used. N-grams, also known as 'lexical bundles' (Biber et al. 1999, p. 993) or 'clusters' (Scott 1997, p. 41) refer to "contiguous words that constitute a phrase, or a pattern of use, and that recur in a corpus" (Cheng et al. 2006, p. 412). Based on the number of words comprising a phrase, n-grams are categorized into bi-grams (e.g., *I see*), tri-grams (e.g., *in order to*), and so on. Although n-grams or lexical bundles can be as long as five or more words, most studies focus on three or four-word bundles, as "there is usually some overlap between the most frequently occurring strings of any given length" (Nesi & Basturkmen 2006). 'N-grams' are limited to only strictly contiguous word combinations. Hence, many instances of collocations or word associations that are non-contiguous (e.g., *not only... but also...*) are not covered. Unlike n-grams, skipgrams include both contiguous and non-contiguous word associations. However, they are restricted to 3-word skipgrams and four 'skips' (Wilks 2005), that is, if two associated words are more than four words apart, they are not found. And like n-gram searches, skipgram searches cannot cover either constituency variation (when the associated words are separated by two or more words, e.g., *this is the case* which can be realized as *this is the most special case* or *this is seldom the case*) or positional variation (when the associated words occur in different position relative to one another, e.g., *'you know... I mean'* which can be realized as *'I mean ... you know'*). A new of way of discovering and categorizing word combinations which is an improvement on n-grams and skipgrams is the concgram. A Concgram is defined by Cheng et al. (2006, p. 414) as "all of the

permutations of constituency variation and positional variations generated by the associations of two or more words". As the concgrams of a corpus are identified and generated with both positional variation and constituency variation being handled and as the concgram search engines are fully automated, that is, they do not need prior input from the user, more possible word combinations or collocational patterns existing in the corpus are likely to be discovered. As Cheng et al. (ibid) point out "many concgrams reveal patterns of collocation which would not have been uncovered, relying on intuition alone or other search engines" (p. 431).

Several attempts have been made to classify collocations. Nesselhauf (2005) adopts a phraseological rather than frequency-based definition of collocation and considers only classifications of restricted collocations. She divides the classifications into three main types. The first type are classifications based on the syntactic characteristics of the collocation, that is according to the word classes of their elements, e.g., adjective + noun (*heavy smoker*) or noun+ noun (*piece of advice*). In the second type, restricted collocations are classified on the basis of the semantic characteristics of the collocator. Cowie (1991, 1992), for example, distinguishes between verbs with a 'figurative' , a 'delexical' and a 'technical' (or 'semi-technical') meaning. *Deliver a speech*, *make recommendations* and *try a case* are examples of the collocations including these verbs. The third type of classification is based on the commutability of the elements of a collocation. Howarth (1996) offers a classification which is limited to verb-noun collocations. For instance, in the collocations *adopt/accept/agree to a proposal/suggestion/recommendation/convention/plan*, the noun has freedom of substitution, but there is some restriction on the choice of verb. As an example of collocations with complete restriction on the choice of both elements, *curry favour* is a typical example.

For the purpose of classification of collocations, some scholars have applied such criteria as fixedness, which refers to the extent to which collocates can be substituted for by other words and suggest a continuum ranging from absolutely fixed to relatively fixed to very free (Lewis 1997,p. 255). Howarth (1998) introduces restricted collocability, semantic specialization, and idiomaticity as his criteria for categorizing collocations. Along his proposed continuum lie free combinations or free collocations, restricted collocations, figurative idioms and pure idioms. Table 2.2 depicts a summary of Howarth's collocational continuum.

	free combinations	**restricted collocations**	**figurative idioms**	**pure idioms**
lexical composites verb + noun	*blow a trumpet*	*blow a fuse*	*blow your own trumpet*	*blow the gaff*
grammatical composites preposition+ noun	*under the table*	*under attack*	*under the microscope*	*under the weather*

Table 2.2 Collocational continuum (Howarth 1998, p.28)

Free collocations are composed of elements used in their literal senses and they can be freely combined and recombined (*make a decision, take a decision*). Restricted collocations, on the other hand, "allow a degree of lexical variation" (Cowie and Mackin 1975, p. xv) and "have one component ... that is used in a specialized, often figurative sense only found in the context of a limited number of collocates" (Howarth 1998, p. 28). Figurative idioms and pure idioms are not considered by some scholars as subclasses of collocations. Hence, they are dealt with under the title of idioms.

In conclusion, collocations which refer to the co-occurrence of words have been defined and classified in different ways. Textual, statistical and psychological

definitions of collocations each emphasize respectively the space between the words which co-occur, the frequency with which they co-occur and the native speakers' intuition as which words co-occur.

2.2.3.2 Idioms

Idioms are another type of formulaic sequences which are sometimes treated as a subclass of collocations by some scholars (Howarth 1998; Nattinger & DeCarrico, 1992; Brown 1974; Jones & Sinclair 1974,1996) and as a type of formulaic expression by others (Moon 1998; Cruse, 1986; Cowie et al. 1983; Bolinger 1976; Aisenstadt 1979).

There is a consensus among scholars that idioms are "phrases or sentences whose figurative meaning is not clear from the literal meaning of their individual constituents" (Abel 2003). A typical example of idioms is '*kick the bucket*' meaning to die. In almost all definitions of idioms the criterion of non-compositionality, i.e., the meaning of the whole not being figured out according to the individual words, is included. However, there are inconsistencies with regard to other criteria in distinguishing an idiom from other formulaic expressions specifically collocations. Grant and Bauer (2004) summarize the criteria employed in several definitions of idioms into three themes of non-compositionality, institutionalization and frozenness.

Compositionality refers to the meaning of the idiom, that is, whether it carries the literal meanings of its individual parts. Some scholars regard idioms as non-compositional (Katz and Postal 1963; Aisenstadt 1979; Makkai 1975; Alexander 1978; Fernando and Flavell 1981), others consider them as partly compositional (Ruhl 1975; Nunberg 1978; Gazdar *et al.* 1985; Langacker 1986), and still others assert that when the meaning is clear they become compositional

(Nunberg *et al.* 1994). Institutionalization involves the fact that the expression is recognizable within a speech community. Frozenness refers to the degree to which an expression allows changes in word order while maintaining its meaning. It also refers to the "ways in which idioms allow neither permutation nor paradigmatic replacement nor addition or deletion of elements" (Grant & Bauer 2004). For example, the idiom '*She kicked the bucket*' loses its meaning if changed to passive '*The bucket was kicked by her*' or if one of its constituents is substituted by another word from the same semantic set as in '*She kicked the pail*'. However, scholars have different reactions with regard to this criterion. Some hold the view that "different degrees of possible fixity or frozenness both syntactic and semantic" (Carter 1998, p. 66) should be allowed.

Grant and Bauer (2004) argue that compositionality is the most distinguishing feature of idioms as when an expression is non-compositional it is, therefore, stored as a whole in memory and is institutionalized. They also claim that non-compositionality "involves some degree of fixedness" (ibid, p. 45). They argue that most of the criteria offered so far are unable to distinguish idioms from among other multi-word expressions because of being too general. In their view, in addition to being non-compositional an idiom has to be non-figurative. They define a non-figurative expression as an expression which cannot be reinterpreted pragmatically. For instance, in the sentence *Jack is hot under the collar*, "we claim the compositional untruth that Jack has increased body heat only in his neck/collar area is analyzed linguistically and proves to be pragmatically unlikely in context" (ibid, p. 50).

In his attempt to define idioms, Cruse (1986) uses the notion of semantic constituent. He argues that in the traditional definition of idiom as "an expression whose meaning cannot be inferred from the meanings of its parts" (p.37), there is an element of circularity and it should be substituted by the

following definition: "an idiom is an expression whose meaning cannot be accounted for as a compositional function of the meanings its parts have when they are not parts of idioms" (ibid, p. 37). Therefore, like other scholars he believes in the concept of non-compositionality as a distinguishing feature of idioms. He then adds that two things must be remembered in dealing with an idiom: first, it is lexically complex having more than one lexical constituent; second, that it is a single minimal semantic constituent and not divisible into semantic constituents. Thus, an idiom is "a lexical complex which is semantically simplex" (ibid, p. 37).

As shown in Figure 1 in the previous section, Howarth (1998) classifies idioms into figurative idioms and pure idioms. He defines figurative idioms as idioms which have 'metaphorical meanings in terms of the whole and have a current literal interpretation' (ibid, p. 28). Pure idioms, on the other hand, "have unitary meaning" (ibid) and their meanings are not inferable from the meanings of their constituent parts. Here again, the criterion which is used to distinguishes the two is the compositionality versus non-compositionality of the expressions.

To sum up, idioms are distinguished from other types of formulaic sequences as being non-compositional, that is, their meanings are not the sum of their constituent parts; frozen, that is, they do not allow variation; and institutionalized, that is, are recognized in a given speech community.

2.2.4 Functions of Formulaic Sequences

Several functions have been identified for formulaic sequences in the literature. Some of these functions have to do with the role they play in language acquisition and some are connected with their contribution to fluent and

accurate language use. Here, the functions they perform in language use are discussed and their role in language acquisition is dealt with in the next part .

The functions that relate to language use have to do with processing or communication and can be subsumed under two main headings: (1) psycholinguistic functions and (2) sociointeractional functions. The psycholinguistic function concerns the role they play in reducing processing. The speakers of language store a large body of formulaic sequences as whole and retrieve them easily from memory whenever they need them without having to construct new ones online. This saves processing effort. In fact it is one of the properties of the human brain to memorize prefabricated patterns. As Nesselhauf (2005) points out, 'psycholinguistic evidence indicates that the human brain is much better equipped for memorization than for processing' (p.2). Wray and Perkins (2000) subcategorize the processing (psycholinguistic) functions of formulaic sequences into three types: (a) short-cut in processing; (b) time-buying; (3) manipulation of information. The first function helps to increase production speed and fluency. The speaker does not have to construct from scratch the sequences which are stored and therefore the amount of new processing is reduced to only new utterances. Based on research such as that by Raichle (1998) and McCrone (1999), Wray and Perkins (2000) state that 'once the brain is familiar with a linguistic task, it is able to by-pass the processing route that was used to learn it' (p.16). The second function has to do with sequences which promote fluency and help the speakers to maintain their turn during planning. Formulaic sequences which are semantically equivalent to single words (e.g. *make a decision; decide*) help to provide rhetorical balance and pace the appearance of novel utterances. Fillers (*you know)*, turn-holders (*yes, but the thing is ...)*, discourse sharp markers (*firstly...secondly...),* and repetitions of preceding input are among time-buyers that help the speaker to

plan time without losing the turn. Manipulation of information is another function classified under the category of psychological functions of formulaic sequences and helps to reduce strain on memory and enables the speakers to gain and retain access to information otherwise unlikely to be remembered. Mnemonics, lengthy texts memorized on purpose and rehearsals like repeating a telephone number to oneself while searching for a pen are examples of formulaic sequences that help the speaker to retrieve information easily.

The socio-interactional function of formulaic sequences is concerned with their role in communication. Manipulation of others, asserting separate identity and asserting group identity are three socio-interactional functions that can be achieved by using formulaic sequences. Of course, this does not mean that these functions cannot be achieved via novel utterances. Manipulation of others is related to the fact that as we are not able to satisfy all our emotional, physical and cognitive needs by ourselves, we sometimes need the action of others to achieve what we want. This can be done through formulaic sequences that imply commands (*No Smoking*); requests (*Could you please pass the salt?)*; politeness markers (*I wonder if you'd mind* ...); or bargains (*I'll give you _for it*) (Wray and Perkins 2000). The second and third functions reflect the personal identity and group identity and of the speaker. To assert individual identity, the speakers may use story-telling skills (*You're not going to believe this, but* ...) or turn claimers (*Yes, but the thing is* ...), for instance, to be taken seriously or they may separate from the crowd using personal turns of phase (*I wanna tell you a story*). To assert group identity, the speaker may use 'in' phrases (*Me be dumb!*), group chants (*We are the champions*), institutionalized forms of words (*Be healthy* (in Farsi, in response to 'say hello to someone'), or ritual (*Our prophet, peace be upon him)* to indicate overall membership or they may use threats (*I wouldn't do that if I were you*), quotations *('I wouldn't want to belong*

to any club that would have me as a member' (Groucho Marx)), forms of address (*Your Majesty*) or hedges (*Well I'm not sure*) to indicate hierarchical relationships (Wray and Perkins 2000).

It is important to note that it is not just the speaker who benefits from the above mentioned functions. Wray (2002b) argues that

> *the less processing entailed in the decoding process, the more likely the hearer is to react appropriately to the message. There are two reasons. First, there is less danger of misunderstanding a familiar form, even if it is subject to distortion or disruption. Second, the hearer associates the particular expression with a pre-agreed meaning, the situational and pragmatic elements of which may not be derivable easily from the composition (p.119).*

In conclusion, formulaic sequences can help the speaker or hearer to produce and comprehend the language with processing effort and time reduced, they help the speakers to remain focused by marking discourse structure and signal the speakers' identity as an individual or as a group member. In other words, they are 'linguistic solutions to single, *non*linguistic (emphasis in the original), problems' (Wray 2002a, p.93). The functions of formulaic sequences are not by no means restricted to those mentioned as their main role is considered to be in language acquisition, both L1 and L2 which is the topic of the next part.

2.2.5 Formulaic Sequences and Language Acquisition

There is general consensus among scholars in the field of language acquisition that formulaic sequences have a significant role in different aspects of L1 and L2 learning. Schmitt and Carter (2004) refer to the role of formulaic language in L1 acquisition and proposes that if for L1 learners, "unanalyzed sequences

provide the raw material for language development ..., it is possible that they serve the same purpose for L2 learners" (p. 12). In case of child first language acquisition, Olson (1973) refers to the child's inability to construct certain utterances without recourse to formulaic phrases and says,

> *Such utterances manifest structures that are non-productive in the child's language at that particular stage, but the utterances are used as a unit for some specific semantic or pragmatic purpose without the child's knowing in some sense the internal structure of the string (p. 156).*

These units may enter the child's system either as unanalyzed strings which are strings of words or morphemes that show grammatical and/or lexical knowledge beyond the child's current generative capacities, drawn from the child's lexicon or repetitions of preceding input, or through a process by which word strings created by the child using its grammatical rules and lexical entries are stored whole (Wray 2002a).

Peters (1983) refers to the storage of combinations as fusion and asserts:

> *children ... also fuse speech sequences, seemingly as shortcuts to avoid having to construct them anew each time. This strategy, of course, makes perfect sense in view of the evidence that small children are, more than anyone, handicapped by short-term processing limitations. .... Thus we find children adopting stereotyped expressions that are neither copied directly from nor even directly reduced from adult usage, giving evidence that some sort of construction process may have gone on before the expression became frozen (p. 82).*

She claims that the child develops strategies to extract the units from heard speech, compares them phonologically with other units and stores them as new units. Then the child puts the knowledge about the stored units to use in picking

up further knowledge about the language. That is to say, the child starts to segment units into smaller ones and store these also as units. The child can then analyze new chunks in the linguistic environment.

There have been several studies of children acquiring their second language supporting the view that children attend to formulaic sequences in language input and adopt them for use and later analyze them (Wong Fillmore 1976, Hakuta 1974, Peters 1977, Clark 1974). Wong Fillmore (1976), for instance, studying five young Mexican children learning English, aged between 5 to 7, found that these children tried to imitate the previously heard input either with or without modifications to be able to interact with others. Then the formulaic sequences were added together, or embedded into novel utterances, to create new utterances. One, for instance, juxtaposed strings, such as *put it* and *right here, to make a new utterance. Finally, they attempted to dissect* sequences into smaller units. Just one of the children "tended to preserve his formulas rigidly in the form he learned them" (ibid, p. 493).

Hakuta (1974) who also conducted a study of a 5-year-old second-language learner finding the same results as Wong Fillmore, maintains that the children participating in the above studies are somehow different from first language learners as they are cognitively more mature and have increased social needs. They, therefore, "may be more prone to memorize long socially appropriate chunks of speech in order to maximize communication with a minimum of language" (Peters 1983, p. 13).

With regard to adult second language acquisition, it is acknowledged that there are similarities between children and adult language learning regarding formulaicity. Nattinger and DeCarrico (1992) state that:

> *There is no reason to think that adults would go about the [learning] task completely differently [from children]. In important ways, the language learning situation is the same for adults as for children, and makes it likely that an adult learner would also find prefabricated language an efficient way to begin to acquire a new language system (p. 27; Cited in Wray 2002b, p. 144).*

Ellis (1994), drawing on evidence from studies on the significance of word combinations in language learning, comes to the conclusion that much of language learning is in fact acquisition of memorized chunks. It is also acknowledged that adult L2 learners make a balance between the use of both analytic and holistic systems. It is indeed impossible to exclude one at the expense of the other. In other words,

> *without the rule-based system, language would be limited in repertoire, clichéd, and, whilst suitable for certain types of interaction, lacking imagination and novelty. In contrast, with only a rule-based system, language would sound pedantic, unidiomatic and pedestrian (Wray 1998, pp. 64-65).*

Unlike Nattinger and DeCarrico, Wray (2000a) takes the position that children and adults follow different patterns of acquisition as they are different regarding their

> *brain plasticity, stage of conceptual development, level of physical dependency, stage of literacy, sense of social identity, quality and quantity of input [they receive], affective factors, and so on (p. 471).*

She maintains that research on formulaic language learning of adults is also different from that of children as adults learn the language either in naturalistic

environments or through instruction. Data on naturalistic language learning, she argues, "tend to be anecdotal rather than reflecting a full record or disciplined periodic collection" (Wray 2002, p. 172) and data gained from class-taught research does not provide in-depth information on particular learners' profile over time, consequently affecting the interpretation of the results and their generalizability.

Studies examining formulaic sequences in adults acquiring L2 naturally have revealed different results. Schmidt (1983) found that his subject, Wes, used formulaic sequences as a communication strategy and that despite his fluency he was not accurate. In a similar vein, Yorio (1989), examining the written English of an adult learner, concluded that he used formulaic sequences like idioms and set phrases as a production strategy. However, Hanania and Gradman (1977) and Shapira (1978) found little evidence of use of formulaic sequences by their subjects.

Irujo (1993) studied the ability of 12 adult learners of English to translate paragraphs including cognate and non-cognate idioms from L1 to L2. She found that although they had achieved a high level of grammar and lexicon, their knowledge of idioms seemed to be lagging behind. This may be partly because the language addressed to learners contains few or no formulaic sequences and partly because learners do not pay attention to these sequences or avoid using them. And although these sequences are used frequently in TV or movies, it is not helpful because input without interaction and negotiation is not enough for learning (Irujo 1986).

Another aspect of formulaic sequences and L2 learning studied is the learners' tendency to overuse and underuse these sequences. Granger (1998) concludes that when learners learn a number of formulaic sequences that they feel confident in using, they tend to stick with these sequences which act as "islands

of reliability" (Dechert 1983, p. 193) for them. Bolander (1998), in an examination of recordings of free speech and a picture description task given to L2 learners of Swedish found that the subjects had a tendency to reproduce formulaic sequences, both where it was correct to do so and where it was not. Foster (2002), on the other hand, found evidence of underuse as the subjects of her study produced fewer formulaic sequences than native speakers. In other words, this implies that non-native speakers are poor users of formulaic sequences.

It has been argued by many scholars that formulaic sequences help L2 learners to communicate at a level beyond their linguistic knowledge. Hakuta (1976) views memorization and reproduction of formulaic sequences as central to successful learning and maintains

> *if learners always have to wait until they acquire the constructional rules for forming an utterance before using it, then they may run into serious motivational difficulties (p.333).*

Similarly, Ellis and Sinclair (1996) consider learning sequences important, for

> *much of language learning is the acquisition of memorized sequences of language (for vocabulary, the phonological units of language and their phonotactic sequences; for discourse, the lexical units of language and their sequences in clauses and collocations) ... . Short-term representation and rehearsal allows the eventual establishment of long-term sequence information for language (pp. 246-247).*

Nattinger and DeCarrico (1992) even go further and believe that prefabs can be used as input by the learner for the analysis of language. However, not everyone agrees that memorization of formulaic sequences might lead to

learning individual words and grammar of the L2 (Krashen and Scarcella 1978, Schmidt 1983, Granger 1998). Granger (1998) considers teaching formulaic sequences for the purpose of learning grammar is useless and asserts, "there does not seem to be a direct line from prefabs to creative language .... It would thus be a foolhardy gamble to believe that it is enough to expose L2 learners to prefabs and the grammar will take care of itself" (p. 157).

Overall, there is considerable amount of evidence both theoretical and empirical showing that children and adults learning their first or second language engage in both analytic and holistic types of speech, initially acquiring prepatterned chunks and later analyzing them and using the knowledge gained this way to facilitate overall language development. Hence, in language learning a balance should be established between creativity and formulaicity.

2.2.6 Formulaic Sequences and Language Processing

In the Chomskyan tradition, the idea that formulaicity and creativity go hand in hand in language acquisition was neglected and formulaicity was placed at the edge of accounts of linguistic processing. However many arguments have been developed over the last few decades for the relationship between formulaic and creative language. Pawley and Syder (1983) while accepting that "a generative grammar is part of what a person must know in order to be competent user of any language" (p.193) propose that "native speakers do *not* exercise the creative potential of syntactic rules to anything like their full extent, and that, indeed, if they did do so they would not be accepted as exhibiting nativelike control of the language" (ibid). Widdowson (1989) takes the position that "there are linguistic environments of a lexical kind which have conditioning effects on the application of syntactic rules, and that knowing this is an essential part of

knowing a language" (pp. 133-134). He then continues with the claim that communicative competence does not mean the ability to construct sentences from scratch by applying rules. It is, he argues

> *much more a matter of knowing a stock of partially pre-assembled patterns, formulaic frameworks, and a kit of rules, so to speak, and being able to apply the rules and to make whatever adjustments are necessary according to contextual demands (p.135).*

In an article on the Emergent Lexicon, Bybee (1998) points to the significant role of memory in language learning and states that

> *actual language seems to rely on memory much more than on abstract analysis....The brain is a powerful categorization device for the efficient sorting and storing of the pieces of our experience, including the units of language use... It is apparently easier to access, produce and comprehend a precompiled chunk than to assemble it part by part for production (p. 434).*

She, then, refers to the duality of linguistic knowledge and makes the distinction between propositional or representational knowledge and procedural knowledge and asserts that

> *a large portion of the stored knowledge that makes language possible is procedural knowledge and stored chunks are procedural chunks, embedded in context not just cognitively and socially, but also embedded physically in the production and comprehension systems along whose paths they run....Evidence for the procedural nature of linguistic chunks is the fact that they are affected by frequency of use (pp. 435-435).*

Wray (1992, 1998, 2002a & 2002b) also suggests a dual-system model in which language is processed both holistically and analytically. Analytic processing entails the construction and interpretation of novel, or partially novel, linguistic material. The holistic processing functions as a remedy to the limitations in short-term memory to create grammatical strings out of small units by rule and as a solution to the problems arising while interacting through language. This system of processing relies on prefabricated units stored and retrieved as a whole in memory. Holistic processing can deal with both prefabricated material and linguistic material for which the application of rules would yield the same results. These prefabricated units reduce processing effort on the part of the both speakers and listeners and as many of them contain pragmatic or situational meanings in addition to their literal meanings, they help the speaker to reduce the danger of the hearer failing to construe the implicatures accurately. Wray (1998) emphasizes that the two systems are connected and "together ensure that we are able to express (and understand) whatever we want, however mundane, poetic or unexpected" (p. 65) and that in

> *adult speakers, the relative balance of the two systems in operation appears to be in favor of the holistic, for we prefer a pragmatically plausible interpretation over a literal one, and we seem able to use with ease formulaic sequences whose internal form we have, apparently, never engaged with (Wray 2002a, p. 18).*

In a similar vein, Sinclair (1991) proposes a dual coding system of language processing. The two modes of construction and interpretation of language, he suggests, are the 'open choice principle' and the 'idiom principle'. The open choice principle involves using grammatical rules to combine words and construct sentences. This principle is consistent with the traditional slot-and-

filler model based on the idea of free choice of individual lexical items, with the main restraint observed being that of grammaticalness. According to this principle, a text is constructed "as the result of a large number of complex choices. At each point where a unit is completed (a word or phrase or a clause), a large choice opens up and the only restraint is grammaticalness" (Sinclair 1991, p. 109). He challenges the idea that the open-choice principle is the only explanation of language use in asserting that "it is clear that words do not occur at random in a text, and that the open-choice principle does not provide for substantial enough restraints on consecutive choices" (ibid, p. 110). The idiom principle, on the other hand, refers to the tendency of language to formulate phrases and idioms and that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (ibid).

He suggests that speakers tend to use both principles, switching from one to the other whenever necessary. He maintains that:

> *The first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle and quickly back again. Lexical choices which are unexpected in their environment will presumably occasion a switch (p. 114).*

In this light, it seems reasonable to say that both native speakers and L2 learners benefit from the formulaic nature of language as they do not need to go through the labor of generating an utterance. This reduces the processing effort as despite the vast storage capacity of memory, the processing of these memories is not done fast (Crick 1979, p. 219).

2.2.7 The Processing of Semantically Opaque Sequences

Among formulaic sequences some are semantically opaque, that is, their individual components do not contribute to their figurative meaning. These expressions which abound in English and are easily produced and comprehended by native speakers cause comprehension problems for non-native speakers (Irujo 1993; Fernando 1996) and have recently regained the attention of psychologists and linguists who are interested in the representation and processing characteristics of these sequences.

Several models have been proposed regarding the processing of these expressions. These models can be characterized as varying in terms of the status they consider for literal and figurative meanings of idioms as they are processed by language users. In general, the theoretical proposals developed for the processing of these expressions fall into two categories: compositional approaches and non-compositional approaches. Within each of these approaches, there are different linguistic and psycholinguistic perspectives on idiom representation and models of idiom processing.

The non-compositional approaches consider idioms as unitary word strings having syntactic and semantic behaviour similar to lexical entries, they are holistically retrieved and the semantic characteristics of their word components are critical in their interpretation. For example, the literal meaning of the idiomatic expression *'to kick the bucket'* as *'to strike a pail with the foot'* and its figurative meaning as *'to die'* do not overlap semantically according to this view.

Non-compositional definitions of idioms have received attention from both linguists and psycholinguists. According to some linguists (Chomsky 1980; Fraser 1970; van der Linden 1992; cited in Titone & Connine 1999), there is a

direct relationship between idiomatic meaning and syntactic function of the idiom. For instance, *'kick the bucket'*, and the intransitive verb 'die' display similar syntactic behaviour (Titone and Connine 1999). It is argued that there is an arbitrary semantic connection between the meaning of an idiom and its parts. However, there are restrictions on syntactic modifications. In other words, the proponents of this view "consider two different syntactic forms of an idiom as comprising two instances of the same idiomatic expression" (Titone & Connine 1999, p. 1657). For example, the phrase *'kicked the bucket'* is syntactically different from the phrase *'kick the bucket'* but has the same idiomatic meaning as it is held that the relationship between an idiom and its meaning is arbitrary. This view, however, is challenged by others like Wasow et al. (1983) as being unable to account for the arbitrariness of the syntactic flexibility of idioms. These scholars maintain that the meanings of some idioms are to some extent the function of their constituents and that some idioms are not as opaque as others (Cacciari & Tabossi 1993). Wasow et.al (1983, p. 109) point out that "the pieces of an idiom typically have identifiable meanings which combine to produce the meaning of the whole". They argue that the elements of an idiom do play a role in its meaning and that is a reason for its flexibility. For example, the idiom *'spill the beans'* can be modified as in *'she didn't spill a single bean'* without impeding its comprehension (Nunberg et al. 1994).

Some of the early psycholinguists have also supported the non-compositional view of idiom representation and comprehension. Several models have been proposed suggesting that idioms are stored and retained as single lexical items. However these models differ in terms of how and when these meanings are accessed. Some like *The Literal Processing Model*, suggested by Bobrow and Bell (1973), posit that the literal meaning of an idiom is first processed, and if this process fails, the idiomatic processing mode is activated and the figurative

meaning is retrieved from the 'idiom word dictionary' in which the idioms are stored as 'idiom words'. In other words:

idiom encountered → literal meaning activated → comprehension
or
no comprehension → figurative meaning activated → comprehension

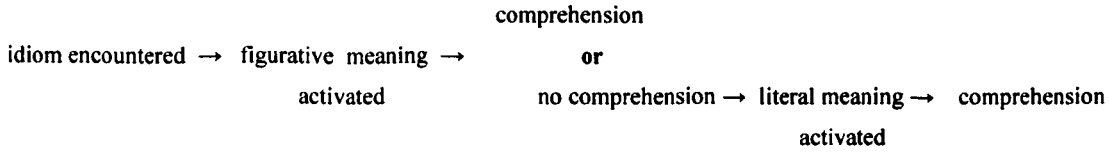
Similarly, Swinney and Cutler (1979) proposed the *Lexical Representation Hypothesis*, arguing for the storage and retrieval of idiomatic expressions as single items but different from the previous model in that they believe in the simultaneous processing of the figurative and literal meanings of idiomatic expressions. They argued that

> *computation of literal and figurative meanings is initiated following presentation of the first word of an idiom. Automatic retrieval of idiomatic meanings should, however, precede a complete specification of literal meaning which involves both the retrieval and compositional analysis of each component word (Titone & Connine 1999 p. 1657).*

In other words:

idiom encountered → literal meaning activated **and** figurative meaning activated → comprehension

In the *Direct Access Hypothesis*, Gibbs (1980) proposed that the literal meaning of the idiomatic expression does not influence its comprehension as the idiom has a strong conventional, figurative meaning having priority over the literal one. Therefore, according to this model, when an idiomatic expression is encountered it is the figurative meaning which is processed first, and then if it fails to be appropriate for the context, the literal meaning is processed. In other words:



The compositional approach to idiom comprehension, on the other hand, argues for the critical role of the constituent parts of idioms in their comprehension and interpretation. It is based on "the notion that the idiomatic meanings are built simultaneously out of literal word meanings and the specific interpretation of these word meanings within a particular context." (Titone & Connine 1999, p. 1661). Therefore, upon encountering an idiomatic expression, the language comprehension device builds the idiomatic meaning out of the literal meanings of the parts of the expression.

This view is, however, challenged by some scholars as being insufficient to account for the interpretation of idioms. It is argued that whether the literal meanings of the parts contribute to the meaning of the whole or not, the possibility of there being a "prepackaged meaning associated with very particular configurations of words" (Titone & Connine 1999, p. 1665) should not be ignored. This leads to quicker and more efficient processing of sequences which are idiomatic than those which are not.

Titone and Connine (ibid), in their *Hybrid Model of Idiom Comprehension*, propose a synthesis of non-compositional and compositional approaches to the interpretation and apprehension of idiomatic expressions based on the findings from psycholinguistic studies of these expressions. Thus, the meaning of an idiomatic expression is processed both holistically as a single lexical item and as a result of the semantic and syntactic analyses of its constituent parts.

There are some similar arguments in relation to the processing of idiomatic sequences by L2 learners which differ with regard to the role they assign to the

literal meanings of these sequences. Some scholars like Nelson (1992) ascribe a critical role to the figurative meaning in the course of idiom comprehension. Others like Liontas (2002) argue for the primacy of literal over figurative meanings. Still others like Abel (2003) argue that idioms should be represented in a dual way. In her proposed model referred to as the Dual Idiom Representation, it is suggested that "there is not only a lexical, but also a conceptual level of representation, and secondly that constituent and idiom entries co-exist at the lexical level" (ibid, p. 334-5). She conducted a study with non-native speakers of English and found that they showed a general tendency to judge opaque, non-decomposable idioms as decomposable. She ascribes a critical role to frequency and suggests that the extent to which the idiom is frequently encountered by non-natives determines the development of an idiom entry. She states,

> *non-natives who read English texts daily and thus come across idioms more frequently, which leads to the development of idiom entries, start to judge idioms as non-decomposable, just as natives do (ibid, p. 349).*

To sum up, as L2 learners receive instruction in idiomatic expressions after they become familiar with the literal meanings of their constituent parts, the figurative meanings become more salient than literal ones and are "likely to be much better established in their mental lexicons than the newly acquired [figurative] ones" (Cieslicka 2006, p. 121). Idiom familiarity, a concept proposed by Schweigert (1986), posits that

> *to be consistent with the assumption that the figurative meanings of idioms are stored in memory, the idioms must be*

familiar enough for the subject to know the idioms' figurative meanings (Schweigert & Moates 1987, p. 283).

2.2.8 Processing of Words and Formulaic Sequences in Reading

How readers access the meaning of words is addressed by de Bot et al (1997). Drawing on Levelt's model of L1 speech processing, they propose a model applicable to both comprehension and production in L1 as well as L2 processing, asserting that at the representational level, the organization and components which apply to comprehension are also applicable to production and that it is just in message realization that the modalities differ. They argue that "understanding both the nature of vocabulary and acquisition of this knowledge requires consideration of the organization of the mental lexicon and the mechanisms of language processing" and consider the Levelt's model "to be the most complete, empirically based language production model available" (ibid, p. 310).

It seems reasonable to provide a brief review of Levelt's model here. Levelt's production model consists of three separate levels of representation: the conceptual level, the lemma level, and the word form (lexeme) level. A lexical item's lemma information or lemma contains the syntactic and semantic information of that item. 'Lexeme' refers to the morphophonological form of the lexical items. Each of these levels has a central role in production and comprehension of language. Another concept which is central to the model is 'lexical hypothesis'. It refers to the mediating role of the mental lexicon in the formulation processes. The mental lexicon which refers to "the store of information about the words in one's language" (Levelt 1989, p. 6) comes between conceptualization and grammatical and phonological encoding. That is,

> *nothing in the speaker's message will by itself trigger a particular syntactic form, such as a passive or a dative construction. There must always be mediating lexical items, triggered by the message, which by their grammatical properties and their order of activation cause the Grammatical Encoder to generate a particular syntactic structure (ibid, p. 181).*

According to the model, the conceptualizer, that is, the "processing system that accomplishes the mental activities of conceiving an intention, selecting the relevant information to be expressed and ordering and monitoring the information" (de Bot et al. 1997, p. 312), is the first system involved in formatting a speakers' intention. Its product is called the preverbal message. Then the formulator which is the next component involved, converts the preverbal message into a phonetic or articulatory plan applying the grammatical and phonological rules. Then the message is articulated. In Levelt's (1989) words:

> *This means that grammatical and phonological encodings are mediated by lexical entries. The preverbal message triggers lexical items into activity. The syntactic, morphological, and phonological properties of an activated lexical item trigger, in turn, the grammatical, morphological and phonological encoding procedures underlying the generation of an utterance. (Levelt 1989, p. 181)*

This model was originally proposed to account for speech production. The updated version of Levelt's model (1993) includes adaptations to account for oral and written modalities and "represents concepts, lemmas, and lexemes as shared in comprehension and production but with different input and output processes"(de Bot et al. 1997, p. 314).

According to de Bot et al (1997), in comprehension both top-down and bottom-up processing are involved and in fact there is an interaction between the two processes while in production just top-down processing is involved and this comprises one of the aspects in which comprehension and production differ. They maintain "in the interpretation of text, both types of information are used, and the learner has the possibility of internalizing at least some features of the new lexeme" (p. 316). They propose that as in language production, in language comprehension, a number of steps have to be taken for a word and its meaning to be matched, however, in the opposite direction from language production. When a string of letters is encountered, it has to activate a lexeme or in other words it has to be matched with it. If the match is sufficient, the activated lexeme activates a lemma, which has to trigger a concept for the comprehension (at the word level) to be successful. In other words, when the lemma is connected with the matching concept, a connection between input and comprehension is made.

But what if the word is unknown? In this case, according to de Bot et al. (1997):

> *various types of information can be used to connect the word form with its meaning: in particular, morphèmes indicating word class or meaning components, syntactic information in the sentence that reveals the argument structure of the lexical items, and semantic information from the surrounding text as it interacts with the learner's knowledge (ibid, p. 316).*

A schematic representation of de Bot et al.'s model is presented in Figure 2.2.

