

T. C.
ULUDAĞ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

FORMULAIC SEQUENCES IN ENGLISH TV SERIES

(YÜKSEK LİSANS TEZİ)

Mustafa AKSAR

BURSA 2010

T. C.
ULUDAĞ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

FORMULAIC SEQUENCES IN ENGLISH TV SERIES

(YÜKSEK LİSANS TEZİ)

Mustafa AKSAR

Danışman

Yrd. Doç. Dr. Meral ÖZTÜRK

BURSA 2010

T. C.
ULUDAĞ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

..... Anabilim/Anasanat Dalı,
..... Bilim Dalı'nda numaralı
.....'nın hazırladığı ".....
....." konulu
..... (Yüksek Lisans/Doktora/Sanatta Yeterlik Tezi/Çalışması) ile ilgili tez
savunma sınavı,/...../ 20.... günü -saatleri arasında yapılmış, sorulan sorulara alınan
cevaplar sonunda adayın tezinin/çalışmasının(başarılı/başarısız) olduğuna
.....(oybirliği/oy çokluğu) ile karar verilmiştir.

Üye (Tez Danışmanı ve Sınav Komisyonu Başkanı)
Akademik Unvanı, Adı Soyadı
Üniversitesi

Üye
Akademik Unvanı, Adı Soyadı
Üniversitesi

Üye
Akademik Unvanı, Adı Soyadı
Üniversitesi

Üye
Akademik Unvanı, Adı Soyadı
Üniversitesi

Üye
Akademik Unvanı, Adı Soyadı
Üniversitesi

...../...../ 2010

ÖZET

Bu çalışma İngilizce dizi filmlerin, söz dizinleri (formulaic sequence) açısından zengin olup olmadığını ve bu filmlerde ne tür söz dizinlerinin daha sık kullanıldığını inceler.

Bu tezde veri olarak, fan sitesinde sitesinde 5.706.789 üyesi bulunan ve ülkemizde de izlenebilen “How I Met Your Mother” adlı dizinin konuşma metinleri incelenmiştir. Dizinin 4 sezonunu oluşturan 82 bölümünün İngilizce alt yazıları metin bankası olarak kullanılmıştır. Dizi metinleri internetten indirilerek bir metin bankası oluşturulmuş ve bu metin bankası üzerinde Thomas Michael COBB’a ait <http://www.lex Tutor.ca/> sitesi kullanılarak sıklık analizi yapılmıştır. Belirlenen liste içerisinde edat, zarf, fiilimsi gibi fonksiyon kelimeleri çıkarılarak geriye kalan içerik kelimelerinden sıklığı en yüksek olan 9 kelime seçilmiştir. Bu kelimelerin eşit sayıda fiil, isim ve sıfat olmalarına dikkat edilmiştir. Daha sonra bu kelimelere bağlamlı dizin analizi uygulanmış ve elde edilen veriler çeşitli yöntemlerle (sözlükler, anadili İngilizce olanlara danışma ve puanlayıcılar) hangi söz dizini türü olduğu tespit edilmiştir.

Bu sonuçlar ışığında, dizi filmlerin söz dizinleri bakımından zengin bir kaynak olduğu anlaşılmıştır. Buna göre, dizi filmler bir eğitim materyali olarak kullanılabilir. Ayrıca, incelenen dizi filmlerde en sık kullanılan söz dizini türünün kalıplar (collocation) olduğu tespit edilmiştir. Hâlihazırda bu çalışma başka diziler üzerinde de uygulanabilir; buna karşın aynı sonuçlar çıkmayabilir.

Anahtar kelimeler : Söz dizini, dizi film, kalıp, metin bankası, bağlamlı dizin analizi

ABSTRACT

This study aims to investigate to what extent formulaic sequences are used in English TV series and which type of formulaic sequence is most frequently used.

This study uses scripts of “How I Met Your Mother” which has 5.706.789 fans in a forum website and can be seen in our country. Scripts of 82 episodes of 4 seasons were used as corpus. Having downloaded the scripts of the episodes, a corpus was formed and a frequency analysis was applied on it by using a website, <http://www.lex Tutor.ca/>, which belongs to Michael Cobb. Function words such as prepositions, pronouns, and auxiliaries were extracted from the list gathered via frequency analysis and the most frequent 9 content words were picked. It was noted that these words included verbs, nouns, and adjectives equally. Afterwards, a concordance analysis was applied the corpus and the target words were analyzed. The results also classified according to FS types by using various check instruments such as dictionaries, native-speaker consultation and inter-rater reliability.

In the light of these results, it has been found out that English TV series are rich in formulaic sequences. Accordingly, it can be asserted that English TV series can be used as an educational material. Furthermore, it has been detected that the most frequent formulaic sequence type in the corpus is “collocation.” The current study might be adapted to different TV series; however, it should be noted that the results might not be the same.

Keywords: Formulaic sequence, TV series, collocation, corpus, concordance

ACKNOWLEDGEMENTS

I owe deepest gratitude to those who had helped me complete this thesis. First and foremost, I would like to give my genuine thanks to my thesis advisor, Asist. Prof. Dr. Meral ÖZTÜRK, for her invaluable suggestions, deep interest, endless assistance, patience, and motivating attitude throughout this thesis process.

I owe much to my colleagues who have been a great support to me throughout this study especially for their invaluable contributions as inter-raters.

For his guidance, assistance and encouragement, I owe special thanks to my wise brother Habib without whom I could never attempt to such a work.

My greatest gratitude is extended to my wife, Nuriye, who has faith in me and has been a great support to me throughout this study from the beginning to the end. I am grateful for her everlasting love, caring, patience, and encouragement. Without her, I would not have completed this thesis successfully.

TABLE OF CONTENTS

Özet	i
Abstract.....	ii
Acknowledgements.....	iii
Contents.....	iv
List of Tables.....	v
List of Figures.....	vi
Abbreviations.....	vii

CHAPTER I

INTRODUCTION

1.1. Background of the Study.....	1
1.2. Statement of the Problem.....	3
1.3. The Aim of the Study.....	3
1.4. Research Questions.....	4
1.5. Assumptions and Limitations of the Study.....	4

CHAPTER II

LITERATURE REVIEW

2.1. Introduction.....	5
2.2. Naming Formulaic Sequence.....	5
2.3. Detection of Formulaic Sequences.....	9
2.3.1. Intuition.....	9
2.3.2. Frequency.....	11
2.3.3. Structure.....	11
2.3.4. Fixedness.....	12

2.3.5. Fluency, stres and articulation.....	12
2.3.6. Corpus and concordance.....	13
2.4. Taxonomy.....	17
2.4.1. Collocation.....	18
2.4.2. Fixed phrase.....	18
2.4.3. Sentence stem.....	19
2.4.4. Idiom.....	20
2.5. Significance of Formulaic Sequences.....	20

CHAPTER III

METHODOLOGY

3.1. Introduction.....	25
3.2. Corpus.....	25
3.3. Target Words.....	25
3.4. Analysis.....	29
3.5. Taxonomy.....	32
3.5.1. Collocation.....	32
3.5.2. Idiom.....	34
3.5.3. Fixed phrase.....	34
3.5.4. Sentence stem.....	35
3.6. Reliability Checks.....	36
3.6.1. Dictionaries.....	36
3.6.2. Inter-rater reliability.....	37

3.6.3. Native speaker consultation.....	40
CHAPTER IV	
RESULTS AND DISCUSSION	
4.1. Introduction	41
4.2. English TV series as a Formulaic Sequence Source.....	41
4.3. The Most Frequent FS Types in English TV series.....	45
CHAPTER V	
CONCLUSION	
5.1. Introduction.....	49
5.2. Conclusions.....	49
5.3. Limitations of the Study.....	49
5.4. Suggestions for Further Research.....	50
REFERENCES	51
APPENDIX	58
Appendix Criteria for FS Types.....	58
VITAE	60

LIST OF TABLES

Table 2.1	Tognini-Bonelli's Contrastive Outline of Text and Corpus	15
Table 3.1	Inter-rater analysts' and researcher's taxonomy means.....	40
Table 4.1	Formulaicity in target word patterns.....	42
Table 4.2	Formulaicity rate of target word types.....	44
Table 4.3	Formulaicity rate of target word types without the words whose rate below 30% (do, new and girl).....	45
Table 4.4	FS type frequency.....	46
Table 4.5	Word type-FS type relationship.....	47

LIST OF FIGURES

Figure 2.1	Terms used to describe aspects of formulaicity	6
Figure 2.2	The functions of formulaic sequences.....	22
Figure 2.3	Schema for the use of formulaic sequences in serving the interests of the speaker.....	23
Figure 3.1	Screenshot of frequency link at the website.....	26
Figure 3.2	Screenshot of choosing language at the website.....	26
Figure 3.3	Screenshot of submitting the corpus.....	27
Figure 3.4	Screenshot of a sample frequency output.....	27
Figure 3.5	Screenshot of a sample of the most frequent 25 words.....	28
Figure 3.6	Screenshot of a sample concordancing at the website.....	29
Figure 3.7	Screenshot of a sample text-based concordances.....	30
Figure 3.8	Screenshot of a sample submitting corpus.....	30
Figure 3.9	Screenshot of a sample concordance output of a text.....	31

ABBREVIATIONS

SLA : Second Language Acquisition

FS : Formulaic Sequence

NS : Native Speaker

NNS : Non-native Speaker

L1 : First Language

L2 : Second Language

etc. : Addecetera

e.g. : Exempli gratia

ibid : Ibidem

et al : Et alii

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Wilkins (1972) states that: “Without grammar very little can be conveyed; without vocabulary nothing can be conveyed.” In order to be able to use target language, even natives, people need vocabulary to be able to build a proper communication. Vocabulary load of an individual makes him/her express him/herself more easily. For that reason, vocabulary acquisition, in recent years, has been specifically emphasized in all SLA methods to a great extent. Wray (2002) asserts that “although we have tremendous capacity for grammatical processing, this is not our only, nor even our preferred, way of coping with language input and output. On the other hand vocabulary knowledge is enough to an extent. Wray (2000) suggests that knowing individual words to know a language is not enough; a learner must also know how they fit together. Researchers found out that vocabulary mostly comprised of fixed and semi fixed recurrent clusters instead of separate words. These recurrent clusters have lots of names in many fields such as linguistics, sociolinguistics, psycholinguistics, applied linguistics, pragmatics, phraseology, lexicography, corpus linguistics, first and second language acquisition, language teaching, neurolinguistics. In this study, formulaic language term will be preferred as Wray (2002) and Schmitt and Carter (2004) preferred in their studies.