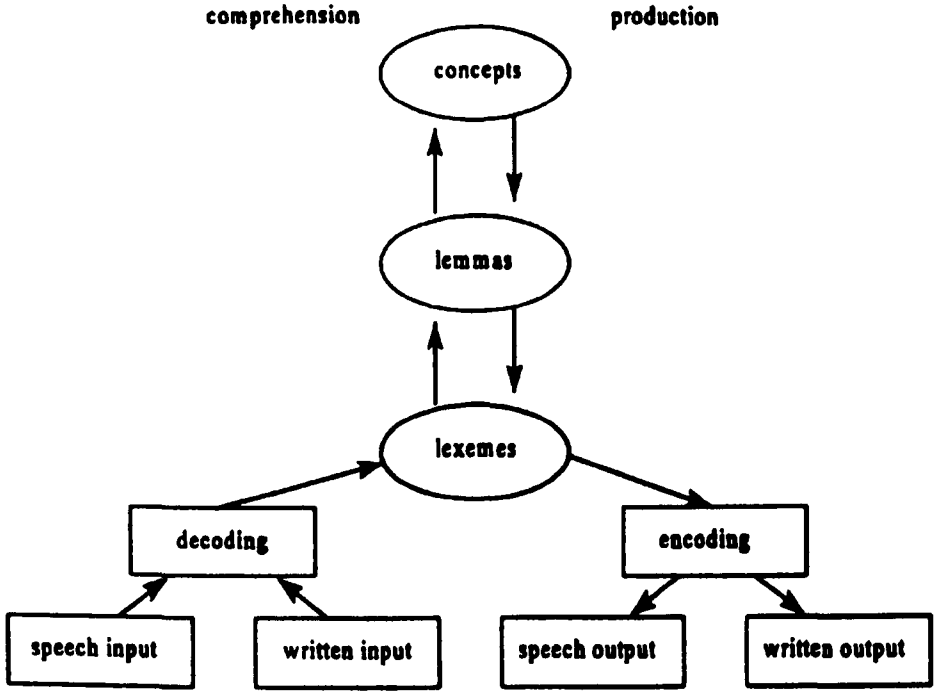


Figure 2.2 Lexical comprehension/production model for oral and written modalities Presented in De Bot et al (1997, p. 315)

In the light of the above discussion, it can be argued that in the comprehension of formulaic sequences the reader goes from the processing of multiple lexemes which activate a single lemma which in turn activate a single concept. On the other hand, in processing a string of words which are not formulaic, multiple lexemes activate multiple lemmas which should be matched with different concepts. For instance, the lemma structure of the idiom *kick the bucket* meaning *to die* has three lexemes, but only one lemma carrying its semantic and syntactic information and a single conceptual meaning, whereas the lemma structure of the word string *kick the bucket* which is not formulaic contains three lexemes, three lemmas and three conceptual meanings for each word. The schematic representations of the processing of formulaic and non-formulaic expressions are as follows (borrowed from Bishop 2004a, p. 63):

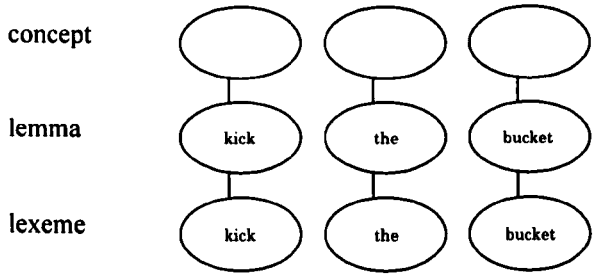


Figure 2.3 Representation of the lemma structure of the phrase *kick the bucket* as nonformulaic expression

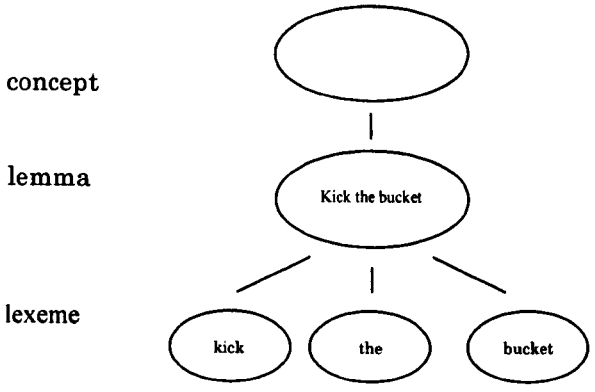


Figure 2.4 Representation of the lemma structure of the phrase *kick the bucket* as a formulaic expression

To summarize, language production and comprehension are believed to have similar organization and components at the representational level and to be different in the realization of a message. Both have three separate levels of representation, that is, the conceptual level, the lemma level, and the lexeme level which contribute in producing and comprehending a message. In language comprehension, when a new word is encountered, it has to activate a lexeme which in turn has to match a concept for the word to be comprehended. In processing a word string which is not formulaic, multiple lexemes have to activate multiple lemmas that have to trigger multiple concepts. In processing

formulaic sequences, by contrast, multiple lexemes, lemmas and concepts are activated.

This section presented some general information related to formulaic sequences, their functions and characteristics. The next section deals with issues related to reading comprehension, and the relationship between formulaic sequences and reading comprehension.

2.3 Reading Comprehension

Reading comprehension can be defined as understanding a text that is read. It therefore refers to the construction of meaning from a text. Readers are required to be actively involved in constructing meaning as "reading is not passive, but rather an active, and in fact an interactive, process" (Carrell 1998, p. 1). In other words, it is the interaction of the readers' purposes and life experiences and the text that make comprehension of a text possible. That is, "comprehension occurs when the reader extracts and integrates various information from the text and combines it with what is already known" (Koda 2005, p. 4).

Reading was once regarded as equal to the mastery of codes and it was widely believed that comprehension was an inevitable product of decoding. As McNeil et al. (1980) point out, "the prevailing notion was that reading consisted of nothing more than the power to utter certain sounds when perceiving certain letters" (p. 129). It was in the 1960s that researchers came to view reading differently and argue that reading without comprehension is limited to decoding symbols rather than extracting meaning from text. In this respect, Dole (2000) maintains, "just because readers decoded a text did not mean they understood it" (p.56).

Koda (2005) classifies approaches taken by scholars to define reading competence into cognitive, developmental and functional views. From a cognitive perspective, reader-text interaction is divided into three processing clusters. *Decoding* is the first step in which the reader directly extracts linguistic information from print. Next, is the *text-information building* in which the reader integrates the ideas extracted from print to uncover text meanings. And then, in the third step, that is, *situation-model construction*, the text information extracted is combined with the previous knowledge (Carpenter et al. 1995, Kintsch 1998, Miller 1988; Perfetti 1994). According to the developmental view, decoding and comprehension develop sequentially. Decoding and comprehension are described as two halves of reading, however, "reading does not equal the sum of decoding and comprehension, for neither decoding in the absence of comprehension, nor comprehension in the absence of decoding, leads to any amount of reading" (Gough et al. 1996, p. 3) The readers first gain sufficient mastery of decoding to help them get information from text and then comprehension develops as a result of decoding competence (Gough & Tunmer 1986; Hoover & Gough 1990). The functional view, supported by Carver (1990, 1997, 2000) posits that the reader's purpose for reading has a determining role in the way in which the information in the text is processed. It is therefore argued that a third factor, that is, the reading purpose, such as locating lexical information or detecting main ideas, is involved in the process of reading.

Over the years, several models have been proposed to account for the processing of text, different studies have been conducted to discover the nature of reading, the factors affecting reading comprehension and to find out about the strategies helpful in extracting meaning from text effectively. Many articles have been written on different purposes for reading, the differences between L1 and L2

reading, and testing the reading skill. In the following sections, a brief review of some of these issues which are relevant to this thesis will be presented.

2.3.1 Models of the Reading Process

Among approaches attempting to explain the nature of reading, the most prevalent in the literature are the Bottom-up, Top-down, and Interactive. Smith (1998) asserts that 'top-down' and 'bottom-up' are terms borrowed from computer technology (p. 218) and that top-down equals what he terms "inside-out, implying that the reader determines how a text will be approached, dealt with and interpreted" (ibid), and the bottom-up view is equivalent to his "outside-in, putting the text in charge, with the letters on the page the first and final arbiters of the reader's responses" (ibid).

2.3.1.1 Bottom-up Processing

Reading was first viewed as a passive process of extracting meaning from print. According to this view, which was referred to as bottom-up model of reading, readers first recognize a number of linguistic signals (letters, syllables, words, etc.) and by means of their linguistic data processing mechanisms, bring some sort of order out of these signals. Then, they build up "a meaning for a text from the smallest textual units at the 'bottom' (letters and words) to larger and larger units at the 'top' (phrases, clauses, intersentential linkages)" (Carrell 1998, p.2). In other words, in this view, which is regarded as data-driven by some scholars (Chun & Plass 1997, Hudson 1998), textual decoding is of primary importance.

2.3.1.2 Top-down Processing

The top-down view of reading challenges the bottom-up view as the latter does not consider the role of the reader and his background knowledge and his expectations about the text which are used as part of the reading process. Goodman is considered to be the pioneer of the view as he proposed for the first time the idea of reading as 'psycholinguistic guessing game'(Goodman 1967, 1971, 1998), although he does not "characterize his theory a top-down model" (Carrell 1998). Goodman (1998) defines reading as "a psycholinguistic process in that it starts with a linguistic surface representation encoded by a writer and ends with meaning which the reader constructs" (p.12). The readers, therefore, do not need to rely on contextual clues to derive the intended meaning. In fact, the readers make predictions regarding the meaning of the text using the graphophonic, syntactic, and semantic systems of the language and verifies these hypotheses using their past experiences and knowledge of the language. In this vein, Urquhart and Weir (1998, p. 42) maintain, "Goodman views reading as a process of hypothesis verification, whereby the readers uses selected data from the text to confirm their guesses". This view is therefore considered to be conceptually-driven.

Goodman (1998) gives a revised model of the reading process and suggests that there are five processes that the brain employs in reading. These processes are recognition-initiation, prediction, confirmation, correction and termination. In recognition-initiation, "the brain must recognize a graphic display in the visual field as written language and initiate reading" (Goodman 1998, p.16). Prediction, the process by which the brain makes predictions, is followed by the confirmation process in which the brain "monitors to confirm and disconfirm with subsequent input what is expected" (ibid). If the predictions are

disconfirmed, they are corrected through the process of correction. When the reading task is completed, the brain terminates reading.

The top-down view has been criticized by some scholars as being unable to give a complete account for the behaviour of a skilled reader. Samuels and Kamil (1998) for instance, maintain that "if the reader is skilled at word recognition but does not know much about the topic, it may be easier to simply recognize the words on the page and rely on bottom-up processes" (p.32). The reader may not have a prior knowledge of the topic as well. And as Eskey (1998) asserts, "poor readers are just as likely as good ones to rely on prior knowledge in deciphering text" (p. 95). It is also argued that in the case of second language learners, the problems arising from such a view are unique as ESL readers may or may not be literate in their first language, they have different views about reading as a social phenomenon in their first language, they come from different orthographic traditions and they start reading English with a different level of knowledge of the language than English-speaking children (Grabe 1998).

2.3.1.3 The Interactive Approach

The problems with top-down approaches to reading led to the emergence of the interactive processing view of reading which emphasizes the interplay between lower-level processing skills and higher-level comprehension and reasoning skills (Chun & Plass 1997). It is argued that successful comprehension involves both processing skills. As Samuels and Kamil (1998) maintain, an interactive model "permits the information contained in higher stages of processing to influence the analysis which occurs at lower stages of processing" (p. 27). This view is based on schema theory which is concerned with the "reader's schemata, or knowledge already stored in memory" (Anderson & Pearson 1988, p. 37). According to this view, when the new information matches the existing

information, comprehension occurs. That is, it is this "interaction of new information with old knowledge that we mean when we use the term comprehension" (ibid). Stanovich (1980) goes a step further and introduces the interactive-compensatory model in reading. As Samuels and Kamil (1998) put it this model is interactive

> *in the sense that any stage, regardless of its position in the system, may communicate with any other stage, and it is compensatory in the sense that any reader may rely on better developed knowledge sources when particular, and usually more commonly used, knowledge sources are temporarily weak (p. 32).*

For example, for a poor reader, knowledge of the topic may compensate for his slow and inaccurate word recognition. By contrast, for a reader who does not have knowledge of the text topic but is skilled at word recognition, reliance on words and phrases may allow for this compensation.

2.3.2 Vocabulary Knowledge and Reading Comprehension

Many research studies, both theoretical and empirical, have supported the idea that vocabulary knowledge is instrumental in and positively correlates with reading comprehension both in L1 and L2 (Anderson & Freebody 1981; Alderson, 2000; Read 2000; Qian 2002). Some scholars even argue that in L2 reading, "some sort of threshold or competence ceiling has to be attained before existing abilities in the first language can begin to transfer" (Alderson, 1984, p. 20). This threshold is widely believed to be lexical (Laufer 1997; Haynes and Baker 1993). Stahl & Nagy (2006, p. 4) maintain that the correlation between vocabulary knowledge and reading comprehension is so high (in the 0.85 to 0.95 range) that some authors have argued that they are psychometrically identical

(e.g., Carver, 2003) and continue that 'the size of one's vocabulary can be used as a predictor of his reading ability' and conclude that this kind of relationship between the two has one obvious interpretation: "having a big vocabulary makes you a better reader" (ibid, p.9). In a similar vein, Nagy (1988) points out that the strong relationship between vocabulary knowledge and reading comprehension is because of the fact that "vocabulary knowledge is fundamental to reading comprehension; one cannot understand text without knowing what most of the words mean" (p.1).

A study on high school students by Anderson and Freeman (1981) revealed "a high correlation between tests of vocabulary and comprehension" (p. 3). It is suggested by the researchers of the study that "a reader's general vocabulary knowledge is the single best predictor of how well that reader can understand text" (ibid). With respect to L2 learners, Laufer (1991) examined the connection between vocabulary knowledge and reading comprehension using a reading test and two different vocabulary tests (the *Vocabulary Level Test* by Nation [1983] and the *Eurocentres Vocabulary Test* by Meara and Jones [1989]) and found significant correlations of .5, at the level of $p < .0001$, and .75, at the level of $p < .0001$, respectively. Similarly, in a study to investigate the influence of *iraab* (a sound symbol in Arabic) and vocabulary on L2 learners of Arabic, Khalideh (2001) found that vocabulary knowledge had a significant main effect and priority over knowledge of *iraab* and argues that "the lexical and semantic knowledge of words is optimal for meaning without requiring the L2 reader to resort to any additional information obtained from lower-level processes such as the syntactic element, *iraab*" (p. 423-4).

Considering vocabulary knowledge as one type of knowledge source strongly related to learner's ability to read and understand texts, Nassaji (2004) asserts that in the context of testing vocabulary, a distinction has been made between

the depth (learner's level of knowledge of various aspects of a given word) and breadth (vocabulary size) of vocabulary knowledge. Qian (1998) investigated the relationship between depth and breadth of vocabulary knowledge and reading ability of second language learners at university level. He found a close and positive relationship between vocabulary size and reading comprehension. A positive relationship between depth of vocabulary knowledge was found as well. In addition, he concluded that "depth of vocabulary knowledge made a unique contribution to the prediction of reading comprehension scores over and above the prediction afforded by vocabulary size, ..." (Qian 1998, p. ii).

To summarize, the strong relationship between vocabulary and reading comprehension has been proved by a large number of studies. However, it is not so clear how. This is dealt with in the next section.

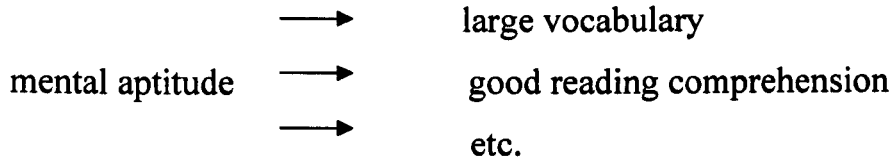
2.3.3 The Hypotheses Regarding Vocabulary-Reading Relationships

Anderson and Freebody (1981) and Nation (1993) provide three hypotheses regarding the relationship between vocabulary and reading comprehension. The first view which is labeled the *instrumental hypothesis* considers vocabulary as a major requirement and a causative factor in comprehension. In other words, it refers to the "idea that it is simply knowing more words that makes you a better reader" (Stahl & Nagy 2006, p. 10). Hsueh-chao & Nation (2000, p. 404) present the diagrammatical representation of the instrumentalist view as follows:

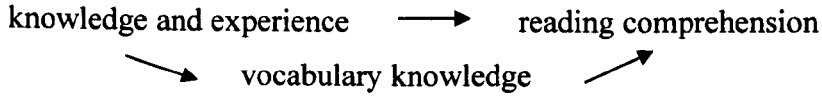
Vocabulary knowledge ⟶ Reading comprehension

According to the *aptitude hypothesis*, one who has a good brain, will easily acquire a large vocabulary. The same is true for reading comprehension and ability to understand oral explanations. Hence, this view sees the relationship

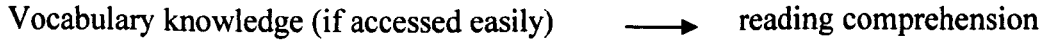
between vocabulary and reading as one which reflects "a more general underlying verbal aptitude" (Stahl & Nagy 2006, p. 11). And the relationship between the two is not a causal one either. That is, reading comprehension is not influenced by vocabulary knowledge. The diagrammatical representation of this hypothesis presented in Hsueh-chao & Nation (2000, p. 404) is as follows:



The third hypothesis referred to as *knowledge hypothesis* considers the knowledge of the concepts that the words represent as the factor responsible for making someone comprehend rather than the words per se. In other words, vocabulary is considered "as an indicator of good world knowledge" (Hsueh-chao & Nation 2000, p. 404). Diagrammatically, this view is represented as follows:



The fourth hypothesis concerning the relationship between vocabulary knowledge and reading comprehension is suggested by Mezynski (1983). The access hypothesis, like the instrumental hypothesis, views vocabulary knowledge as a causative factor in comprehension if it is accessed easily. "Fluency of lexical access, speed of coping with affixed forms, and speed of word recognition" (Hsueh-chao & Nation 2000, p. 404) are among the factors that affect this access.



To sum up, whatever the hypothesis, it is shown once more that there is a close relationship between vocabulary and reading comprehension directly or indirectly.

2.3.4 Automatic Word Recognition and Reading Comprehension

Stahl and Nagy (2006) distinguish knowledge of word meanings from word recognition "which involves recognizing the written form of words and sight vocabulary, which is recognition of words 'by sight' or automatically" (p.3). It is argued that it is not just the knowledge of word meanings that affects constructing meaning out of a text, but "efficient visual-information processing" (Koda 2005, p. 30) should also be deemed "as one of the vital competencies for successful comprehension" (ibid). A long-standing controversy in this vein is whether automatic word recognition affects reading comprehension.

On the one hand, constructing the meaning of a text is "highly demanding and requires high amounts of cognitive control" (Fukkink et al. 2005). On the other hand, the limited capacity of the human mind prevents the reader attending to more than one thing (letter identification, word recognition, accessing word meanings) at the same time. It is therefore suggested that efficient reading is achieved if resources needed for higher level processing are not used for lower level processes such as word recognition. In the same vein, the *verbal efficiency hypothesis*, suggested by Perfetti (1985) refers to the fact that constraints on memory and attention reduce text comprehension unless the time and effort used for recognition of single words are used for higher-order comprehension tasks. In other words, Perfetti argues that "the core proposal is that comprehension is limited by efficient operation of local processes" (ibid, p, 120).

Automatic processing, according to Segalowitz (2000, 2003, cited in Fukkink et al. 2005), refers to the "reorganization, routinization, or bypassing of serial execution of component processes" (Fukkink et al. 2005, p. 56) and is different from fast processing. In reading, on the one hand, many processes have to be activated at the same time and on the other, a limited number of mental resources can be activated simultaneously in the working memory. Automated word recognition, therefore, can reduce the processing burden on memory and in L2 word recognition, "automatic processing could mean that word recognition proceeds directly from the printed word to meaning activation without passing through stages of phonological recoding or translation into the L1" (ibid).

In summary, as Fukkink et al. (2005) point out, "it may be not only familiarity with the meaning of words but also, and more specifically, the speed with which readers can retrieve the meaning of words from their mental lexicon that may play an essential role in text comprehension" (p. 54).

2.3.5 Lexical Inferencing and Reading Comprehension

One issue which is relevant to both vocabulary and reading comprehension is lexical inferencing. There is a consensus among scholars that readers use a number of context cues or knowledge sources to overcome the frustration caused by unfamiliar words while reading (Nassaji 2004, Parel 2004). In other words, when readers confront new or unknown vocabularies or word combinations, they resort to their lexical inferencing strategies which involve "guessing the meaning of an unfamiliar word using available linguistic and other cues" (Bengeleil and Paribakht 2004, p. 225).

Empirical research supports this view and indicates a high correlation between reading comprehension and lexical inferencing (Bengeleil and Paribakht 2004,

Parel 2004). Several attempts have been made to identify contextual cues and knowledge sources which help readers to guess the meanings of unknown words and expressions while reading. Some classify these clues into interlingual, intralingual and extra-lingual cues (Carton 1971; Haastrup 1991; cited in Paribakht 2005). Others like Paribakht and Wesche (1999) distinguish two main types, linguistic, e.g., knowledge of discource and word morphology and extralinguistic, e.g., topic and word knowledge.

Research has also shown that several factors affect readers' inferencing during reading. Paribakht (2005) summarizes these factors as contextual and learner-related factors. She includes the number of occurrences, the density and the importance of the unknown word and text properties like length among contextual factors and background knowledge of the learner including procedural knowledge and knowledge of vocabulary, learner's previous learning experience and language proficiency as learner-related factors.

In sum, when L2 readers do not know the meanings of some words or expressions in a text, they try to extract the meanings by using available cues. Their success in doing so depends to a large extent on their prior knowledge including their L2 proficiency and the characteristics of text, including its difficulty level.

2.3.6 Formulaic Sequences and Reading Comprehension

Formulaic sequences have recently attracted the attention of scholars in the field of EFL/ESL learning and teaching and despite plenty of evidence indicative of the significant role of formulaic sequences in language acquisition and language production, little research, either theoretical or empirical, to date, has directly addressed the significance of these sequences in language comprehension,

particularly reading comprehension. This is partly due to the lack of an established method for testing one's receptive knowledge of these sequences, and partly because "comprehension is inherently less easy to investigate than production and, in particular, it would be difficult to tell whether a breakdown in comprehension had occurred within, or at the boundaries of, a sequence" (Wray 2002a, p. 232). That is the reason why most research studies are devoted to finding the relationships between subjects' knowledge of formulaic sequences and their fluency in speech or writing.

There are, however, some views on the reading process that can be employed to get some insights into the plausible effects of word strings on reading comprehension. It is a consensus among scholars investigating reading that it is an active process which involves the readers' background knowledge or schemata. This background knowledge is, according to Carrell (1998, p. 4), of three types, that is, 'linguistic schemata' which involve the knowledge of words and phrases; 'content schemata' involving knowledge of content area of the text; and 'formal schemata' which are concerned with the knowledge of the rhetorical structure of the text. Carrell (1998), refers to research studies conducted within the general framework of schema theory, for instance Carrell & Wallace (1983), that report a significant role played in ESL reading by schemata. Linguistic schemata are related to the concept of formulaic sequences in that they are part of this knowledge. It, therefore, makes sense to say that the knowledge of these sequences helps the reader to make correct predictions, particularly in respect of sequences for which the meaning of the whole is not the sum of the parts. Schema theory is related to the top-down versus bottom-up processing in reading. It is widely acknowledged among reading experts that reading involves both processes functioning interactively. In essence, word strings which are

processed as wholes, if recognized by the reader, can compensate for lack of background knowledge and spare the reader

> *some of the processing effort of parsing the language stream into individual words, recognizing each of the individual words, and extracting meaning from the composite of the individual word meanings and the related grammatical structuring (Schmitt & Carter, 2000).*

In a similar vein, Nattinger and DeCarrico (1988) maintain that having a considerable number of frequently occurring phrases in mind increases comprehension as this knowledge "relieves learners of having to attend to each individual word as it is used and allows them to focus attention on the larger structure of discourse" (p. 92). In the same vein, Schmitt (2000) refers to the role of formulaic sequences in comprehension and asserts that recognition of lexical phrases as whole units "spares the listeners [or reader] some of the processing effort required to interpret an utterance [or a sentence] word by word" (p. 4).

In addition, some formulaic word strings are believed to function as "macromarkers", a term defined by Chaudron and Richards (1986) referring to discourse markers that indicate the overall organization of discourse and the relations within it. Nattinger and DeCarrico (1992) classify lexical phrases into three groups, that is, social interaction markers, necessary topic, and discourse devices based on their functions in structuring discourse. Of the three, discourse devices correspond to the macromarkers proposed by Chaudron and Richards (1986). They define discourse devices as "lexical phrases that connect the meaning and structure of discourse" (Nattinger and DeCarrico 1992, p. 64) and categorize them as logical connectors (*as a result*); temporal connectors (*the next is Y*); Spatial connectors (*on the corner*); Fluency devices (*by and large*);

exemplifiers (*in other words*); relators (*the (other thing X is Y*); Qualifiers (*it depends on X*); evaluators (*there is no doubt that X*); summarizers (*to make a long story short*). These "represent different categories of meaning and pragmatic characteristics of discourse and conversational structure that exist in many different types of situations" (ibid, p. 59) and help the language user mentally organize the message in the text as it goes along and evaluate it. As Chaudron and Richards (1986) maintain, "prior knowledge of this sort helps top-down processing by initiating expectations and predictions" (p. 116) about the text.

Another issue slightly related to formulaic sequences is that of the relationship between eye-movement and reading comprehension. One of the earliest findings from eye-movements research was that they indicate the processes that occur in the readers' mind. It is based on the assumption that the greater the number of regressions and forward fixations and the longer their duration, the more difficult it is for the reader to proceed and comprehend the text. This shows that our eyes are under the control of the cognitive processes involved in sentence comprehension. Ehrlich and Rayner (1981), found that readers had fewer eye fixations on words predictable on the basis of the preceding context and the words predicted and fixated were focused for shorter duration. Underwood et al (2004) took advantage of these findings and attempted to study eye-movements during the reading of texts to find out how formulaic sequences are processed by native and non-native readers. The results suggest that while reading a passage both groups fixated less often on words which were part of a formulaic sequence, than on the same words when they were part of non-formulaic text. For the natives, the fixation duration was shorter on formulaic sequence words but for the non-natives the fixation duration was the same for both formulaic and non-formulaic words. To summarize, as Underwood et al (2004) put it:

> *we could speculate that mastering the recognition of formulaic sequences in written texts is an incremental process, and early partial mastery is rewarded mainly by not needing to fixate on the vocabulary in a text as much, but it is only with fuller mastery that the requirement for a 'full duration' fixation lessens (p.163).*

To investigate whether formulaic sequences lead to processing ease or advantage, Schmitt and Underwood (2004) conducted a study to explore the processing of these sequences through a self-paced reading task to identify the recognition point for the sequences (at which point in a sequence recognition occurs). The results were, however, inconclusive, "with native speakers showing no difference in recognition time" (Schmitt and Underwood 2004, p. 187) and nonnatives taking longer time with later words in a sequence.

In a study on the possible processing advantage for formulaic sequences while reading, Conklin and Schmitt (2008) found that formulaic sequences were read more quickly than their nonformulaic counterparts by both native and nonnative speakers no matter whether these sequences were used literally or idiomatically. Their results "support the assertion that formulaic sequences are involved in more efficient language processing" (Conklin and Schmitt, 2008, p. 86).

Bishop (2004a,b) investigates still another type of connection between formulaic sequences and reading. His research addressed the question of whether the use of typographical highlighting of words and formulaic sequences stimulates nonnative learners to click on those items to access glosses of their meaning in a computerized reading task. It was assumed that an increase in glossing of unknown salient formulaic sequences would result in an increase in the number of items comprehended compared to an identical text containing no salient formulaic sequences. The results revealed that the treatment group did better

than the control group which means making formulaic sequences typographically salient increases their chance of being noticed and looked up, which in turn influences the overall comprehension of the text. He also found that unknown non-salient formulaic sequences were glossed less frequently than words. This means that "unknown formulaic sequences are less easily recognizable as holistic entities than words, because unlike words, it is not clear, a priori, where the boundaries of unknown formulaic sequences lie" (Bishop 2004a, p. 239).

To summarize, both bottom-up and top-down process are involved in reading as the reader needs the knowledge of words and phrases to compensate for his unfamiliarity with the topic. Studies on eye-movement and typographical saliency of formulaic sequences support the view that formulaic sequences may help the reader to bypass the effort to process words one by one to derive meanings from a text.

2.4 Summary

The first section of this chapter was a review of issues related to formulaic sequences which would help the reader to form a framework of the nature of these sequences and their properties. It was argued that formulaic sequences are multi-word units which are stored and retrieved as wholes and spare the language user time and effort. These sequences are placed along different continua of fixity vs variability, continuity vs discontinuity, and transparency vs opacity. There are different types of multi-words, the most common are collocations like *rancid butter*, idioms like *kick the bucket*, phrasal verbs like *to turn out* and poly-words like *by the way*. Different functions of formulaic sequences were examined and it was concluded that the most important

functions of these sequences are of sociolinguistic, psycholinguistic and pragmatic nature.

The role of formulaic sequences in adult second language learning was also summarized and relevant studies were presented. The approaches to language processing which include multi-words in their accounts were summarized and then different models concerning the processing of opaque formulaic sequences were presented.

The second section of this chapter dealt with the models of L1 and L2 reading, the role of vocabulary knowledge in reading comprehension and studies concerning their relationship. Research has shown that the knowledge of vocabulary and automatic word recognition have considerable impact on L2 reading comprehension. The last part of this section addressed the relationship between formulaic sequences and reading comprehension. It examined different studies related to formulaic sequences and reading comprehension including the connection between the frequency and duration of eye fixation on words and formulaic sequences and reading comprehension, the processing advantages of formulaic sequences while reading and the relation between typographically salient formulaic sequences and reading.

As the above review of the literature indicates, in the context of EFL reading comprehension researchers have addressed the questions of whether the readers benefit from the knowledge of vocabulary during reading and whether they can use available clues to guess the meanings of unfamiliar words; however, little or no research has been done in the area of formulaic sequences. Meanwhile, in the context of formulaic sequences, most researchers have addressed the production skills, speaking and writing, and the contribution they receive from formulaic sequences and no research has directly studied the contribution of

these sequences to EFL reading comprehension or the cues that the readers employ to extract the meanings of unknown expressions while reading. These issues are addressed in this study using a methodology which is discussed in the next chapter, **Methodology.**

Chapter III
Methodology

3.1 Introduction

This chapter provides a detailed account of the overall design of the present thesis. As mentioned in Chapter 1, this thesis is composed of a main study which is quantitative, dealing with tests and their scores and a follow-up study which is qualitative, dealing with subjects' recalls and interviews.

Both quantitative and qualitative methods of data collection and analysis are employed in this study as research has shown that when both methods are used a more complete analysis is yielded and one method can compensate for the shortcomings of the other. For example, quantitative analysis is more objective and generalisable but may miss details while qualitative analysis is less objective but reveals more.

The first part of this chapter presents the methodology used for the main study. It starts with the research hypotheses and is followed by a description of the design, including the variables and the statistical procedures used to analyse data. Then, details are provided on the participants and settings. This is followed by information about the selection and construction of instruments, which include a test of TOEFL, a test of reading comprehension, a test of understanding formulaic sequences plus a test of recognition (awareness) of these sequences. The results of the pilot studies conducted to test the reliability of the tests which are used in the study are also given.

The second part of this chapter presents the methodology employed for the follow-up study and starts with stating the purpose of the study followed by information about the participants. Then the procedures of data collection will be presented

3.2 The Main Study

3.2.1 Research Hypotheses

As reviewed in the previous chapter, on the one hand, the majority of L2 studies on formulaic sequences focus on the productive skills, that is, speech fluency and writing ability rather than receptive skills like reading comprehension, and on the other hand, the large number of studies which have been done on the relationship between different aspects of vocabulary knowledge, including vocabulary size, depth of vocabulary and word recognition and reading comprehension have not considered the connection between the knowledge of multi-word units and L2 reading ability. This study brings these two lines of research together and attempts to determine the contribution of recognizing and understanding of formulaic sequences to EFL students' reading comprehension. Based on the research questions presented in chapter I, the following null hypotheses were formulated:

Main Hypotheses:

1. There is no significant relationship between Iranian EFL learners' recognition of formulaic sequences and their performance on a test of reading comprehension.

2. There is no significant relationship between Iranian EFL Learners' understanding of formulaic sequences and their performance on a test of reading comprehension.

3. The interaction of understanding and recognition of formulaic sequences does not affect the learners' performance on a test of reading comprehension.

Secondary Hypotheses:

1. There is no significant relationship between frequency of formulaic sequences and recognition of these sequences.
2. There is no significant relationship between frequency of formulaic sequences and understanding these sequences.
3. There is no significant relationship between Iranian EFL learners' recognition of formulaic sequences and their overall proficiency.
4. There is no significant relationship between Iranian EFL learners' understanding of formulaic sequences and their overall proficiency.
5. There are no differences between male and female subjects' performance on the recognition tests.
6. There are no differences between male and female subjects' performance on the understanding test.

In the main study, two designs were employed: an ex post facto design and a factorial design. Hatch & Farhady (1982) refer to ex post facto designs as designs that are used "when the researcher does not have control over the selection and manipulation of the independent variable" (p. 26). When employing ex post facto designs, researchers are concerned about the type and/or degree of relationship between pairs of two or more variables rather than a cause-and-effect relationship. Correlational designs are the most commonly used subset of ex post facto designs. In this study, in order to find answers to the first two questions, correlation analyses are conducted to determine whether there are relationships between variables (meaning and reading and awareness and reading) and to determine the types and degrees of relationships, if any.

The second design is a factorial design as there are two independent variables, each having two levels and the effect of the interaction between the independent variables and the dependent variable is considered. To investigate the relationship between these variables and their interaction effect, a two-way ANOVA (Analysis of Variance) is utilized. Two-way ANOVA is a statistical analysis in which the effect of two or more independent variables (e.g., understanding and recognizing formulaic sequences) each with two levels (e.g., high and low) on one or more dependent variable(s) (e.g., reading comprehension) is investigated.

As in correlational designs, there is no causal relationship between the variables, one variable is referred to as the dependent and the other as the independent variable arbitrarily. In this study, the dependent variable (in both designs) is students' performance on a test of reading. The independent variables are (a) understanding of formulaic sequences, and (b) recognition of these sequences. Students' level of English proficiency, nationality, L1 and educational stage are the control variables which were controlled "to neutralize the potential effect" (Hatch & Farhady, 1982, p. 16) they might have on students' performance.

3.2.2 Participants

The participants who took part in the quantitative study fall into two groups: those participating in the pilot studies and those participating in the main study. 51 students, 23 male and 28 female, participated in different pilot studies which were conducted to measure the reliability indices and other test characteristics of the tests developed for the purpose of this study. 48 students were native speakers of Farsi studying English Language and Literature. They were in their third or fourth year. The students who participated in the pilot studies were not from the same university as those participating in the main study, but were

similar to them as all universities in Iran have to comply with the curriculum approved of by the ministry of Science, Research and Technology. Different groups of these students took different tests or different versions of the same test. Three native speakers of English studying at undergraduate level participated in the pilot studies by answering two of the tests developed for the purpose of this study.

For the main study, 53 students, 24 males and 29 females, took part. They were Iranian students within the age range of 21 to 30 who were also studying English language and Literature at Islamic Azad University. They were from among junior and senior students with an approximately upper-intermediate proficiency level of English. Initially this group had a greater number of students, but some were excluded from the study as either they failed to take one or two of the four tests administered or did not meet the criterion set to establish the homogeneity of the subjects in terms of their overall proficiency. The reason for selecting subjects at similar proficiency level was that research (Spöttl and McCarthy 2003, 2004) has shown proficiency "to be a significant factor in learner ability to decode multi-word strings" (Spöttl and McCarthy 2004, p. 201). The subjects had all studied English as a foreign language in school for six years and had studied at least three years of English at university. In Iran, students majoring in English take courses which help them improve their language, including pronunciation practice, conversation practice, reading, writing and grammar at the first two years of studying at university. From the second year on, they take some courses on English literature, teaching English as a foreign language, linguistics, and translation. Involvement in the study was voluntary.