However, it is steadily becoming more difficult to rule the whole vocabulary. Therefore the vocabulary to be taught should be practical and easy to teach and learn beside being easily storable in the brain. Assuming that FSs are “glued together” and stored as a single “big word” as defined by Ellis (1996), it can be asserted that a formulaic sequence covers the same space in brain as a single word. Miller (1956), Bower (1969) and Simon (1974) argue that chunking information into single complex units increases the overall quantity of material that can be stored in short-term or working memory. Ellis and Sinclair (1989) note that a person’s phonological memory

span correlates with his or her language learning capacity. Formulaic sequences are ubiquitous in language use (Nattinger and DeCarrico, 1992) and they make up a large proportion of any discourse. There have been several studies calculating the proportion of FS in language. Erman and Warren (2000) figured out formulaic sequences of various types make up 58,6% of the spoken English discourse they analyzed and 52,3% of the written discourse. Likewise, Foster's raters calculated that 32,3% of the unplanned native speech they analyzed were formulaic (Foster 2001). The results prove that FSs are really widespread in language but the exact number is still unknown and it is probably difficult to be known as Wray (2002) states: "store of formulaic sequence is dynamic and is constantly changing to meet the needs of the speaker. Additionally, Nattinger and DeCarrico (1992) believe that the research is too thin on the ground to truly know the extent of their use. On the other hand Pawley and Syder (1983) claims that sentence-length expressions familiar to the ordinary, mature English speaker probably amounts, at least, to several hundreds of thousands.

Although FSs have a great influence in language, they have their own usage. As Bishop (2004) points it is difficult to learn FSs due to some reasons. Therefore many scholars did their studies on native speakers to calculate the proportion of FS in language since natives are good at them. In this respect, native speaker-origin source will be helpful for nonnative speakers. So, it can be asserted that corpora formed by using native speaker language are a good source of exposition of nonnative speakers to the authentic language. With the rising technology corpus studies have become a good source for researchers despite some scholars' critics. Today, there are a number of corpora which were formed for different reasons such as British National Corpus (BNC) indicating how often the sequences occurred in general English, Cambridge and Nottingham Corpus of Discourse in English (CANCODE) indicating how frequent sequences were in spoken discourse and Michigan Corpus of Academic Spoken English (MICASE) indicating sequence frequency in academic spoken discourse. Also, all of the major international ESL dictionaries are now corpus-based.

1.2. Statement of the problem

Formulaic language incorporating idioms, proverbs, and sayings, constitute a very significant portion of communication in English language (Schmitt and Carter 2004). While this situation doesn't create a problem for NSs, due to the deficiency of exposition to the language it creates a problematic situation for NNSs. Since FSs constitute a very significant portion of communication in English language (Wray 2000) and L2 learners are less exposed to English, they notice FSs less and fail to learn them as efficiently as single words (Adolph and Durov 2004). According to Wray (2002), a non-native can only learn to prefer those which are the usual forms in a given speech community by observation and imitation. In the same vein, the formulaic language of L2 learners tends to lag behind other linguistics aspects (Irujo 1993). This may be partly due to a lack of rich input: Irujo (1986) suggests that idioms are often left out of speech addressed to L2learners. Kuiper (2004) asserts that full mastery often take years of NNSs.

Today media is in the center of people's lives. It can somehow shape world people. As argued by Connell, Bridgley, & Edwards (1996/1999) "no generation has a bigger media history because no previous generation has had access to so many different kinds of media and such a range of media products" Among the mass media, it is clear that television has the greatest significant and continued impact on our present culture (Signes, 2001). When considered from this point, native media products might be a good source for learners who lack exposure to the target language. In that case, English TV series are a good formulaic source for NNSs. The convenience of English TV series as a FS source will be tested in this study.

1.3. The Aim of the Study

This study aims to determine whether English TV series are a good source for Formulaic Sequence. In other words, it is aimed to detect how frequent formulaic sequences are used in English TV series. Furthermore, the study also explores what types of FS are most frequently used in English TV series.

1.4. Research Questions

This study aims to answer these following research questions:

- 1) Are English TV series rich in formulaic sequence?
- 2) What type of FS is frequently used in English TV series?

1.5. Assumptions and Limitations of the Study

In addition to the general critics for corpus-derived studies, it can also be criticized with this study that the corpus of this study is neither natural nor artificial. Furthermore, this study's replication won't give the same results due to some subjective reasons explained in the study.

CHAPTER II

LITERATURE REVIEW

2.1. Introduction

This chapter is intended for reviewing previous research and accumulated literature on all aspects of formulaic sequences including their definitions, significance in English language and FS verifying tools. Additionally, corpus-driven studies and tools and some critics about these are also described in this chapter.

2.2. Naming Formulaic Sequence

Formulaic speech traditions may well be as old as storytelling and doing politeness (Brown and Levinson, 1987; Ferguson, 1976). One of the first studies on formulaic sequences was conducted by Milman Parry and Albert Lord in the 1930s and 1940s as they searched for explanations as to how Homer, blind and illiterate, could have created two of the great founding texts of Western literature. Lord's (1960) book *Singer of Tales*, the result of this pioneering field work, made a considerable impact in literary scholarship because it opened a new way of looking at oral traditional literature. It was even suggested that whole cultures might be influenced by the ways in which linguistic traditions are carried: either orally, or both orally and by means of writing (Ong, 1982). This way of thinking has been influential in many areas of research such as folklore (Foley, 1990; Jackson, 1988), cultural anthropology (Edwards and Sienkewicz, 1990), and literary studies (Foley, 1995), but it has had little impact on linguistics (Kuiper 2004). Lord deals with formulaicity in his study in psycholinguistic and socio-linguistic senses. However he didn't intend to draw a clear picture of FSs.

With the advent of vocabulary studies, many scholars began to study on "glued together" (Ellis 1996) words. However, each of them used various terms for these word strings. They adopted a term just for their study and defined it roughly as George Miller (1956) did." A "chunk," *roughly speaking*, is a structured set of information that has a

single address in memory. So, this field owned a huge amount of terminology which often creates ambiguity among researchers. Wray sums up this situation as:

Both within and across subfields such as child language, language pathology and applied linguistics, different terms have been used for the same thing, the same term for different things, and entirely different starting places have been taken for identifying formulaic language within data. (Wray 2002)

This plethora of terms sometimes results in cross-fields ambiguities. This is because while some terms define the same thing, some of them might refer to a different concept belonging to a different field. That is, while labels vary, it seems that researchers have very much the same phenomenon in mind (Weinert 1995). Wray (2002) criticizes her colleagues so: "...all of which have something useful to say, but none of which seems fully to capture the essence of the wider whole." Schmitt and Carter (2004) states the same situation as "With this diversity in mind, it is little wonder that different researchers have looked at formulaic sequences and seen different things, resulting in a variety of terminology to express various perspectives." (Schmitt and Carter 2004).

amalgams – automatic – chunks – clichés – coordinate constructions – collocations – complex lexemes – composites – conventionalized forms – F[ixed] E[xpressions] including I[dioms] – fixed expressions – formulaic language – formulaic speech – formulas/formulae – fossilized forms – frozen metaphors – frozen phrases – gambits – gestalt – holistic – holophrases – idiomatic – idioms – irregular – lexical simplex – lexical(ized) phrases – lexicalized sentence stems – listemes – multiword items/units – multiword lexical phenomena – noncompositional – noncomputational – nonproductive – nonpropositional – petrifications – phrasemes – praxons – preassembled speech – precoded conventionalized routines – prefabricated routines and patterns – ready-made expressions – ready-made utterances – recurring utterances – rote – routine formulae – schemata – semipreconstructed phrases that constitute single choices – sentence builders – set phrases – stable and familiar expressions with specialized subsenses – stereotyped phrases – stereotypes – stock utterances – synthetic – unanalyzed chunks of speech – unanalyzed multiword chunks – units

Figure 2.1. Terms used to describe aspects of formulaicity (Wray 2002)

Until Wray's book, *Formulaic Language and Lexicon*, no scholar could make a comprehensive definition of formulaic sequences. Observing the mess in FS terminology field, Wray (2002) needed to a term which doesn't carry previous baggage, and which can be clearly defined. In this respect, Wray created a definition which is more comprehensive than earlier definitions although it has still some problematic issues which will be discussed later. He preferred using formulaic sequence as a term. Wray (ibid) explains formulaic is a term carrying its associations of "unity" and of "custom" and "habit", while sequence indicates that there is more than one discernible internal unit, of whatever kind.