3.2.3 Tests

In this section, the rationale and the procedures followed in the development of tests and their features as well are discussed. A battery of tests were conducted to gather data in this study.

Several months were devoted to the development and pre-testing the tests. Three tests were developed for the purpose of this study: a reading comprehension test, a test measuring students' ability to recognize (their awareness) formulaic sequence and a test of understanding of these sequences.

3.2.3.1 Proficiency Test

An earlier version (1997) of the Test of English as a Foreign language, TOEFL, was used to tap the subjects' overall proficiency to be used as a criterion to establish homogeneity of the participants and their comparability as well. The test was used in its original form and it was not adapted. It was composed of three sections, i.e., listening comprehension containing 50 items, structure containing 40 items and reading comprehension having 50 items. The items in all subtests were in multiple-choice format.

3.2.3.2 Purpose-built Reading Comprehension Test

A test of reading comprehension was developed to test the subjects' reading ability. It was composed of two passages with 10 items each. To select the passages for the reading test, several criteria had to be met. First, they had to be authentic, that is, written for native speakers and not for ELT purposes. Therefore, a number of magazines and newspapers were consulted as the articles in these publications are not usually simplified and the criterion of

authenticity could be met. Attempts were also made to choose passages of appropriate length. A long-standing controversy regarding reading tests is whether the texts used have to be short or long. In this vein, Chastain (1988, p. 234) asserts that

> *language teachers usually favor short reading passages. This tendency may have its roots in a desire not to overload the students, but it probably is an outgrowth of the idea of reading as a laborious process of deciphering a complex and sometimes unfamiliar linguistic code for which the students were required to know all grammar and vocabulary. The result has been that students focus too much attention on language as opposed to meaning, and they read much too slowly.*

Another criterion which was taken into account while selecting passages was the familiarity with the genre of the texts. As the university students studying English in Iran pass courses on reading journalistic materials, they were familiar with the genre of the passages taken from newspapers. Care was also taken to select passages having topics about which the subjects had previous knowledge. As Allen et al. (1988) put it, "knowledge of topic is clearly *a* if not *the* determining factor in comprehension" (p. 170, emphasis in the original). In the same vein, Bernhardt (1986) maintains that "comprehension or 'proficiency' is text dependent. Readers may appear to be proficient on topics with which they have familiarity and non-proficient on topics with which they do not have familiarity" (p. 26). Attempts were also made to select texts compatible with the culture and religion, and the political beliefs of the test takers. The passages were also examined for their level of difficulty to be compatible with the students' level of English proficiency. There are different ways to examine the difficulty level of texts. Allen et al. (1988) assert that deciding on the difficulty level of a text can be done qualitatively, for example, using intuition and

experience or quantitatively using readability formulas. They add that readability formulas based on word and sentence length are not employed these days as they are not reliable. In this case, the researcher used her intuition and experience to decide on the difficulty level of the passages. The last criterion to be considered was the presence of a number of formulaic sequences in the passages as formulaic word strings were the focus of the study.

After examining a large number of passages from newspapers, having the above criteria in mind, two texts were chosen from 'The Guardian'. The first passage was entitled *'Resist the Official Pol-speak of Bush's War on Terror'* and was about the power of words in the world of politics. It was a topic that the subjects were familiar with as it dealt with war in Iraq and Iran's nuclear activities. The second passage was about whaleburgers offered in Japanese restaurants with the title *'Savouring the Whales'*. The selected passages were about 1000 words each. The reason for selecting passages of this length is that short passages do not necessarily equal easiness of the text and as Allen et al. (1988) point out "lengthier texts may well be more cohesive and, hence, more interesting, for learners" (p.170). The second reason was that the passages had to contain a number of formulaic sequences.

Once the reading passages were selected, they had to be tested for their difficulty level. In this study the students' feedback was used for this purpose, drawing upon Edwards's (1996) suggestion that the "communicative intent of the author" (p. 351) is the most important factor to determine the difficulty of text. Before test construction, that is, writing the comprehension questions, a pilot study was conducted with 13 Iranian university students. They were asked to read the passages silently and then provide a self-report on their comprehensibility and difficulty level. There was a consensus among the readers that they could get the writers' messages and that the passage on 'whaling' was

much easier than the passage on 'word power'. This meant that it (the passage on 'whaling') was not suitable and had to be replaced by another one.

This time around, a passage from 'The Daily Telegraph' was selected which was about Darwin's natural selection theory entitled *'Their Future Isn't Orange'*. Having almost the same length as the passage on word power, it was a scientific article on a topic accessible to all students as they had studied about it at high school. 19 Iranian students, different from the first group, read the passages and as in the previous pilot study provided self-reports. It turned out that the passages had almost the same level of difficulty.

For each passage, 10 cued recall questions were written. The items were designed to detect the degree of readers' understanding of the main ideas and the supporting details. It was decided to use cued recall questions over recall protocols for several reasons. Although recall protocols are believed to be the most straightforward procedure for testing reading comprehension, they have some disadvantages. Recall protocol analysis is demanding. Moreover, they give an account of what has been stored, but they do not give information of what is not remembered. They are dependent on memory and it is therefore difficult to make a distinction between what is extracted from the text and what is retrieved from stored knowledge bases (Koda 2005). Free oral recall was not used as well because the assumption was made that the students would be more likely to produce more if there was not the time pressure of oral recall and if they were prompted by questions. Multiple-choice items were not also considered for a number of reasons. Multiple-choice questions are objective but constructing distractors should be done with care as they may alter stored information extracted from the text or provide the reader with passage-related information. Therefore, "performance variation could be more attributable to task-specific skills than to reading ability" (Koda 2005). Cued recall questions, however,

required short answers to questions asking for specific information. They were, therefore, easier to score although challenging to construct, and at the same time did not have the shortcomings of multiple-choice questions.

The questions constructed for this study were based on the ideas in the text and also were mostly based on the parts of the text containing the lexical phrases being questioned in the tests of formulaic sequences. The first and the last questions for each passage dealt with the main ideas. Some of the items required the readers to make inferences. The questions were designed in a way that they could not be answered based on subjects' background knowledge and required knowledge of the text. The instruction and the questions were given in Farsi as research has shown that it is sometimes hard to decide whether difficulties are because of poor reading ability or to a faulty understanding of the task itself. The readers were also required to write the answers in Farsi to eliminate the effects of writing ability on scores. In this vein, Lee (1986a, 1986b) maintains that "when recall is done in the L1 as versus the target language, better results are gained as the learners are given a chance to "more fully demonstrate their comprehension" (Lee 1986a, p. 353).

The test including two passages followed by the questions was piloted on 16 students. The pilot test served in two ways in this phase of the study. First, it was used to determine the length of time the participants would need. Second, it was used to measure the reliability index of the test. The students took the tests in their regular class time. One of the students did not finish the test and was therefore excluded. To estimate the reliability of the test, the KR-21 formula of estimating reliability was employed. The reliability of the test turned out to be .81 which is well above the acceptable reliability level introduced in the table of acceptable reliability by Lyerly (1959, p. 75). In the table, one can find that for a group of 15 subjects and a test of 20 items, the acceptable level of reliability is

.408 at the .05 and .540 at .01 level. The reliability index of this test is therefore acceptable. See Appendix A for a sample of the reading comprehension test.

It should be mentioned that the passages may seem to be long. However, as in real life the subjects have to read news articles which might be at times longer, it would not cause problems for the readers. In addition as the readers are not supposed to deal with the specific pieces of information such as particular dates or places in the questions and as the questions seek for answers based on their general understanding of the text, the length of the passage would not be problematic.

3.2.3.3 Test of Recognition of Formulaic Sequences (TRFS)

The second test was a test of recognition (awareness) of the formulaic sequences. It was composed of 40 items assessing subjects' ability to recognize target sequences as they appeared in separate sentences. The target sequences were selected from the passages employed in the reading test. First, a total number of 75 opaque, semi-opaque and transparent formulaic sequences was extracted. Then, the target sequences were selected taking into account a couple of criteria. An attempt was also made to include different types of sequences in the test including, collocations, idioms, phrasal verbs, etc. The selected sequences had to be compatible with the subjects' level of English proficiency. Hence, some easy sequences such as *'in turn'* were excluded as too easy and some like *'rapid-fire lingo'* as too tricky. The decision regarding the difficulty level of the sequences was based on the intuition and experience of the researcher. Thus 62 candidate sequences were selected, 31 from each passage. The next step was to consult reference materials including *Collins Cobuild English Dictionary* and *Oxford Dictionary of Collocations*, which listed and discussed formulaic sequences of various kind, to determine if they were

counted as formulaic sequences. Where a sequence was not included in the above dictionaries, the British National Corpus was consulted using the following web address: www. natcorp.ox.ac.uk/corpus/index.xml. The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language and a search result shows whether a word or phrase exists in the corpus along with its total frequency and up to 50 examples. If the expression in question was counted as a formulaic sequence, it was added to the candidate list. Once the list of candidate formulaic sequences was compiled, the next task was to decide which one to choose. Thus, the candidate sequences were subjected to frequency analysis in the BNC. Candidates with relatively low frequencies (less than 5 occurrences) were deleted from the list. As mentioned before, in composing questions for reading comprehension, attempts were made to write questions whose answers were in the sections of the texts which included formulaic sequences. In selecting target sequences for the tests of formulaic sequences (Recognition and Understanding), care was taken to choose sequences occurring in parts of the passage which carried main or supporting ideas on the basis of which reading questions were constructed.

After applying the predetermined criteria to the list, 40 word strings were selected, some of which were word combinations such as *'when it comes to'*, that are used by the native speaker without necessarily having to use grammatical rules to produce them, and some were opaque or semi-opaque sequences, the meanings of which were not the sum of their constituent parts and could not be translated into Farsi on a word for word basis as well. See Appendix B for a list of target formulaic sequences along with their frequencies in BNC.

To construct the test, several test formats were considered. First, it was decided to use a c-test in which the target formulaic sequence is replaced by a blank and the initial letter of each word of the sequence is provided. The testee, then, has

to fill the blanks using his knowledge of formulaic sequences. An example follows:

*Radio One is **u**... **s**... from all sides.*
[answer: under siege]

However, as the test involved production rather than recognition this was ruled out.

Then, it was decided to use a test format in which the students had to decide which one of the four words or word groups in italics forms a common sequence with a word typed in bold. This format, although it had an advantage over the first as involving recognition, was rejected for several reasons. First, the sequences were decontextualized. Second, it was mostly suitable for testing only particular types of formulaic sequences, e.g., collocations and therefore was not applicable to all types. An example follows:

under .... *definition/siege/value/turn* [answer: siege]

To contextualize the target sequences in tests, a multiple-choice test was designed in which the sequence was taken from a sentence or two and the subjects would be required to fill in the blanks with the appropriate choices. However this one also was ruled out as for some sequences it was not possible to find distracters as every choice would have been a possible answer or the test would turn out to be a test of prepositions or phrasal verbs. An example is given below:

Radio One is ..... siege from all sides.

a. on b. under c. at d. over

As a consequence, a fourth option was tried. In this test, the sequences were inserted in sample sentences and the students were required to separate the words and lexical phrases in the sentences using slashes. The purpose of this test was to find out if the formulaic sequences in question were recognized as whole by the testees, which in turn could be used as an indicator of their awareness of the sequences. It is argued by the scholars that "the mind stores and processes these chunks as individual wholes" (Schmitt 2000, p. 400). It should, therefore, follow that they would be recognized as individual wholes. An example of the resulting format follows:

Radio One/ is /under siege/ from all sides.

An undergraduate student native speaker of English did the test and it turned out to be acceptable with respect to face validity. The test was then piloted on the 16 Iranian students who had answered the test of reading comprehension two weeks before. Their scores were used to test the reliability of the test using the KR-21 formula and it turned out to have an acceptable index of .73. Only items were scored correct where their formulaic sequences were identified and separated from other words. See Appendix C for a sample of the test.

The researcher made every possible attempt to design a test tapping subjects' recognition of formulaic sequences in a reliable and valid way. Many formats were considered, however, many others would have been possible. What seems necessary to be mentioned here is that this test was the first attempt to date to assess subjects' recognition of formulaic sequences and that sometimes a test that seems to have a simple format can yield as reliable and consistent results as a test having a complicated format.

3.2.3.4 Test of Understanding Formulaic Sequence (TUFS)

The third test was a test of formulaic sequences assessing the subjects' knowledge of the meaning of the word strings extracted from the reading passages. 20 sequences were selected from among the sequences employed in the previous test. The target sequences ranged in length from two to five words. As the meaning of the sequences was the main concern for the construction of this test, only those sequences having idiomatic meaning or in other words, semantically opaque or semi-opaque sequences were counted. As a consequence, the subjects could not guess the meaning of the whole from the parts.

Once the target formulaic sequences were selected, the items were written. As the previous test, in the process of test construction, several test formats were decided before reaching a final decision including translation and the Vocabulary Knowledge Scale from Wesche and Paribakht (1993). But they were ruled out as involving production in terms of the psychological processes involved in answering the questions. Since the purpose of this research is to examine the effect of understanding the formulaic sequences found in the passages included in the reading test and since reading comprehension is deemed as a receptive skill, the test of meaning of formulaic sequences needed to be a recognition test in terms of the cognitive process the testees would use to answer the items. As a consequence, it was decided to construct a multiple choice test so that the subjects had to recognize the correct answer, that is, the meaning of the sequence, from among choices provided. The target sequences were put in bold font employing the same format for presenting the target sequence as used in the *Collins Cobuild English Dictionary* (CCED). In CCED, the words' or phrases' definitions "(or explanations, as we often call them) are written in full sentences, using vocabulary and grammatical structures that occur

naturally with the word being explained" (CCED, p. xviii). In other words, the phrases are explained giving information about their collocates and structures and their context of use. Sometimes these pieces of information are given by paraphrasing the word or phrase in context. Hence, in this test, the information about each sequence distinguished the sequence as a whole from a word string whose meaning can be inferred from the sum of its constituent parts. The sequences were not contextualized so as to make sure that the subjects knew the meanings of the sequences and did not use their inferencing skills to guess the correct answer. This is in line with the findings of some studies such as Schmitt (1999) which show that while subjects can sometimes give the meaning of words included in a context, they cannot provide answers to a test measuring the same words presented in isolation. In addition, unlike single words which may have different meanings in different contexts, most formulaic sequences have one fixed meaning. Therefore, it was not necessary to include them in sentences to distinguish their particular meanings.

Each item was followed by 4 options. The distractors were written to be semantically similar to the correct option, and as similar in form and length as possible. Because the options for each item all had a similar meaning and form, it would be relatively difficult to guess correctly unless the subject had some intuition about the correct form of the target formulaic sequence. Instruction on how to do the test was given along with examples in Farsi as well as in English to eliminate misunderstanding. An example of the resulting format follows (item No. 6 in the test):

6. If you say that something is **a case in point**, you mean that ..... .
   a. you are not sure about something mentioned
   b. a crime or mystery is being investigated
   c. it is a good example of something mentioned
   d. you are going to start talking about it [answer: c]

The draft test was first piloted on two undergraduate students who were native speakers of English. They completed the test 100% correctly. This showed that the test had face validity and appeared coherent to native speakers. Then, 16 Iranian students, the same students who had answered the reading comprehension test and the recognition test, answered the questions and the results were used to examine the test characteristics of the instrument including the item facility (IF) and item discrimination (ID) indices. The presumption in using these indices is that the reliability of the test may be improved for future use by removing items with low indices (those either too easy or too difficult and those least able to discriminate levels of overall attainment) and replacing them with others. Then the reliability indices of the tests were measured. As there were no tests developed prior to this test to assess subjects' knowledge of the meaning of formulaic sequences, the test could not be examined in terms of validity. However an acceptable index of reliability is necessary, albeit not sufficient, for a test to be considered as valid.

The results of the item analysis suggested that all items had acceptable IF indices ranging from 0.47 to 0.80. ID indices were also acceptable with the exception of one item which had a value of 0.32, the least among the items, showing that it was selected more by lower level students than by their higher level counterparts. The item was therefore replaced by another one. To calculate the reliability of the test, Kuder-Richardson formula 21 was employed. The reliability index of the test was 0.67 suggesting an acceptable reliability level based on the table of acceptable reliability mentioned in the previous section. See Appendix D for a sample of the test.

Like the previous test, the researcher tried to develop a test having the characteristics of a good measure as far as possible. An acceptable reliability

index as well as a subjects' familiarity with the test format led the researcher to consider the test as suitable for the purpose for which it was developed.

3.2.4 Procedure

This section provides information about the procedures followed to collect data including test administration and scoring procedures.

The main study was conducted at one of the branches of Islamic Azad University. About a week prior to the testing date, the participants were informed about the study and that their participation was voluntary. In order to encourage the participants to take part in the study and answer the tests with full attention, they were informed that the first three with the highest scores would be given awards. Their instructors were also asked to encourage them to participate in the study. All juniors and seniors studying English were invited to take part. 79 out of 98 potential subjects registered to participate in the study but just 74 turned up at the time of testing. The testing sessions were set in advance. All the testing was administered by the researcher to ensure the uniformity of instructions and to resolve potential confusion or misunderstanding. The regular course instructors helped the researcher in doing her job. Due to time constraints, data collection took place in two sessions with a four days interval. Both testing sessions were set in the morning, starting from 8:30 to minimize the effect of fatigue.

On the first day of testing, the researcher gave a brief description of the nature of the study and its aims to the participants. Students were informed that all the information collected during the study would be kept confidential, their scores would not be shared with their teachers and would not affect their final grades. They were also told that they were free to withdraw from the competition at any

time. Upon entering the testing site, each participant was assigned a code number and all the test materials were pre-coded with numbers as well. Care was taken to ensure that the same subject was given the test bearing the same number code and that on the second day the students were given the same code number.

As only two sessions were available and as the two tests of the formulaic sequences had to be administered with some time interval, these tests were given to the subjects before and after the TOEFL in the following order.

First, the test of awareness of formulaic sequences was administered. 30 minutes were allotted to the subjects to complete this test. The reason why this test was administered before the TOEFL and the test of the meaning of the formulaic sequences was given after the TOEFL was because it was thought that as the target sequences were typed in bold in the second test (TUFS), they would provide the subjects with clues and affect their performance on the recognition test. Then, after a short break, without leaving their seats, the subjects were given the TOEFL in the following order: first the listening comprehension (55 minutes) section was administered, structure and written expression (25 minutes) was next, followed by the reading comprehension section (55 minutes). It lasted approximately 140 minutes. Then, the answer sheets and the test materials were collected.

It may be thought that by this time the students might have got tired. However, as all the tests required recognition or comprehension (in the case of listening and reading comprehension in TOEFL) of language and not production, they would have not been under much psychological pressure in the process of answering the tests.

The subjects were allowed to take a five-minute-break and then the last test, the multiple–choice test of understanding formulaic sequences was given. Students had to mark the correct answers on the answer sheets attached to the test material. It took the students participating in the pilot study a maximum of 27 and a minimum of 14 minutes to do the same test. The students in the main study were therefore allowed a time limit of 20 minutes to do the test. Caution was taken to ensure that the students did not take notes and did not copy the questions as this would affect their performance on the reading test.

There was a four-day interval before the reading comprehension test was given to make sure that any direct memory of the tests of formulaic sequences and any clues from those measures were minimized. The time taken to read the text, and answer the questions for the students participating in the pilot study ranged from 26 minutes to 53 minutes. Hence, in the main study, the first passage was given first and the subjects had 40 minutes to read it thoroughly. Then the questions were distributed. They had 10 minutes to answer the questions. Then the students took a break for five minutes and the second passage was distributed following the same procedure as the first passage. The reason for giving the passages first and then the questions was because the students would not have a chance to look at the questions prior to reading them to get clues and would not have extra time to go back to the passages to find the possible answers in the texts after reading them. Half of the students received the passage on 'Bush's war on terror' first and the passage on 'Their future isn't orange' second and the other half received the passages in the opposite order. This would minimize the effects of topic familiarity and fatigue.

A number of problems were encountered during the data collection period. Firstly, one student left the testing site after completing the TOEFL and did not continue taking part in the experiment. Secondly, three students were absent on

the second day of the experiment and failed to take the reading test. Third, one student arrived late on the second day and could not take the test under similar conditions as others. He was therefore excluded from the list of participants and his score was not counted. Lastly, and most importantly, there were no volunteers for an interview and recall session which was intended to gather data for the qualitative analysis. This was due to several reasons. Most of the subjects had to attend their regular classes after the reading test was administered and there was no other session available for the researcher to do the interview. However, this problem was tackled by employing some Iranian ESL students at a later date. This is dealt with under the heading 'The Follow-up Study'.

The tests were then scored using different methods for each test. First, the students' performance on TOEFL was measured as it was used as a criterion to choose the potential subjects of the study. Students' raw scores on each section were converted into score ranges and then the total (converted) score range was obtained for each subject. Then the obtained scores were arranged from the highest to the lowest. The scores ranged from 497 to 584. It was decided to include the subjects whose scores ranged from 540 to 570 (based on the mean score and standard deviation, that is, those whose scores were 2 standard deviations above and 2 standard deviations below the mean score were included). Eleven subjects with scores lower than 540 and five subjects with scores above the criterion were excluded from the study and their tests were not scored.

The test of awareness of formulaic sequences was scored using a two point scale. First the number of correct answers was counted and scored. Here, the sequence had to be recognized as a whole, with all the constituent parts. If a sequence was identified and separated with extra words or with some words

missing, it was awarded two points provided that the meaning of the sequence was not affected. For example, in the case of the sequence '*a multitude of sins*', if '*a*' was not included it was considered as correct; however, if '*sins*' was not included, it was not regarded as a complete answer since the meaning of the sequence would change. In these cases, if the phrase identified was a formulaic sequence although not the one intended by the researcher one point was given. The maximum possible score was 80 for the 40 items.

The test of understanding formulaic sequences was scored giving one point to each correct answer. As it was in multiple-choice format, the answers were either wrong or right and there was no partial credit. As the test was composed of 20 items, the maximum possible score was 20. There was no penalty for providing incorrect answers.

To score the reading comprehension test, the researcher wrote down the answers to the questions and then identified the idea units. This provided the researcher with a set criteria to be used for scoring the test. The reading comprehension test was scored on a two-point scale. The subjects who provided a totally correct response to an item got two points. If a partial answer was provided, one point was given. If a question was composed of two parts, one point was devoted to each part. Where a question required an answer containing more than two idea units, any two correct idea units were scored two. The maximum possible score was, therefore, 40 for the 20 items. To establish inter-rater reliability, a second rater who was an ELT university teacher was asked to score 15 randomly selected reading tests independently using the same criteria as the researcher. The correlation between the two sets of scores (by the researcher and the second rater) was estimated and an agreement of 96% was achieved suggesting that the scores given by the researcher were reliable.

3.3 The Follow-up Study

As mentioned at the beginning of this chapter, a follow-up study which was qualitative was conducted to gain more insights into the role of formulaic sequences in L2 reading comprehension. In addition, the results of this analysis were used to find out about the strategies the EFL/ESL readers employ to overcome the problems they encounter with respect to unfamiliar formulaic sequences during reading. In this part information regarding the participants, the test and the procedure employed to gather data for the follow-up study is presented.

3.3.1 Participants

Unfortunately, due to administrative problems, it was not possible for the researcher to gather data for the qualitative study from the same subjects who participated in the main study. Therefore, Iranian students studying at the University of Liverpool were invited to participate in the study. Seven subjects volunteered to take part, however the data from only six (2 females, 4 males) were used in the analysis because one of the subjects withdrew midway as he found the reading passage 'very challenging'. They were graduate students (science and social sciences) and all had been in the UK for at least two years. They had also passed the university English language entrance requirement of TOEFL 550 (CBT 213) or IELTS 6.0. They were aged between 26 and 42.

3.3.2 Test

To gather data the passage entitled *'Resist the Official Pol-speak of Bush's War on Terror'* included in the reading comprehension test used in the main study was employed. The subjects had to read it and retell what they remembered.

3.3.3 Procedure

The subjects attended individual interview sessions lasting for about 2 hours per subject. A combination of free and cued recall, interview and introspective methods of data collection was employed to gather data for the qualitative analysis. All tasks were performed in the subjects' native language (Farsi); however, they were allowed to use either language. Each session was tape-recorded.

First, the subjects were interviewed to obtain information about their reading habits and the strategies they usually employ to infer the meanings of unknown words and phrases while reading. Then they were given instructions on what they had to do and the recall procedure. After that they were asked to read the passage and retell everything they remembered. It is widely agreed that free recall is "the most straightforward procedure for assessing the outcome of reader-text interaction" (Koda 2005, 236). As the reader is asked to retell whatever they remember from the text, additional task requirements such as answering questions in multiple-choice format do not contaminate the data. One of the disadvantages of free recall, however is its reliance on memory. To compensate for this shortcoming, the researcher asked questions to direct their recall and prompted them whenever needed. In some cases where the subjects did not mention some of the target idea units (the idea units selected to be tested in the main study), related questions were asked to see whether it was due to lack of comprehension or simply retention difficulty. As the purpose of the study was to determine whether knowledge of formulaic sequences would help L2 readers, the subjects' recall were directed so that the researcher could identify the possible connection between comprehension and knowledge of the sequences and the knowledge sources the subjects used to extract the meanings of unfamiliar sequences while reading.

Then the subjects were asked to reread some selected paragraphs of the passage and talk about the parts they had difficulty understanding. By doing so, the researcher could determine whether their problems were grammatical complexity of the sentences or lexical in nature; and if lexical, whether they were limited to single words or word combinations.

Each interview and recall session was then transcribed and relevant parts were translated into English by the researcher. Care was taken for the translations to be as close as possible to their Farsi versions. Subsequent analyses were based on the transcriptions and the notes the researcher had taken during the recall sessions.

It is worth noting that in qualitative designs external variables are inevitable as the methods for gathering and analysing data are not as objective as in quantitative designs. In this study, it was hoped to control for external factors and to gather as much data as possible by using a combination of interview, recall and introspective methods.

3.4 Summary

In this chapter the overall design of the study was described. It was hypothesized that understanding formulaic sequences and the ability to recognize them in a text would not influence Iranian EFL/ESL students' performance on a reading comprehension test. The main study involves ex post facto and factorial designs to analyse the data and the follow-up study involves qualitative analysis of data gathered through recalls. The independent variables, i.e., understanding and recognition of formulaic sequences, and the dependent variable, i.e., reading comprehension were also introduced. The subjects who

took part in the study were depicted as well. Finally, the procedures used to develop and score the tests were explained.

In the following chapter, **Results and Analyses**, an account of the statistical analyses employed to analyse data and their results will be given. The results of the follow-up study will also be presented.

Chapter IV
Results and Analyses

4.1. Introduction

The first part of this chapter presents the results of the statistical analyses employed to test the null hypotheses formulated on the basis of the research questions presented in chapter 1. The results will be presented in three parts in accordance with the two sets of hypotheses of the quantitative analyses - the main hypotheses of the study and the secondary hypotheses which are relevant but not the main focus of the study - and the qualitative analysis.

The main research questions addressed the relationships between recognition and understanding of formulaic sequences and Iranian EFL readers' reading comprehension. They also sought to determine if the interaction of recognition and understanding of these sequences affect reading comprehension. The data consisted of scores from four tests: a Test of English as a Foreign Language (TOEFL), a test of reading comprehension, a test of recognition of formulaic sequences and a test of understanding these sequences. The study employed correlational and factorial designs. To analyze the data, SPSS (Statistical Package for the Social Sciences) version 15.0 was used. An .05 level of significance (5% or less chance of being wrong) for rejecting the null hypotheses was selected for both main and secondary hypotheses.

The secondary research hypotheses addressed the relationships between the subjects' overall proficiency level and their performance on the recognition and understanding tests. The research assumption is that learners who score high on the overall proficiency (not just their reading) test get better scores on these tests. The study also examines whether there exists a relationship between the frequency of occurrences of formulaic sequences and the subjects' performance

on the recognition and understanding tests and if there is any difference in male and female subjects' performance on the tests of formulaic sequences.

The second part of this chapter is devoted to the results of the follow-up study which will be used to gain more in-depth information regarding the relationship between knowledge of formulaic sequences and L2 reading comprehension. In addition, it attempts to find out what kinds of knowledge sources and contextual cues L2 readers employ to deal with unknown formulaic sequences while reading.

4.2.1 The Main Study

4.2.1.1 Main Hypotheses

In this part, first a summary of the descriptive statistics of the test results will be given. Then each research question and its corresponding null hypothesis will be investigated under separate headings.

4.2.1.2 Descriptive Analysis

The first step in analyzing the data was performing descriptive statistics. Before that, the accuracy of data entry was examined. Then the data were checked for missing values. Descriptive statistics showed that none of the variables had missing values. The data were also checked for outliers (i.e., an extreme value on a particular item). No outlier was identified, which suggests reasonable homogeneity of the subjects. Floor and ceiling effects (the extent to which the scores pile up at the high end or low end) were considered as well. There was no sign of such effects which suggests that the tests were neither too easy nor too difficult.

The results of the descriptive analysis including the mean scores, standard deviations, the minimum and maximum scores on each test, are shown in Table 4.1 This provides a general profile of the full sample and describes the basic features of the data in the study.

	Min.	Max.	Max. Possible Score	Mean	SD
Recognition	44	72	80	61.83	5.51
Understanding	7	16	20	11.55	2.36
Reading	24	36	40	29.90	2.84

Table 4.1 Descriptive Statistics of variables N = 53

An examination of the mean score on the tests shows that on average the subjects answered 77% of the questions on the recognition and 58% of the questions on the understanding tests correctly. The subjects did relatively better on the recognition test which suggests that either the test was easier or the subjects were better at recognizing formulaic sequences. In terms of the reading test, the subjects answered 75% of the questions correctly.

4.2.1.3 Hypothesis 1

Hypothesis 1 stated that the higher the students' scores on the recognition test, the better their performance would be on the test of reading comprehension. The null hypothesis was formulated as: there is no relationship between Iranian EFL learners' recognition of formulaic sequences and their performance on a test of reading comprehension. To see whether there was a relationship between these variables and if so, what the direction was, a correlational analysis was performed. As the data sets followed a normal distribution (using a histogram Figure 4.1) and the relationship was linear (using a scatter graph, Figure 4.2), a

Pearson Product Moment Correlation Coefficient was used.

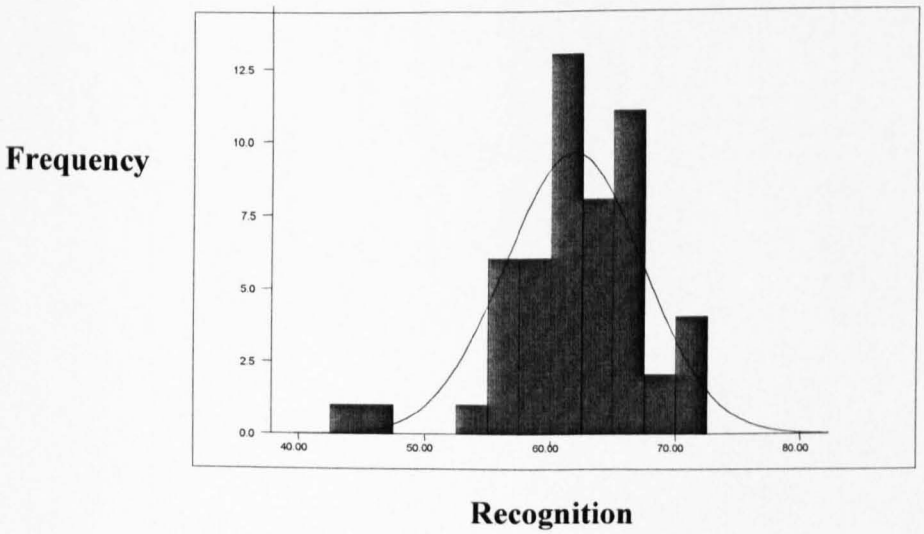


Fig. 4.1 Graph showing the normal distribution of scores on recognition test

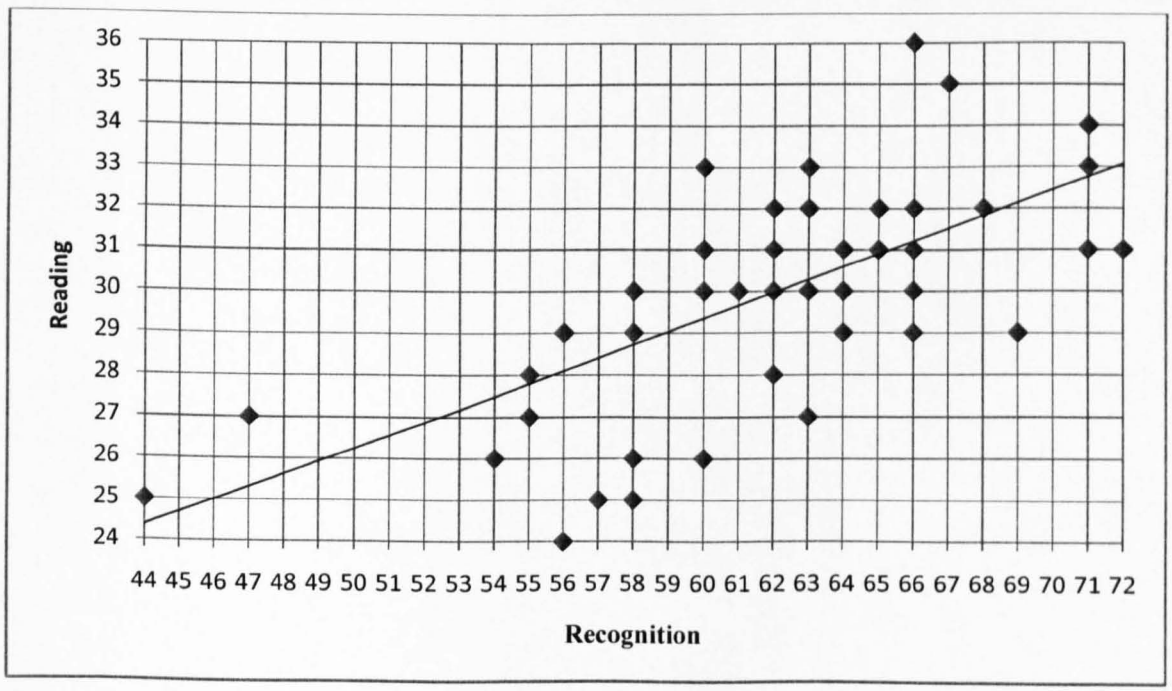


Fig. 4.2 Scatter graph showing the linear relationship between recognition and reading

A value of .667 for the correlation coefficient indicated that at the $p<0.05$ level of significance, a positive relationship between recognition of formulaic sequences and reading comprehension exists. The null hypothesis is therefore rejected.

With respect to the strength of the relationship, Hinkle et al.'s (1979, p. 85) rule of thumb for interpreting the size of correlation coefficient was used as the criterion:

.90 to 1.00 (-.90 to -1.00) very high positive (negative) correlation
.70 to .90 (-.70 to -.90) high positive (negative) correlation
.50 to .70 (-.50 to -.70) moderate positive (negative) correlation
.30 to .50 (-.30 to -.50) low positive (negative) correlation
.00 to .30 (.00 to -.30) little if any correlation

Thus, according to the above guide, a value of .667 shows a moderate positive relationship between the variables. This relationship is shown in Figure 4.7.

The correlation coefficient was then converted into variance overlap (common variance) between the two variables to be able to see how much variance in one measure could be accounted for by the other. As Hatch & Farhady (1982) put it,

> *the magnitude of common variance between two tests reflects the degree to which a S's score on one test can be predicted from his or her score on the other test or the degree to which the ratings of one judge can be predicted from the ratings of a second judge (p. 203).*

The variance overlap between the sets of scores turned to be .445, indicating that almost 45% of the variance in the reading test scores can be explained by variation in the recognition test scores or vice versa.

4.2.1.4 Hypothesis 2

The second null hypothesis stated that: there is no relationship between Iranian EFL understanding of formulaic sequences and their performance on a test of reading comprehension. In other words, according to this hypothesis, the scores on the test of understanding formulaic sequences and the scores on reading comprehension test do not go together or are not correlated. In order to reject the null hypothesis and show that these two variables are co-related, a Pearson Product Moment Correlation Coefficient was calculated and a value of .180 was found between the two variables. This correlation between the variables is not significant at the .05 level of significance. The null hypothesis is not therefore rejected, leading to the conclusion that the two variables do not change correspondingly. And according to the guide given above, values between .00 to .30 show little if any correlation. This relationship is shown in figure 4.3.

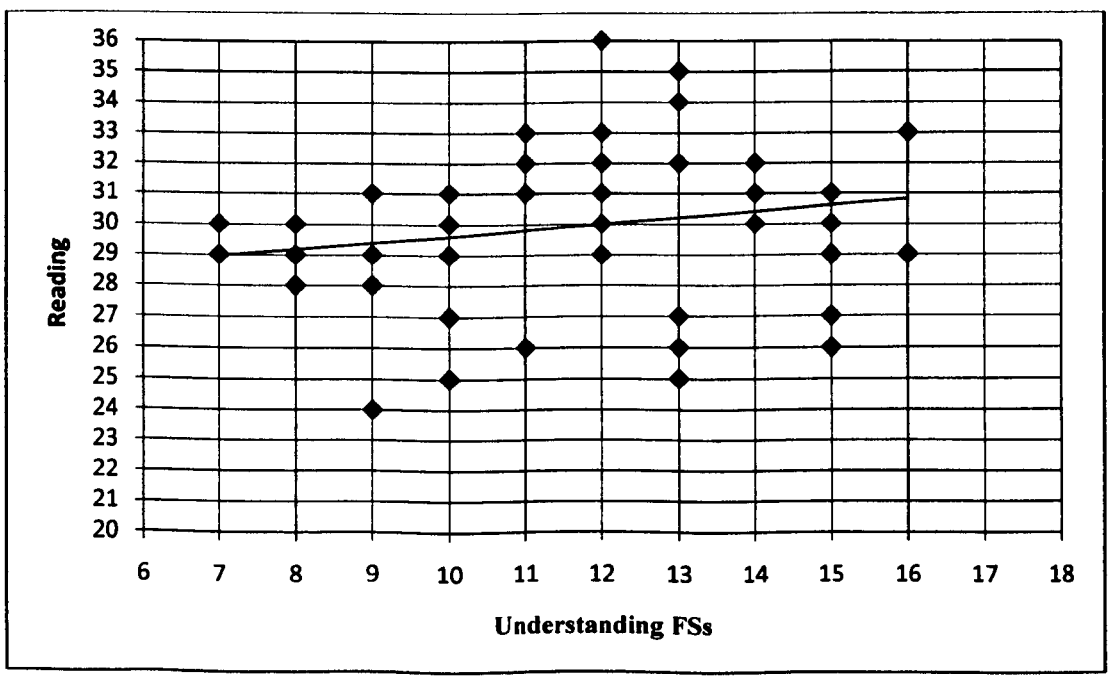


Fig. 4.3 Correlation between understanding FSs and reading

As Figure 4.3 shows, no visible trend of correlation can be seen, though if the scores scattered in the centre are circled, a very slight positive correlation can be seen. The common variance between the two variables was .032 showing that only 3.2% of variance in reading comprehension scores could be accounted for by the ability to understand formulaic sequences.

4.2.1.5 Hypothesis 3

The third null hypothesis stated: the interaction of understanding and recognition of formulaic sequences does not affect the learners' performance on a test of reading comprehension. To test this hypothesis a two-way analysis of variance (Two-way ANOVA) was performed.

Let us recall that the independent variables, recognition and understanding, each have two levels, high and low. This is shown in Table 4.2 along with the mean scores of the four groups on the reading test.

		Understanding (factor B)	
		High	**Low**
Recognition (factor A)	**High**	31.73	30.70
	Low	29.27	28.24

Table 4.2 Reading Means of High and Low groups of Independent Variables

A comparison of the means of the four groups indicates that regardless of their scores on the understanding test, whether high or low, those who scored high on the recognition test got slightly higher scores on the reading test than those who scored lower scores on the test.

To see whether the differences in the means due to the recognition factor are statistically significant, the results of the two-way ANOVA, performed on the reading scores by 'recognition' and 'understanding' (shown in Table 4.3) were considered. As the table shows, no significant interaction between the two variables ($F = 0.095$ ($p = 0.759 > 0.05$) is found, denoting that recognition of formulaic sequences affected the reading scores of the subjects who scored high and those who scored low on the understanding test almost similarly.

Dependent Variable: Reading

Source	Mean square	F	Sig.
Understanding x Recognition	.544	.095	.759

Table 4.3 Two-way ANOVA for Understanding x Recognition and Reading

In order to better understand what has happened between the levels of the factors, the means of the groups are plotted in Figure 4.4.

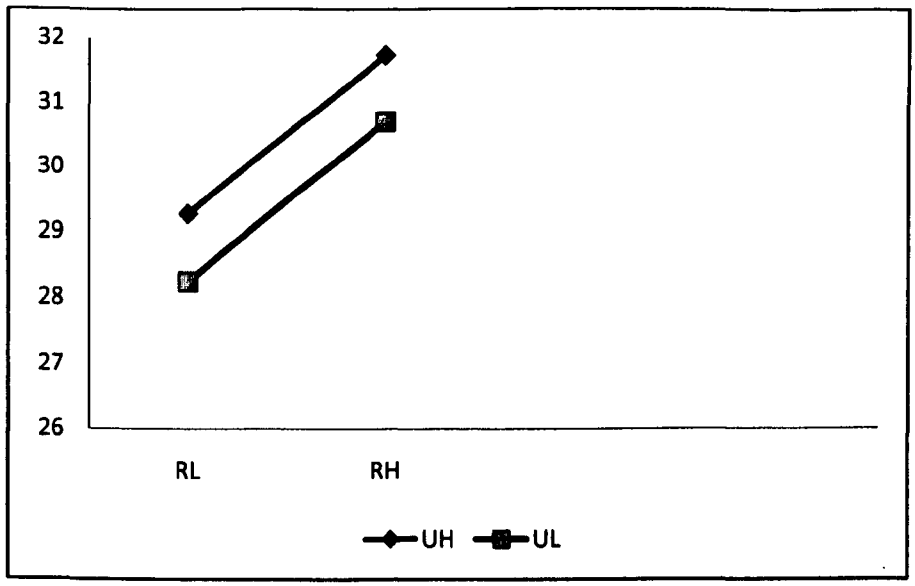


Figure 4.4 Recognition x Understanding
RL: Recognition Low RH: Recognition High
UL: Understanding Low UH: Understanding High

It is clear from the figure that the main effect for recognition is stronger than that for the interaction between recognition and understanding, as both high and low understanding groups got higher scores on the recognition test. In other words, we cannot claim that those who got higher scores on the understanding test got also higher scores on the recognition test or vice versa.

4.2.2 Secondary Hypotheses

This section deals with the statistical analyses performed to test the null hypotheses formulated on the basis of the following secondary questions using the existing data:

7. Is there any relationship between frequency of formulaic sequences and recognition of these sequences?

8. Is there any relationship between frequency of formulaic sequences and understanding these sequences?

9. Is there any relationship between Iranian EFL learners' recognition of formulaic sequences and their overall proficiency?

10. Is there any relationship between Iranian EFL learners' understanding of formulaic sequences and their overall proficiency?

11. Is there any difference between male and female subjects' performance on the recognition tests?

12. Is there any difference between male and female subjects' performance on the understanding test?

To find answers to the first two questions, the frequency of occurrences of the 40 formulaic sequences used in the recognition and understanding (20 sequences) tests was estimated by consulting the British National Corpus. Next, the number of correct answers given to each item in each test was counted. The data were then analysed and two sets of correlation coefficients were calculated (once for the relationship between frequency and recognition and once for frequency and understanding). As the data was not normally distributed (using a histogram, e.g., Fig. 4.5), degrees of relationships between the variables were calculated using rank correlation. The Spearman rank-order correlation was therefore employed.

The first null hypothesis stated that there is no relationship between the frequency of occurrence of formulaic sequences and recognition of these sequences.

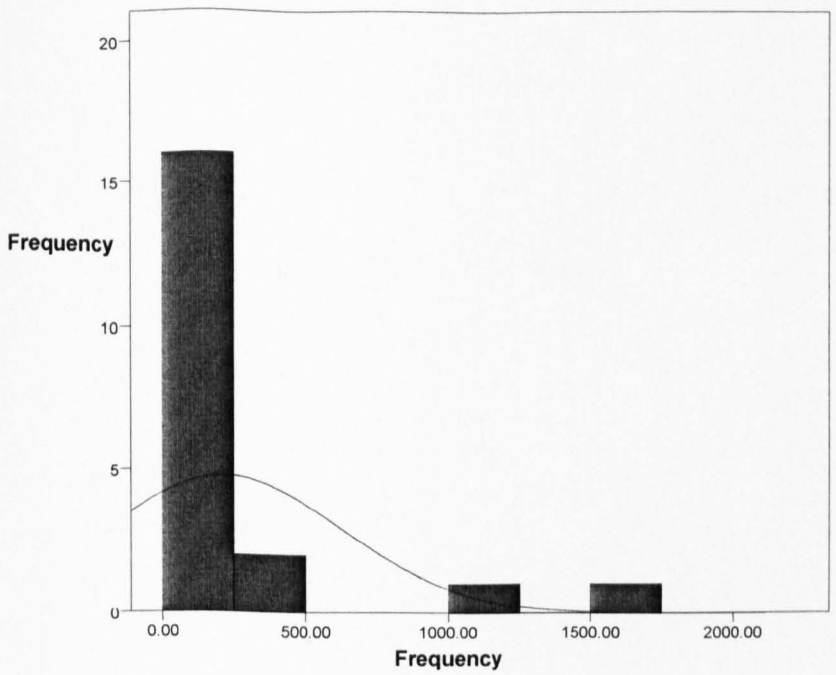


Fig. 4.5 Graph showing the distribution of frequencies of formulaic sequences

To test the hypothesis, the frequency of occurrences of the 40 opaque and transparent sequences used in the recognition test and the corresponding correct answers given to each item in the test of recognition were considered as the variables. Then, the data was rank ordered and the Spearman correlation coefficient was calculated.

The value of the correlation coefficient did not indicate a relationship between the variables ($\rho = 0.137$, $p > 0.05$). The null hypothesis is therefore supported. In other words, no evidence was found to suggest that these EFL learners recognized highly frequent formulaic sequences better than those sequences that occur less frequently in a corpus of 100 million words.

Among the target formulaic sequences, those which are recognized by greater number of subjects are '*when it comes to*', '*it is little wonder that*', '*survival of*

the fittest', '*go against*', '*in collaboration with*', '*natural selection*', '*far beyond*', '*turn out*', '*as a rule*' and '*in one's mind's eye*'. As Appendix B shows, some of these sequences are not necessarily highly frequent sequences in the BNC. For example '*it is little wonder that*' appears in the BNC 24 times while '*when it comes to*' occurs 1418 times and both score highly, 49 and 48 respectively. Sequences which are not recognized easily by students were '*invoke God's help*', '*realize the full potential of sth*', '*not a bit of it*', '*in the making*', and '*spin doctor*'. A look at the frequencies of these sequences in the BNC shows that some of these sequences like '*not a bit of it*' and '*in the making*' are repeated 42 and 225 times in the BNC respectively which are far more than 6 which is the frequency of occurrence of '*invoke God's help*' or '*realize the full potential of sth*', however all are answered correctly by just a small number of subjects (ranging from 27 to 32). A more detailed discussion regarding the apparent lack of relationship between recognition of formulaic sequences and their frequencies in the BNC is given in Chapter IV section 5.2.2.1.

The corresponding null hypothesis to the second question stated that: there is no relationship between frequency of formulaic sequences and understanding these sequences. To test the hypothesis, the frequency of occurrences of the 20 opaque formulaic sequences used in the understanding test and the corresponding correct answers given to each item in the test of understanding were considered as the variables. Here again, the data was rank ordered and the Spearman correlation coefficient was calculated. The results of the analysis ($\rho = .061$, $p > 0.05$) did not show a significant correlation between the two variables suggesting that the number of times a sequence occurs in the corpus is not correlated with understanding the sequence (for the formulaic sequences studied and the subjects in this study). The null hypothesis was therefore supported.

A look at the answers given by the subjects shows that the formulaic sequences that received the most correct answers on the test of understanding are *'turn out'*, *'ever since'*, *'do one's worst'* and *'as a rule'*. While *'Turn out'* occurs in the BNC 1050 times, *'do one's worst'* occurs just 6 times; however, both are among highly scored sequences. And while *'dead end'* and *'a case in point'*, which occur in the BNC 105 and 125 times respectively, are among sequences that are given the lowest correct answers (the former is answered correctly by 29 subjects and the latter by 26), *'stumble across'* is answered by 36 subjects although it occurs only 8 times in the BNC. This is discussed in more detail in Chapter IV section 5.2.2.1. See Appendix E for a list of problematic and unproblematic sequences.

Table 4.4 summarizes the results of correlation analyses for the first and the second hypotheses.

	Frequency & Recognition	Frequency & Understanding
Spearman's rho	.137	.061

Table 4.4 Correlations between frequency of FSs and Recognition and Understanding

The third and the fourth hypotheses dealt with the relationships between the subjects' overall proficiency and their recognition and understanding of formulaic sequences. To establish the homogeneity of the subjects for the main study, a TOEFL test was used and students falling between the score range of 540-570 were selected. To test the hypotheses related to proficiency, the scores of all subjects who participated in the study regardless of their TOEFL scores

were considered in order to have a wider range of proficiency level (497-584). Thus, the scores of 16 more students were included, yielding a total number of 69 subjects. To test these hypotheses a Pearson Product Moment Correlation Coefficient was employed as the scores were normally distributed.

With respect to the third hypothesis stating that the higher the subjects' overall proficiency, the higher their scores on the recognition test, the results showed a positive relationship between the two variables at $P = < 0.05$ ($r = .427$). According to the guide given by Hinkle et al (1979), a value of correlation between .30 to .50 indicates a low relationship. It can therefore be concluded that there is a low but significant positive correlation between recognition and overall EFL proficiency.

The common variance between the two variables is .194 showing that 19% of the variance in recognition test scores could be accounted for by the subjects' overall proficiency level or vice versa.

The fourth hypothesis investigated the relationship between the students' overall proficiency and their scores on the test of understanding formulaic sequences. The results of the correlation coefficient showed a value of .441, significant at 0.05. Using the guide for interpreting the size of correlation given by Hinkle et al (1979), it can be concluded that there is a low but significant positive correlation between the two variables. Therefore, the null hypothesis cannot be supported, leading to the conclusion that the higher the subjects' proficiency level, the better their understanding of formulaic sequences. The common variance between the two variables is .182 suggesting that 18% of variance in the understanding test scores could be accounted for by the subjects' overall proficiency level or vice versa. Table 4.5 summarizes the results of correlation analyses for the third and the fourth hypotheses.

	Understanding	Recognition	Proficiency
Understanding		.438*	.441*
Recognition	.438*		.427*
Proficiency	.441*	.427*	
N=69			

* Correlation is significant at the 0.05 level (2-tailed).

Table 4.5 Correlation between Understanding and recognition of FSs and Overall Proficiency

The fifth and the sixth questions addressed the difference between male and female students' performance on the recognition and understanding test. The corresponding null hypothesis to the fifth question is: there is no difference between males and females in terms of recognition of formulaic sequences. Table 4.6 presents the descriptive statistics, including the mean scores and standard deviations of males' and females' scores on the recognition test.

	N	Mean	SD
Male	24	61.54	6.20
Female	29	62.07	4.98

Table 4.6 Descriptive statistics for male and females' performance on recognition test

As Table 4.6 shows, there is a very slight difference in the mean scores of males and females on the understanding test. The question is: is it statistically significant? To find the answer to this question and thereby test the hypothesis,

an independent samples t-test was performed. The results of the t-test are presented in Table 4.7.

	F	Sig.	t	Sig. 2- tailed
Equal variances assumed	.068	.796	-.344	.733

Table 4.7 Independent samples t-test for males and females scores on Recognition Test

As the table shows, there is not a statistically significant difference between the mean recognition score for males and females ($t = -.344$, $p = .733 > .05$). The null hypothesis is therefore supported, indicating that there is no difference between males and females in terms of recognition of formulaic sequences.

The null hypothesis corresponding the sixth question states that there is no difference between males and females in terms of understanding formulaic sequences. Table 4.8 presents the descriptive statistics, including the mean scores and standard deviations of males and females on the recognition test. As the Table 4.8 shows, the difference between the means of males and females is not much.

	N	Mean	SD
Male	24	11.13	2.23
Female	29	11.89	2.32

Table 4.8 Descriptive statistics for male and females' performance on the understanding test

To prove empirically the similarity of the two groups' performance on the understanding test, and thereby to test the hypothesis, an independent samples t-

test was performed. The results indicate (Table 4.9) that there is no statistically significant difference between the mean scores of males and females on the test of understanding formulaic sequences.

	F	Sig.	t	Sig. 2-tailed
Equal variances assumed	-1.23	.610	-.344	.226

Table 4.9 Independent samples t-test of males and females on Understanding Test

The null hypothesis is therefore supported ($t = -1.23$, $p = .226 > .05$). In other words male and female subjects performed similarly on the understanding test.

4.2.3 Summary of the Quantitative Analysis (The Main Study)

The quantitative part of the study presented the results of the statistical analyses to test the main and the secondary hypotheses of this study. The results of the three main hypotheses of the study along with the statistical analysis employed to test them are as follows:

Hypothesis 1: the results of the correlational analysis showed a moderate and statistically significant relationship between recognition of formulaic sequences and reading comprehension.

Hypothesis 2: the results of the correlational analysis did not show any relationships existing between understanding formulaic sequences and reading comprehension.

Hypothesis 3: the results of the two-way analysis of variance showed a

significant main effect for recognition; however, no main effects for understanding formulaic sequences and interaction effect for recognition x understanding were found.

The results of the secondary hypotheses are summarized as follows:

Hypotheses 1 and 2: the results of the correlational analysis showed that there is no relationship between recognition of and understanding formulaic sequences and their frequency of occurrence.

Hypotheses 3 and 4: the results of the correlational analysis showed low positive relationships between overall proficiency level and both recognition and understanding of formulaic sequences.

Hypotheses 5 and 6: the results of the independent samples t-test failed to show any significant differences between males and females in terms of recognition or understanding of formulaic sequences.

4.3 The Follow-up Study

4.3.1 Subjects' Recall

In this part, the results of the recall and interview sessions are presented. The results pertaining to each subject are given in separate sections. Only translations of the excerpts of the transcriptions that were relevant to this study are presented. For the sake of brevity, parts of the transcripts where the subjects knew the meanings of formulaic sequences or where they did not know the meanings of some sequences and did not try to use available cues to figure out the meanings the transcripts are not given. Pauses, false starts and other types of fillers are not included in the transcripts either.

In presenting the translated transcriptions, normal type is used for the

researcher's explanations and italics are used for the interviews and recalls. The formulaic sequences used in the text and addressed by the subjects are typed in bold.

Transcription Conventions:

I = Interviewer

S = subject

[ ] = researcher's words

( ) = words or formulaic sequences spoken in English by the subjects

Subject One:

The pre-reading interview with this subject showed that he likes to skim through articles in online newspapers. In addition, he does not bother to look the words up in a dictionary as far as the unknown words do not hamper his comprehension of main ideas. He prefers to guess the meanings of unknown words and phrases using his background knowledge of the topic of the text. It took him twenty five minutes to read the passage.

S: *It's about the politicians who don't mind their language.* [Repeats the same idea in other words]

I: *How did you come to that conclusion?*

S: *In the first paragraph, it says that words are important and how Bush's talk about al-Qaida caused problems.*

I: What kind of problem?

S: War. Bush asked everyone to be against them[to do one's worst]. [Looks through the text and reads aloud] *'An **open-ended**, global, **no-holds-barred war on terror**'. The outcome was war against terrorism. People were **locked up**.*

*I: What does **no-holds-barred** mean?*

S: جنگ تمام عیار. Doesn't it? We say a جنگ تمام عیار in Farsi.

I: How did you know that?

S: When I was reading the sentence, the first meaning that came to my mind was this. (Global), all over the world and تمام عیار.

[Continues] *S: America celebrates it independence while Palestinians are trying hard to find freedom.* [In the passage it goes 'US ***declaration of independence'***...'***engaged in a struggle*** for ...']

I: What does it imply?

S: Words.

I: What about words?

*S: How they are used by politicians. I mean, American politicians talk about **declaration of independence** but they don't care about Palestinians.* [The writer in fact talks about how words can define how a nation sees itself.]

[Continues] *S:What does **WMD** stand for? I know that it is something used by the West to justify their actions but, I don't know what exactly it refers to.*

I: How do you know they are used for justification?

S: [He reads the related part of the text aloud] *'**WMD** is a vague... that can be used to **cover a multitude of supposed sins'**.*

I: What does it mean?

S: I don't know exactly. I mean I don't know what it means in Farsi. I think it means to hide their wrongdoings.

I: What made you to think so?

S: 'Cover' means پوشاندن. I don't know 'multitude'. I know multi means چند. [He thinks] *I don't know but 'sins' means گناه. So, 'to cover sin' means they try to hide their wrongdoings.*

*I: What do you think **WMD** stands for?*

S: I told you before. I don't know.

I: Can you guess?

S: 'W' may stand for word as the topic is on words. 'M' may [Thinks]. *I don't know.*

[At the end of recall session] *I: Did you get all parts of the text?*

S: I think I got the main points. There were some parts I didn't. It was because I didn't know the words. For example, I didn't know the meaning of [points to 'emancipation' already circled] *I don't even know how to pronounce it, Or,* ***'D-day'*** *in 'Sorry ducks, it's* ***D-day****'. What is it? I guessed it might be a secret code.*

I: How come?

S: Because it talks about the ***second world war*** *and Nazis. It says they were afraid of being overheard.*

He did not have much difficulty in getting the main points of the text. Of the idea units determined for the scoring of the reading comprehension test, half were covered fully in his first attempt to recall the passage. The remaining idea units which were not mentioned first were addressed, some partially, when the researcher directed his recall. All in all, he could be considered an above average reader. Although he had wrong impressions regarding some specific ideas, in general, he did get the main ideas. This can be attributed to his good command of English, or his reading habits (He is used to reading online news articles) or his large vocabulary as there were just a few words he had circled while reading. Although it is difficult to say with certainty whether his good reading was due to his knowledge of formulaic sequences, it can be said that he used several sequences in his recall (their translations or implied meanings). This shows that these sequences have been important in understanding the text and that he knew these sequences or at least managed to guess their meanings. In some cases, he used a concept which was implied by a sequence although he did not use the exact sequence. For example, instead of using the Farsi equivalent of the expression '*to cover a multitude of sins*' he defined it as 'to

justify their wrongdoings'. Later when asked about the meaning of the expression, it was shown that just the words 'cover' and 'sins' had been enough for him to guess the meaning of the expression. He also resorted to Farsi in addition to context clues as a strategy to guess the meanings of sequences as in '*no-holds-barred war on terror*' although his guesses were not accurate in some cases.

Subject Two:

The interview with this subject showed that she devotes much of her time reading texts in her subject and keeps up with the news via TV. She tries to get the meaning of unknown words from context or the structure of words but eventually uses a bilingual dictionary to look up the words and check her guesses. While reading she would regress several times to get a sentence or a part of a sentence. It took her forty minutes to finish reading the text.

S: That ordinary people can't talk freely and should watch their language at war time but politicians can, even if it is not sensible. [She confuses sensitive for sensible]

I: What made you come to that conclusion?

S: First paragraph. That words can cause problems if not used properly. [Gives the Farsi equivalent of the expression '**careless talk costs lives**'] *During the* ***second world war*** *careless use of words caused lots of people to die.*

I: Is it in the text?

S: It says 'costs lives'. It also says that terrorism is defined differently. It depends on who talks about it. It was about Bush's ***war on terror****. Well, he* ***declared war against*** *terrorism and persuaded others to believe that Muslims are terrorists.*

I: How did you come to that conclusion?

S: Because the passage is all about terrorism and ***war on terror****. In the first paragraph, I remember it said that bush asked for a* [Thinks] ***war against al-Qaida****.* [Looks for it.] *Here* [Reads aloud] *'****crusade against*** *al-Qaida'. It says terrorism is increasing* ***[growth industry]****. It can be seen everywhere from Afghanistan to Palestine.*[Not mentioned in the text]

*I: Do you know what '****growth industry****' means?*

S: [Thinks]. *I don't know.* [Reads the sentence aloud] *I mean I don't know its Farsi equivalent but I thought it meant increasing, something that grows and increases. It asks 'why does it grow'? It was my first impression.*

S: It also talks about ***(WMD)****. That different countries define it differently. I mean developed and developing countries.*

I: What do you mean?

S: I mean for different purposes. Developed countries use it for a different purpose. Developed countries use it to do whatever they want, call it war with Satan.

I: What does it refer to?

S: I thought it was an abbreviation that meant weapon, (arsenal) I'm not sure. But I think it is a kind of weapon.

I: How did you know that?

S: Because it says 'developed countries have their own ***WMD****, of course, but their arsenals are ... .'*

I: What does 'arsenal' mean?

S: Weapon.

I: What do you remember about the word 'occupation'?

S: Occupation? America occupied Iraq and says it is freedom. They avoid using it.

I: Why?

S: Because Bush says they have released Iraq.

*I: What does '**barged uninvited into**' mean?*

S: [Reads it aloud] *I thought it meant go uninvited into a place, a country , for example. Because it says 'uninvited'.*

I: Did you understand all parts of the passage?

S: I think so. The first part was difficult to understand. Very difficult indeed. I had to go back and read some sentences twice. But it was ok. I didn't know some words and phrases. It seems difficult at first but as you proceed through the text, it gets easier.

I: Can you give examples of words you didn't know?

S: [Looks at the underlined words for a while.] *For example, '**fifth columnists**' I know the meaning of 'fifth' and I know what 'columnists' means, but I can't see the connection* [Gives the Farsi equivalent of fifth plus columnist as journalist] *what has it got to do with Nazis. I didn't know the meaning of 'crusade' but I guessed it might have something to do with war because it say '**crusade against al-Qaida**'.*

This subject was satisfied with her reading as she had thought the passage would be difficult based on her first impression from the title of the text. In comparison to the first subject, she could recall fewer idea units. For example, when asked about the concluding part of the passage, she said she had a vague impression. However, despite the fact that she had underlined a relatively large number of words, she had got half of the target idea units of the text.

Like the first subject, she tried to guess the meaning of unfamiliar expressions from the context and co-text. For example, she guessed 'W' in '*WMD*' stands for weapon because she knew what the word 'arsenal' which was used in the same sentence meant. She also used the Farsi equivalents of some expressions like '*careless talk costs lives*'. In some cases, she used the meaning of one part of a sequence to get the whole meaning. For instance, to get the meaning of

'growth industry', she used her knowledge of the meaning of growth. One interesting finding about this subject was that she used her background knowledge about the topic to overcome her difficulties in comprehension. Her guesses regarding the meanings of words and expressions were not always successful, though, as was shown in the case of *'fifth columnist'*. What is interesting is that when referring to this phrase she actually used it as a sequence. This shows that she was aware that it was a phrase and tried to come up with the meaning of the whole using her knowledge of its parts.

Subject Three:

The pre-reading with this subject showed that he is not contented with his English although he says since commencing his PhD his English has improved a lot as a result of working with native speakers. He thinks he does not have a good knowledge of vocabulary and has problems in understanding and using some grammatical features like 'tense'. He looks the meanings of the unknown words up using a bilingual dictionary. When he started reading he decided to withdraw from the experiment as he felt the text was too difficult for him. Then he changed his mind and continued reading. Although he had problems understanding particular sentences, he managed to get the main ideas of the text plus some of the idea units specified by the researcher. However, he got some parts wrong. It took him about half an hour to read the text.

S: Playing with words. You do something and by using words you try to deceive others. For example, liberation instead of occupation. They use good labels to prevent people from showing negative reactions. US occupied [barge into] *Iraq* ***in the name of freedom*** *and refuses to go away. Occupation is not a good word.*
I: What do you mean?
S: Americans don't want to use it. [Reads] *No good word, oh,* [Laughs] ***no-go***

word. Why did I think it was 'no good word'?

*I: Do you know what '**no-go**' means?*

S: [Thinks] *It means 'don't go'. Maybe.*

I: Do you know what this means? **[barge uninvited into]**

S: Not exactly. I guess it means to go uninvited. Somebody else's country.

[Continues] *S: It is not clear what exactly the word 'terrorism' means. Different interpretations. US kills people and **locks them up** in Guantanamo and does all this **in the name of (war on terror).***

*I: What was the outcome of Bush's '**war on terror**'?*

*S: He encouraged everyone to do whatever he wants, his worst intentions, **in the name of** (security). They put people into prisons.*

I: His worst intentions? What do you mean?

S: Yes, here [Reads aloud: 'every aspiring autocrat ***to do his worst*** ...].

[Continues] *S: They* [US] *want to bring freedom for Palestinians but occupy their lands. They say one thing and do something different.*

*I: What does '**lock up**' mean?*

S: I didn't know but I guess it meant to put in prison because it says [reads aloud: '... or locked up in Bagram or Guantanamo or ...].

*I: What does '**no-holds-barred**' mean?*

S: I think it means a war that has no end.

I: How did you come to this conclusion?

S: Because (no) means it doesn't have. [Thinks] (holds) means to stop something. So, (no stop).

*S: **Semantic (paradox).***

*I: What does '**semantic paradoxes**' mean?*

S: Concepts? Meaning. Something that. When what you do contradicts what you say. We have the word in Mathematics.

[Continues] *S: It is also about talking carelessly. Nonsense talk causes*

problems, disaster. Like the ***second world war.***

I: You mean ***the second world war*** *broke out due to loose use of language?*

S: No. Bush's talk caused war. In the second world war, they were careful not to be heard by the enemy. They used codes.

I: What Enemy?

S: ***(Fifth columnist)*** *ستون پنجم دشمن, Nazis.*

I: What does ***fifth columnist*** *mean?*

S: ستون پنجم دشمن.

I: Did you know it?

S: I guessed, as it is about Nazis and war.

I: Just because it was about war?

S: No. I knew the expression ستون پنجم دشمن in Farsi and when I saw fifth I guessed it might refers to it.

I: What does the writer recommend?

S: That politicians should watch their language. That ***free expression*** *should not be ignored.*

I: Do you know what ***spin doctor*** *means?*

S: Where is it? [Reads aloud] *It says 'as every spin doctor knows,... '. (Stupid)? Every stupid doctor knows that.*

I: What made you think so?

S: [Laughs] *I don't know. It just came. Wants to say everybody knows that. Something which is axiomatic.*

I: What do you think was your main problem in understanding the text?

S: So many proper nouns. I couldn't even read them. Some sentences were long. I couldn't find the connection between the first part of a sentence and the second part of it. Some words were difficult as well .

Although he was not as competent as other participants (based on their IELTS or TOEFL scores and the number of words underlined), he had no difficulty

recalling the main points of the text as he used context clues to process through the text. However, in some cases, he could not talk about the specific idea units that the researcher asked about. He tried to extract the meanings of unfamiliar expressions using a number of knowledge sources, although in some cases, he was not successful. He tried to guess the meanings of some expressions using his background knowledge. For instance, in the case of '*semantic paradoxes*', his knowledge of maths came to his help as the word 'paradox' is not translated but used in its original form. He also guessed the meanings of some expressions by summing the meanings of their parts. e.g., '*no-go*', although in his first attempt he had seen the word as 'no good'. For him, as for previous subjects, co-text has been helpful as in the case of '*barge into a place*'. He used the words 'uninvited and somebody else's country' to guess the meaning of the sequence.

Subject Four:

She is not keen on reading news articles but enjoys reading her own subject. She believes she is a slow reader even in Farsi. She looks every word up in a bilingual dictionary. She does not trust her guesses. However, she believes knowledge of the topic sometimes is helpful to guess some words. She managed to finish reading the text in forty five minutes.

S: ***Second world war*** *and that careless talk would cause problems. In general, it says that politicians should mind their language. And also about Bush. He* ***declared war against*** *al-Qaida.* ***War on terror****. That terrorism is a word that is defined differently by different countries. And that they do whatever they want* ***in the name of*** *freedom. And that politicians should be careful. For example, Bush did not use the word 'war' after* ***[ever since]*** *his use of the word caused problems for Muslims.*

I: What do you remember about ***WMD****?*

S: I didn't know what the abbreviation referred to. I really didn't get that paragraph. 'Totemic' for example what does it mean? Totemic ***WMD*** [Repeats]. [Reads the paragraph once more] *I don't get this part just that* ***WMD*** *is used to cover faults* [**to cover a multitude of sins**]. *Maybe it refers to a specific use of language. I mean a strategy to cover one's faults. And developed countries have different strategies to cover their faults than developing countries.*

*I: What does '****under siege****' mean?*

S: Under siege [Thinks]. *I don't know what siege means. Pressure? Under pressure.*

I: How did you come to that?

S: [Reads aloud] *'****civil liberties*** *are everywhere* ***under siege'****. Just because liberty or freedom is under pressure. They don't want people be free.*

*I: What does '****civil liberty****' mean?*

S: Freedom of human beings?

I: What words do politicians avoid? And why?

S: Occupation of Iraq. Resistance. [Reads the paragraph aloud] *They don't want it to look like Palestine.*

I: What was your main problem in understanding the text?

S: Words. I skipped the parts that had unfamiliar words. If I try to understand difficult parts, I lose the plot. Then I have to start from scratch. It takes me lots of time to read a passage and understand every part of it. I look each unknown word up in a dictionary. I don't trust my guesses.

She could talk about the main points of passage. However she has difficulty talking about specific idea units even when addressed. She failed to use context clues to guess the meanings of words and phrases. She used avoidance strategies whenever she was directed to parts that she had problems understanding. For example, when asked about occupation, she did not mention the real reason for

politicians avoiding using the word and when asked about the meaning of the expression 'barged into', she did not even try to guess using context cues. This may also be due to her learning style. She might be more reflective than impulsive because she herself pointed out that she did not trust her guesses. Unlike other subjects, she did not try to use analogy as a strategy to get the meaning of the parts she failed to get.

Subject Five:

This subject uses online dictionaries and tries to find the meanings of unfamiliar words from the context. He likes reading news articles as far as he is familiar with the topic. It took him about thirty five minutes to read the passage.

S: Care should be taken in using words. Media and politicians. Words carry concepts that may have negative effects if not used properly. Different interpretations of a word lead to war and unprecedented outcomes. Politicians, media should mind their language and think about their effects on the societies.
I: How did you come to this?
S: I think everything is said in the first paragraph. Yes. In fact, I had this information in mind throughout the passage [The information in the first paragraph]. *I mean I tried to interpret the difficult parts by resorting to what I remembered from the first paragraph. It says* ***'Careless talk costs lives'*** [In English]. *It is repeated a couple of times. Once here and ... here* [points to the expressions]. *American equivalent* ***'loose lips sink ships'*** *and I remember I saw it in the last part of the text as well.*
[Continues] *S: Bush* ***declared war on terror.*** *Terrorism has changed into an issue that everyone tries to avoid. It has a wide application. They use it as a cover term to justify themselves.*
I: Justify what?

*S: **War against terrorism**. They **lock up** others. They say they want to suppress violence. They call them terrorists, evil ...*

*I: What does '**by definition**' mean?*

S: Natural.

I: What do you mean?

I: I mean everybody knows them that way. It is defined this way.

*I: Do you know the meaning of '**growth industry**'?*

S: (Growth) means [Farsi equivalent]. *In economics we have **(growth economy)**. (Growth industry). It means an industry which is growing. Something which is growing. Maybe the industry of advertisement, or military or everything connected with these.*

*I: What do you remember about **WMD**?*

S: [Struggles to find out what it stands for] *I thinks W stands for* [Thinks] *word or* [Thinks] *war.*

I: What is it about?

S: As far as I remember, it is about reviving a school of thought.