Wray's comprehensive definition which is also still valid today is as follows:

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

Although Wray makes an effort to create a definition as inclusive as possible, his definition is criticized in some respects. Some scholars even criticize the necessity for a definite definition of formulaic sequence. They think it is too difficult to define formulaic sequence due to their diversity. It is still problematic among researcher what criteria should be considered identifying a formulaic sequence. First of all, it is still a problematic issue: Which come first: definition or identification? Wray (2002) suggests that identifying something obviously relies on how you define it. However, the relationship between definition and identification is circular: in order to establish a definition, you have to have a reliable set of representative examples, and these must therefore have been identified first. Wray's definition is criticized from two aspects. One of them is that how we can know a sequence is stored and retrieved whole (holistic) from memory at the time of use. Read and Nation (2004) find it challenging because the means of storage and retrieval of the same sequence can differ from one individual to another, and can differ from one time to another for the same individual depending on a wide range of factors such as changes in proficiency, changes in processing demands, and changes in communicative purpose. On the other hand, Underwood, Schmitt and Galpin (2004), in their study calculating eye movement while

reading formulaic sequence-oriented text, came to that conclusion: “We now have evidence that the terminal in formulaic sequences are processed more quickly than the same words when in nonformulaic contexts. This provides evidence for the position that formulaic sequences are stored and processed holistically. Schmitt and Underwood (2004), in another study, stand behind their assertion by adding that the mind is able to predict the end words of the sequence from the previous words in the sequence. Due to the need for verifying that FSs are stored holistically, researchers conducted many studies on this issue. From these researchers, Spöttl and McCarthy (2004), in their study comparing knowledge of formulaic sequences across L1, L2, L3, and L4, conclude that where formulaic sequences are processed holistically, it seems that they can be transferred holistically across L1, L3 and L4, albeit by individually determined strategic and linguistic routes. A second critic for Wray’s definition is how it is known that a sequence is prefabricated. Some researchers tried to explain this deficiency via fluency and hesitations (Underwood, Schmitt and Galpin 2004; Schmitt and Underwood 2004).

Kuiper (1996) tries to define formulaic sequences with their properties. According to him they are not merely strings of words, but phrases and that they are lexical items like words. Schmitt and Carter (2004) also list properties of formulaic sequences:

‘Formulaic sequences appear to be stored in the mind as holistic units, but they may not be acquired in an all-or nothing manner’.

‘Formulaic sequences can have slots to enable flexibility of use, but the slots typically have semantic constraints’.

‘Formulaic sequences can have semantic prosody.’

‘Formulaic sequences are often tied to particular conditions of use.’

Conklin and Schmitt (2007) argue that formulaic sequences are more important than being strings of words attached together with collocational ties.

To sum up, in order to be able to define or determine formulaic sequences, clearly identification of formulaic sequences plays a vital role.

2.2 Detection of Formulaic Sequences

Bishop (2004) asserts that formulaic sequences have no clearly delineated boundaries. While native speakers acquire formulaic sequences inductively, the situation gets worse for non native speaker due to some reasons, mainly lack of exposure. According to Schmitt *et al* (2004), it is probable that most of the less transparent formulaic sequences were acquired through exposure. Another drawback of NNSs is similarity of formulaic sequences to non formulaic words. Howarth (1998) argues L2 learners' problems with formulaic sequences are attributable to "a lack of awareness of the phenomenon". Wray (2000) suggests that learners are at a disadvantage when "trying to express ideas idiomatically" because of difficulty of distinguishing formulaic sequences from the plethora of nonformulaic word combinations which can be generated from individual words. Because of such reasons, a framework for formulaic sequences, at least for the ones in target, is needed.

As Wray (2002) stated before, in order to establish a definition, a reliable set of representative example must be identified first. Wray (*ibid*) determines two basic ways in which formulaic sequences can be collected. One is to use an experiment, questionnaire or other empirical method to target the production of formulaic sequences as data. The other is to collect general or particular linguistic material and then hunt through it in some more or less principled way, pulling out strings which, according to some criterion or group of criteria, can justifiably be held up as formulaic. Since the first is problematic in the view of authenticity, the latter will be explained.

2.2.1. Intuition

In spite of being rather subjective and quite little scientific, many scholars accept intuition as a tool for detecting formulaicity (Wray 2002; Read and Nation 2004; Moon 1998, Erman and Warren 2000; Schmitt, Grandage and Adolphs 2004). Wray asserts that formulaicity might vary from person to person. "There are strings which are formulaic for a particular speaker, because he or she has fused and stored them: these are not necessarily a chunk in anyone else's lexicon" (Wray 2002). Accepting that

intuition is dubious from modern “scientific” perspective, Read and Nation (2004), set some conditions to accept “intuition” scientific.

- a definition of what is meant by a formulaic sequence is carefully formulated in advance, as previously discussed.
- the investigator communicates the definition to a second person, who then attempts to replicate the investigator’s identification of the formulaic units.
- instead of relying on the researcher’s judgement, a panel of judges is formed to analyse the database and a multiword unit is accepted as formulaic only when most, if not all, the judges identify it as such.

Wray accepts intuition as a scientific research; at least as a guide, even she herself criticize its being used as only criteria. Even where some other measure is primarily in use, intuition still tends to guide the design of experiments, the interpretation of results and the choice of examples used in the published reports (Wray 2002).

There are surely so much critic on accepting intuition scientific even by Wray. Corpus linguists such as Sinclair (1991) argue that their research reveals intuition to be a very fallible means of investigating the facts of language use, with regard to the relative frequency of linguistic features, typical meanings of lexical items, characteristic patterns of collocation, and so on. Secondly, in the context of second language acquisition research, the native speaker intuitions of the researcher are often brought to bear to account for the language production of learners, who may or may not have an intuitive basis for what they say or write in the second language. This means that the formulaic status of sequences in learner language is even more difficult to establish by means of intuition than in the case of native speaker production. A third difficulty identified by Wray (2002) is that recognition of formulaic language may depend on the shared knowledge which comes from membership of a particular speech community rather than being universal among users of the language concerned. This represents just one more limitation on the value of intuition as an investigative procedure.

2.2.2. Frequency

Frequency-based data can be gathered from the corpora which were specially formed. In this field, corpus linguistics, computer searches are conducted to establish

the patterns of distribution of words within text. This is done on the basis of frequency counts, which reveal other words a given target word most often occur with. This frequency is, consistently, far from randomness. This might be assessed as a criterion for calling a string “formulaic”. Wray (2002) explains its reasoning “...the more often a string is needed, the more likely it is to be stored in prefabricated form to save processing effort, and once it is so stored, the more likely it is to be the preferred choice when that message needs to be expressed”. It is also inferred from a frequent word string that the node of the sequence is inclined to collaborate with other words. Sinclair and Renouf (1998) observe “the more frequent a word is, the less independent meaning it has, because it is likely to be acting in conjunction with other words, making useful structures or contributing to familiar idiomatic phrases”

On the other hand, just in preface and acknowledgment of the her book, Wray declares that “those who place their faith in frequency counts as the only valid arbiter of formulaicity will not welcome my call for the reinstatement of native-speaker intuition as the best witness to the part of our lexicon which we use with most creative flexibility.” Another criticism on frequency criterion is about infrequent formulaic sequences. Some formulaic sequences might be infrequent and show no flexibility. According to Moon (1998) *kith* occurs only in *kith and kin*, and *dint* only in *by dint of*.

2.2.3. Structure

Some researchers studied on formulaic sequences on the basis of their form. There are two possible ways to detect FS on form basis. The first and the least useful of them is to define formulaic sequences as the set of multiword strings listed in a particular dictionary (e.g. Kerbel and Grunwell 1998; Moon 1998). More productive is criteria deriving from empirical investigation which is partly applied in this study while determining the criteria of FS types. Butler (1997), on the basis of his frequency-based exploration of Spanish text, notes that “the majority of the longer repeated sequences ...begin with conjunction, articles, pronouns, prepositions or discourse markers. Such an analysis was conducted in this study on detecting sentence stems. For instance,

according to a noteworthy frequency of “*time*”, it has a formulaic form having slots which can be used with various different words:

It *be (is,was)* (*high*) *time* *to... .*

2.2.4. Fixedness

If any component of the sequence cannot be extracted or substituted, it means that that sequence is fixed. This criterion has been used for fixed phrases and idioms in this study. However, this fixedness is limited to some extent. Wray (2002) argues that only a small subset of formulaic sequences are entirely fixed: those which are not, legitimately permit insertions. Indeed, the fixedness criterion does not sit well with the existence of semi-fixed sequences, which contain slots for a variety of compulsory and optional material to be inserted. Read and Nation (2004) accept non-compositionality and fixedness as two mostly recognized structural criteria.

2.2.5. Fluency, stress and articulation

Features such as overall fluency intonation pattern and changes in speed of articulation are all potential pointers to a stretch of prefabricated material (Pawley & Syder 2000). The identification of FSs is, of course, restricted to spoken language. In written language, punctuation might be criteria to some extent, though. Pawley asserts that pauses with lexicalized phrases are less acceptable than pauses within free expressions, and after a hesitation the speaker is more likely to restart from the beginning of the expression.

Fluency might be a well criterion for detecting FS. Wray (2002) suggests that the speaker could directly benefit from using prefabricated material as a means of reducing his or her processing load. Pawley and Syder (1983) assert that formulaic sequences offer processing efficiency because single memorized units, even if made up of a sequence of words, are processed more quickly and easily than the same sequences of words which are generated creatively. This assertion is supported by evidence from Kuiper (1996) and his colleagues (Kuiper and Haggio, 1984), who show that ‘smooth talkers (auctioneers, sportscasters) use formulaic language a great deal in order to

fluently convey large amounts of information under severe time constraints. Underwood, Schmitt and Galpin (2004) demonstrate that words, when they are part of formulaic sequences, are read more quickly than the same words when embedded in non-formulaic text. Hesitations and pauses, of course, aren't unnecessary. Moreover they are less acceptable within formulaic sequences. Wray (2002) states that "if a formulaic string is treated as a single, holistic unit, it ought to be relatively resistant to internal dysfluency and inaccuracy. Therefore, we can make the prediction that there would be far fewer pauses and errors within formulaic strings than between them". Additionally, some researchers (Lord 1960; Kuiper 2004) studies on race callers', story tellers', auctioneers' and radio sport commentators' speech in order to detect formulaic sequence influence on their fluency. The results have shown that almost all of them gain a great skill at formulaic sequences.

2.2.6. Corpus and concordance

Corpus research has been immensely useful in applied linguistics in numerous ways. It has added a powerful new tool to the range of procedures available for the study of formulaic sequences. It has allowed the compilation of dictionaries which better represent the way words are used, and all of the major international ESL dictionaries are now corpus-based (Schmitt, Grandage and Adolphs (2004). Francis and Sinclair (1994) have argued that "corpus data provides us with incontrovertible evidence about how people use language" and that the corpus data gives the opportunity to examine very quickly more language than one is likely to encounter in a lifetime. Lewis (2000) agrees with Francis and Sinclair, "Corpus linguistics and computer corpora are powerful tools, and regularly produce new, and unquestionably better, descriptions of English than we have ever had before. Corpora not only demonstrate that non-canonical forms abound in language; they also allow these forms to be analyzed and classified (Moon, 1998; Philip 2003). Corpora have been consulted to provide *descriptive* rather than *prescriptive* grammars of English (Biber *et al.*, 1999; DeCarrico and Larsen-Freeman, 2002; Carter and McCarthy, 1988). Corpus analysis has also done much to increase our understanding of the phenomenon that, in English (and perhaps most/all languages?), speakers tend to use the same clusters of words over

and over again (e.g. Sinclair, 1991; Cowie, 1998; Moon, 1998). Also it should be noted that nearly all corpora are compiled from authentic language of various types, which real people have produced.

Sinclair (1991) defined corpus as a collection of texts of naturally-occurring language compiled to identify the characteristics of a state or variety of a language. According to Kennedy (1998), corpus linguistics is “based on bodies of text as the domain of study and the source of evidence for linguistic description and argumentation”. Conrad (2000) defines corpus linguistics as “the empirical study of language relying on computer-assisted techniques to analyze large, principled databases of naturally occurring language”. Tognini-Bonelli (2001) takes corpus as “a computerized collection of authentic texts, amenable to automatic or semi-automatic processing or analysis”. In her view, since the texts included in a corpus do not lose their textual identity and the original source of the given language is accessible upon demand, issues such as text typology and register can easily be studied. Partington (1998) lists style and authorship, historical, translation, register, lexis, syntax, text, and spoken language studies as well as lexicography as some of the main uses to which corpora are put.

Corpora aren't unique tool to detect formulaic sequence. The quantitative evidence supplied by the software needs to be evaluated by the application of human judgment to determine which of the word sequences are formulaic — and if a classification system is involved, which ones fit in which categories.(Read and Nation 2004)

According to Frankenberg and Garcia (2005), a concordance is a list of occurrences of a given word, part of a word, or combinations of words, together with their contexts, within a corpus of text. Concordance software can be used to find collocational clusters in corpus data. The most flexible software allows the researcher to specify a search word or words and to gather and count the occurrences of collocates for several positions on either side of the search node. Such software is an extremely valuable tool for research on formulaic language. However, it is essential for the

researcher to examine each instance of the data to make sure that it is relevant. Clearly, valid cluster analysis requires manual checking of the data.

Another limitation of concordance software is that it can automatically locate only contiguous sequences. In order to locate non-contiguous ones, it is necessary for the researcher to enter in the search request either a contiguous subpart of the whole sequence or at least one key lexical component of it (Read and Nation 2004).

Another criticism directed at corpus research is that it is deprived of the original context of the communication. Aston (2001) also agrees that concordancing analysis cannot be said to address learners' need to be involved in the negotiation of meaning in its pragmatic aspects, which are deemed by Widdowson to be of vital importance to language acquisition, since instances of language are indeed decontextualized from their original communicative settings.

Tognini-Bonelli (2001) provides a useful outline of the series of contrasts between her understanding and text:

Table 2.1 Tognini-Bonelli's Contrastive Outline of Text and Corpus

A TEXT	A CORPUS
Read whole	Read fragmented
Read horizontally	Read vertically
Read for content	Read for formal patterning
Read as a unique event	Read for repeated events
Read as an individual act of will	Read as a sample of social practice
Instance of <i>parole</i>	Gives insight in <i>language</i>
Coherent communicative event	Not a coherent communicative event

Hunston (2002) classifies modern corpora in the electronic medium into eight discernible types according to purpose of building corpora: *specialized corpora* which contain texts that aim to be representative of a specific type of text such as newspaper editorials, lectures, academic articles in a particular subject, student essays or conversations [e.g. the Michigan Corpus of Academic Spoken English (MICASE) or Cambridge and Nottingham Corpus of Discourse in English (CANCODE)]; *general corpora* that contain a wide variety of texts, in often greater number of words than specialized corpora, and that are often exploited to produce reference materials for language learning or translation as well as reference purposes in comparative studies (e.g. the 400-million word Bank of English or the 100-million word BNC); *comparable corpora* mainly used by translators and learners to identify differences and equivalences in two or more languages or varieties of the same language based on the same proportions of certain text types (e.g. the International Corpus of English); *parallel corpora* containing texts that have been rendered from one language into another (e.g. a collection of European Union regulations published in all the official languages of the Union); *historical* or *diachronic corpora* used to study the course of development a language has followed over time (e.g. the Helsinki Corpus); *learner corpora* containing any collection of texts produced by learners and used to identify the aspects in which their languages differ from each others' or native-speakers' as in the International Corpus of Learner English (ICLE) that can be studied in comparison to the Louvain Corpus of Native English Essays (LOCNESS); *pedagogic corpora* that contain the languages learners have been or will be exposed to in their programs to be used for language awareness.

It is necessary to remind that identification can't be based on a single criterion as Wray (2002) emphasizes, researchers will generally need to apply more than one form of analysis in order to obtain valid results. The researcher has to determine certain criteria for the target data. This criterion may vary even among FS types; that is, while fixedness criterion is applied for fixed phrases, intuition and fluency might be applied for collocation.

2.3. Taxonomy

Trough the information in literature regarding to formulaic sequence, it is difficult to create a clear cut-off categorization. Moreover it is reasonable not to find such information since each researcher use different criteria to detect formulaic sequence. This will, of course, result in sub categories whose boundaries are woolly.

This is probably due to lack of a joint definition. However, while there are still more to study on formulaic sequences, it is highly difficult to agree on a certain definition. In that case, it is normal to meet different criterion, terms, and categorization. Even Wray who did the unique nearly inclusive definition accepts the necessity of a fragmented definition. “Another solution is to accept a fragmented definition or, to put it another way, establish a bundle of features, any or all of which a formulaic sequence may possess, but none of which is individually necessary” (Wray 2002).

The researchers do their classification according to whether the sequences are form-based or functional. For instance, Becker (1975) offers the following six-way division:

- polywords (e.g., *(the) oldest profession; to blow up; for good*)
 - phrasal constraints (e.g., *by sheer coincidence*)
 - meta-messages (e.g., *for that matter . . .* (message: ‘I just thought of a better way of making my point’); . . . *that’s all* (message: ‘don’t get flustered’))
 - sentence builders (*(person A) gave (person B) a (long) song and dance about (a topic)*)
 - situational utterances (e.g., *how can I ever repay you?*)
 - verbatim texts (e.g., *better late than never; How ya gonna keep ’em down on the farm?*)
- (adapted from Becker 1975).

Such approaches throw light upon the taxonomy of this study’s taxonomy even though it differs in terms. In this part, it will be tried to compare this study’s FS categories with the same categories in the literature. This study deals with both functional and formal aspects of FSs by taking into consideration most of the identification tools, sometimes more than one tool for a sub-category, and categorizes FSs into four: Collocations, Fixed Phrases, Sentence Stems and Idioms.

2.3.1 Collocation

Collocation is probably the most problematic topic in FSs as Wouden (1997) stated, “what goes under the header of “collocation” is very heterogeneous. As stated before most researchers consider collocation as the main topic of formulaic sequence, in other word, they use *collocation* term instead of *formulaic sequence*. Similarly, the meaning of the term varies depending on one’s purpose and theoretical orientation (Liu 2010). Therefore it is rather difficult to understand what they meant in “collocation”. However, collocation is just a subset of formulaic sequence in this study.

According to Carter (1988), collocation is an aspect of lexical cohesion which embraces a ‘relationship’ between lexical items that regularly co-occur. Liu (2010) argues that collocation terms mainly concentrates on two major meanings. The first is “the tendency for certain words to occur together.” However they don’t have to be recurrent, there might be short space of each other in a text (Sinclair 1991). Hunston (2002) illustrates it: “...the words toy co-occurs with children more frequently than with women and men (because) toys belong to children, on the whole, rather than adults”. The second meaning is “habitual combinations of words such as “do (not make) laundry” (Liu 2010). Firth (1957) explains, “You shall know a word by the company it keeps”. Wray (2002) argues that “In the context of “collocation” we find that some words seem to belong together in phrase, while others, that should be equally good, sound odd.

The fixedness of collocation is questioned and it is divided into scaled subcategories: “strong”, “medium strength” and “weak” (Crowther, Dignen and Lea (2002) or “fixed” strong” and “weak” (O’Dell and McCarthy 2008).

2.3.2. Fixed phrase

Fixed phrase is usually considered as the ones which never allow any change of the sequence components. Therefore, they are entitled as “non-compositional phrases” or “idiom-like or semi-idiomatic units”. This kind of consideration might result an ambiguity between fixed phrases and idioms. Nevertheless, the most important difference between fixed phrases and idioms is that idioms have metaphorical meaning. In other words, idioms are semantically opaque. On the other hand many scholars don’t

mind naming these idiom-like phrases. They just occupy with the properties of FSs and explains there are also fully-fixed sequences. Bateson, for instance, (1975) asserts some formulaic sequences are fully fixed (e.g. fancy seeing you here; Nice to see you) and can bypass the entire grammatical construction process. As it can be inferred from Bateson's samples, such phrases aren't a component of the sentences. That is, they are used separately from the sentence. Schmitt and Carter (2004) points an advantage of fixedness "...sometimes "fixedness" is an advantage in that it can be easily recognized and learned." However, Wray (2002) cannot accommodate fixed expressions in any categories.

For fixed phrases, insertion of new elements into the sequence is rather difficult. Pawley (1986), compares *first (and only) attempt* with *first (and only) aid*. It can be concluded from Pawley's comparison, any components of fixed phrases cannot be replaced by a new one. That is, no insertion or extraction can be made on fixed phrases. Wray (2002) argues that this fixedness is limited.