I: What do you mean?

S: [Reads the sentence 'This is a universally understood term... proliferating'] *It is put forth and even intellectuals agree with it* [Not mentioned in the text] *and raise their military budgets. Take advantage of it.*

*I: What does '**cover a multitude of supposed sins**' mean?*

S: [Repeats] *I don't know what sin means. For example, a (nonspecific) term is covered.*

I: What do you mean?

S: The meaning of something has been changed. [He is wrong]

[Continues] *S: Politicians avoid using some words. For example 'resistance and occupation'. They are afraid to encourage them to fight.*

I: Whom?

S: Iraqis. Palestinians. They use military forces to set them free. They enter their country without being asked for. **[barge uninvited into]** *Occupation, freedom. There is a (paradox) in their words.* **[semantic paradoxes]**
I: Was the passage difficult for you?
S: No, not really. There were some words I didn't know. But in general I think it was fine. No, I didn't have much difficulty.

Like other subjects, he got the main ideas of the text right. In his recall, he provided additional information not mentioned in the text. In other words, he tried to find a link between what he read and his background knowledge about the topic. For him, the first paragraph played a key role to the comprehension of the whole text. He believed that as he got the point in the first step, he was not confused and could connect ideas mentioned throughout the text to the ideas in the first paragraph and proceed through the text. He had circled just a few words or phrases as unknown or at least problematic. He also used his knowledge of technical words (in the case of '*growth industry*') and structure of the words (proliferating, although wrong) to get the meanings of unknown expressions or words.

Subject six:

The pre-reading interview showed that this subject enjoys reading articles on politics and tries to guess the meanings of unknown words using his background knowledge of the topics. He also uses a bilingual dictionary whenever he fails to guess the meanings of the unknown words and expressions. This subject was very enthusiastic to read the passage as he believed the title was interesting and as he proceeded through the text he showed more interest and used to repeat, 'Oh this is all about us. That's why you have selected it!' He finished the text in twenty five minutes and said he could get the message of the text as he was

familiar with the topic. He was very quick in guessing the meanings of the target expressions although in some cases his guesses were wrong.

S: This is about the politicians, Muslims, war, ***WMD****, any many other things.*

I: Can you talk about them one by one?

S: It says politicians should not tell whatever comes to their mind. Bush says Muslims are dangerous and ***declares war against*** *al-Qaida after September 11.*

S: [Continues] *He asked every (autocrat) to do everything they could* **[to do his worst]** *against terrorism* ***in the name of*** *Security.*

I: What does ***no-hold-barred*** *mean?*

S: I thought it meant ***open-ended****. A war that has no boundaries.*

I: How did you come to that?

S: Because it says ***Open-ended****, global and a war that has no end. It does not hold limits.*

S: [Continues] *Terrorism is used as an explanation by politicians to justify what they do and say. Everybody hates it and it is still the topic of every conversation.*

*I: What does '****in the abstract****' mean?*

S: In sum?

I: What made you come to that meaning?

*S: I know abstract means summary. Therefore, I think '****in the abstract****' means 'in sum'.*

*I: What does '****growth industry****' refer to?*

S: Terrorism. That it is increasing.

I: What does it mean?

S: An industry which is growing.

I: Did you know its meaning?

S: I don't know. I thought it meant an industry or something which is growing or becoming worse.

S: [Continues] *Occupying other countries and war. Bush* ***declared an open-***

ended war *against terrorism and imprisoned* **[lock up]** *Muslims. Bush is more careful these days, though.*

S: (Occupation) makes Iraqis believe that their country is occupied **[barge into]** *and try to free it.*

*I: What does **WMD** mean?*

S: It is about weapons. I think. I remember it said they think their weapons are acceptable.

I: Who?

S: The developed countries. They use weapons to explain what they do. **[to cover a multitude of sins]**.

I: Did you have any problems in understanding the text?

S: It was easy to follow but I didn't know a few words.

This subject had a good understanding of the text and in comparison to other subjects knew more words and expressions (the number of words circled). His success in getting the text can be attributed to his better knowledge of words or his interest in the topic. Like other subjects, he used several knowledge sources to guess the meanings of unfamiliar expressions. For example, he tried to get the meaning of '*in the abstract*' by summing the meanings of the parts. He also used co-text to guess the meaning of '*WMD*' as he said he remembered the sentence talking about weapons. Like other subjects, his main problem was understanding some words.

4.3.2 Analysis of Recalls

What the recall and interview sessions of the subjects yielded can be summarized as follows:

It was found that not all formulaic sequences that appeared in the text seem to be

problematic. Some like '*loose lips sink ships*' and '*to do one's worst*' are among sequences that did not cause much problem as their meanings were clear to the subjects or they could guess their meanings easily using available clues. However, some like '*rapid fire lingo*' could not be easily guessed while reading. A list of examples of problematic and unproblematic sequences is presented in Appendix E and a detailed discussion as why some sequences caused problems and some did not is given in Chapter IV section 5.3.

It was also found that in their recalls, the subjects used formulaic sequences, either transparent or opaque, in English or Farsi to express their point. This is of interest as in cases where the researcher asked directed questions, the subjects use of the target sequences as well as sequences which were not selected was an indication of their recognition of these sequences, and in the case of opaque ones understanding them.

The analysis of these subjects' recall and interview suggests that whether or not the subjects knew the meanings of target formulaic sequences (and other sequences which had not been selected for the test of understanding formulaic sequences employed in the main study), they could identify the main ideas of the text. All subjects were able to talk about the topic of the passage and their differences in performance were found to be in their answers given to questions asking for specific ideas. This was also seen in the answers given to the reading comprehension test in the main study. Almost all subjects who took the test had answered the items related to the main ideas of the passages correctly. It was found that the subjects who participated in the qualitative study used their knowledge of the component parts of the sequences to extract the meanings of unknown phrases and to interpret the sentences and paragraphs in which they appeared. In addition to the topic, the students resorted to a number of other clues to guess the meanings of these sequences.

It was also seen that understanding of some target formulaic sequences played a key role in the subjects' understanding of specific ideas and misunderstanding them led to either misinterpretation or lack of full comprehension while lack of awareness of the meanings of some other sequences did not affect the subjects' general understanding. Subject five, for instance, did not know the meaning of the sequence '*cover a multitude of sins*', and tried unsuccessfully to guess its meaning using his knowledge of its components parts . His understanding of the paragraph in which the sequence was used was to a great extent affected by his inaccurate guess. It can therefore be said that the contribution of formulaic sequences to subjects' comprehension depended to a large extent on their nature and their function in the text. That is, these sequences can be divided into two groups, those which have a key role in comprehension and those which do not play a vital role and are used just for emphasis or more explanation. For example, it seems to be the fact that some formulaic sequences like '*spin doctor*' and '*by definition*' do not have a critical meaning bearing function, at least in this text, as they did not hamper comprehension in cases where the subjects did not know or could not guess their meanings or guessed them wrong. Unlike these sequences, expressions like '*careless talk costs lives*', '*to do one's worst*', '*to cover a multitude of sins*', '*to barge into*' and even '*WMD*' seemed to have important functions and lack of comprehension or wrong guesses about these sequences caused problems and led the subjects to wrong conclusions about the paragraphs in which they were used. There were also some sequences which were not among the target formulaic sequences but proved to be crucial. For example, '*semantic paradoxes*' was mentioned by two subjects in their recall. In the text, this sequence summarizes a whole paragraph in one sentence. This may be an indication that the subjects already knew the sequence or at least could recognise it and easily remember and use it in their recalls.

As the subjects did not have a chance to look the words or expressions up in a

dictionary or make appeals for assistance directly to the interviewer, in cases where they did not know the meanings of sequences, some tried to extract the meanings of these sequences using available cues in the context and co-text and their own background or they were asked by the interviewer to do so. Paribakht (2005) and Bengelali and Paribakht (2004) have studied the knowledge sources used by different groups of L2 readers to infer the meanings of unfamiliar words while reading and have presented a taxonomy ranging from linguistic to extra-linguistic sources. In this part, a taxonomy of knowledge sources used by the subjects of this study to infer the meanings of unfamiliar sequences will be presented, drawing upon the classification given by Paribakht (2005).

As Figure 4.6 shows, the same general knowledge sources (linguistic and non-linguistic) found for lexical inferencing by Paribakht (2005) were also found to be used by the subjects of this study for inferencing the meanings of formulaic sequences. However, the subcategories differ from Paribakht's classification as in this research formulaic sequences are involved and the number of subjects are relatively fewer than she employed in her research.

I. Linguistic sources
- *A. L2-based or Intralingual sources (L2)*
  - *1. Level of target sequence*
    - *a. Meanings of component parts (MCP)*
    - *b. Association (A)*
  - *2. Sentence level or co-text (CT)*
  - *3. Discourse level (DL)*
- *B. L1-based or Interlingual sources (L1)*

II. Non-linguistic sources
- *A. Knowledge of topic (KT)*
- *B. Knowledge of technical terms (KTT)*

Figure 4.6 The taxonomy of cues used in inferencing formulaic sequences

As Figure 4.6 shows, the subjects of this study have used both linguistic and extralinguistic cues to infer the meanings of unknown expressions. Examples of each knowledge source and their subcategories are given below.

Some readers tried to guess the meaning of formulaic sequences using their knowledge of L2 at the level of the sequence itself, the sentence in which it was used or the whole discourse. In using target sequence cues the readers use their knowledge of the meanings of the components of the sequences and the association between them. For example:

*I: What does '**in the abstract**' mean?*
S: In sum?
I: What made you come to that meaning?
*S: I know 'abstract' means summary. Therefore, **in the abstract** means 'in sum'.*

Association refers to the readers association of the unknown sequence with a word or sequence which is familiar (Paribakht 2005, p. 232). For example:

*I: What does '**no-holds-barred**' mean?*
*S: I thought it meant **open-ended**. A war that has no boundaries.*
I: How did you come to that?
*S: Because it says **Open-ended**, global and a war that has no end. It does not hold limits.*

In some cases the whole sentence or some components of it helped the readers extract the meanings of unknown sequences. For example:

*I: What does '**barged uninvited into**' mean?*
S: I think it means go uninvited into a place, a country , for example. Because it says 'and refuses to go away'.

Discourse level cues involve understanding the meanings and relationships between several related sentences. For example:

*'**D-day**' in 'Sorry ducks, it's **D-day**'. What is it? I guessed it might be a secret code.*

I: How come?

*S: Because it talks about the **second world war** and Nazis. It says they were afraid of being overheard.*

Sometimes it is the readers' mother tongue that comes to their help in deriving the meanings of unfamiliar sequences, though in some cases the guesses are not accurate. For example:

*I: What does **no-holds-barred** mean?*

S: جنگ تمام عیار. Doesn't it? We say a جنگ تمام عیار in Farsi.

Knowledge of topic and knowledge of technical terms are among non-linguistic sources that the readers employed to infer the meanings of sequences. Examples are as follows:

1. Knowledge of topic:

*I: What do you think **WMD** stands for?*

S: I told you before. I don't know.

I: Can you guess?

S: 'W' may stand for word as the topic is on words. 'M' may [Thinks]. *I don't know.*

2. Knowledge of technical terms:

*I: Do you know the meaning of '**growth industry**'?*

S: (Growth) means [Farsi equivalent]. *In economics we have* ***growth economy.*** *(Growth industry). It means an industry which is growing.*

To gain more detailed information about the relationship between understanding formulaic sequences and reading comprehension, four parts or paragraphs of the text containing main idea units were selected. Each part contained some transparent and opaque formulaic sequences. Two to five opaque formulaic sequences that seemed to affect understanding those paragraphs were selected. Then, each subjects' recall was analysed to see whether understanding the sequences relevant to the selected parts affected comprehension of those parts or not and what knowledge sources or strategies were used by each subject to cope with the difficulties arising from unknown sequences. Table 4.10 presents the idea units which were selected along with the opaque sequences which occurred in the paragraphs or parts containing the target idea units. The number of the corresponding paragraph/s talking about each idea unit is also given.

Idea Units	Paragraph Number	Target formulaic Sequences
Terrorism	6 to 8	*In the abstract, growth industry, a case in point*
Bush's war on terror	9 & 10	*No-holds-barred, to do one's worst, lock up, rapid-fire lingo, by definition*
WMD	15	*To cover a multitude of sins, under siege, WMD*
Occupation	17	*No-go, barge into*

Table 4.10 Idea units and accompanying formulaic sequences

Table 4.11 summarises subjects' knowledge of the formulaic sequences, either prior to inferencing or after it along with different strategies used by each subject to infer the meanings of unknown sequences. This table also helps to get

a better understanding of the role of these sequences in comprehension of the selected parts.

Target FSs	1	2	3	4	5	6
In the abstract				√		MCP
A case in point			√		√	
Growth industry		MCP	√	MCP	KTT	MCP
No-holds-barred	L1		MCP	CT & A		MCP & A
To do one's worst	√	√	√	√	√	√
Lock up	√	√	CT	√	√	√
By definition	√				MCP	
Rapid-fire lingo						
Under siege	√			MCP		
Cover a multitude of sins	MCP	√	√	√		√
WMD	MCP,CT & KT	CT & A		CT	MCP & DL	CT & A
Barge into	√	CT	CT		√	√
No-go			MCP	√	√	√

Table 4.11 Subjects' knowledge of sequences and their inferencing Strategies

√= previously known Blank cell = not known (or unsuccessful inferencing)

Key to the Abbreviations:

MCP= Meanings of Component Parts

CT= Co-text

KTT= Knowledge of Technical Terms and Expressions

KT= Knowledge of Topic

DL= Discourse Level

A= Association

The first selected idea addressed terrorism. The sequences used in the corresponding paragraphs (6 to 8) were '*in the abstract*', '*a case in point*' and '*growth industry*'. As Table 4.11 shows, the first sequence was known by just one reader prior to inferencing and just one reader could guess its meaning, albeit not its full meaning, using his knowledge of the meanings of its components. Two readers (subjects 3 & 5) knew the meaning of '*a case in point*' in advance and the others either could not or did not try to guess its meaning. Although these sequences were not known by many subjects, this did not affect their comprehension as they were able to talk about their corresponding paragraphs easily. In the case of '*growth industry*' all but one subject either knew in advance or could guess its meaning . This shows that unlike '*in the abstract*' and '*a case in point*' this sequence was critical in conveying the writer's message, as the readers had felt the need to use the available sources to guess its meaning to get the intended message of the writer and they either used its Farsi equivalents or other words or expressions representing the same concept as '*growth industry*' in their recalls. With regard to the use of knowledge sources to guess its meaning, three out of four readers used their knowledge of the component parts of the sequence.

The second idea unit was about Bush's war on terror. '*No-holds-barred*', '*to do one's worst*', '*to lock up*', '*rapid-fire lingo*' and '*by definition*' were included in the paragraphs (9 & 10) talking about it. Here again, it was found that all but two subjects (subjects 2 & 5) used the Farsi equivalent of the sequence '*no-holds-barred*' in their recalls. However those who did not know its meaning and could not guess it did not have much problem understanding the sequence as it was accompanied by 'open-ended' and 'global' which were easier for the readers to understand and could convey more or less the same concept as '*no-holds-barred*'. In the case of '*To do one's worst*' and '*to lock up*', their Farsi equivalents were also used in the subjects' recalls and as the table shows all

subjects knew the meanings of the sequences prior to inferencing apart from one who was able to use the co-text of '*to lock up*' to guess its meaning. Understanding these sequences helped the readers to a great extent to get the writer's message. As the table shows, none of the readers either knew the meaning of '*rapid-fire lingo*' or managed to guess its meaning. The same is true for '*by definition*'. Only one reader tried to guess its meaning using his knowledge of its component parts. However, their lack of understanding did not cause much difficulty in understanding the paragraphs containing these sequences.

The third idea unit concerned '*Weapons of Mass Destruction*' or '*WMD*'. Almost all subjects had at least a partial understanding of the paragraph (15) addressing this idea. None knew what the abbreviation '*WMD*' referred to and what is interesting is that unlike other sequences whose component parts were used to arrive at their meanings, this sequence was guessed using contextual cues, either at sentence level or at discourse level. Unlike '*WMD*', the sequence '*cover a multitude of sins*', was either known or its meaning was inferred using the meanings of its components. The Farsi equivalent of this sequence was used by all but one reader in their recalls and it also helped readers guess what '*WMD*' stood for. One reader tried to guess the meaning of the sequence using his knowledge of its component parts. However, he was not successful and his inaccurate guess led to misinterpretation.

The last idea unit referred to 'occupation'. '*To barge into a place*' and '*no-go*' were the sequences which were used in the paragraph (17) talking about occupation. '*To barge into a place*' had a major role in conveying the message of the writer and was either known or could be guessed. All readers used its Farsi equivalent or its definition in their recalls apart from one (subject 4) who did not even try to guess. This did not however hamper her comprehension of

the paragraph as she tried to talk about the paragraph using other pieces of information available in the text. In other words, although this subject could not guess the meaning of '*to barge into*', there were some other cues in the paragraph including examples of occupied countries like 'Palestine' which helped her to achieve partial comprehension. '*No-go*', however, did not have a major role as the readers' failure to guess its meaning did not result in failure in comprehending the whole paragraph.

What seems necessary to be mentioned here is that, when addressing specific ideas, for example, terrorism, the subjects did not limit their recalls to what was said in the particular paragraph/s addressing the idea. They used whatever they could remember from the text and also its topic. This reflects the fact that in reading comprehension a combination of factors including contextual resources are involved. In addition, one's failure or success in comprehension cannot be attributed to just one factor, e.g., his/her understanding of words and sequences.

The analyses of the data showed that although the subjects were different with regard to their reading habits, their knowledge of the target formulaic sequences and their strategies to infer the meanings of unfamiliar sequences, they did not differ much with regard to their performance on the reading task. Almost all subjects either knew or could guess the meanings of the key target sequences using different cues. This had to a large extent a positive effect on their understanding of the core meaning or the gist of the related parts. For example, the subjects either knew or could use cues like the meanings of the component parts or the co-text of the sequence '*to cover a multitude of sins*' to guess its meaning. This sequence has an important role in comprehension of the paragraph which talks about '*WMD*', albeit not a fully determining role. Thus, in some cases, this partial comprehension based on the readers' guesses helps them guess the meaning of another unfamiliar sequence, for example, '*WMD*'. This

results in a cycle. In other words, sometimes the components of sequences or the readers' general understanding of the text help them to guess the meanings of unknown sequences. This helps them to get the parts of the text in which the sequences are used, albeit not full comprehension. This partial comprehension based on the readers' guesses helps them to guess the meanings of other words or sequences. That is, all of the components of reading comprehension including phrase identification and understanding of sequences go hand in hand to help the readers build up a picture of the text which eases further processing of the text. In cases where a gap is created by unknown sequences, all other components of reading come to help and at least a partial or a vague comprehension is achieved. This, of course, depends on the first hunches. In cases where the hunches are not accurate, the following attempts in the process of meaning construction fail. All in all, it can be said that whether prior unfamiliarity with a given sequence affects comprehension depends on several factors including the textual cues, the function and features of the sequence and how each subjects uses available cues to solve comprehension problems due to unknown sequences while reading.

It was also found that readers seem to be selective with regard to sequences the meanings of which need to be guessed. Analyses of the recalls showed that in some cases the readers tried to guess the meanings of sequences while in others it was the researcher who asked them to do so after they had finished reading the passage. Upon closer inspection, it was found that the readers had tried to guess the meanings of the sequences which they thought could have helped them in comprehension. This might be an indication that while reading the readers skip sequences which do not play important roles in understanding. In other words, as long as the flow of meaning is not interrupted, readers do not feel the need to use available cues to guess the meanings of unfamiliar sequences or to look them up in a dictionary. Only in cases where a word or sequence hampers

comprehension do the readers try to solve it using context cues and other knowledge sources. It was also found that in cases where the readers were not familiar with a concept represented by a sequence or the sequence did not have a one-to one Farsi equivalent such as '*by definition*' or '*spin doctor*', fewer attempts were made to infer the meanings of the sequences. It seems to be easier for the readers to infer the meanings of words or expressions which convey familiar concepts and for which they can find equivalents in their L1.

Analysis of the data also indicated that the subjects used linguistic knowledge sources such as knowledge of the component parts of the sequences and co-text more often than non-linguistic cues like knowledge of the topic to seek the meanings of unknown sequences. The frequencies of use of each cue were then calculated. To do so, the number of times each knowledge source was employed by the subjects to guess the meanings of the sequences were counted. Of 28 inferencing attempts, in 12 cases (43%), the subjects used their knowledge of the component parts of the unfamiliar sequences. Co-text or the information at the sentence level was used 8 times (28% of cases). Association was used 4 times (14% of cases) and L1, knowledge of technical words or expressions, discourse level sources and knowledge of topic were employed just once each. This leads to the conclusion that among the knowledge sources available, the subjects of this study used the meanings of components of the unknown sequences more often to infer the meanings of whole sequences and knowledge of technical terms and L1 sources were the least often used sources. This also shows that in inferring the meanings of unfamiliar sequences L2-based knowledge sources were used most frequently by the participants.

It was also noticed that in some cases the subjects employed single sources and in others a combination of knowledge sources were used to come up with a meaning. This shows that in the process of inferencing the readers try to use as

many knowledge sources and cues as possible to arrive at least at a partial meaning of the sequence. What is interesting is that it looks as if in most cases the subjects start inferring from the components of the sequence and move to the immediate context and then the larger context to arrive at the meaning of the sequence. This indicates that in the case of formulaic sequences it is the components parts of the sequences that the readers resort to first as in most cases the parts are familiar to the reader and it seems to be easier to solve a problem using the known to arrive at the unknown.

In conclusion, the results of the analyses of the recalls revealed that being able to recognize a sequence helps the readers get the ideas of the text as in the case of '*careless talk costs lives*'. It was also found that in some cases knowledge of the meanings of the sequences does not play a determining role in understanding the text as many other factors such as the available contextual cues, subjects' strategies to infer the meanings of the sequences and the nature of the sequences are involved.

4.3.3 Summary of the Qualitative Analyses (The Follow-up Study)

The qualitative analysis of the follow-up study showed that for the subjects who participated in this part of the study, knowledge of specific formulaic sequences which appeared in the text which they had to read and recall did not have much effect on their general understanding of the text. However, to understand specific ideas, knowledge of some sequences which had major functions in expressing those ideas seemed to be important but did not have a determining role in some cases. The results also showed that the subjects used several cues including knowledge of the component parts of the sequences and context to extract the meanings of unfamiliar sequences.

In the next chapter, **Discussion and Conclusion**, the results of both quantitative and qualitative analyses will be discussed in detail.

Chapter V
Discussion and Conclusions

5.1 Introduction

This study was designed to examine the contribution of formulaic sequences to EFL reading comprehension. In particular, the study sought to determine if recognition of formulaic sequences and understanding of these sequences have any relationships with Iranian EFL learners' reading comprehension. A follow-up study was designed to further examine the relationship between understanding formulaic sequences and reading comprehension. In addition, it tried to find out what kinds of cues the readers use to infer the meanings of unknown sequences.

In the first section of this chapter, the results of both the studies presented in the previous chapter will be interpreted in the context of the research questions. The second part is devoted to a discussion of the limitations of the study followed by the implications of the findings and how the findings would help improve EFL/ESL teaching syllabi and materials. Then, some suggestions for further research will be given followed by a general conclusion.

5.2 Discussion: The Main Study

5.2.1 Main Hypotheses

5.2.1.1 Hypothesis 1:

The first goal of this study was to examine the relationship between recognition of formulaic sequences and reading comprehension in Iranian EFL readers. The basic premise was that if the participants could recognise the formulaic sequences which appeared in the texts they would have a better performance on

the reading comprehension test. This was based on the findings of previous research on the relationship between word recognition and reading comprehension on one hand and the role of formulaic sequences on language production on the other. The results of the correlational analysis revealed a moderate but significant connection between the two variables. This means that the students' scores on reading comprehension rise as their scores on the recognition test increase or vice versa.

These results are in agreement with previous studies (Rudell & Hua 1997; Swanson & Berninger, 1995; Perfetti & Hogaboam 1975), although it is difficult to directly compare the findings because the previous studies mostly investigated the role of single word recognition in reading comprehension and this study is the first attempt to investigate the relationship between recognition of formulaic sequences and reading comprehension. In addition, every study differs from the others in the tasks or tests, the design, and the analyses used.

Several explanations can be put forth for the findings of this study. First, they are consistent with the claims some scholars have made about the contribution of lower level skills such as word recognition to reading comprehension (Koda 1996, 2005; Hulstijn 2001; Qian 2002). It is established that "deficiency in word recognition is directly linked to poor comprehension performance" (Koda 1996, p. 451). In this vein, Grabe & Stoller (2002) maintain that

> *the most fundamental requirement for fluent reading comprehension is rapid and automatic word recognition (or* ***lexical access****- the calling up of the meaning of a word as it is recognized. Fluent L1 readers can recognize almost all of the words they encounter(p. 20).*

It is argued that while it is necessary for the reader to be familiar with the topic of the text or grasp the context, this would not be possible unless the reader is to some extent able to recognise the words constructing the text. In fact, it is the interaction of higher level and lower level constructs that makes understanding possible and it is the lower order processes that pave the way for the higher order ones. If word recognition reduces processing effort and time, it follows that the recognition of formulaic sequences would reduce these processes to a greater extent, as the reader does not need to process the words one by one to come up with the meaning of the whole. Research has shown that processing of formulaic sequences takes less time than non-formulaic phrases suggesting that the former are holistically retrieved from memory without the need to be syntactically analyzed or parsed. (Jiang & Nekrasova 2007, Conklin & Schmitt 2008, Underwood et al. 2004). Parsing time is therefore saved giving the language processer, in the case of this study, the reader, enough time to attend to other higher –order sub-skills of the reading process such as comprehension.

Another explanation that seems to be useful in the interpretation of the results has to do with the results of the studies on the eye-movement and formulaic sequences. It is argued that readers fixate longer on uncommon or unfamiliar words and that they skip highly predictable ones. Research (Underwood et al. 2004) has shown that readers show fewer, shorter fixations on the terminal words in a sequence which is formulaic. Underwood et al. (2004) maintain that

> *mastering the recognition of formulaic sequences in written texts is an incremental process, and early partial mastery is rewarded mainly by not needing to fixate on the vocabulary in a text as much, but it is only with fuller mastery that the requirement for a full duration fixation lessens (p. 163).*

Less and shorter eye fixation is considered as evidence highlighting the predictability of formulaic sequences and how it helps readers. The more predictable the words in a sequence, the less often the readers have to fixate and the shorter the duration of each fixation is. This helps decrease processing time and effort as the readers do not need to fixate their eyes on all the words of the sequence one by one.

In other words, when the reader is familiar with the phrases which are formulaic, he or she can process larger meaningful units of language at each eye fixation. This issue is best addressed by Eskey and Grabe (1998):

> *Good readers, by definition, read fast, which means that with each fixation of their eyes they very quickly process good-sized chunks of text and thereby good-sized chunks of meaningful discourse in building up a meaning for the text as a whole . By contrast, many readers of a second language... try to read word by word, a strategy that effectively destroys their chances of comprehending very much of the text. Taking in, at each fixation, so little information, and fragmenting sense units, such readers place an intolerable strain on their memory systems, so that by the time they have made their painful way to the bottom of the page they have long since forgotten what the top was about (p. 233-4).*

Dolch (1949) has also considered the ability to fixate the eyes on meaningful groups of words as one of the characteristics of good readers and maintains that "a good reader, it appears, perceives, at each fixation, groups of two or three words which may be called 'phrases'" (p. 341). In his article entitled "*Phrase Perception in Reading*" he gives practical suggestions as how to develop readers' phrase perception ability, and argues that through practice the readers produce a mental set toward the larger units than single words, which in turn aid them in real reading (ibid, p. 343).

The results of the study can also be explained in the context of working memory theories. It is argued that in complex activities like reading which involve several cognitive processes, there is a limited number of mental resources that can be activated simultaneously (Koda, 1996). In essence, the limited capacity of the human mind hampers the efforts both to pay attention to several things at a given point in time and to process incoming information on the basis of knowledge and expectations. It is argued that the "more attention required, the more resources are consumed and the slower the processing" (McLaughlin et al, 1983, pp. 135-158) and a task which is handled routinely proceeds automatically and requires little processing time and energy. As Smith (1971) states

> *Since sentence meaning cannot be determined on a sequential word-by-word basis, it is obvious that information from several printed words has to be held in short-term memory at any time. This load on short-term memory can be reduced by "chunking" information in larger units (p. 78).*

In the case of reading, some readers use much of their attentional capacity for lower level processes and they can hardly switch their attention to comprehension of the text. In this vein, Swanson and Berninger (1995) maintain that the difference between less skilled and skilled readers is in their ability to manage the limited capacity of the working memory. They assert that skilled readers assign more working memory resources to higher-order processes such as text comprehension while less skilled readers assign more resources to low-order processes such as word recognition.

This is in line with Stanovich's (1990) claim that cognitive demands in reading reduce as the skills to recognize words are accessed automatically. Automaticity is believed to be essential for higher-order processing, such as reading, as it

requires quick, accurate and effortless performance of several sub-skills. Automatic performance of these sub-skills makes the effective performance of higher-order skills such as comprehension and metacognitive functions possible (Samuels & Flor 1997). It would therefore be beneficial to the readers if they could retrieve the routinized sequences from their mental lexicon without much effort to be able to deal with the higher order processes at the same time. This view has received support from a number of psychologists such as Anderson (1983) who developed the ACT-R theory and Logan (1988) who introduced the Instance Theory. Both of these theories view "automatic processing as effortless and fast, reflected in a decrease of latencies of the task at hand" (Fukkink et al. 2005, pp. 54-75).

Nattinger and DeCarrico (1992) have also supported the idea of the working memory and its effect on language processing, and claim that in the context of language learning, the number of units that short-term memory holds is constant, and that the meaning-bearing units of language tend to be phrasal rather than single words. They, then, continue that

> *this prefabricated speech units has both the advantage of more efficient retrieval and of permitting speakers (and hearers) to direct their attention to larger structure of the discourse, rather than keep it focused narrowly on individual words as they are produced (p. 32).*

This can also be looked at from a different perspective. As readers decode words and phrases, they "assemble them into propositions" (Perfetti 1985 , p. 37). As this assembly takes place within a mechanism which is limited in capacity (memory), successful recognition of phrases leads to successful assembly of propositions and their integration into representations that can survive in long-

term memory. The limited-capacity memory has therefore enough resources to deal with the ongoing process of meaning construction (ibid).

One possible explanation on the findings relevant to the first hypothesis could be provided by the notion of "grammar of expectancy" (Oller and Richards 1973; Oller and Streiff 1975) which suggests that it is easier to process elements which are predictable. In essence, it is based on that fact that "cognitive processes are dependent to a great extent on the human capacity to anticipate elements in the sequence of experience" (McErlain 1999). Nattinger (1980) also refers to Oller's expectancy grammar and maintains that prediction of a sequence of words which allows less variation is easier and this explains to a great extent how we process language. He continues that

> *the degree to which words constrain those around them, and the assurance we have that certain words are going to follow certain others, are the facts we use to make sense of language and to create all sorts of subtle variations and surprises (ibid, p. 339).*

This predictability allows the sequences to be easily and quickly retrieved from memory which in turn frees the working memory resources for higher-order processes such as text comprehension. The predictive nature of formulaic sequences is also acknowledged in the literature (Wray 2002, Frisson et al. 2005). Frisson et al. (2005), for instance, assert that "a word is predictable when the likelihood that a word will follow a given context fragment is high" (p. 682).

It is argued that by knowing a formulaic sequence, the reader makes predictions about the whole sequence using the information from the context and the first or the key element of the sequence. So, it is sometimes the context that helps the reader predict the sequence already known or a constituent part of the sequence

or both. Whatever the situation, the outcome is reduced processing effort and time.

One of the functions of formulaic sequences is manifested in their use as discourse markers. The speaker or the writer uses these sequences to "signal the relationship between a piece of text and what precedes or follows, and the speaker's attitude towards the surrounding text" (Wray 2002a, p. 87). That is, these sequences do provide additional information which would definitely ease the readers' job to identify the attitude of the writer (whether he/she is for or against the stated ideas) and his/her intended meaning, and to keep track of the mental map of the text (*in the first place, in the second place*,..). The result is again, less struggling and more time and free cognitive resources to deal with other aspects of the text and reading process.

It is important to note that while word recognition and recognition of formulaic sequences both involve identification of familiar items and their retrieval from memory, recognition of formulaic sequences (in this study) is different from word recognition in that word recognition involves identification of a word and its meaning but by recognition of formulaic sequences is meant the familiarity of the subjects with the target formulaic sequences. Therefore, it is not just the pronunciation, orthography or meaning of a sequence which is the main concern but also awareness of the fact that some words are grouped together and appear to be formulaic. Hence, a sequence may be recognized as formulaic but the meaning of the sequence may not be known or remembered as is the case with some idioms. In addition, words are easier to identify in a text, as they are singled out from other words by means of a white space at the beginning and at the end of a sequence of letters, but formulaic sequences have no identifiable boundaries when they appear in a text. In the case of formulaic sequences, when the readers recognise a sequence, they may retrieve its meaning as well,

that is they may know the meaning. Or, they may use the context to guess its meaning. In either case, any successful recognition of the phrase helps the readers as they do not use their cognitive resources and their time to process each word separately to come up with the meaning of the sequence. In other words, when the sequence is recognized, two different situations may arise. Either the sequence is transparent (e.g., to conceal a truth) and its meaning is the sum of its parts or it is opaque (e.g., to cut the mustard) and its meaning is not derived from its components. In the case of transparent sequences, recognition and retrieval of meaning are both likely to occur. In the case of opaque sequences, the reader may or may not retrieve their meaning. When an opaque formulaic sequence is recognised as a whole, it is less demanding cognitively and requires less time to process its meaning (e.g., using context) as the readers process the sequence as a single lexical item rather than attending to each word separately and as Moon (1998) puts it, "accessing a list [of readymade items from memory] is likely to be faster than generating a word-by-word interpretation" (p. 33).

In sum, the findings of this study shows that recognition of formulaic sequences while reading reduces strains on the working memory and spares time and energy. This, in turn, helps readers to take shortcuts in decoding the messages of the writers.

5.2.1.2 Hypothesis 2:

The second goal of this study was to investigate whether understanding formulaic sequences which appeared when reading texts would affect the readers' performance. The basic assumption underlying this hypothesis was that if knowledge of vocabulary meaning affects reading comprehension, it would follow that understanding opaque formulaic sequences might help the readers in

constructing text meaning. To test the hypothesis, a correlational analysis was performed to assess the degree of relationship between the two variables. The results of the analysis supported the null hypothesis suggesting that understanding formulaic sequences does not correlate highly with the scores on the reading comprehension test.

This certainly conflicts with the common-sense notion that understanding formulaic sequences and reading comprehension should be highly correlated. It seems reasonable to say that successful apprehension of a text requires understanding words and phrases which appear in the text and as opaque formulaic sequences bear meanings which are not sometimes the same as the sum of their constituents parts, knowing the meaning of these sequences would benefit EFL/ESL readers. In fact, literature on the relationship between word meaning and reading comprehension abounds with conflicting views and findings. In this regard, Beck et al (1982) maintain that

> *unknown words can create gaps in the meaning of a text; if too many gaps occur, the student may not be able to construct meaning. Words that are familiar but accessed with difficulty allow comprehension in principle; however, the difficulty of accessing their meaning may interfere with the processing of the text, since attention must be diverted from constructing meaning of the discourse to searching for the words' meanings. Comprehension is impeded by not knowing enough of the words (p. 507).*

Some studies report strong correlations between word meaning and reading comprehension (Laufer 1991, Koda 1989) while others report a lack of correlation between the two (Stahl & Fairbanks 1986). Studies which investigated the effect of vocabulary instruction on reading comprehension have also reported conflicting results. Some (Draper & Moeller 1971; Kameenui &

Carnine 1982; Stahl 1983) suggest that teaching vocabulary improves reading comprehension. Others (Tuinman & Brady 1974) report no improvement in comprehension.

There are several explanations which could be used to justify the lack of relationship between the two variables in this study. One reason speculated for the lack of relationship or low relationship between understanding formulaic sequences and reading comprehension in this study has to do with the fact that the subjects might have used their inferencing skills to guess the meaning of unknown sequences while reading. As the test of formulaic sequences was in multiple-choice format and no context was provided for target formulaic sequences, the subjects could not use inferencing skills to guess the meanings of sequences when they took the test. In contrast, in the actual process of reading the texts of the reading test they were provided with the context, either at the sentence level or at the discourse level, which could well have affected their performance on the reading test. There is plenty of research (Fraser 1997, 1999; Paribakht & Bengeleil 2004; Paribakht 2005) reporting on the role of inferencing as one of the strategies used by second language readers to overcome the problems arising from unknown words. Cooper (1999), for example, asked students to report the strategies they employed during the course of reading and concluded that "use of context was the major strategy employed by the participants to arrive at the meaning of the expressions" (p. 258). Some scholars have also noted the contribution of context in providing additional information about 'the finer aspects of meaning' (Nation 2001, p. 242 cited in Steinel et al. 2007, p. 481) which would result in better comprehension.

This is in line with what Sinclair (2004) refers to as 'semantic reversal'. He claims that it is not always the parameters of a lexical entry, either a word or a

phrase, that determine its precise meaning but its verbal environment. In other words, "the flow of meaning is not from the item to the text but from the text to the item" (ibid, p. 135). This reflects the important contribution of the textual environment in supplying the precise meanings of given sequences. Sinclair summarizes it through the following example:

> *If you are reading a book that is outside your normal area of expertise, and come across items that the author assumes you know the meanings of – like abbreviations – you can either break off in your reading and consult a specialized glossary, or plough on with whatever understanding you can glean from reversal techniques, After a while either your interpretation will break down altogether or you will survive the particular passage because none of the unresolved meanings are critical to your overall understanding (p. 136).*

In other words, the relationship between formulaic sequences and the context in which they are used is reciprocal. That is, it is sometimes the formulaic sequence that provides additional information (prosody, pragmatic, semantic) to the reader towards a better understanding of the writer's intention and in some cases, it is the context that helps the reader guess the meaning of the sequence. With regard to the role of context, drawing upon research by Bobrow and Bell (1973), Ortony et al. (1978), and Gibbs (1980) Moon (1998) states that "appropriateness of context, 'biasing' contexts, or contextual clues help to resolve ambiguity thus speeding up the processing of ambiguous FEIs [fixed expressions and idioms]" (p. 32).

An alternative explanation is that these results might be due to the nature of formulaic sequences selected for this study. Some target opaque formulaic sequences used in this study do not have a significant role in conveying meaning and are used just to perform an emphatic function or to help the writer to

develop his ideas further. That is, these sequences add to the meaning already established or are used to illustrate a point presented earlier. In addition, in some cases additional information is given in the preceding and following sentences that will help the reader to guess the meaning of the sequence. '*A case in point*' is a case in point. The preceding paragraph in the text and the following sentence in which this sequence is used help the reader guess the meaning and the fact is that even if the reader does not guess, it would not hamper the comprehension as it does not have a main meaning bearing function in the text. The same is true for the phrase *to barge into a place*. This sequence is used in the following context: *Occupation makes it sound as if the US has barged uninvited into somebody else's country....* Here, the words *occupation, uninvited* and *somebody else's* are the key elements in guessing the meaning of the sequence.

Another possible explanation for the weak relationship between the variables (understanding FSs and reading comprehension) is the order of administration of the tests. On the first day of the test administration, the students answered the understanding test after they took the recognition test and TOEFL. It seems possible that the subjects might have got tired by the time they took the test and might not have answered the questions with care. In addition, as the test was in multiple-choice format, it is possible that the subjects' true scores could be affected by their attempts to guess the answers rather than trying to answer with full attention or on the basis of their knowledge.

To sum up, several factors might have been responsible for the lack of relationship between understanding FSs and reading comprehension. It might be due to a real lack of connection between the two variables or due to testing factors including the difficulty level of the understanding test or its format. It

might especially be because of the contributory role of context in guessing the meanings of unknown sequences.

5.2.1.3 Hypothesis 3:

The third hypothesis of this study investigated the possible interaction between recognition of formulaic sequences and understanding these sequences. The basic assumption underlying this hypothesis was that the learners who gain higher scores on both recognition and understanding tests outperform those who do not. The results of the two-way analysis of variance (ANOVA) showed no interaction effect between the independent variables (recognition x understanding). That is, those who gained higher scores on the recognition test performed similarly no matter whether they got better scores on the understanding test or not. This suggests a significant main effect for recognition. On the interpretation of interaction results Hatch and Farhady (1982) state that "if the interaction(s) are not significant, then much stronger claims can be made about the effect of the independent and moderator variables" (p. 160).

Obviously, it is tricky to give a clear explanation of these findings mainly because reading comprehension is a very complex task involving the simultaneous activation of different subcomponents. In addition, as Koda (2005) states

> *vocabulary processing during reading necessitates interrelated processing skills, including constructing a context, accessing stored information through visual word displays, selecting a relevant meaning based on contextual information, and evaluating the appropriateness of the chosen meaning in subsequent sentences (p. 48).*

Readers' background knowledge of the text topic and the strategies they use to overcome problems (e.g., unknown words or phrases) while reading are among other factors that influence readers' performance. Furthermore, the testing procedure, the tests, and many other factors would be responsible for the findings of this study. It may simply be that the recognition test was easier than the understanding one. One could argue that the superiority of those who got higher scores on the recognition test in doing the reading test would be less marked if a different test format was used to assess the subjects' both recognition of and understanding formulaic sequences.

5.2.2 Secondary Hypotheses:

Although the secondary hypotheses were not central to the study, they provided some valuable information regarding the relationships between the frequency of formulaic sequences and recognition and understanding these sequences, the proficiency level of L2 readers and their ability to recognize and understand these sequences and the difference between males and females in recognition and understanding these sequences. In this section the results of the analyses to investigate the hypotheses dealing with these issues are presented.

5.2.2.1 Secondary Hypotheses 1 & 2:

The first and the second hypotheses which are addressed in this section are concerned with the frequency of occurrence of formulaic sequences and the learners' ability to recognize and understand them. The assumption underlying these hypotheses is that if there is a significant relationship between word frequency and word knowledge as shown by previous research (Monsell et al.

1989; Morton 1969), it would follow that frequent word combinations could be recognized and understood by more subjects. In other words, it is based on the idea that "high frequency (i.e. common) words require less perceptual information to raise their activation to threshold, hence are recognised more quickly than low frequency words" (Milton 1994). The results of the correlational analysis suggest that unlike vocabulary, there is no connection between frequency of these sequences and the subjects' performance on the recognition and understanding tests.

The results of this study are in line with what Bishop (2004a) reported. In his thesis, he studied the relationships between frequency of occurrence of formulaic sequences and the number of participants knowing them and found no significant relationship between the two variables. In a similar vein, but on single words, Macleod and Kampe (1996) found that low-frequency words were recognized better than higher-frequency ones. In their literature review, they cite the results of studies by Balota & Neely (1980); Gorman (1961) and Shepard (1967) which also showed the same results. There are, however, some conflicting views and research results reported in the literature. A number of arguments are in favour of the relationship between frequency and word recognition (Balota & Chumbly 1984; Forster & Chambers 1973; Kirsner 1994) and many research studies have also reported significant relationships between the two variables (Kichun et al. 2004, Monsell et al. 1989). Ellis (2002a) asserts that "the recognition and production of words is a function of their frequency of occurrence in the language" (p. 152).

In the case of the results of this study, one possible explanation for the lack of relationship can be that higher frequency of occurrence in the BNC (British National Corpus) does not equal higher frequency in the materials that EFL learners study or the number of times they experience particular sequences.

They may encounter some formulaic sequences more than others not because they occur in the BNC more frequently but because they occur more in their textbooks or classroom interactions. A sequence that occurs in the BNC less than 5 times and is considered as having a low frequency of occurrence may be one of the sequences that the learners have to learn, among other sequences and words, to pass, for instance, a reading course. It should also be noted that the BNC is not 100% perfect as a source providing information about the frequencies of formulaic sequences. This is because the BNC contains only 90 million words of written text; this is really quite little, e.g., in comparison with the text available on the Internet. In addition, it is sampled largely according to what was available without copyright problems, not truly representative, and not complete texts. Moreover, the frequency distribution of words and formulaic sequences (with a short head of high-frequency items and an enormous long tail of very low-frequency items) is such that except for strings occurring say 50 to 100 times or more, it is really difficult to be confident about the likely true frequency in a much larger representative corpus. With regard to the connection between frequency and familiarity, Abel (2003) maintains "the relationship between frequency and familiarity is a unidirectional one: a frequently occurring idiom will be judged as familiar, but a familiar idiom is not necessarily a frequent one" (p. 346). It seems therefore important to state that what should be considered as an important factor affecting the subjects' knowledge of these sequences is the frequency of exposure to these sequences rather than frequency of occurrence in the BNC or any other corpus.

In an article on frequency effects on language processing, Ellis (2002b) introduces "seven major factors that modulate the effect of frequency" (p.307), among which failing to notice cues, pedagogical input and practicing seem to be important in interpreting the results of this study. Ellis maintains that second language learners fail to acquire a particular pattern or feature of high frequency

due to lack of noticing and suggests that instruction should be in a way that these features become salient. He also asserts that the classroom environment "can distort the patterns of exposure" (ibid, p. 312) and that for a pattern to be internalized, there should be enough rehearsal from the language learner. In another article he expands on this idea and maintains that

> *frequency is a necessary component of theories of language acquisition and processing. In some guises it is a very rudimentary causal variable. Learners analyse the language input that they are exposed to; practice makes perfect (Ellis 2002a, p.178).*

To sum up, the ability to understand formulaic sequences or recognize them may not correlate with their frequency of occurrence in the different corpora, because it is the number of times a sequence is exposed to the students, practicing and noticing them that improves the chances of them being known by learners.

5.2.2.2 Secondary Hypothesis 3 & 4:

The third and the fourth hypotheses investigated the relationship between recognition and understanding of formulaic sequences and overall English proficiency. The results of the correlational analyses showed low but significant positive relationships between the variables. It could therefore be concluded that the higher one's knowledge of formulaic sequences (recognition and understanding), the higher his/her overall proficiency or vice versa.

These results are consistent with what can be found both theoretically and empirically in the literature. It is acknowledged that knowledge of formulaic sequences helps EFL/ESL learners come across as proficient speakers (Pawley and Syder 1983, Wray 2002a, Schmitt 2004). The empirical studies conducted

in this regard have also confirmed these claims (Schmitt et al. 2004, Boers et al. 2006). In a study on the role of formulaic sequences on oral proficiency, Boers et al. found that students who received instruction of formulaic sequences outperformed the control group significantly. They not only were more fluent, they were also more accurate. Similar results were found by Schmitt et al. (2004). Their study showed that relatively proficient English for academic purposes learners knew a considerable number of target formulaic sequences. Gyllstad (2007) has also found the same results with collocations. In his doctoral thesis, he investigated the difference between higher and lower proficiency subjects in their receptive knowledge of collocations and found that the higher proficient language learners had a better command of English collocations. In the same vein, Alzahrani (1998), studied the relationship between collocational knowledge and overall proficiency of 81 advanced ESL learners whose native language was Arabic. He also found a strong relationship between the two. In another study, Keshavarz and Salimi (2007) found the same results regarding collocational knowledge and overall proficiency with Iranian EFL students.

In interpreting these findings, the possible explanation that comes to mind for these results is that vocabulary in general and formulaic sequences in particular as one of the aspects of vocabulary are among the most vital elements of language that affect overall proficiency (Laufer and Nation, 1999) as they are units of language that carry meaning and in the case of formulaic sequences, they have additional merits of reducing the pressure on the language learner as they are stored holistically in the memory and are therefore retrieved as such requiring less time and energy.

It should be added that the results of the hypothesis addressing recognition of formulaic sequences and overall English proficiency can be linked to the results found for the first main hypothesis of the study which dealt with the relationship

between recognition of formulaic sequences and reading comprehension. On the one hand, it is found that recognition of formulaic sequences and reading comprehension have a significant connection. On the other hand, it is found that knowledge of formulaic sequences and overall language proficiency are related. Research (Clapham 1996, Keshavarz et al. 2007) has also shown a close relationship between reading comprehension and EFL proficiency. It can therefore be concluded that formulaic sequences make reading comprehension which is one aspect of proficiency less demanding (Nattinger and DeCarrico 1992, Schmitt 2004, Wray 2002).

5.2.2.3 Secondary Hypotheses 5 & 6

The fifth and the sixth hypotheses investigated the role of the subjects' sex and their performance on the recognition and understanding test with this basic premise that research in language learning has indicated differences in gains between males and females. The results of the independent sample t-tests showed no statistically significant differences between males and females in recognizing and understanding formulaic sequences.

Despite the large body of research studying the role of sex in language acquisition and learning different aspects or components of language such as grammar and vocabulary, its contribution to the knowledge of formulaic sequences has not been investigated as much. It is therefore not possible to compare the results of this study with any previous ones. However, it can be said that in general, these results are not consistent with what is reported in the literature in terms of the differences between males and females in language achievement (Kaylani 1996) and application of language learning strategies (Oxford 1990). In addition, research has shown that males and females differ in the range and variation of strategies used to learn vocabulary (Catálan 2003),

however, it seems the outcome is not very different as this study shows. In an study on the knowledge and acquisition of formulaic sequences, Schmitt et al. (2004) investigated the role of individual differences such as aptitude, age and gender and found no significant correlations between the individual difference variables and acquisition of formulaic sequences. They attribute the lack of correlation to "other factors related to the learning context" (p. 69). It should be noted that the reported differences between males and females in the literature would indeed be the effects of other differences that the learners bring to the learning situations, e.g., their learning styles and the differences in learning situations or the methodologies.

One possible explanation for the lack of difference in recognition and understanding of formulaic sequences by males and females in this study can be the fact that they have been exposed to more or less the same materials taught by the same teachers and through the same methodologies.

One issue that should not be ignored in studying the role of learner variables in EFL learning is that these factors might be affected by other external factors such as learning and testing situations. As Schmitt et al. (2004) maintain, investigating such variables as sex and other individual variables and their effect on the acquisition or knowledge of formulaic sequences requires longer and more focused studies.

5.3 Discussion: The Follow-up Study

The qualitative analysis was performed to gain more in-depth information on the relationship between understanding formulaic sequences and reading comprehension. In other words, it sought to decide whether or not similar results as the main study could be obtained on the relationship between formulaic

sequences and reading comprehension using a different methodology. It also attempted to discover what cues readers employ to overcome their problems due to unknown expressions while reading.

The results of the analyses of six subjects' recalls and interviews were to some extent in line with what was found in the main study, that is a low relationship between understanding formulaic sequences and reading comprehension. In other words, not knowing the meaning of sequences appearing in a text does not seem to affect readers' apprehension. However, this depends to some extent on the nature and function of the sequences and their surrounding context. In essence, the results of the qualitative analysis confirm the interpretations given in section V.2.1.2. That is, not all sequences play a major role in understanding a text or parts of it and misinterpretation of them does not affect the overall understanding of a message. In some cases, the unknown sequence which happens to have a key meaning-bearing role may occur in a context which eases the process of extracting its meaning by the reader. This is the case with some of the sequences that appeared in the text used in the qualitative study. For example, the expression '*careless talk costs lives*' was one of the key sequences which almost all the subjects mentioned in their recalls, either in Farsi or in English. It seems that this expression has given the subjects a general impression of the text and even one of the subjects attributed his successful comprehension of the text to this expression and said that it helped him build up a picture of the text which he could cling to throughout the text. Scott and Tribble (2006) refer to this property of words in context as keyness which is "a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail" (p. 55-6).

The notion of 'keyness' applies to the components of a sequence as well. That is, in some sequences some words function as key words which help the reader extract the meaning of the sequence. For example, most subjects tried to guess the meaning of '*growth industry*' by summing the meanings of its components parts. It is therefore not possible to generalize the findings to all sequences and all types of contexts.

It was also found that the readers could get the gist of the text whether or not they knew the meanings of sequences. That is, unfamiliar formulaic sequences would not cause much difficulty in understanding the main ideas of the passage. Even when it comes to understanding a sentence containing an unknown sequence and conveying specific ideas readers are able to get a rather vague impression of it as the sequence is not isolated and its surrounding verbal environment helps them in the process of extracting the meaning of the sentence. However, in some cases the readers are not able to infer the meaning of a given sequence using the immediate context and use other knowledge sources available.

What should be taken into account in the interpretation of the results is the number of opaque and semi-opaque formulaic sequences which were used in the text. The text was composed of around 1000 words 31 of which were opaque or semi-opaque sequences. That is just 2.5% (rough statistics) of the text was composed of these sequences. It can be said that in a text of this size this proportion of opaque sequences would not affect understanding to a great extent if at all. It could therefore be concluded that with regard to this text understanding of formulaic sequences would not be a major factor influencing understanding the whole text except for a few sequences which had key functions in conveying the message of the writer.

Another issue has to do with extracting meanings of unknown sequences using available clues. A number of linguistic and non-linguistic knowledge sources and cues were employed by the subjects of this study to guess the meanings of unfamiliar formulaic sequences. Among the cues, using the meanings of the components of a sequence was most frequently used. This contradicts the results obtained by Bengeleil and Paribakht (2004). In a retrospective study to examine the effect of EFL learners' reading proficiency on their lexical inferencing while reading, they found that sentence-level meaning was the major source of inferencing for their subjects. In the case of formulaic sequences, however, the subjects mostly resorted to what was familiar, that is, the meanings of the component parts of the sequences, to arrive at the meaning of the whole. This is in line with what scholars who work on the processing of idioms have found (Titone and Connine 1999). Titone and Connine, for example, conclude that

> *given that word meanings are always activated during idiom processing ... , component words of idiomatic sequences may contribute substantially to the construction of idiomatic meanings (as they would for less compositional or transparent idiomatic phrases) (p. 1671).*

It should be noted that although on the surface it looks as if the subjects use their knowledge of the words comprising the sequences to arrive at their meanings, their final decision might to a great extent be affected by their overall knowledge of the topic and the context either at the sentence level or at the discourse level. This is because the actual meaning of a word or sequence is determined by its surrounding context. In addition, the second most often used cue was the co-text or the immediate context of the unknown sequence. This in fact verifies the suggestion that the meaning of a word or sequence should be considered in the context in which it occurs. In this regard, Nation (2001)

maintains "it seems that most of the clues for guessing word meanings from context will come in the immediate context, that is within the same sentence as the unknown word" (p. 245-6). It is also established that reading is a complex activity involving several cognitive resources at a given point in time. It is therefore rational to say that while in the process of inferencing the meanings of formulaic sequences readers try to extract the meanings of sequences using their knowledge of the component parts of the sequences most often, their background knowledge and other higher level cues are also involved although the reader may not be aware of it, as in reading both higher level and lower level skills are activated simultaneously.

Several other factors seem to affect readers guessing such as keyness of the sequence or familiarity of the concept. It is evident that readers try to guess the meanings of sequences which are not key to understanding less often than those which are critical. In addition, it is believed that "if the concept is already known, guessing is easier" (Nation 2001, p. 245, citing Nagy, Anderson and Herman 1987). Therefore in cases where the concept is strange or there is no one to one equivalent for the sequence representing it in readers' L1, guessing gets difficult. Another factor affecting readers' inferences has to do with their overall proficiency level. Bengeleil and Paribakht (2004) found that their advanced L2 readers made "more correct and particularly correct inferences than the intermediate-level readers" (p. 240). A possible reasoning for this conclusion may be that advanced readers know more words and can therefore comprehend a text better. This, in turn, can lead to more resources being available both at text level and at word level. Still another reason for different inferences with respect to a given sequence can also be attributed to the different learning styles that L2 learners adopt in learning the language. Some are risk takers and use whatever clues available to make guesses about the meanings of unknown words or sequences while others do not take risks and

prefer to reflect on their inferences or are simply not good at making hunches or are more tolerant of ambiguity.

It should also be mentioned that not all guesses were successful. In some cases unsuccessful guesses led to misinterpretation. This depended in fact on the nature of the sequence and its function and its surrounding context. In cases where the sequences did not have a crucial role in the text, wrong guesses did not affect understanding of the parts in which they occur. For example, one subject made an unsuccessful attempt to guess the meaning of '*the fifth columnist*'. However, this did not affect her general understanding of the paragraph in which it occurred.

To some up, the results of the follow-up study showed that recognition of some key formulaic sequences helps reader to get the gist of the text to a great extent. It also showed that whether or not readers' understanding of opaque sequences affects their comprehension of a text depends to a large extent on the nature of the sequences and their function in a text, their verbal environment, and some external factors including readers' reading habits and strategies they employ to overcome problems arising while reading.

5.4 Limitations of the Study

It is important to mention some of the limitations of this study before going on to the pedagogical implications that the results suggest for EFL/ESL teaching. The first three limitations have to do with the subjects of the study and the remaining two concern the tests.

First, this study was limited to language learners at upper-intermediate level of English proficiency. If subjects at different levels of proficiency in English had been employed, different results might have been achieved. Second,

unfortunately none of the students who participated in the main study turned up for the interview as they had to attend their regular classes and the researcher was not given extra time by the cooperating university to arrange an interview. Therefore, the subjects who participated in the follow-up study were from among postgraduate students who were studying in the University of Liverpool in the academic year 2007-2008. It would have yielded more consistent results if the same students as in the main study had taken part in the follow-up study. In addition, it would lead to more generalisable and consistent results had a larger number of subjects participated in the follow-up study. Third, the subjects who participated in the main study were limited to the students at one of the branches of Azad University which is run by the private sector. The interpretations of the results must therefore be done with care as it is not easy to generalise the results beyond these subjects. Fourth, the tests which were used to assess recognition of formulaic sequences and understanding these sequences were not validated as there were no existing tests assessing the same constructs against which to validate the new tests and finally, in the main study only a single measure of reading comprehension-namely, open ended questions was used. If different measures of reading comprehension such as multiple-choice questions or recall protocols had been used, the study might have yielded different results.

5.5 Pedagogical Implications

The present study is an initial step to take the body of knowledge about the relationship between knowledge of formulaic sequences and language comprehension in general and reading comprehension in particular a step ahead. Several implications for second language learners and teachers and materials developers arise from the findings of this study. The first part of this section deals with the implications for ESL teaching and training and the second part is concerned with the implications for materials development.

5.5.1 Implications for Teaching

One of the important results of this study is the fact that a close connection between recognition of formulaic sequences and reading comprehension exists. The implication for language teaching is that it is incumbent on teachers to help ESL readers to enhance their reading ability by encouraging them to read by phrases rather than by words. One of the advantages of this is that it helps the learner to overcome the notion that they have to understand every word in a text to be able to grasp the meaning of the whole text. In this vein Eskey & Grabe (1998) maintain that "the teacher must also induce students to abandon the word-by-word approach to reading by introducing exercises that, for similar reasons, force students to read in meaningful 'chunks'" (ibid, p.229). They then continue to assert that in a language like English in which "sequences of words frequently function as single lexical items (*in spite of, put up with*)" (ibid, p. 233), this has a crucial role in helping L2 learners read better. Providing students with exercises that improve their ability to automatically process some reading subcomponents, such as word identification and phrase recognition is advantageous as well, as they provide opportunities for the students to practice maintaining a balance between the working memory capacity and their time on the one hand and the demanding subcomponents of reading activity on the other.

This requires the teachers to raise the students' awareness of the nature and importance of formulaic sequences and help them to recognize the sequences and collocations for themselves while reading to develop a sense of understanding and awareness of what forms the meaning units. This includes not only instruction on what lexical phrases and collocation are but also designing tasks and exercises to provide ample opportunities for students to practice recognizing, understanding and using these sequences. In this regard, Arnaud & Sauvignon (1997) assert that "learners should know a complex unit

when they see or hear one, and this in turn implies that we teach them some lexicological facts, such as the nature and categories of complex units, and their behavior on discourse" (p. 168) In a similar vein, Lewis (2000) recommends teachers to help raise learners' awareness of chunks and "develop an understanding of the kinds of chunks found in the texts they meet" (p. 161) as this "increases the chance of them noticing useful language, rather than many other features that are irrelevant from an acquisition point of view" (p. 162).

This awareness-raising in formulaic sequences is especially important in the context of English as a foreign language as in these situations, on the one hand, learners are mostly text-oriented and on the other, it is difficult for them to recognise formulaic sequences in text. This is because "unlike speech, text does not contain conspicuous word boundaries in a hefty portion of the input the learners encounter" (Wible 2008, p. 166). In this regard, Wible (2008) states that

> *basically, for them [EFL learners] the task is not one of unbundling what they first considered to be unanalyzed wholes without parts and discovering the word boundaries and the words inside, but the reverse: discovering that some sequences of the discrete units occurring between white spaces in text are in some respects best considered as bundled wholes despite the lack of typographical evidence that this is so (p. 167).*

The results of the study would be of particular significance in teaching English in Iran. As most teachers teaching EFL are graduates from domestic universities and as most of them have not even had a chance to visit an English speaking country, their English is not native-like, and they are not familiar with the majority of formulaic sequences that are used by native speakers in everyday conversation. Their English is therefore simple, lacking these sequences, and the students usually do not experience these sequences in an interactive

situation. This requires teachers to improve their knowledge of these sequences, for example through reading original texts to be able to introduce them to their students.

In addition, it would be both interesting for students and informative if teachers introduce problematic formulaic sequences such as, '*by definition*', which do not have a one to one equivalent in Farsi, and non-canonical sequences such as '*as it were*', as students often get confused when they come across these sequences in texts. Moreover, as these sequences are usually omitted from textbooks designed for Iranian students, and students encounter them in non-interactive situations outside classroom such as films, it is the responsibility of the teachers to design activities to enhance students' awareness of these sequences so that they can receive feedback on usage.

Another important implication of the study concerns the teaching of vocabulary. It will be highly beneficial if teachers adopt a 'lexical approach' to teaching vocabulary as "much of our language [is] stored in units larger than individual words. It is not certainly enough that learners are 'learning new words'; an adequate lexicon is also about the quality of the lexicon" (Lewis 2002, p. 56). In this regard McCarthy (1990) suggests that "over-concentration on learning single words may hinder the development of the L2 phrasal lexicon" (p. 45). Choosing this approach leads to more helpful recording of new vocabulary and more accurate use of the new items later and reinforces "the fact that we both read and speak in 'chunks' of language, giving flexible variety, and letting the students get the 'feel' of useful groups of words" (Brown 1973, p. 2). It is worth noting that this applies to both explicit and implicit vocabulary learning. It is beneficial for the learners to become aware of the advantages of learning new vocabulary along with the words they accompany.

Teachers should also be trained in techniques for introducing new formulaic sequences and helping students internalize them. Just encouraging the learner to notice or recognize and memorize these sequences would not lead to desirable outcomes because a requirement to learning is exposure and practice. The teachers should be able to devise activities and tasks that present and encourage retrieval, receptive or productive, of these sequences. Lewis (1997) presents a number of activities including summarizing a text based on the sequences, reading aloud to improve student's sense of intact intonation contour and categorizing sequences using structural and semantic criteria.

Finally, based on the results of the qualitative analysis, it seems reasonable to make students aware of the advantages of improving their inferencing skills. In doing so, it would be ideal to set appropriate tasks to train the students in identifying and using knowledge sources and cues available in the surrounding text and their background knowledge to infer the meanings of unfamiliar sequences while reading.

5.5.2 Implications for Materials Development

One possible implication for materials developers and syllabus designers is to incorporate explicit teaching of formulaic sequences in their EFL/ESL syllabi and materials. It would be highly desirable if reading textbooks were designed in a way that introduced the most frequent and useful sequences first and then the less frequent ones. One way to bring these sequences to the students' attention is to highlight them in bold. This increases their salience and the chance of them being noticed easily. In the case of opaque sequences their synonyms or definitions can be presented in the margins. Then exercises can be designed in such a way as to encourage the learners to retrieve what they have noticed in the main text to understand or to produce a piece of discourse.

Suggestions include: recognizing phrases in sentences, fill in the blanks exercises in which a situation is described and the subjects have to fill the blanks with appropriate expression, and listening activities in which they listen to a piece of discourse and identify sequences they hear from among the choices given. "It is in doing exercises like these that second language readers are first exposed to the notion that they can process English in meaningful phrases, that is, in units larger than the word" (Eskey and Grabe 1998, p. 233).

Another useful way of including formulaic sequences in syllabi is the use of concordances as they provide multiple encounters with phrases or lexical items in different contexts and reveal many aspects of these items through authentic data. These aspects include the grammatical structures in which they occur or their typical usage, collocations, semantic prosody, and meaning. In this regard, Jones and Heywood (2004) maintain that

> *the use of concordance lines as a way of studying lexis, as well as other linguistic features, is advantageous in that it requires a deep and thoughtful level of mental processing as students become involved in investigating for themselves the typical patterns of use of the target items (p. 272).*

It should be added that all this has to be done systematically so that the learners become familiar with the importance of these sequences and how they work in association with one another first and then be provided with exercises that help them develop their competence of formulaic sequences so that they gain independence to improve their knowledge of them when on their own (Wei, 1999). As Kennedy (1990) puts it:

> *just as the teacher of botany does not take students into the jungle and expect them to learn about all the plants by simply being exposed to them, so the language curriculum designer and*

> *classroom teacher can facilitate learning by systematic presentation of the role of important language items and their linguistic ecology- the company words keep (p. 227, cited in Wei 1999).*

The target sequences have also to be presented adopting a spiral syllabus in which the same sequences are presented repeatedly so that the students do not forget them. In this regard, Lewis (2002) states that

> *research evidence shows new items need to be recycled if they are to be fully acquired. This may occur naturally through reading and listening ... . What is important is to encourage learners to look back at the language they have recorded and do something – perhaps anything – with it (p. 49).*

Lewis (1993) does not seem to be satisfied with just practicing sequences that appear in the textbooks. He suggests that "a resource book of lexical phrases, including sentence heads and institutionalized utterances, should be an important priority for one of the major publishing houses" (p. 132).

A further implication for materials development relates to lexical inferencing. Exercises specially designed to train inferencing skills can help L2 readers' learn how to extract the meanings of unknown sequences and words by employing available cues, within and outside the text while reading.

5.6 Suggestions for Further Research

This section is concerned with the suggestions which evolve out of the results of this study. This study has provided some insights into the relationship between formulaic sequences and reading comprehension and overall English proficiency. It also considered the role of one of the individual differences factors, namely, the subjects' gender in knowing formulaic sequences. In

addition, the relationship between frequency of occurrence of formulaic sequences and the number of correct answers given to the tests assessing their recognition and understanding was investigated. Future studies would consider different aspects of formulaic sequences as well as different factors affecting their acquisition.

In order to be able to generalize these findings, future research should examine a larger group of subjects. As the syllabus and the materials as well as the teaching strategies used in teaching EFL students varies across institutions, it seems reasonable to broaden the scope of the study to incorporate subjects with different educational backgrounds. Furthermore, this study was performed on students at upper-immediate level of English proficiency. A study employing subjects at different levels of proficiency would yield more generalizable results.

In this study relationships between reading and two aspects of formulaic sequences were examined. That is, as this study is correlational, it does not speak directly to the issues of causality in instruction. It would be interesting to investigate the possible effects of general knowledge of formulaic sequences on reading comprehension in future studies. This involves two groups of subjects, control and treatment, with the treatment group receiving instruction on target sequences. Then, the effect of instruction on the treatment group can be tested employing statistical techniques such as ANOVA or t-test.

This study involved reading texts which were taken from newspapers. In future studies, interesting results might be achieved if texts having different genres are selected to see if formulaic sequences affect reading texts having different genres differently. For instance, whether reading narrative texts would benefit from knowledge of formulaic sequences more, or reading journal articles. In addition, different genres may include different types of formulaic sequences. This would contribute to the assessment of a wider range of sequences. As

Glaser 1998 puts it, "phraseological units are unevenly distributed in texts, and their occurrence depends on the text type and the given genres" (p. 130). In the same vein, Moon (1998) states, "some genres are marked by relatively high densities of FEIs, others are not and may even seem to block the use of certain kinds of FEIs" (p. 68).

In addition, it should not be ignored that reading is composed of several sub-skills only one of which is comprehension. In this study this aspect of reading was investigated. In the future studies, it would be interesting to study the relationships between other sub-skills of reading such as reading speed and knowledge of formulaic sequences.

This study was mainly product oriented with the results of students' processing assessed through tests; future studies could be more process oriented with the focus on the online processing of formulaic sequences. The students can be also asked to report the strategies they employ to overcome their problems with regard to the unknown formulaic sequences while reading a text. This requires recall protocols in which the subjects' reports can be used to analyse the processes involved in reading.

Finally, further research is needed with special reference to the pragmatic or discoursal functions of formulaic sequences in a text and their effect on reading comprehension. For instance, how formulaic sequences which function as discourse organizers like relators (e.g., *it has to do with*) or qualifiers (e.g., *it is worth noting that*) facilitate EFL reading comprehension.

5.7 Conclusion

The findings of this research study are interesting for at least two reasons. On the one hand, they contribute to the literature on formulaic sequences in general

and their contribution to reading comprehension in particular. They also support the previous research on multi-word units with respect to their relationship with overall proficiency. As all language skills are inter-related, it is not surprising to say that competency in formulaic sequences would affect both language comprehension and production and all language components simultaneously. On the other hand, a significant aspect of the findings is their potential practical implications for designing syllabi and methodologies employed in teaching EFL/ESL learners.

Given that there was virtually no empirical research concerning the relationship between formulaic sequences and reading comprehension, this study was an initial attempt to find out whether these sequences can be added to the factors contributing to better reading by EFL/ESL learners. It has empirically shown a significant positive relationship between recognition of formulaic sequences and reading comprehension. Obviously, recognition of formulaic sequences is not the only factor having connection to reading comprehension. Reading is a complex process in which many subcomponents including "linguistic knowledge, metacognitive skills, and strategic coordination of lower-and higher-order processes are all involved simultaneously" (Fukkink et al. 2005, p. 72) and word recognition or phrase identification is just one of the subcomponents of this process. The use of preformed word combinations supports comprehension, as the reader can understand the meaning of a passage or text without having to attend to every word. Especially, in the case of non-native-like expressions, unfamiliarity with these expressions is problematic for the readers as these combinations "can irritate the recipient and draw the attention away from the message" (Nesselhauf 2005, p.2). And knowing them help the readers "anticipate and remember information" (Wray 2002a, p. 87).

Recognition of formulaic sequences and understanding these sequences during reading are considered as lower-level processes which if done automatically without the expenditure of the cognitive resources, spares the working memory for the execution of higher level processes needed to overcome textual demands. As Chun and Plass (1997) maintain, "when more words are recognized immediately or known, more attention can be paid to top-down processing and to overall comprehension, resulting in higher scores on the comprehension measure" (p. 516). Nassaji (2003) encourages increasing readers' sensitivity to textual relationships and recommends using reading exercises that help readers "to process words and their relationships in texts as efficiently as required for fluent processing and understanding of text" (p. 271).

Automatic processing of these sequences requires ESL/EFL learners to keep them in mind holistically so as to be able to retrieve them as such whenever needed. This involves hard work from the learners themselves and the teachers and the materials developers. Teachers are responsible to provide opportunities for the learner to understand the importance of these sequences in language acquisition and learn how to recognize, use and internalize these sequences. Being aware of these sequences during reading saves the readers time and energy to process the higher level tasks of reading such as comprehension.

Understanding of formulaic sequences did not show a close relationship with reading comprehension in the main study. Moreover, the results of the qualitative analysis in the follow-up study showed that whether understanding of these sequences affects reading comprehension depends on several factors including their contextual environment, their nature and their function in a text. It was also found that whenever readers feel that an unknown sequence plays an important role in understanding a text, they try to extract its meaning using available cues and knowledge sources.

With regard to the results of this study, as we are dealing with students in an academic setting – in this case adult learners- it is not surprising that our participants could have brought with themselves a host of experiences, aptitudes, background knowledge and metacognitive strategies to the interpretation of the text. So, in interpreting and generalising the results of the study all these factors should be taken into consideration. In other words, these findings are unique to this study with these particular subjects, tests and methodology. In order to understand the complete picture and be able to generalize these findings, the study has to be replicated, employing subjects with different backgrounds, adopting different designs and using different test formats to assess recognition and understanding of formulaic sequences and a variety of test formats to assess reading comprehension as "the outcome of each individual assessment task provides a limited representation of reading comprehension" (Brantmeier 2005, p. 22).