Some discourse markers can be included into fixed phrases such as "you know" and "in fact". Moder and Martinovic (2004) explain discourse markers: "In linguistics, a discourse marker is a word or a phrase that is relatively syntax-independent and does not change the meaning of the sentence, and has a somewhat empty meaning."

2.3.3. Sentence stem

Schmitt and Carter (2004) explain sentence stems by giving an example even if they don't name it as *sentence stem*, "The underlying structure to these sentences is '____ thinks nothing of ____', which allows the flexibility to express the 'unexpected' notion in a wide variety of situations". This variety is, of course, restricted to semantic constraints. Wray names sentence stems as *semi-preconstructed phrases* and formulates "...such as NP_i set + tense $POSS_i$ sights on (V) NP_j , require the insertion of morphological detail and/or open class items, normally referential ones (giving, for instance, *The teacher had set his sights on promotion; I've set my sights on winning that cup*) (Wray 2002).

On the other hand there is a challenge for sentence stem identification. The problem is that sentence stems are difficult to identify using current concordancing packages. “Modern concordancers are good at identifying contiguous sequences automatically sequences, but we do not yet have software which can identify flexible formulaic sequences automatically from corpora” (Schmitt and Carter 2004).

2.3.4. Idiom

In the case of idioms, their meaning could not be derived from the sum of meanings of the component words and they did not always follow the rules of grammar. Semantically-opaque formulaic sequences, such as idioms, where the meaning of the sequence cannot be derived from knowledge of the component words. The only way to know the meaning of the idiom is to have learned it as a sequence. Wood’s (1986) definition of the ‘true’ idiom is “a complex expression which is wholly non-compositional in meaning and wholly non-productive in form”. Flavell and Flavell (1992) state that idioms “break the normal rules” either syntactically or semantically. Nattinger and DeCarrico (1992) define idioms as “complex bits of frozen syntax, whose meanings cannot be derived from the meaning of their constituents, that is, whose meanings are more than simply the sum of their individual parts”. Williams (1994) uses the term *idiom* to refer to “any defined unit whose definition does not predict all of its properties”

2.4. Significance of Formulaic Sequences

Recognizing the role of formulaicity is fundamental to understanding the freedoms and constraints of language as a formal and functional system. Specifically, it is proposed that formulaic language is more than a static corpus of words and phrases which we have to learn in order to be fully linguistically competent (Wray 2002).

The significance of formulaic sequence is explained in two lines. One is the contribution of formulaic sequence to the learner. The second one is related to lack of formulaic sequence. Principally both of them amount the same thing. The contribution formulaic sequence will be explained firstly.

Bolinger (1976) asserted that “our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs” and Charles Fillmore (1979) argued that “a very large portion of a person’s ability to get along in a language consists in the mastery of formulaic utterances”. Wray (2002) concludes that It is more efficient and effective to retrieve a prefabricated string than create a novel one. Formulaic sequence reduces processing efforts as Kuiper (1996) stated:

Formulae make the business of speaking (and that of hearing) easier. I assume that when a speaker uses a formula he or she needs only to retrieve it from the dictionary instead of building it up from its constituent parts.

Formulaic sequences have also a social role in that it helps individual belong to a certain group, social context and have a style. Wray (2002) states some social functions of formulaic language:

Speakers seemed able to express their identity as an individualizing deliberately memorized strings and stylistic markers, and their identity as a group member by adopting customary ritualistic utterances, idiomatic turns of phrase and collocations...as individuals both imitate the preferred forms of others and also contribute to the pool of idiomatic material from which others draw. This suggests that the formulaic material plays a central role in maintaining the identity of the community.

The necessity of integration of the individual to the community is one of the most effective ways of learning/acquiring FSs anyway. Kuiper (2004) claims that formulaic performance takes place where speakers are under pressure. Dörnyei, Durow and Zahran (2004) support Kuiper (1996), “it cannot be learnt effectively unless the learner integrates, at least partly, into the particular culture. For example, the context-appropriate application of colloquial phrases cannot be learned from textbooks, but only through participation in real-life communicative events. Wray makes a good summary for the functions of formulaic sequence in her book *Formulaic Language and the Lexicon*. According to her, Formulaic discourse markers seem able to support both the speaker’s and the hearer’s processing simultaneously and another major role for formulaic sequences was found to be that they signaled the speaker’s identity as an individual or as a member of a group (Wray 2002). Wray (ibid) thinks that formulaic

sequences actually serve single goal: the promotion of the speaker's interests. These interests include

- having easy access to information (via mnemonics, etc.);
- expressing information fluently;
- being listened to and taken seriously;
- having physical and emotional needs satisfactorily and promptly met;
- being provided with information when required;
- being perceived as important as an individual;
- being perceived as a full member of whichever groups are deemed desirable (Wray 2002).

The interests explained above are given in a figure by Wray in figure 2.2., below. In this model, the discourse functions are subsumed into the main functions of supporting speaker and hearer processing, both of which they do simultaneously.

Wray explains the processes speaker's choices in using novel and formulaic language to achieve a goal in another figure, figure 2.3. In this schema, three primary aims are identified as the underlying motivations for speaker output. The majority of text is either referential or manipulative, with only mnemonics falling into the category 'access information', which leads directly to fully fixed formulaic sequences. Both reference and manipulation can draw on both formulaic and novel utterances, and the processes by which this can happen are an indication of why the relationship between form and function seemed so complex. Each route through the schema represents a large set of possible formulaic and nonformulaic strings, with the outcome determined by the speaker's priorities and ability to anticipate the hearer's knowledge. Where the speaker aims to be referential, there is most chance that novel constructions will be needed. However, there are opportunities for reducing the processing load by using preassembled polymorphemic words and fixed and semi-fixed formulaic word strings. When the speaker wishes to manipulate the speaker, be that by inciting an action or a perception, or by indicating the text structure so that the hearer can more easily map the shape of the discourse, the priority in selecting the form of words is the anticipation of the hearer's own formulaic inventory. Often this will coincide with sequences that are formulaic for the speaker too, but where it does not, the speaker will take the route of novel construction in an attempt to create a string that is easy for the hearer to decode,

even though it is effortful to encode. In such instances, there is a direct conflict between the processing costs to the hearer and the speaker (Wray 2002).

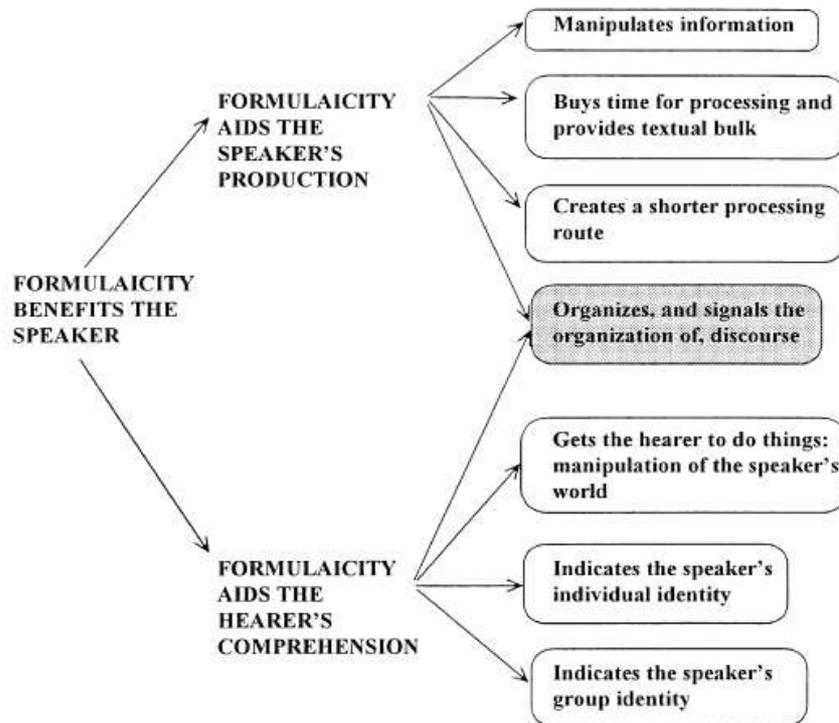


Figure 2.2. The functions of formulaic sequences. Reprinted from *Applied Linguistics*, vol. 21(4), A. Wray, "Formulaic sequences in second language teaching: principles and practice", p. 478, copyright 2001, with permission from Oxford University Press.

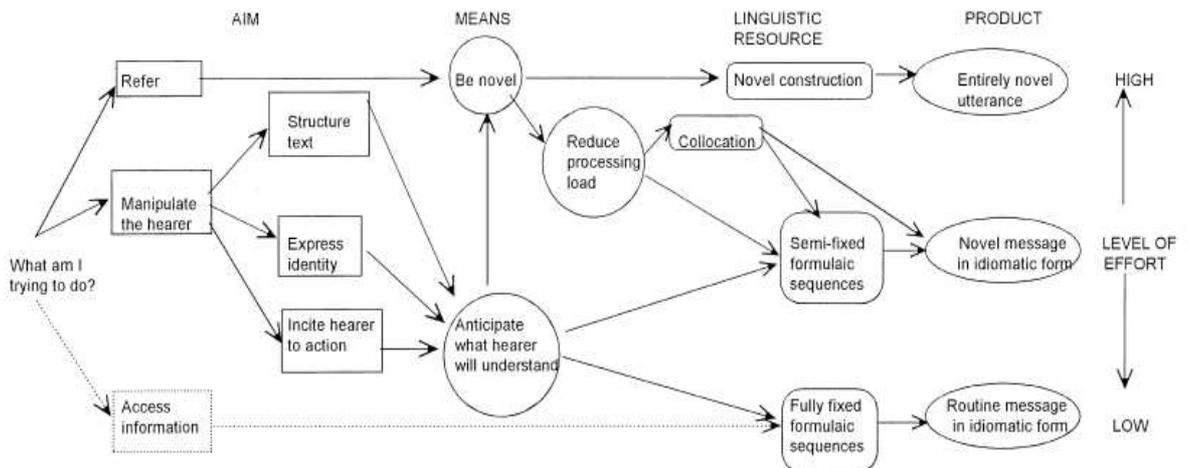


Figure 2.3. Schema for the use of formulaic sequences in serving the interests of the speaker (Wray 2002).