A final note is that this study can lay the foundations for future studies on the role of formulaic sequences in comprehension using different subjects, materials and procedures as no single approach can provide the whole picture and it is just by examining these sequences from different perspectives that a comprehensive account of them can be provided.

Bibliography

Abel, B. (2003) 'English idioms in the first language and second language lexicon: a dual representation approach' *Second Language Research,* 19: 329-58

Aisenstadt, E. (1979) 'Collocability restrictions in Dictionaries' ITL: Review of Applied Linguistics 45, 6: 71-74. [5.1]

Alderson, J.C. (2000) *Assessing Reading.* Cambridge, England: Cambridge University Press

Alexander, R.J. (1978) 'Fixed expressions in English: A linguistic, psycholinguistic, sociolinguistic and didactic study (part 1)' *anglistik & englischunterricht*, 6:171-88

Alexander, R. J. (1984) 'Fixed expressions in English: reference books and the teacher' *ELT Journal*, 38: 127-32

Alexander, R. J. (1987) 'Problems in understanding and teaching idiomaticity in English' *Anglistik und Englischunterricht*, 32:105-122

Allen, A., Bernhardt, E., Berry, M. and Demel, M. (1988) 'Comprehension and text genre: an analysis of secondary school foreign language readers' *The Modern Language Journal*, 72, 2: 163-172

Altenberg, B. (1990) 'Speech as linear composition' in G. Caie, K. Haastrup, A. L., Jakobsen, Nielsen, J. E., Sevaldsen, J., Specht, H. and Zettersten, A. (eds.) *Proceedings from the Fourth Nordic Conference for English Studies*, Copenhagen University Press, Copenhagen pp 133-143

Al-Zahrani, M. S. (1998) 'Knowledge of English lexical collocations among male Saudi college students majoring in English at a Saudi university' *Unpublished doctoral dissertation*, Indiana University of Pennsylvania, Pennsylvania, USA

Anderson, J. R. (1983) *The Architecture of Cognition* Mahwah, NJ: Erlbaum

Anderson, R. C., and Freebody, P. (1981) Vocabulary knowledge' in J.T. Guthrie (ed.) *Comprehension and Teaching: research review* Newark, DE: International Reading Association pp 77-117

Anderson, R. C. and Pearson, D. P. (1988) 'A schema-theoretic view of basic processes in reading comprehension' in P. L. Carrell *Interactive Approaches to Second Language Reading* pp 37-55

Arnaud, P. J. L. and Savignon, S. J. (1997) 'Rare words, complex lexical units and the advanced learner' in J. Coady & T. Huckin (eds.) *Second Language Vocabulary Acquisition* Cambridge: CUP pp 157-173

Balota, D.A., and Neely, J.H. (1980) 'Test-expectancy and word-frequency effects in recall and recognition' *Journal of Experimental Psychology: Human Learning and Memory, 6*: 576-587

Balota, D. A., and Chumbley, J. L. (1984) 'Are lexical decision a good measure of lexical access? The role of word frequency in the neglected decision stage' *Journal of Experimental Psychology: Human Perception and Performance* 10, 340-35

Bates, E., Bretherton, I. and Snyder, L. (1988) *From First Words to Grammar* Cambridge University Press, Cambridge

Bates, E. and MacWhinney, B. (1989) 'Functionalism and the composition model' in B. MacWhinney & E. Bates (eds.) *The Cross-Linguistics Study of Sentence Processing* Cambridge University Press Cambridge pp 3-73

Bauer, L. (1983) *Word formation* Cambridge: Cambridge University Press

Beck, I. L., Perfetti, C. A. and McKeown, M. G. (1982) 'Effects of long-term vocabulary instruction on lexical access and reading comprehension' *Journal of Educational Psychology* 74, 4: 506-521

Becker, J. (1975) 'The phrasal lexicon' Bolt Beranek & Newman Report no. 3081, Al Report no. 28. Reprinted in R. Shank & B.L. Nash-Webber (eds.) *Theoretical Issues In Natural Language Processing* Cambridge, MA: Bolt Beranek & Newman, 60-63

Bengeleil, N. F., Paribakht, T. S. (2004) 'L2 reading proficiency and lexical inferencing by university EFL learners' *The Canadian Modern Language Review*, 61, 2: 225-49

Bernhardt, E. B. (1986) 'Proficient texts or proficient readers?' *ADFL Bulletin* 18: 25-28

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999) *Longman Grammar of spoken and written English* Harlow Longman

Bishop, H. (2004a) *The Noticing of Formulaic Sequences by Second Language Readers* Unpublished PhD Thesis University of Wisconsin-Madison

Bishop, H. (2004b) 'The effect of typographic salience on the look up and comprehension of unknown formulaic sequences' In N Schmitt (ed.) *Formulaic Sequences: Acquisition, processing, and use* John Benjamins pp 227-248

Bobrow, S., and Bell, S. (1973) 'On catching on to idiomatic expressions' *Memory and Cognition,* 1: 343-346

Boers, F. Eyckmans, J. Kappel, J, Stengers. H. and Demecheleer, M. (2006) 'Formulaic sequences and perceived oral proficiency: putting a lexical approach to test' *Language Teaching* Research 10, 3: 245-261

Bolander, M. (1989) 'Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish in K. Hyltenstam & L.K. Obler (eds.) *Bilingualism across the Lifespan* Cambridge: Cambridge University Press pp 73-86

Bolinger, D. (1976) 'Meaning and Memory' *Forum Linguisticum, 1: 1-14*

Brantmeier, C. (2005) 'Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish' *The Modern Language Journal,* 89, i: 37-50

Brantmeier, C. (2006) 'Advanced L2 learners and reading placement: Selfassessment, CBT, and subsequent performance' *System,* 34: 15–35

Brown, D. F. (1974) 'Advanced vocabulary teaching: the problem of collocation' *RELC Journal,* 5,2: 1-11

Brown, R. (1973) *A First Language: The Early Stages* London: Allen and Unwin

Bybee, J. (1998) 'The emergent lexicon' *Chicago Linguistics Society,* 34:421-435

Cacciari, C. and Tabossi, P. (1993) *Idioms: Processing, structure, and interpretation* (eds.) Hillsdale, NJ: Erlbaum

Carpenter, P. A., Miyake, A., & Just, M. A. (1995) 'Language Comprehension: Sentence and discourse processing' *Annual Review of Psychology*, 46, 91-120

Carrell, P. L. (1998) 'Interactive text processing: Implications for ESL/second language reading classrooms' in P. L. Carrell, J Devine & D E Eskey (eds) *Interactive Approaches to Second Language Reading* Cambridge: Cambridge University Press, pp 239-259

Carrell, P. L., and Wallace, B. (1983) 'Background knowledge: Context and familiarity in reading comprehension' in M. Clarke & J. Handscombe (eds.) *On TESOL '82: Pacific perspectives on language learning and teaching* Washington, DC: TESOL pp 295–308

Carrell, P. L. & Grabe, W. (2002) 'Reading' in N. Schmitt (ed.) *An Introduction to Applied Linguistics* London: Arnold pp 233-250

Carter, R. (1988) 'Vocabulary, Cloze and discourse: an applied linguistic view' in R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching* Harlow: Longman pp 161-180

Carter, R (1998) *Vocabulary: applied linguistic perspective* London; New York: Routledge

Carter, R. (2004) *Language and Creativity: The art of Common Talk* London: Routledge

Carter, R. and McCarthy, M. (1988) *'Vocabulary and Language Teaching'* Pearson Education Limited

Carton, A. (1971) 'Inferencing: A process in using and learning language' in P. Pimsleur & T. Quinn (eds.) *The Psychology of Second Language Learning* Cambridge: Cambridge University Press pp 45-58

Carver, R. P. (1990) *Reading Rate: A review of research and theory* New York: Academic Press

Carver, R. P. (1997) 'Reading for one second, one minute, or one year from the perspective of Rauding theory' *Scientific Studies of Reading*, 1: 3-45

Carver, R. P. (2000) *The cause of high and low reading achievement* Mahwah, NJ: Erlbaum

Carver, R. P. (2003) 'The highly lawful relationships among pseudoword decoding, word identification, spelling, listening, and reading' *Scientific Studies of Reading,* 7: 127–154

Catalán, R. M. J. (2003) 'Sex differences in L2 vocabulary learning strategies' *International Journal of Applied Linguistics,* 13,1: 54-77

Chastain, K. (1988) *Developing second Language skills: Theory and practice* (3rd ed.) New York: HBJ Publishers

Chaudron, C. and Richards, J. (1986) 'The effect of discourse markers on the comprehension for lectures' *Applied Linguistics,* 7:113-127

Cheng, W., Greaves C, and Warren, M. (2006) 'From n-gram to skipgram to concgram' *International Journal of Corpus Linguistics* 11:4 411-433

Chomsky, N. (1980) *Rules and Representations* Columbia University Press

Chun, D.M. and Plass, J. L. (1996) 'Facilitating reading comprehension with multimedia' *System,* 24, 4: 503-519

Chun, D. M. and Plass, J. L. (1997) 'Research on text comprehension in multimedia environments' *Language Learning and Technology,* 1, 1: 60-81

Cieslicka, A. (2006) 'On-line processing of idiomatic expressions by second language learners' *Second Language Research,* 22, 2:115-144

Clapham, C. (1996) *The development of IELTS: A study of the effect of background knowledge on reading comprehension.* Cambridge: Cambridge University Press

Clark, E. V. (1974) 'Performing without competence' *Journal of Child language,* 1: 1-10

Clark, H. and Clark, E. V. (1977) *Psychology and language: An Introduction to Psycholinguistics*, New York: Harcourt Brace Javanovich

Conklin, K. and Schmitt, N. (2008) 'Formulaic Sequences: Are they processed more quickly than nonformuliac language by native and nonnatives speakers?' *Applied Linguistics,* 29, 1: 72-89

Cooper, T. C. (1999) 'Processing of idioms by L2 learners of English' *TESOL Quarterly* 33, 2: 233-262.

Coulmas, F. (1979) 'On the sociolinguistic relevance of routine formulae' *Journal of Pragmatics*, 3: 239-266

Coulmas, F. (1981) 'Introduction: conversational routine' in F Coulmas (ed.) *Conversational routine* The Hague: Mouton pp 1-17

Coulmas, F. (1994) 'Formulaic language' in R.E. Asher (ed.) *Encyclopedia of Language & Linguistics* Oxford: Pergamon, 1292-1293

Cowie, A. P. (1978) 'The place of illustrative material and collocations in the design of a learner's dictionary' in P. Strevens (ed.) *In Honour of A. S. Hornby*, Oxford University Press

Cowie, A. P. (1988) 'Stable and creative aspects of vocabulary use' in R. Carter & M. MacCarthy (eds.) *Vocabulary and Language Teaching London* New York: Longman pp 126-139

Cowie, A. P. (1991) 'Multiword units in newspaper language' in Sylviane Granger (ed.), *Perspectives on the English Lexicon.* A Tribute to Jacques van Roey Louvain-la-Neuve: Cahiers de l'Institut de Linguistique de Louvain pp. 101– 116

Cowie, A. P. (1992) 'Multiword lexical units and communicative language teaching' in Pierre J. L. Arnaud & Henri Béjoint (eds.), *Vocabulary and Applied Linguistics* (pp. 1–12). Houndsmills: Macmillan.

Cowie, A. P. and Mackin, R. (1975) *Oxford Dictionary of Current Idiomatic English*, Vol. I- Phrasal Verbs, Oxford University Press

Cowie, A. P., Mackin, R. and McCaig, I. R. (1983) Oxford *Dictionary of Current Idiomatic English. Vol. 2: Phrase, Clause and Sentence Idioms* Oxford: Oxford University Press

Crick , F. H. C. (1979) 'Thinking about the brain' *Scientific American* 241: 219-232; reprinted in 'The Brain' *Scientific American* special issue pp 130-137

Cruse, D. A. (1986) *Lexical Semantics* CUP

De Bot, K., Paribakht, T. S., and Wesche, M. B. (1997) 'Towards a lexical processing model for the study of second language vocabulary acquisition' *Studies in Second Language Acquisition*, 19: 309-329

DeCarrico, J. S. and Nattinger, J. R. (1988) 'Lexical phrases for the comprehension of academic lectures' *English For Specific Purposes,* 7:91-102

Dechert, H.W. (1983) 'How a story is done in a second language' in C. Faerch & G. Kasper (eds.) *Strategies in Interlanguage Communication* New York: Longman, pp 175-195

Dolch, E. W. (1949) 'Phrase perception in Reading' *The Elementary School Journal* 49, 6: 341-147

Dole, J. A. (2000) 'Explicit and Implicit Instruction in Comprehension' in B. Taylor, M. F. Graves and P. W. van den Broek *Reading for meaning : fostering comprehension in the middle grades* New York : Teachers College Press ; Newark, DE : International Reading Association pp 52-69

Draper, A. G., and Moeller, G. H. (1971) 'We think with words (therefore, to improve thinking, teach vocabulary)' *Phi delta Kappan,* 52: 482-484

Edwards, A. L. (1996) 'Reading Proficiency Assessment and the ILR/ACTFL Text Typology: A Reevaluation' *The Modern Language Journal,* 80, 3: 350-361

Ehrlich, S. F. and Rayner, K. (1981) 'Contextual effects on word perception and eye movements during reading' *Journal of Verbal Learning and Verbal Behavior* 20: 641- 655

Ellis, N. C. (1996) 'Sequencing in SLA: phonological memory, chunking and points of order' *Studies in Second Language Acquisition,* 18: 91-126

Ellis, N. C (2002a) 'Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition' *Studies in Second Language Acquisition* 24: 143-188

Ellis, N. C (2002b) 'Reflections on frequency effects in language processing' *Studies in Second Language Acquisition* 24: 297-339

Ellis, N.C. and 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

Ellis, R. (1994) *The Study of Second Language Acquisition* Oxford: Oxford University press

Erman, B. and Warren, B. (2000) 'The idiom principle and the open choice principle' *Text,* 20,1: 29-62

Eskey, D. E. (1998) 'Holding in the bottom: an interactive approach to the language problems of second language readers' in P.L. Carrell *Interactive Approaches to Second Language Reading* Cambridge: Cambridge University Press pp 93-100

Eskey, D. and Grabe, W. (1998) 'Interactive models for second language reading: Perspectives on instruction' in P. Carrell, J. Devine, & D. Eskey (eds.), *Interactive approaches to second language reading* New York: Cambridge University Press pp 223-239

Farghal, M. and Obiedat, H. (1995) 'Collocations: A neglected variable in EFL' *International Review of Applied Linguistics in Language Teaching*, 33(4):315-331

Fernando, C. (1996) *Idioms and idiomaticity* OUP

Fernando, C. and Flavell, R. (1981) 'On Idiom: critical views and perspectives' *Exeter Linguistic studies*, Volume 5. Exeter: University of Exeter

Firth, J. R. (1951) 'Modes of meaning' in Firth, J. R. (ed.) (1957) *Papers in Linguistics*, Oxford University Press

Firth, J.R. (1957) *Papers in Linguistics* OUP

Firth, J. R. (1968) 'A synopsis of linguistic theory 1930-1955' in F. R. Palmer (ed.): *Selected Papers of J R Firth 1952-1959*. Bloomington: Indiana University Press pp 1-32.

Forster, K., and Chambers, S. (1973) 'Lexical access and naming time' *Journal of Verbal Learning and Verbal Behaviour* 12: 627-635

Foster, P. (2002) 'Rules and 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 and Testing* London, new York; Longman pp 75-94

Fraser, B. (1970) 'Idioms with a transformational grammar' *Foundations of language,* 6: 22-42.

Fraser, C. (1997) 'The impact of lexical processing strategy instruction on L2 readers' strategy use, reading rate, reading comprehension, and vocabulary learning' Unpublished doctoral dissertation, Ontario Institute for Studies on Education of the University of Toronto

Fraser, C. (1999) 'Lexical processing strategy use and vocabulary learning through reading' *Studies in Second Language Acquisition,* 21: 225-242

Fries, C. C. (1945) *Teaching and Learning English as Foreign Language* University of Michigan Press, Ann Arbor

Frisson, S., Rayner, K. and Pickering, M. (2005) 'Effects of Contextual Predictability and Transitional Probability on Eye' *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31, 5: 862–877

Fukkink, R. G., Hulstijn, J. and Simis, A. (2005) 'Does training in second-language word recognition skills affect reading comprehension? An Experimental study' *The Modern Language Journal,* 89, 1: 54-75

Gates, A. (1949) 'Character and purposes of the yearbook' in N Henry(ed.) *The Forty eighth Yearbook of National Society for the Study of Education: Part II Reading in the elementary school* Chicago, University of Chicago Press pp 1-9

Gazdar, G., Klein, E., Pullam, G., and Sag, I. (1985) *Generalized Phrase Structure Grammar* Oxford: Blackwell

Gibbs, R. W. (1980) 'Spilling the beans on understanding and memory for idioms in conversation' *Memory and Cognition,* 8: 149-156

Gitsaki, C. (1996) *The Development of ESL Collocational Knowledge.* Unpublished Ph.D. Thesis. Brisbane, Australia: The University of Queensland.

Gläser, R. (1998) 'The stylistic potential of phraseological units in the light of genre analysis' in A. P. Cowie Phraseology: Theory, Analysis, and Applications OUP pp 125-145

Goodman, K. (1967) 'Reading: A psycholinguistic guessing game' *Journal of the Reading Specialist,* :126-135

Goodman, K (1971) 'Psycholinguistic universals in the reading process' in P. Pimsleur and T. Quinn (eds.) *The Psychology of second language learning*, 135-142. Cambridge: Cambridge University Press

Goodman, K (1998) 'The reading process' in P.L. Carrell, J. Devine and D. E. Eskey (eds.) *Interactive Approaches to Second Language Reading* Cambridge; New York : Cambridge University Press pp 11-21

Gorman, A.M. (1961) 'Recognition memory for nouns as a function of abstractness and frequency' *Journal of Experimental Psychology,* 61: 23–29

Gough, W. and Tunmer, W. (1986) 'Decoding, reading, and reading disability' *RASE: Remedial and Special Education,* 7: 6-10

Gough, P. B., Hoover, W. A., and Peterson, C. L. (1996) 'Some observations on a simple view of reading' in C. Cornoldi & J. Oakhill (eds.) *Reading comprehension difficulties: Processes and intervention* Mahweh, NJ: Lawrence Erlbaum Associates pp. 1-14

Grabe, W. (1998) 'Reassessing the term "interactive"' in P.L. Carrell, J. Devine and D. E. Eskey (eds.) *Interactive Approaches to Second Language Reading,* 56-70

Grabe, W. and Stoller, F. L. (2002) *Teaching and Researching Reading* Harlow: Pearson Education

Granger, S. (1998) 'Prefabricated patterns in advanced EFL writing: collocations and formulae' in A P Cowie (ed.) *Phraseology: theory, analysis and applications* Oxford: Clarendon Press, 145-160

Wible, D. (2008) 'Multiword expressions and the digital turn' in S. Granger & F. Meunier (eds) *Phraseology in Foreign Language Learning and Teaching* John Benjamins Publishing Company pp. 163-181

Grant, L. and Bauer, L. (2004) 'Criterion for Re-defining Idioms: Are we Barking up the Wrong Trace?' *Applied Linguistics,* 25, 1: 38-61

Gyllstad, H. C. (2007) *Testing English Collocations: Developing receptive tests for use with advanced Swedish learners* Unpublished Doctoral Thesis, Lund University

Haastrup, K. (1991) *Lexical inferencing procedures, or talking about words: Receptive procedures in Foreign Language Learning with Special Reference to English* Tübingen, Germany: Gunter Narr

Hakuta, K. (1974) 'Prefabricated patterns and the emergence of structure in second language acquisition' *Language Learning,* 24, 2: 289-297

Hakuta, K. (1976) 'A case study of a Japanese child learning English' *Language Learning* 26: 321–351

Halliday, M. A. K. (1966) 'Lexis as a Linguistic Level' in Bazell, C. E., Catford, J. C., halliday, M. A. K. & Robins, R. H. *In memory of F. R. Firth*, Longman

Hanania, E. A. S. and Gradman, H. L. (1977) 'Acquisition of English structures: a case study of an adult native speaker of Arabic in an English-speaking environment' *Language Learning* 27,1: 75-91

Hatch, E. M. and Farhady, H. (1982) *Research Design and Statistics for Applied Linguistics* Rowley, Massachusetts: Newbury House

Harley, T.A. (2001) *The Psychology of Language: From Data to Theory* Psychology Press

Haynes, M., and Baker, I (1993) 'American and Chinese readers learning from lexical familiarization in English texts' in T. Huckin, M. Haynes, & J. Coady (eds.), *Second Language Reading and Vocabulary Acquisition* Norwood, NJ: Ablex pp. 130-152

Hickey, T. (1993) 'Identifying formulas in first language acquisition' *Journal of Child Language,* 20: 27-41

Hinkle, D. E., Wiersma, W. and Jurs, S. G. (1979) *Applied Statistics for the Behavioral Sciences* Houghton Mifflin Company

Hockett, C. F. (1959) 'The objectives and process of language teaching' Reprinted in D. Byrne (ed.) *English Teaching Extracts* London: Longman, 1969

Hoey, M. (1991) *Patterns of Lexis in Text* Oxford: OUP

Hoey, M. (1997) 'From Concordance to Text Structure: New Uses for Computer Corpora' in Melia, J. & Lewandoska, B. (eds.) *Proceedings of PALC 97.* Lodz: Lodz University Press.

Hoey, M. (2000) 'A World Beyond Collocation: New Perspectives on Vocabulary Teaching' in Lewis, M. (ed) *Teaching Collocations.* Hove: Language Teaching Publications pp 130-152

Hoey, M. (2003) 'Why grammar is beyond belief' in J-P. van Noppen, C. den Tandt and I. Tudor (eds.) *Beyond: New Perspectives in Language, Literature and ELT* Special issue of Belgian Journal of English Language and Literature, new series 1: 183-96

Hoey, M. (2005) *Lexical priming: a new theory of words and language.* London: Routledge

Hoover, W. A., and Gough, P. B. (1990) 'The simple view of reading' *Reading and Writing: An interdisciplinary Journal*, 2: 127-160

Howarth, P. (1996) *Phraseology in English academic writing. Some implications for language learning and Dictionary making* Tubingen: Niemeyer

Howarth, P. (1998) 'Phraseology and second language proficiency' *Applied Linguistics,* 19, 1: 24-44

Howatt, A. P. R. (1984) *A History of English Language Teaching.* Oxford: Oxford University Press

Hsueh-chao, H. M. and Nation, P. (2000) 'Unknown vocabulary density and reading comprehension' *Reading in a Foreign Language*, 13 (1) 403-430

Hudson, T. (1998) 'Theoretical perspectives on reading' *Annual Review of Applied Linguistics,* 18: 43-60

Hulstijn, J. (2001) 'International and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity' in P Robinson (ed.), *Cognition and Second Language Instruction* Cambridge: Cambridge University Press pp. 258-286

Hunston, S., Francis, G. and Manning, E. (1997) 'Grammar and vocabulary: showing the connections' *ELT Journal,* 51, 3: 208-216

Irujo. S. (1986) 'Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language' *TESOL Quarterly*, 20,2: 287-300

Irujo. S. (1993) 'Steering clear: avoidance in the production of idioms' *International Review of Applied Linguistics in Language teaching,* 31, 3: 205-219

Jespersen, O. (1924) *The Philosophy of Grammar* London, New York

Jiang. N. and Nekrasova, T. (2007) 'The processing of formulaic sequences by second language speakers' *The Modern Language Journal,* 91, 3: 433-445

Jones, S. and Sinclair, J. M. (1974) 'English lexical collocations – a study in computational linguistics' reprinted in J. A. Foley (ed.) *J. M. Sinclair on Lexis and Lexicography* Singapore: National University of Singapore pp 21-54

Jones, M. and Haywood, S. (2004) *Facilitating the acquisition of formulaic sequences* N. Schmitt (ed.) Amsterdam/Philadelphia: John Benjamins Publishing Company, pp 269-300

Kammeenui, E.J., Carine, D. W., & Freschi R. (1982) 'Effects of text construction and institutional procedures for teaching word meanings on comprehension and recall' *Reading Research Quarterly,* 1982, 17: 367-388

Katz, J and Postal, P. M. (1963) 'Semantic interpretation of idioms and sentences containing them' MIT Research laboratory of Electronics, *Quarterly progress Report,* 70: 275-82

Kaylani, C., (1996) 'The influence of gender and motivation on EFL learning strategy use in Jordan' in R. Oxford (ed.) *Language learning strategies around the world: Cross-cultural perspectives* Honolulu: University of Hawai'i: Second language teaching and curriculum center, pp 75-88

Kennedy, G. (1990) 'Where grammar and vocabulary teaching meet' *Language teaching Methodology for the Ninties,* pp 215 228

Keshavarz, M. H. and Salimi, H. (2007) 'Collocational competence and cloze test performance: A study of Iranian EFL learners' *International Journal of Applied Linguistics,* 17, 1: 81-92

Keshavarz, M. H., Atai. M. R. & Ahmadi, A. (2007) 'Content schemata, linguistic simplification, and recall' *Reading in a Foreign Language,* 19, 1: 19-33

Khalideh, S. A. (2001) 'The relationship between knowledge of *iraab*, lexical knowledge, and reading comprehension of nonnative readers of Arabic' *The Modern Language Journal,* 85, iii pp 416-431

Kichun, Nam, Yoonhyong, L. and Chang, H. L. (2004) 'The Locus of Word Length and frequency Effect in Comprehending English Words by Korean-English Bilinguals and Americans' *ICONIP*: 306-315

Kintsch, W. (1998) *Comprehension: A paradigm for cognition* New York: Cambridge University Press

Kirsner, K (1994) 'Implicit processes in second language learning' in N. Ellis (ed.) *Implicit and Explicit Learning of Language* San Diego, CA: Academic Press pp 283-312

Koda, K. (1989) 'The effects of transferred vocabulary knowledge on the development of L2 reading proficiency' *Foreign Language Annuals,* 22, 529-540

Koda, K. (1996) 'L2 word recognition research: a critical review' *The Modern Language Journal,* 80, 4: 450-460

Koda, K. (2005) *Insights into Second Language Reading* Cambridge University Press

Krashen, S. and Scarcella, R. (1978) 'On routines and patterns in language acquisition and performance' *Language Learning* 28, 2:283-300

Langacker, R. W. (1986) *Foundations of Cognitive Grammar* Volume 1. Standford, California: Stanford University press

Laufer, B. (1991) 'How much lexis is necessary for reading comprehension?' in P. J. L. Arnaud & H. Bejoint (eds.) *Vocabulary and applied linguistics* Basingstoke: Macmillan pp 126-132

Laufer, B. (1997) 'The plight in second language reading: words you don't know, words you think you know, and words you can't guess' in J. Coady & T. Huckin *Second Language Vocabulary Acquisition* Cambridge University Press pp 20-34

Laufer, B. and Nation, P. (1999) 'A productive-size test of controlled productive ability' *Language Testing*, 16: 33-51

Lee, J. F. (1986a) 'On the use of the recall task to measure L2 reading comprehension' *Studies in Second Language Acquisition,* 8: 201–212

Lee, J. F. (1986b) 'Background knowledge and L2 reading, *Modern Language Journal,* 70: 350–354

Lee, S. K. (2007) 'Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form' *Language Learning,* 57, 1: 87–118

Levelt, W. (1993) *Speaking: from Intention to Articulation* Cambridge, MA: MIT Press

Levelt, W. J. M., Roelofs, A. and Meyer, A. S. (1999) 'A theory of lexical access in speech production' *Behavioral and Brain Sciences*, 22:1-75

Lewis, M. (1993) *The Lexical Approach: the state of ELT and a way forward* Hove: Teacher training Publications

Lewis, M. (1997) 'Pedagogical implications of the lexical approach' in J. Coady & T. Huckin (eds.) *Second language vocabulary acquisition* Cambridge: Cambridge University Press pp 255-270

Lewis, M. (2000) 'Language in the lexical approach' in M. Lewis (ed.) *Teaching Collocation: Further Developments In The Lexical Approach* Hove: Language Teaching Publications pp126-154

Lewis, M. (2000) 'learning in the lexical approach' in M. Lewis (ed.) *Teaching Collocation: Further Developments In The Lexical Approach* Hove: Language Teaching Publications 155-185

Lewis, M. (2002) *Implementing the Lexical Approach: Putting Theory into Practice* Heinle: Thomson

Liontas, J. (2002) 'Context and idiom understanding in second languages' *EUROSLA Yearbook* John Benjamins pp 155-85

Logan, G. D. (1988) 'Toward an instance theory of automatization' *Psychological Review,* 95: 492-527

Lyerly, S. B. (1959) 'Significance Levels for the Kuder-Richardson (21) Reliability coefficient' *Educational and Psychological Measurement,* 19, 1: 73-75

Lyons, J (2002) *Introduction to Theoretical Linguistics* Cambridge University Press

McCarthy, M. (1990) *Vocabulary* OUP

McCone, J. (1999) 'States of mind' *New Scientist,* 2178: 30–33

McErlain, T. (1999) 'The nature of listening: the need for listening in English for academic purposes' *IBÉRICA,* 1: 77-81

MacKay, D. M. (1951) 'In search of basic symbols' *Cybernetics: transactions of the 8th conference* pp 181-221

Makkai, A. (1975) 'The cognitive organization of idiomaticity rhyme or reason?' *Georgetown University Working Papers of Language and Linguistics,* 11: 10-29

McLaughlin, B, Rossman, T and McLeod, B. (1983) 'Second Language Learning: An information-processing perspective' *Language Learning,* 33, 2: 135-158

Macleod, C. M and Kampe, K. E. (1996) 'Word frequency effects on recall, recognition, and word fragment completion tests' *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1: 132-142

Meara, P. M. (1980) 'Vocabulary acquisition: a neglected aspect of language learning' *Language Teaching and Linguistics: Abstracts,* 13: 221-46

Meara, P. M. (1983) *Vocabulary in a Second Language* CILT, London

Meara, P., and Jones, G. (1989) *Eurocentres Vocabulary Test 10 KA* Eurocentres, Zurich

Mel'cuk, I. (1998) 'Collocations and lexical functions' in A P Cowie (ed.) *Phraselology: Theory, Analysis and Applications* Oxford: Clarendon Press pp 23-53

Mezynski, K. (1983) 'Issues concerning the acquisition of knowledge: effects of vocabulary training on reading comprehension' *Review of Educational Research*, 53, 2: 253-279

Miller, G.A. (1988) 'The challenge of universal literacy' *Science*, 241: 1293-1299

Milton, N. (1994) 'Word recognition'
http://www.epistemics.co.uk/staff/nmilton/papers/word-recognition.htm

Monsell, S., Doyle, M. C, and Haggard, P. N. (1989) 'Effects of frequency on visual word recognition tasks: Where are they?' *Journal of Experimental Psychology: General,* 118: 43-71

Moon, R. (1992) 'Textual aspects of fixed expressions in learners' dictionaries' in P.J.L. Arnaud & H. Bejoint (eds.) *Vocabulary & Applied Linguistics* Basingstoke: Macmillan pp13-27

Moon, R. (1998) *Fixed Expressions and Idioms in English* Oxford: Clarendon Press

Morton, J. (1969) 'The interaction of information in word recognition' *Psychological Review,* 76: 340-354

Nagy, W. (1988) *Teaching Vocabulary to Improve Reading Comprehension* Newark, DE: International Reading Association

Nagy, W. E., Anderson, R. C. and Herman, P. A. (1987) 'Learning word meanings from context during normal reading', *American Educational Research Journal,* 24: 237-270

Nassaji, H. (2003) 'L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing,' *TESOL Quarterly*, 37, 4: 645–70.