Formulaic sequences can contribute to the establishment and maintenance of an appropriate style for a particular genre. A writer can use structures and turns of phrase

to suggest a relaxed or a formal style, and there are sets of formulaic sequences which belong together in achieving such effects (Wray 2002).

Wray (2002) paraphrases the functions and contributions of formulaic sequences as:

- establish a culture of interaction with carers;
- supplement gesture and other nonlinguistic behaviours in conveying the most important manipulative messages before the production of rule-governed language is possible;
- represent the entry of the child into the group of those who know this or that rhyme or song and expect certain linguistic behaviour;
- provide the child with material for analysis; and
- reduce the child's processing load once novel construction is possible.

Considering the lack or inadequate or wrong use of formulaic sequences, there might be dysfluent and clear nonnative-like speeches. Beneke (1981) asserts that failing to use a native-like expression can create an impression of brusqueness, disrespect or arrogance. Lack of formulaic sequence means generating every speech and this brings a huge burden to the learner. Jespersen (1924/1976) observed that "a language would be a difficult thing to handle if its speakers had the burden imposed on them of remembering every little item separately". Schmitt and Carter (2004) conclude that if one kind of lexeme produces a learning burden, there is no reason to believe that other types of lexeme (i.e. formulaic sequences) are any different in this respect. In a similar vein, such a burden will have psychological matter. 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 (Hakuta 1976). Incompetence in formulaic sequences emerges in several shapes: avoiding using FS, under-use (Dagut and Laufer 1985; overuse (Granger 1998; De Cook 2000) or misuse (Yorio 1998; Howarth 1998).

CHAPTER III

METHODOLOGY

3.1. Introduction

This chapter incorporates the corpus, the selection of the target words, instrument analysis including the formulaic sequences categories and analysis reliability checks as dictionaries, inter-rater reliability and native-speaker consultation.

3.2. Corpus

The corpus, used as data, was formed by using about 85 episodes' scripts of a TV series called "How I Met Your Mother". The scripts were downloaded from the internet and formed as a unique text which includes 236.813 words on approximately 2000 pages. The reason that why this TV series was chosen as corpus is its compatibility with real spoken language. Despite the fact that the dialogues were not formed spontaneously and arbitrarily, the dialogues can be observed in everyday conversations since the plot isn't based on an extraordinary issue. This situation made the corpus more compatible with authentic language.

The scripts were converted to a word file. However, there were still timing numbers in the text. They were deleted before analyzing to enable corpus to consist only sentences.

3.3. Target Words

The corpus provided only non-analyzed data. In order to get the necessary words, the corpus was downloaded on a website called <http://www.lex tutor.ca/>. A frequency analysis (<http://www.lex tutor.ca/freq/>) was applied to the text to acquire the most frequent nouns, verbs, and adjectives in the text. English link (<http://www.lex tutor.ca/freq/eng>) was chosen and the corpus was submitted. (<http://www.lex tutor.ca/freq/eng/freqout.pl>). The screenshots of frequency analysis process have been given in below in figure 3.1., 3.2., 3.3., 3.4.

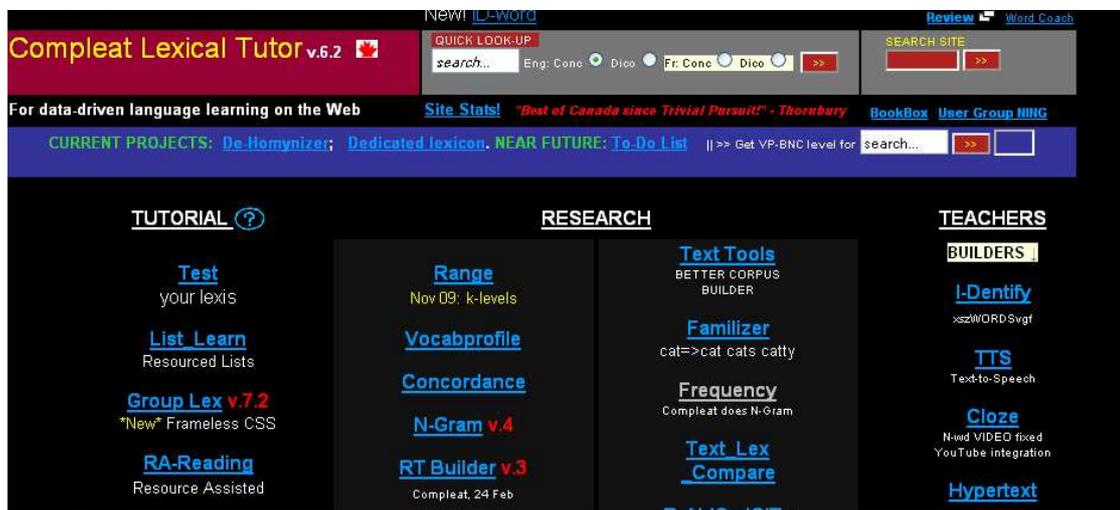


Figure 3.1. Screenshot of frequency link at the website

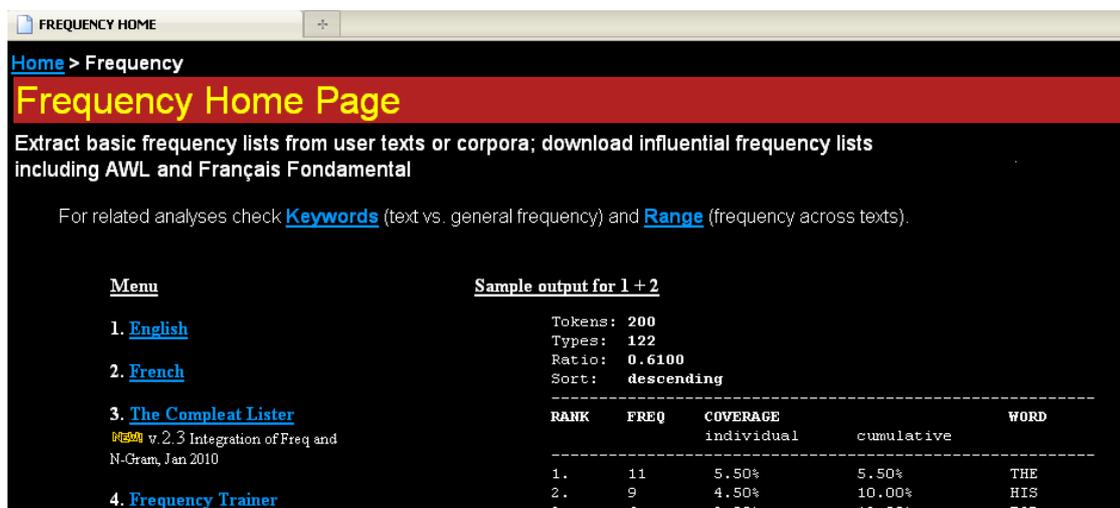


Figure 3.2. Screenshot of choosing language at the website

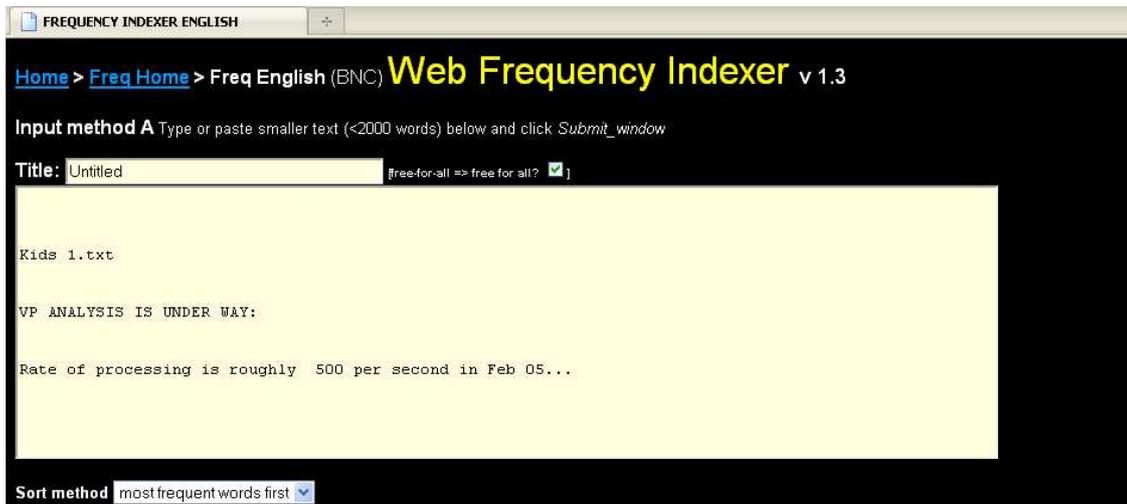


Figure 3.3. screenshot of submitting the corpus



7.	3252	1.38%	17.27%	IT
8.	2851	1.21%	18.48%	THAT
9.	2650	1.12%	19.60%	OF
10.	2347	0.99%	20.59%	IN
11.	2315	0.98%	21.57%	IS
12.	2282	0.97%	22.54%	THIS
13.	2127	0.90%	23.44%	I'M
14.	2081	0.88%	24.32%	WHAT
15.	2000	0.85%	25.17%	MY
16.	1802	0.76%	25.93%	ME
17.	1757	0.74%	26.67%	NO
18.	1744	0.74%	27.41%	JUST
19.	1744	0.74%	28.15%	SO
20.	1733	0.73%	28.88%	WE
21.	1703	0.72%	29.60%	IT'S
22.	1675	0.71%	30.31%	WAS
23.	1576	0.67%	30.98%	OH
24.	1575	0.67%	31.65%	HAVE
25.	1573	0.67%	32.32%	NOT

Figure 3.5. Screenshot of a sample of the most frequent 25 words.

Nevertheless, as it can be seen in the list above in figure 3.5., there are still function words in the target word file. The first content word in the list is in line 24. This list assisted to choose target content words.

The results were copied on an excel file and word scores were identified as shown in figure 3.5. The words to be analyzed have to be content words. The function words don't form formulaic sequences on their own. On the other hand, they can form formulaic sequences provided that there is a content word next to it. According to linguists, content words are those whose meaning is best described in a dictionary and which belong in open sets so that new ones can freely be added to the language while function words described as those ones with little inherent meaning but with important roles in the grammar of a language (Lightfoot 1979). The content words include nouns, adjectives, verbs and some adverbs. Whereas the function words are pronouns, conjunctions, prepositions, auxiliary verbs and some adverbs. The most frequent adjectives, verbs and nouns were chosen from the corpus and subjected to analysis process. These are "have", "know", "do" as verbs; "good", "great", "new" as adjective and "time", "night", "girl" as nouns.

3.4. Analysis

The content words don't form formulaic sequence themselves. They need to collaborate with other content or function words. Therefore the content words chosen from the list aren't still formulaic sequence. In order to understand whether they form formulaic sequence or not, a concordance analysis has been applied to corpus. In order to do that, the corpus was converted into text file since the website used for concordance doesn't accept other file types. Then concordance link was clicked on <http://www.lextutor.ca/>. Since a certain corpus is to be used, "English" link was clicked in "text-based concordances" topic. Nevertheless, since the website is able to analyze the files below 300 kb (about 50.000 words), the corpus required to be analyzed in 7 parts. All parts, one by one, uploaded and submitted. (http://www.lextutor.ca/concordancers/text_concord/). The whole process has been shown below in figure 3.6., 3.7., 3.8, 3.9.

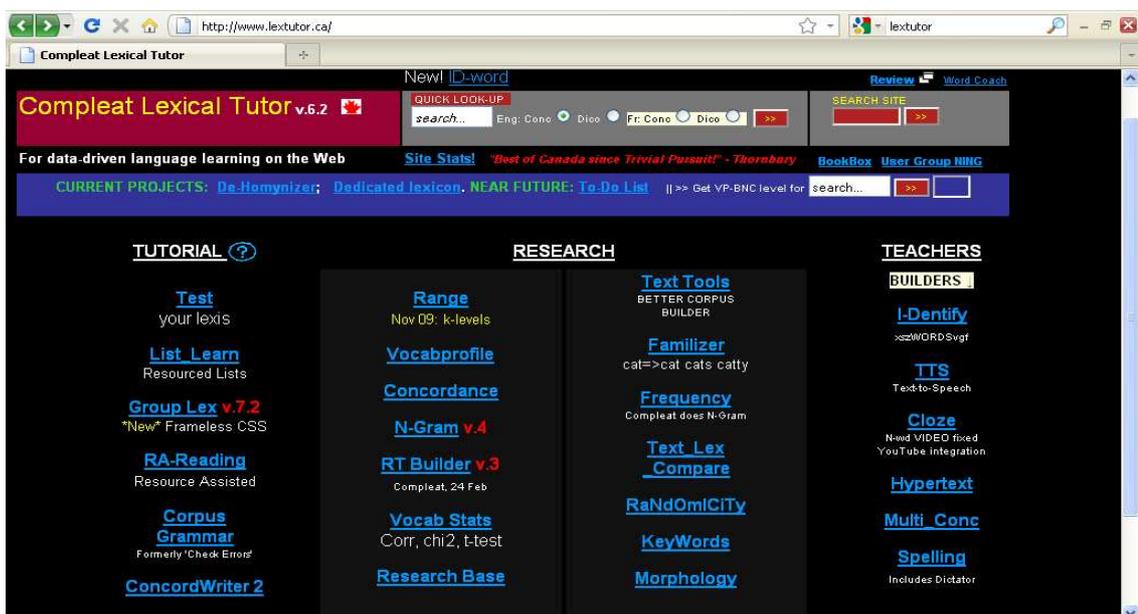


Figure 3.6 Screenshot of a sample concordancing at the website

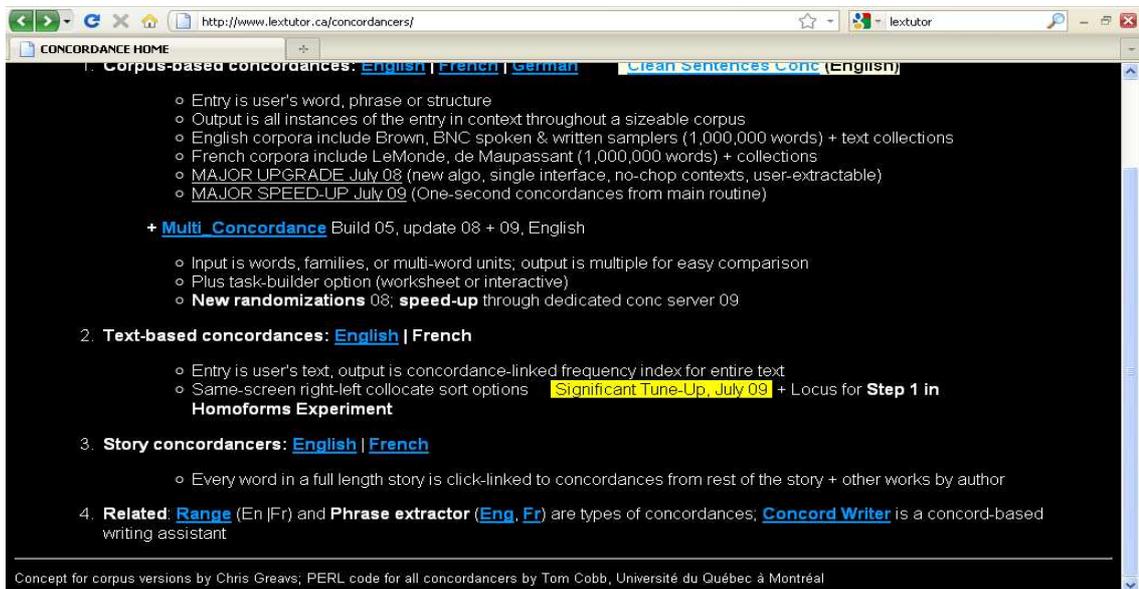


Figure 3.7. Screenshot of a sample text-based concordances

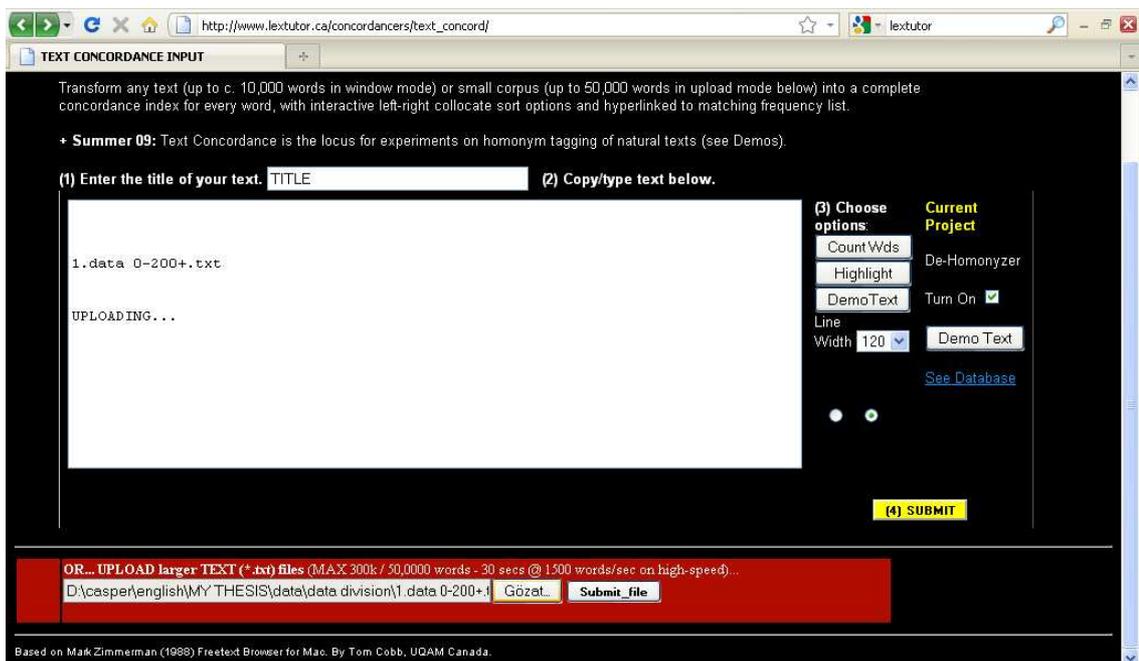


Figure 3.8. Screenshot of a sample submitting corpus



Figure 3.9. Screenshot of a sample concordance output of a text.

Target words (have, do, night...) were copied and pasted in a word file. But verbs “have” and “do” still didn’t include target content words completely. There were still auxiliaries among them as can be seen below. Therefore these function words extracted from the corpus and it has been ready to categorize into formulaic sequence types. However, it is better to explain what categories are and how they are structured.