Nassaji, H. (2004) 'The relationship between depth of vocabulary knowledge and L2 Learners' lexical inferencing strategy use and success' *The Canadian Modern Language Review,* 61,1: 107-134

Nation, I. S. P. (1980) 'Strategies for receptive vocabulary learning' Guidelines: *RELC Supplement,* 3: 171-5

Nation, I. S. P. (1993) 'Vocabulary size, growth and use' in R. Schreuder and B. Weltens (eds) *The Bilingual lexicon* Amsterdam/Philadelphia: John Benjamins pp 115-134

Nation, I. S. P. (2001) *Learning Vocabulary in Another Language* Cambridge: CUP

Nattinger, J. R. (1980) 'A lexical phrase grammar for ESL' *TESOL Quarterly,* 14, 33: 337-344

Nattinger, J. R. and DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching* Oxford: Oxford University Press

Nelson, E. (1992) 'Memory for metaphor by nonfluent bilinguals' *Journal of Psycholinguistic Research,* 21:111-25

Nelson, K. (1981) 'Individual differences in language development: implications for development and language' *Developmental Psychology,* 17, 2: 170-187

Nelson, M. (2000) *A Corpus-based study of the Lexis of Business English and Business English Teaching Materials* Unpublished Thesis. University of Manchester, Manchester, UK

Nesi, H. and Basturkmen, H. (2006) 'Lexical bundles and discourse signaling in academic lectures' *International Journal of Corpus Linguistics,* 11,3: 283-304

Nesslehauf, N. (2005) *Collocations in a Learner Corpus* Philadelphia, PA. USA: John Benjamins Publishing Company

Nunberg, G. (1978) *The Pragmatics of Reference* Bloomington: Indiana University of linguistics

Nunberg, G., Sag, I. A., and Wasow, T. (1994) 'Idioms' *language,* 70: 491-534

Nuttall, C. (2000) *Teaching Reading Skills in a Foreign Language* Macmillan Publishers Limited

Oller, J. W., Jr. and Richards, J. C. (1973) *Focus on the Learner: Pragmatic Perspectives for the Language Teacher* Rowley, Mass.: Newbury House

Oller, J. W., Jr. and V. Streiff (1975) 'Dictation: a test of grammar based expectancies' in R. L. Jones and B. Spolsky (eds.) *Testing language Proficiency* Arlington, Virgnia: Centre for Applied Linguistics pp 71-88

Olson, G. M. (1973) 'Developmental changes in memory and the acquisition of language' in T. E. Moore (ed.) Cognitive development and the acquisition of language. London: Academic Press, pp 145–157

Ortony, A., Schallert. D. L., Reynolds, R. E. and Antos, S. J.(1978) 'Interpreting metaphors and idioms: Some effects of context on comprehension' *Journal of Verbal Learning and Verbal behavior*, 17, 4: 465-477

Oxford, R.L. (1990) *Language Learning Strategies: What every teacher should know* Boston: Heinle & Heinle

Palmer, H. E. (1933) *Second Interim Report on Collocations* Tokyo: Kaitakusha

Parel, R. (2004) 'The impact of lexical inferencing strategies on second language reading proficiency' *Reading and Writing: An Interdisciplinary Journal* 17: 847–873

Paribakht, S., and Wesche, M. (1993) 'Reading comprehension and second language development in a comprehension-based ESL program' *TESL Canada Journal, 11,* 9-29

Paribakht, S., and Wesche, M. (1999) 'Incidental vocabulary acquisition through reading: An introspective study' *Studies in Second Language Acquisition*, 21: 203-220

Paribakht, T. S. (2005) 'The influence of first language lexicalization on second language lexical inferencing: a study of Farsi-speaking learners of English as a foreign language' *Language Learning*, 55, 4: 701-748

Partington, A. (1998) *Patterns and Meanings* Amsterdam: John Benjamins

Pawley, A. and Syder, F.H. (1983) 'Two puzzles for linguistic theory: native-like selection and native-like fluency' in J.C. Richards & R. W. Schmidt (eds.) *Language and communication* New York: Longman, pp 191-226

Perfetti, C.A. (1985) *Reading Ability* New York: Oxford University Press

Perfetti, C. A. (1994) 'Psycholinguistics and reading ability' in M. A. Gernsbacher (ed.) *Handbook of Psycholinguistics* New York: Academic Press pp 849-894

Perfetti. C. A. and Hogaboam, T. (1975) 'The relationship between single word decoding and reading comprehension skill' *Journal of Educational Psychology,* 67, 4: 461-469

Peters, A. M. (1977) 'Language learning strategies: does the whole equal the sum of the parts?' *Language,* 53, 3:560-573

Peters, A. M. (1983) *Units of Language Acquisition* Cambridge: Cambridge University Press

Prendergast, T. (1864) 'The Mastery of Languages, or, the Art of speaking Foreign tongues idiomatically' London: R. Bentley. (Mastery Series for French, 1868; German, 1868; Spanish, 1869, Hebrew, 1971; and Latin 1872)

Qian, D. D. (1998) *Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language*, Unpublished PhD Thesis, University of Toronto, Ontario, Canada

Qian, D. D. (2002) Investigating the relationship between vocabulary knowledge and Academic reading performance: An assessment perspective *Language Learning,* 52, 3: 513-536

Raichle, M. E. (1998) 'The neural correlates of consciousness: an analysis of cognitive skill learning philosophical' *Transactions of Royal Society of London, Series B* 353: 1889-1901

Read, J. (2000) *Assessing Vocabulary* Cambridge, England: Cambridge University Press

Richards, J. (2000) 'Series Editor's Preface' in N. Schmitt *Vocabulary in Language Teaching* Cambridge: Cambridge University Press

Richards, J. C., and Rodgers, T. S. (2001) *Approaches and Methods in Language Teaching* New York: Cambridge University Press

Robins, R. H. (1967) *A Short History of Linguistics* Longman, London

Rudell, A. p. and Hua, J. (1997) 'The recognition potential, word difficulty, and individual Reading ability: on using event-related potentials to study perception' *Journal of Experimental Psychology,* 23, 4: 1170-1195

Ruhl, C. (1975) 'Kick the bucket is not an idiom' *Interfaces,* 4: 2-4

Samuels, S. J. and Flor, R. F. (1997) 'The importance of automaticity for developing expertise in reading' *Reading and Writing Quarterly,* 13: 107-121

Samuels, S. J. and Kamil, M. L. (1988) 'Models of the reading process' in Carrell, P., Devine, J. & Eskey, D. (ed.) *Interactive Approaches to Second Language Reading* Cambridge University Press

Schmidt, R. W. (1983) 'Interaction, acculturation, and the acquisition of communicative competence: a case study of an adult' in N. Wolfson & E. Judd (eds,) *Sociolinguistics and Language Acquisition,* Rowley, MA: Newburry House pp 137-174

Schmitt, N. (1999) 'The relationship between TOEFL vocabulary items and meaning, association, collocation, and word-class knowledge' *Language Testing,* 16: 189-216

Schmitt, N. (2000) 'Key concepts in ELT' *ELT Journal,* 54:4

Schmitt, N. and Carter, R. (2000) 'Lexical phrases in language learning' *The Language Teacher Online*

Schmitt, N. (2004) *Formulaic Sequences: Acquisition, Processing, and Use* (ed.) John Benjamin Publishing Company

Schmitt, N. and Carter, R. (2004) 'Formulaic sequences in action: An introduction' in N. Schmitt (ed.) *Formulaic sequences: Acquisition, Processing, and Use* John Benjamin Publishing Company pp 1-22

Schmitt, N., Dornyei, Z., Adolphs, and Durow, V. (2004) 'Knowledge and acquisition of formulaic sequences: A longitudinal study' in N. Schmitt (ed.) *Formulaic Sequences: Acquisition, Processing, and Use* Benjamin Publishing Company pp 55-86

Schmitt, N. and Underwood, G. (2004) 'Exploring the processing of formulaic sequences through a self-paced reading task' in N. Schmitt (ed.): *Formulaic*

Sequences: Acquisition, Processing, and Use Benjamin Publishing Company pp 173-189

Schweigert, W. A. (1986) 'The comprehension of familiar and less familiar idioms' *Journal of psycholinguistics Research*, 15; 33-45

Schweigert, W. A. and Moates, D. R. (1987) 'Familiar idiom comprehension' *Journal of psycholinguistics Research,* 17: 281-296

Scott, M. (1997) 'PC Analysis of Key Words and Key Words' *System*, 25, 2: 233-245

Scott, M. (1998) *WordSmith Tools 3* Oxford: Oxford University Press

Scott, M. (1999) *WordSmith Tools Users Help File* Oxford: OUP

Scott, M. and Tribble. C. (2006) *Textual Patterns: keyword and corpus analysis in language education*, Amsterdam: Benjamins

Segalowitz, N., Watson, V., and Segalowitz, S. (1995) 'Vocabulary skill: single-case assessment of automaticity of word recognition in a timed lexical decision task' *Second Language Research*, 11, 121-136

Segalowitz, N. (2000) 'Automaticity and attentional skill in fluent performance' in H. Riggenbach (ed.) *Perspectives on Fluency* Ann Arbor, MI: University of Michigan Press pp 200-219

Segalowitz, N. (2003) 'Automaticity and second language learning' in C. Doughty & M. Long (eds.) *The Handbook of Second Language Acquisition* Oxford, UK: Blackwell pp. 382-408

Shapira, R.G. (1978) 'The non-learning of English: case study of an adult' in E.M. Hatch (ed.) *Second language acquisition: a book of readings* Rowley, MA: Newburry House pp 246-255

Shepard, R. N. (1967) 'Recognition memory for words, sentences, and pictures' *Journal of Verbal Learning and Verbal Behavior,* 6: 156-163

Sinclair, J. McH. (1966) 'Beginning the study of lexis' in C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins *In memory of F. R. Firth* Longman

Sinclair, J. McH. (1987) *Looking Up: An Account of the COBUILD Project in Lexical Computing* London: Collins

Sinclair, J. McH. (1991) *Corpus, concordance, collocation* Oxford: Oxford University Press

Sinclair, J. McH. (1996) 'The search for units of meaning' *Textus IX*: 75-106

Sinclair, J. McH. (2004) *Trust the Text* Routledge: Taylor & Francis Group

Singleton, D. (1999) *Exploring the Second Language mental lexicon* Cambridge University Press

Smith, F. (1971) *Understanding Reading* Lawrence Erlbaum Associates, Publishers Hillsdale, New Jersey

Smith, F. (1998) *The Book of Learning and Forgetting* New York: Teachers. College Press

Spöttl, C. and McCarthy, M. (2003) 'Formulaic utterances in the multi-lingual context' in J. Cenoz, B. Hufeisen, and U. Jessner (eds) *The Multilingual Lexicon* 133-151 Dordrecht: Kluwer

Spöttl, C. and McCarthy, M. (2004) 'Comparing knowledge of formulaic sequences across L1, L2, L3, and L4 in Schmitt, Norbert(ed.) *Formulaic Sequences : Acquisition, Processing, and Use* Philadelphia, PA, USA: John Benjamins Publishing Company pp 191-225

Stahl, S. (1983) Differential word knowledge and reading comprehension. Journal of reading Behaviour, 15, 33-50

Stahl, S. A., & Fairbanks, M. M. (1986) 'The effects of vocabulary instruction: A model-based meta-analysis' *Review of Educational Research*, 5: 72-110

Stahl, S. A. & Nagy, W. E. (2006) *Teaching Word Meanings* Mahwah, NJ. USA: Lawrence Erlbaum Associates, Incorporated

Stanovich, K. E. (1980) Toward an interactive-compensatory model of individual differences in the development of reading fluency *Reading Research Quarterly,* 16, 1: 32-71

Stanovich, K. E. (1990) 'Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity' *Developmental Review*, 10: 72-100

Steinel, M., Hulstijin, J. Steinel, W. (2007) 'Second language idiom learning in a paired-associate paradigm' *Studies in Second Language Acquisition,* 29: 449-484

Stubbs, M. (1995) 'Collocations and semantic profiles: on the cause of the trouble with quantitative studies' *Functions of Language, 2: 1-33*

Swanson, H. L. and Berninger, V., (1995) 'The role of working memory in skilled and less skilled readers' comprehension' *Intelligence,* 21, 83-108

Swinney, D.A., & Cutler, A. (1979) 'The access and processing of idiomatic expressions' *Journal of Verbal Learning and Verbal Behaviour*, 18: 523-534

Tan, A., Moore, D. W., Dixon, R. S., and Nicholson, T. (1994) 'Effects of training in rapid decoding on the reading comprehension of adult ESL learners' *Journal of Bahavioural Education,* 4: 177-189

Titone, D. A. and Connine, C. M. (1999) 'On the compositional and noncompositional nature of idiomatic expressions' *Journal of Pragmatics,* 31:1655-1674

Tuinman, J. J. and Brady, M. E. (1974) 'How does vocabulary account for variance on reading comprehension tests? A preliminary instructional analysis' in P. Nacke (ed.) *Twenty-third National Reading Conference Yearbook*, Clemson, S.C.: The National Reading Conference

Underwood, G., Schmitt, N. and Galpin, A. (2004) 'The eyes have it' in N. Schmitt (ed.) *Formulaic sequences: acquisition, processing and use* Philadelphia. PA, USA: John Benjamins Publishing Company pp 153-169

Urquhart , S., and Weir, C. (1998) *Reading in a foreign language: Process, Product and Practice* New York: Longman

Van der Linden, E. (1992) 'Incremental processing and the hierarchical lexicon' *Computational Linguistics,* 18: 219-238

Van Lancker, D. (1987) 'Nonpropositional speech: neurolinguistic studies' in A. W. Ellis (ed.) *Progress in the Psychology of Language*, vol. 3, Hillsdale, NJ: Lawrence Erlbaum pp 49-118

Van Lancker, D. and Kempler, D. (1987) 'Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients' *Brain and Language* 32: 265-277

Wasow, Thomas, Ivan Sag and Geoffrey Nunberg (1983) 'Idioms: An interim report' in S. Hattori and k. Inoue (eds) Proceedings of the XIIIth International Congress of Linguistics Tokyo

Webelhuth, G. (1995) 'X-bar theory and case theory' in Webelhuth, G. (ed.), *Government and Binding Theory and the Minimalist Program* Blackwell, Cambridge, MA, pp 15-95

Wei, Y. (1999) 'Teaching collocations for productive vocabulary development' Paper presented at the annual meeting of the teachers of English to Speakers of other languages, New York pp 9-13

Weinert, R. (1995) 'The role of formulaic language in second language acquisition: a review' *Applied Linguistics*, 16, 2:180-205

Weinreich, U. (1969) 'Problems in the analysis of idioms' in J. Puhvel (ed.) *Substance and Structure of Language* Berkeley and Los Angeles University of California Press

Wible, D (2008) 'Multiword expressions and the digital turn' in Phraseology in F. Meunier and S. Granger (eds.) *Language Learning and Teaching* john Benjamins Publishing Company pp 163-186

Widdowson, H. G. (1989) 'Knowledge of language and ability for use' Applied Linguistics, 10, 2:128-137

Wilks, Y. (2005) 'REVEAL: the notion of anomalous texts in a very large corpus' Tuscan Word Centre International Workshop. Certosa di Pointignano, Tuscany, Italy, 31 June-3 July 2005

Williams, A. (1998) *Prefabricated Chunks in the Teaching of the Language of Negotiations*. Unpublished M.Ed Thesis. Manchester: University of Manchester

Williams, G. (2002) 'In search of representativity in specialised corpora: categorisation through collocation' *International Journal of Corpus Linguistics,* 7, 1: 43-64

Willis, D. (1990) *The lexical Syllabus* London: Harper Collins

Wong Fillmore, L. (1976) *The second time around: cognitive and social strategies in second language acquisition* Unpublished Doctoral Dissertation, Stanford University

Wong Fillmore, L. (1979) 'Individual differences in second language acquisition' in C.J. Fillmore, D. Kempler & S-Y.W.Wang (eds) *Individual differences in language ability and language behavior* New York: Academic Press pp 203-228

Wood, D. (2006) 'Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency' *The Canadian Modern Language Review/La Revue canadienne des langues vivantes,* 63,1: 13–33

Wray, A. (1992) *The focusing hypothesis: the theory of left hemisphere lateralized language re-examined* Amsterdam: John Benjamins

Wray, A. (1998) 'Protolanguage as a holistic system for social interaction' *Language and Communication,* 18: 47-67

Wray, A. (2000) 'Formulaic sequences in second language teaching: principles and practice' *Applied Linguistics,* 21, 4: 463-489

Wray, A. (2002a) *Formulaic language and the lexicon.* CUP

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' in N. Schmitt (ed.) *Formulaic Sequences: Acquisition, Processing, and Use* Publisher: John Benjamins Publishing Company pp 249-268

Wray, A. & Perkins, MR. (2000) 'The functions of formulaic language: an integrated model', *Language and Communication* 20,1:1-28

Yorio, C.A. (1980) 'Conventionalized language forms and the development of communicative competence' *TESOL Quarterly,* 14, 4: 433-442

Yorio, C. A. (1989) 'Idiomaticity as an indicator of second language proficiemcy' in K. Hyltenstam & L. K. Obler (eds.) *Bilingualism across the lifespan* Cambridge: Cambridge University Press pp 55-72

Zimmerman, C. B. (1997) 'Historical trends in second language vocabulary instruction' in J. Coady & T. Huckin *Second Language Vocabulary Acquisition* Cambridge University Press

Appendices

Appendix A: Reading Comprehension Test

Appendix B: List of Target Formulaic Sequences and their Frequencies in the BNC

Appendix C: Test of Recognition of Formulaic Sequences

Appendix D: Test of Understanding Formulaic Sequences

Appendix E: Examples of Problematic and Unproblematic Opaque and Semi-Opaque FSs

Appendix A Reading Comprehension test (See p. 86)

Read the following passages carefully and answer the questions in *Farsi*.

Passage 1

Resist the Official Pol-speak of Bush's 'War on Terror'
From 'Civilization' to 'WMD', Words are Weapons in the Global Crusade
Simon Tisdall

Second world war posters warning that "careless talk costs lives" embodied an enduring truth. Then the fear was that fifth columnists might overhear conversations of value to the Nazis. Elsie: "Can you take the kids to the park?" Herbert: "Sorry, ducks, it's D-day. I'll be invading Normandy all week." The equivalent US slogan was "loose lips sink ships". Sixty years on, in another era of conflict, the careless talk comes more often from politicians - but it is potentially just as deadly. When George Bush, soon after September 11, referred to a "crusade" against al-Qaida, he helped persuade Muslims that they were under renewed attack from Richard the Lionheart in a US navy bomber jacket. In the context of a mooted "clash of civilizations", Bush's loose use of language was not only insensitive. It was unthinkingly reckless.

Bush has avoided the word "crusade" ever since. But he still regularly talks about the need to defend "'Civilization'" and "the civilized world" against "dark forces". He never quite says which part of the planet is the "uncivilized" or "dark" bit. Perhaps he means Kandahar. Or Eastbourne. It is unclear. But the unspoken implication is deeply divisive, even racist, not to say insulting.

Given that words, spoken, written or broadcast, are our main form of communication, and given that words have such inherent potency, it is a wonder that today's sound-biting leaders are not more careful what they say. In short, they should mind their language.

Words can define how a people or a nation sees itself: the US declaration of independence is one obvious example. Yet modern-day Palestinians also see themselves engaged in a struggle for "independence" and "freedom" from external oppression. The US ignores such semantic paradoxes.

Or take last month's nuclear arms talks in Tehran. What mattered to the Iranians as much as enriching uranium, it transpired, was that national "sovereignty" and "dignity" be upheld. In such cases, words become symbols and benchmarks, as important, if not more so, than the "actions" (not words) that are ostensibly more significant.

Words such as "imperialism", "emancipation", "self- determination" and "liberation" define how history is scripted, how the future will be shaped, how contemporary conflicts are perceived and thus how they may or may not be resolved.

Terrorism is a salient case in point. In the abstract, "terrorism" is a terrible thing; everybody deplores it; nobody supports it. Why then is terrorism such a growth industry? Because its definition is not agreed. It depends where you stand. Terrorism has, thus, become a much abused, highly fungible term.

For Donald Rumsfeld, for example, the weekend helicopter attack at Fallujah was simply the work of "terrorists". That statement conceals a larger, unpalatable truth.

To the oppressed of the world, the men of violence are, variously, militants, freedom-fighters, guerrillas, insurgents, heroes, martyrs. The real terrorists belong to the "other side". Yet "state terrorism" is a concept that is barely recognized by the ostensible oppressors.

Which brings us back to Bush. By declaring an open-ended, global, no-holds-barred "war on terror", Bush invited every aspiring autocrat to do his worst in the name of "security" (another much-scandalized word).

In this loose-lipped, rapid-fire lingo, such people, whether killed or locked up in Bagram or Guantanamo or a thousand other hell-holes, are by definition "evil". Here, you might think, is another trip-wire for the unwary, to be sidestepped by sensible politicians in the secular, rational west. Not a bit of it.

Ronald Reagan denounced the "evil empire" of Soviet times. But Bush, author of the "axis of evil", Tony Blair and others have gone for linguistic broke. Not for them, it seems, any deep reflection on the moral connotations of their language. The world divides into biblical good and bad, black and white. Blessed are the peacemakers; battered are the "evil-doers".

Little wonder that General William Boykin, a leading Pentagon Christian soldier, could declare that the US was at war with Satan, that Muslims worshipped idols, and the only true god (not Florida or the supreme court) had picked Bush for president.

Italy's Silvio Berlusconi has strayed into the same Islam-denigrating territory. Malaysia's Mahathir Mohamad let slip his own counter-bigotry last month, claiming that Jews "rule the world by proxy". Such crass slurs are enough to make a philologist pine.

The latest addition to pol-speak, to the modern leader's essential lexicon, is WMD. This is now a universally understood term, or so you might think. WMD is proliferating, it's deeply frightening, and it's coming to a cinema or tube near you.

Yet totemic WMD is also a reason why civil liberties are everywhere under siege, why military budgets are rising, why the developing world is not developing, and why your opinion is ignored. In fact, WMD is a vague, non-specific term that can be (and is) used to cover a multitude of supposed sins. Developed countries have their own WMD, of course, but their arsenals are somehow deemed acceptable. Not so the WMD that may or may not exist in developing countries or "rogue states" (whatever that means). This species of unauthorized WMD is deemed destabilizing.

There are certain words, conversely, that the west's leaders gingerly eschew. These include "resistance" - too encouraging a label for the rag-taggle "remnants" opposing Iraq's emancipators, especially when used with a capital "R", as in French.

And then there is "occupation". Occupation, as in Iraq, is a no-go word; liberation is far preferable. Occupation makes it sound as if the US has barged uninvited into somebody else's country and refuses to go away. It makes Iraq sound like Palestine, Tibet, Afghanistan or, heaven forbid, Vietnam. That really is careless, ship-sinking talk.

Greater sense and sensitivity in use of language is required of politicians - and indeed the media. The urge to suppress arguably loaded words or phrases should as a rule be resisted as inimical to free expression and better understanding. As every spin doctor knows, acceptance of "official" terminology and definitions can amount to implicit endorsement of official policy. But the search for the right, exact mot juste requires constant awareness of possible ambiguity and politically and culturally charged, multiple meanings.

As ever, in all human discourse, there is truth and there is propaganda. As ever, it is important to be able to tell the difference. Before passing the ammunition, pass the word.

لطفا سوالات زیررا به فارسی پاسخ دهید.

1.موضوع اصلی پاراگراف اول چیست؟

-

-

2. سیاستمداران چگونه برای رسیدن به اهداف خود از واژه ها استفاده میکنند؟

-

-

3. در تعریف "تروریسم" چه مشکلی وجود دارد؟

-

-

4. اعلام جنگ بر علیه تروریسم از سوی بوش چه نتیجه ای داشت؟

-

-

5. تفاوت "WMD" در کشورهای پیشرفته و ممالک در حال پیشرفت چیست؟

-

-

6. سیاستمداران چگونه زندانی کردن افراد را توجیه می کنند؟

-

-

7. سیاستمداران در استفاده از واژه ها چه محدودیتی دارند؟ چرا؟

-

-

8. چرا سیاستمداران غربی از بکار بردن واژه هایی مثل "resistance" پرهیز می کنند؟

-

-

9. در مورد عراق از چه واژه های استفاده نمی شود؟ چرا؟

-

-

10. موضوع اصلی این متن چیست؟

English translation

1. What is the main point in the first paragraph?
2. How do politicians use words to achieve their goals?
3. What is it about the definition of 'terrorism'?
4. What did Bush's 'war on terror' cause?
5. What is the difference between 'WMD' in developed and developing countries?
6. How do politicians justify imprisoning people?
7. When using words, what should politicians do? why?
8. Why do western politicians avoid using words like 'resistance'?
9. What word is avoided in the case of Iraq? Why?
10. What is the general theme of the passage?

Passage 2

Their future isn't orange

While they were exploring the foothills at the border between south-west France and Spain, Enrico Coen and his friend Christophe Thébaud stumbled across a pretty problem, one that has created a mountain for evolutionary scientists to climb.

While driving in a valley, they passed an abundance of snapdragons, all coloured magenta. But near the town of Ripoll they spotted something fascinating, a natural floral boundary. In a hybrid zone hundreds of metres across, orange and whitish snapdragons thrived. Beyond them stretched acres of snapdragons of another colour, all yellow.

European snapdragons can be crossed to make hybrids and, as anyone who mixes paints will tell you, yellow plus pink is orange. Were the scientists witnessing the recent birth of a new, orange species, a blend of the genetic palette to create another member of the 28 or so Antirrhinum species that can be found in southern Europe? Or were the orange snapdragons the equivalent of mules, sterile crosses that could be born but not thrive?

This is just the kind of simple question that leads to new scientific insights. Prof Coen, Whibley and Langlade at the John Innes Centre near Norwich and Prof Bangham and Hannah of the University of East Anglia decided to study the snapdragon problem in collaboration with Dr Thébaud of the Université Paul Sabatier, Toulouse.

In the journal Science, they provided the answer with an explanation that ranges far beyond the world of florists and botanists. The work sheds new light on the mechanics of evolution and undermines an argument often used by Creationists. But there is bad news for evolutionary scientists, too. Their favourite metaphor for describing evolution is primitive and misleading.

The team found something remarkable around the floral boundary: there was plenty of mixing between genes of the yellow and magenta plants but much less mixing when it comes to the genes controlling pigment colours. The scientists concluded that the orange-flowered snapdragon was not a new species in the making but the equivalent of a mule, the dead-end cross of a donkey and a horse.

What was it about this particular colour? Orange varieties of flowers do, of course, exist: plant breeders have selected them. But it turns out that although an orange snapdragon can make seed, this colour is not very visible to bees, the insects which pollinate the plants and enable them to reproduce. This is not a problem for other orange flowers which appear bright at ultraviolet frequencies (which bees are sensitive to, unlike us), as is the case with poppies; or which are pollinated by humming birds (which can see orange.) But this is a problem for the snapdragons, which depend on bees.

Scientists have a powerful visual metaphor they use to express the problem facing orange blooms. They draw an abstract landscape, using a method that was unveiled in 1932 by an American geneticist, Sewall Wright. In his mind's eye he saw mathematical panoramas which reveal how thousands of genes in the genetic recipe of a plant or animal influence its success, an "adaptive landscape" of genetic possibilities.

One of the landscape's three dimensions could represent how much a snapdragon plant uses the magenta gene. The second could be the yellow gene. And the third, the vertical dimension, could represent how well plants with each combination of colours reproduce. That would mean pure magenta or yellow would produce towering mountains of success (or fitness, as scientists like to say), while blends that are doomed by lack of bee interest (orange) create troughs.

But this simple answer leads to another question. For a snapdragon to evolve between sporting the magenta flowers of Antirrhinum majus pseudomajus and the yellow ones of Antirrhinium majus striatum, as it undoubtedly did in the past, it would first have to produce less-attractive orange flowers.

That would mean the path of evolution taking the low road, through an unfit intermediate that bees would often ignore. That would go against the idea of the survival of the fittest, which says that only

the plants with the most attractive flowers are able to compete thrive. And that, in turn, would be great news for Creationists. If Darwin's theory of natural selection fails because critical intermediates – such as these elusive orange snapdragons – perform no function for selection to preserve then God Himself would have to intervene to help yellow become magenta and vice versa.

But in the case of snapdragons, colour is controlled by at least three genes: Rosea and Eluta, which affect the magenta pigment anthocyanin, and Sulfurea, which affects the yellow pigment, aurone. That means we need four dimensions (one for each gene and one for how well the various gene combinations reproduce), when Wright's metaphor is usually kept to three because that is all the human brain can easily cope with.

Today, there is a way to explore more dimensions. "There are computational methods for understanding and visualising high-dimensional problem," said Prof Bangham. Visualised in higher dimensions, a clear high road from magenta to yellow opens up, via paler varieties, without having to evolve through the lower orange forms that bees are less interested in, and without invoking the help of God.

"This is a totally different way of picturing evolution. The evolutionary possibilities are enormous and much richer than traditionally depicted," said Prof Coen. "We now understand how these plants can evolve to produce different colours whilst staying attractive to pollinating insects – we've found that colour is variable but constrained to a defined path." Beyond the snapdragon, this kind of study will also help understand how the process of evolution creates new species.

The message complements insights into evolution that have come from closer to home, on the St Kilda archipelago, the most remote part of the British Isles. There a team led by Dr Josephine Pemberton is counting, measuring and DNA fingerprinting its brown, white and black residents, the primitive Soay sheep.

The goat-like stocky animals are the descendants of a Bronze Age breed that has been resident on St Kilda for some 4,000 years. St Kilda was abandoned by the 36 remaining human residents in 1930, but not by the sheep. Research on them began in 1959 and now amounts to one of the most detailed population studies in the world, providing profound insights into evolution.

There are many examples of how, when isolated on islands, species shrink. And indeed the Soay sheep, which currently number about 1,800, are becoming smaller. The usual theoretical explanation is that islands have limited resources, so for the numbers of sheep to grow, the animals shrink.

But it is the bigger, fast growing, sheep that tend to survive St Kilda's bleak winters, which can see the population crash by up to 60 per cent (there are no natural predators.) "We know bigger sheep do better," said Dr Alastair Wilson, one of the team. "Evolution predicts they should be getting larger, yet they are not."

The effects of growth boosting genes are being masked in some way, perhaps because the sheep are changing their environment and food supply. But if that explains what is happening to the sheep, what about local mice? They are getting bigger, now half way towards being rat sized.

This does not mean Darwin got it wrong. What it does suggest is that Darwin's theory will only realise its full potential when translated from simple stories, metaphors and pictures into a mathematical form that can capture all the nuances, richness and complexity of the real world.

متن 2

لطفا سوالات زیر رابه فارسی پاسخ دهید.

1. منظور نویسنده از مسئله زیبا چیست؟

-

-

2. نتایج یافته ها چه تاثیری بر سایر علوم داشت؟

-

-

3. نتیجه تحقیقات دانشمندان چه بود؟

-

-

4. چرا "گونه های نارنجی " مساله ساز بودند؟

-

-

5. دانشمندان چگونه می خواستند مساله گونه های نارنجی را حل کنند؟

-

-

6. روش پیشنهادی چه مشکلی داشت؟

-

-

7. دانشمندان چطور مشکل را حل کردند و مزیت روش جدید چه بود؟

-

-

8. یافته های ضد و نقیض در رابطه با "گوسفند Soy" چه بودند؟

-

-

9. دانشمندان چه توجیهی برای این یافته ها دارند؟

-

-

10. موضوع اصلی این متن چیست؟

-

TEXT BOUND CLOSE TO THE SPINE IN THE ORIGINAL THESIS

English translation of comprehension questions

Passage 2 Their future isn't orange

1. What does the writer mean by the 'pretty problem'?
2. How did the results of the study on the 'pretty problem' affect other fields?
3. What did the scientists find as a results of their study?
4. Why did the 'orange varieties' cause problem?
5. How did the scientists try to solve the problem in the case of the 'orange varieties of flowers'? Explain.
6. What was wrong with their proposed method?
7. How did the scientists try to solve the problem associated with their method and what was the advantage of the new method?
8. What are the contrasting findings regarding Soay sheep?
9. What are the possible explanations for these findings?
10. What is the general theme of the passage?

Appendix B List of Target Formulaic Sequences and their Frequencies in the BNC (See p. 88)

1. When it comes to	1418	25. Turn out	1050
2. It is little wonder (that)	24	26. Do one's worst	6
		27. Dead end	105
3. Not to say	681	28. By definition	478
4. Survival of the fittest	66	29. Closer to home	96
5. Go against	100	30. Ever since	1651
6. As is the case with	57	31. A case in point	128
7. Not a bit of it	42	32. In the abstract	8
8. In the context of	1955	33. It is a wonder that	292
9. Invoke God's help	6	34. As a rule	15
10. Realize the full potential	6	35. Barge into	6
		36. Spin doctor	59
11. Halfway towards	5	37. Shed light on	15
12. In the case of	4720	38. Stumble across	6
13. Mountains to climb (a mountain to climb)	17	39. In one's mind's eye	128
14. In collaboration with	208	40. Lock up	357
15. In some way	1102		
16. Resolve the conflict	1330		
17. Declaration of independence	72		
18. It is this combination (of)	5		
19. Natural selection	516		
20. Far beyond	393		
21. Under siege	98		
22. A multitude of sins	16		
23. In the making	225		
24. Growth industry	26		

Appendix C Test of Recognition of Formulaic Sequences (See p. 90)

در جملات زیرکلمات و عبارات را از هم جدا کنید.

مثال: هنگام /عبور از خیابان/ جانب احتیاط/ را/ رعایت کنید.

In the following sentences separate words or lexical phrases using slashes.

Example: *Complete /the front of your script book / and / seat card /in block letters.*

1. But knowing such things isn't much help when it comes to shopping and eating.
2. Under such circumstances, it is little wonder that they experience difficulties.
3. To those who've never received million dollar royalty cheques, this sounded a little odd, not to say offensive.
4. Survival of the fittest was a superb natural mechanism which we damaged.
5. Changes are being made here which go against my principles and I cannot agree with them.
6. This would require farmers to apply for planning permission as is the case with other commercial or industrial developments.
7. Did he give up? Not a bit of it.
8. We are doing this work in the context of reforms in the economic and social spheres.
9. The great magicians of old always invoked their gods with sacrifice.
10. The school strives to treat pupils as individuals and to help each one to realize their full potential.
11. You are now half way towards real success.
12. Surgical training takes at least nine years, or eleven in the case of obstetrics.
13. 'We had a mountains to climb after the second goal went in', said Crosby.
14. Rock musicians are working in collaboration with an orchestra to create a new opera.
15. She's worried in some way about him, but I can't get anything out of her.
16. The United Nations are hoping to resolve the conflict quickly.
17. Drafted by Thomas Jefferson, the declaration of independence is at once the nation's most cherished symbol of liberty.
18. It is this combination of wit and political analysis that makes his article so readable.
19. What you did went far beyond what can be acceptable to the rule of law.
20. The reason this seems paradoxical is that we place so much emphasis on natural selection as the driving force of evolution.
21. Curtains around a bed can also cover a multitude of sins in patient care.
22. Meanwhile the British remained under siege in their base at Suez.

23. And who knows, there could be a future star in the making here.
24. Can mechanical engineering be a growth industry when few can name one significant UK- owned example?
25. Being back in the basement didn't turn out to be any good after all.
26. Nina does her worst to destroy the investigation.
27. But, despite the talent on show, the journey finished in a dead end in terms of goals.
28. Notes and coin by definition are totally liquid.
29. For this reason, we will not attempt an international analysis but focus closer to home on the situation in UK schools.
30. People have been telling me that ever since I was a small boy.
31. The US is once again a case in point, but the situation also arises in the majority of the decolonised third-world states.
32. I don't think if you just consider products or a market in the abstract that you can be ahead of your time.
33. There weren't many secrets in that neighbourhood and it is a wonder that Mud's own was kept for so long.
34. Contesting articulations of musical practices could as a rule now arise only at the level of consumption.
35. How dare you think you can just barge into my property!
36. Officially director of campaigns, image-maker or spin doctor he is credited with transforming the Party's fading image.
37. But a better understanding of how it works could also shed light on other questions concerned with the development of sensory nerve cells.
38. As often happens with medical advances, the scientists stumbled across the vaccine by chance.
39. How close was the picture that you painted in your mind's eye to the actual object?
40. However, he was not thrown out, he was taken back to the station and locked up for the night.

Appendix D Test of Understanding Formulaic Sequences (See p. 93)

In the following items, choose the best answer.

1. **If someone does their worst,** they ..... .
a. do something which they didn't intend to do
b. make lots of mistakes in doing something
c. think about unpleasant things that might happen
d. do everything unpleasant that they possibly can

2. A **spin doctor** is someone who is..... .
a. skilled in bone surgery
b. specialized in behavioral disorders
c. skilled in public relations
d. specialized in scientific research

3. If something has been the case **ever since** a particular time, it has been the case..... .
a. at regular intervals
b. increasingly
c. all the time from then until now
d. occasionally

4. A **dead end** job or course of action ...
a. does not finish before a specified date
b. leads to serious problems
c. does not lead to further developments
d. lead to success at the expense of others' lives

5. If someone or something **sheds light on** something, it ..... .
a. makes it start to shine
b. makes it easier to understand
c. makes it lighter to lift and carry
d. makes it start burning

6. If you say that something is **a case in point**, you mean that ..... .
a. you are not sure about something mentioned
b. a crime or mystery is being investigated
c. it is a good example of something mentioned
d. you are going to start talking about it

7. If you say that something covers or hides **a multitude of sins**, you mean that it conceals ..... .
a. the difference between a number of things
b. things by appearing to be much better than it really is
c. a behavior that is believed to break the laws of God
d. an action or behavior that people approve of

8. If someone or something is **under siege**, they are being ..... .
a. helped by other people
b. severely criticized by others
c. governed by others
d. kindly supported by their friends

9. If you **stumble across** something, you ..... .
a. you find it or discover it unexpectedly
b. destroy it on purpose
c. are confused as it is difficult to understand
d. tread on it unintentionally

10. When you talk about something **in the abstract**, you talk about it ..... .
a. in general
b. in secret
c. vaguely
d. in detail

11. If a topic of discussion is **close to home**, it is ..... .
a. accurate or connected with you in a way that makes you embarrassed
b. about an issue in which success or failure are equally possible
c. about something that seems to be obviously true
d. so interesting that you are very likely to be influenced by it

12. If something **turns out** to be a particular thing, it ..... .
a. is discovered to be that thing
b. stops being that thing
c. is widely believed to be that thing
d. stops moving towards it

13. if you see something **in your mind's eye**, you ..... .
a. stare at it
b. stop looking at it
c. imagine it
d. notice it

14. To **lock** someone **up** in a place means to ..... .
a. hide them there
b. prevent them from entering that place
c. put them in an unpleasant situation
d. put them in prison

15. When you say **it is a wonder that** something happens, you mean that .....
a. it is extremely good
b. you are satisfied with it
c. it is unexpected
d. you are not pleased

16. If someone **barges into** a place, they
a. rush into it in a rough way
b. enter it by force with an army
c. enter it in large numbers
d. discover it unexpectedly

17. If you say that something has a particular quality **by definition**, you mean that … .
a. it has this quality simply because of what it is
b. its quality is clear and distinct
c. its quality will not change
d. it has achieved this quality gradually

18. If you say that something happens **as a rule**, you mean that ….. .
a. it happens accidentally
b. it usually happens
c. it is possible in principle
d. it is controlled by regulations

19. If you describe a person or thing as something **in the making**, you mean that …….. .
a. they are going to be successful
b. they are going to be replaced by others
c. they are going to become known as that thing
d. they are going to improve

20. A **growth industry** is one which …… .
a. helps organizations develop their size and activity
b. is getting a better reputation
c. helps in the industrialization of a country
d. is increasing in size or activity

APPENDIX E Examples of Problematic and Unproblematic Opaque and Semi-Opaque FSs (See p. 113)

Unproblematic Sequences	Problematic Sequences
careless talk costs lives loose lips sink ships no-holds-barred lock up cover a multitude of sins barge into a place growth industry	fifth columnist in the abstract rapid-fire lingo trip-wire WMD spin doctor mot juste rag-taggle (wraggle-taggle) by definition