4459. I know at least... Im not done. She likes dogs Otis Redding **DOES**
the crossword. Shes into sports but not so much so that
4198. your mom doesnt want you to have kids with. Hooray? I mean **DID** she
really think that your college girlfriend Cheryl wou
4229. t guy. I actually set them up. You remember Dr. Muchnik. He **DID**
your braces. Okay my heads about to explode and I d ont
6765. aving one of the best nights of my life your uncle Marshall **HAD**
been having one of the worst nights of his. So thats it?
6783. forever. It never is. Well we just had our first fight. We **HAD** our
first fight. Oh no whatd you fight about? Oh no what
6938. at one week from today I will be leaving Metro News One. It **HAS**
been an honor to bring you the news each night and read
6940. k? Why do you talk?! He needs her password. Its not like he **HAS** her
password. Ooh this is good he has her password. Ther
6977. believe that she gave him her number. Maybe she really does **HAVE** a
brain injury. Thank you everyone. Thank you everyone.
6980. es since I was God you were just a kid. Well look lets just **HAVE** a
few drinks. Well relax and... Yeah that sounds great.
HAVE a good time? Oh yeah yeah. Sorry you couldnt join us. Y
6984. a rain check. And well check in with you tomorrow New York. **HAVE** a
great night. And were clear. Bummer. I was hoping to

6985. course whatever I say youre just going to do the opposite so **HAVE** a great weekend. Good luck screwing up your life. Come
 7005. things up for Lily. Yeah youre right. Besides Marshall and I **HAVE** been planning on having a guys night out before the wed
 7006. lame! I dont even want to go anyway. Uh I dont know. Things **HAVE** been so weird between us. Whatd she say? Whatd she say?

3.5. Taxonomy

The criteria, used to determine the type of formulaic sequence, were composed of general points of literature. Moreover, some criteria were composed by analyzing the target words, i.e. they aren't universal.

3.5.1 Collocation

The collocation term has got really vague boundaries. While some scholars (Pawley and Syder 2000; Lewis 2000, 1997; Jackson 1988; Leech 1982) define it as a sub-category, some define it as a main category (Liu 2010). In this study, collocation will be considered as a sub-category of formulaic sequences.

Collocation is word patterns which are frequently used together. In spite of some scholars who believe there is no arbitrariness in forming collocations, there is a fact that there is a habitual usage in collocations. A collocation should consist at least two words and at least one of them should be a content word. What makes a word pattern collocation is its restriction of using any other word instead of the one used in sequence. In other words, while there is a synonym or a close-meaning to be used instead of the content word in the sequence, it is not preferred. An example from the corpus will explain the situation better.

4088. I'm gonna win her back. I can trust you guys. i... we would **COME to regret** it. We just didn't know it yet. i You have it?

The verb “*come*” has several synonyms or other words having close meaning in Oxford Thesaurus (2004 Edition) like: *approach, advance, proceed, arrive, happen, occur, reach*. However, any of these words don't replace the verb “*come*”. Moreover, it shouldn't be thought that all the formulaic sequences in the corpus are formed because of the target words. Sometimes, target words can be a component of the formulaic sequence. That is, while the target word in “look good” sequence is “good”, “look” collocates with “good”. That is, the other word, “look” instead of the target word is the base in this sequence.

The criteria for collocation detection are listed below:

1. Collocations are pairs or groups of words that are frequently used together
2. They have predictable relationship with each other
3. Generally, one of the words is flexible but this flexibility is restricted with certain words
4. If the word which takes place in thesaurus dictionary as synonym of the word in the sequence doesn't match the same meaning and use in original sequence, it is collocation.
5. Each word constituting collocation should be a content word. This is called lexical collocation.
6. The collocation should take place in a collocation dictionary

There are a few examples from the corpus below:

4821. babysitter I spent weeks with round like a ball. Mom! Just **DO me a favor**. If you know someone special not go as I did.
 7444. e Tyty you have done? Funny story. Another night that silly **HAD a nightmare**. And when he was telling us about which was
 7045. he role Tyler went to the Grant. What is the Grant me **looks GOOD** actor. Really? Look at this. Hey Tyler! Tyler! Tyler! G
 7242. ll "home". Thank you and welcome home. Everything was **going GREAT** until one day in the elevator yeah its ted no no no no
 21244. lways wanted to say that. I mean in a bar. I say it **all the TIME** at home. Dude someday we should buy a bar. And when we

The sequences in bold in the examples above are formulaic because; “*make*” instead of “*do*”, “*experience*” instead of “*have*”, “*seem*” instead of “*look*”, “*progress*” instead of “*go*”, “*occasion*” instead of “*time*” cannot be used. This restriction can sometimes soften. The word can be replaced by another word. This replacement creates a new formulaic sequence with the new word. As in the example below, the verb “have” can be replaced by “take”, but it cannot be replaced by “keep” or “sit”.

7683. Frozen Snowshoe. Please were friends. Just call me "Shoe." **HAVE a seat**. So Shoe... How would you like to give me... an
 10380. authority was compromised. Oh hey Marshall. **TAKE a seat**. I know how much you love stools. Than

There are two kinds of collocations, namely “lexical” and grammatical”. Grammatical collocations consist of a lexical item and a preposition or a grammatical word whereas lexical

collocations include word combinations, and don't include prepositions or grammatical words (Bahns 1993; Lewis 2000). This study deals with the lexical collocation because of the fact that the word collocating with the base word should be a key to predict the meaning as collocations are partly transparent. On the other hand, in grammatical collocations, since the word collocating with the base word is rather difficult to enable the meaning predict. For that reason, the sequences formed with grammatical words, especially with prepositions like phrasal verbs (give up, look down on etc.) ,adjective phrases (interested in, afraid of, etc.) and verbal phrases (depend on, consist of, etc.), haven't been taken into consideration.

3.5.2. Idiom

Idiom is defined as the expression which has metaphorical meaning; that is, it is often impossible to guess the meaning from the individual words put together. In other words, in order to call an expression "idiom", there should be a different meaning beyond the words constituting it. This meaning is usually metaphorical. A metaphor is a way of describing something by comparing it to something else that has similar qualities. Idioms are usually cultural statements and they have a cultural background.

The words in an idiom cannot be replaced by any synonym or close-meaning word. They are fixed phrase and except from grammatical changes, they can't be changed. Some idiom samples are shown below. In this study, the expressions which have different meaning from its individual words have been accepted as idiom in beside checking idiom dictionaries.

6457. wanted to diversify so... here I am. So **Im in GOOD hands**. You found me a match There she is. You
 19507. bout nine months now. Closer to ten I think. **TIME flies**. So ten months ago you just up and deci
 7628. ext three seconds and you snap its neck with your legs that **HAS no business** being on your reel. No no I get it I get it.
 14686. es alone doesnt mean shes a les... Okay lets **HAVE a big hand** for Robin the reporter. All right

The criteria for idiom are listed below:

1. Idioms have a special meaning that is different from the meanings of all individual words put together
2. It's often impossible to guess what they mean.
3. They are fixed sentences; that is, their components cannot be replaced with a synonym or any other word.

4. Idioms have metaphorical meaning.
5. The idiom should be verified by an idiom dictionary.

3.5.3. Fixed phrase

“Fixed phrase” isn’t a widely used term. Some scholars entitle formulaic sequence as fixed phrase. However, in this study, it will be defined as strict word patterns. Fixed phrases are the expressions which are usually used in the same way and construction and cannot be changed. Additionally and most importantly, they are always used separately from the sentence; that is, they are not syntactic component of the sentence. Briefly, they can be defined as simple fillers and short expressions. They might be used at the beginning of the sentence, after subject or at the end of the sentence or completely apart from sentence as shown in the examples from the corpus below.

The criteria used for detecting fixed phrases are given below:

1. They aren’t usually flexible that is they don’t allow any changes. They are used in the same way in almost every sentence
2. They are apart from sentence. They aren’t a syntactic component of the sentence.
3. They are simple, short fillers for the sentences.

4062. want to get a taco? A taco? You love them remember? Right. **COME on**
 Im hungry. iSo Robin and I went back to just being r
 9704. ting serious between us and that scares you. **GOOD night** Ted. Okay
 Randy now tell me without loo
 9825. sale and youre the boots baby. Oh its Robin. **GOOD luck.** Bye Mom.
 Thats my mom. Hello Bring me a12113. ght on New Years Eve.
 It was a New Years tradition. But **you KNOW** what we did this year? Crazy
 monkey sex? No. Okay all t

3.5.4. Sentence stem

Sentence stems are flexible devises that structure the sentence. They are basic sentence of a clause. They form syntactic frame of the sentence and they often have some certain grammatical structure like subject+ verb + clause. The structure with sentence might be a subordinate clause or verb. Moreover, sentence stems aren’t as strict as fixed phrases. They are short sentences which have one or more slots to be filled in a limited number of ways. Here are some examples of sentence stems from the corpus;

41563. this other party. It's no big deal. **The first TIME** I rode in a limo I was five. I was on the way
 7080. m bringing him a new pair. Marshall forgot his pants. **It's a GOOD thing that** you came by because he has a really important
 30927. s anymore Ted Lily and I are married now **Its TIME** Were getting our own place Actually I left the
 7509. e. There was a surprise party that night **How COME** nobody told me
 People think I can't keep a sec

As stated before one challenge of sentence stem is its identification since sentence stems' components aren't contagious and today's concordances don't yet such a function. Therefore, human judgment and manual checking of the data are necessary to get a healthy output.

Considering the samples above, sentence stem occur so:

The first time Sentence (simple past)

It is a good thing Noun Clause

It is time Sentence

How come Sentence

The criteria of a sentence stem can be seen below:

1. They are basic sentence of a clause. They form syntactic frame of the sentence and they often have some certain structure subject+ verb + clause.
2. The rest of the sentence after the sentence stem can be filled with anything that is it has slots to be filled.
3. The sentence stem has to be linked to a sentence structure.

It should be noted that a sentence stem may contain a collocation. However this doesn't mean that collocations are subset of sentence stems. For example:

The first time sentence (simple past) is a sentence stem since it has a slot to fill. On the other hand, it would be a collocation if it didn't have a slot and require a grammatical structure as occurred in: "for the *first time*".

"Last time", for instance, is collocation if it is individually used. Besides, it is usually used as "for the *last time*". However, if there is a sentence after "last time" sequence and it refers to the final time that activity placed in the sentence, on its right, it is "sentence stem". The following examples will make the issue clear.

19487. soul mate. That'll be bucks. No way. The **last TIME** I did this the girl turned out to be engaged.

21255. tered but technically Im not a celebrity. Plus for the **last TIME** I put the mask on as a joke. OK not you Barney. OK who

The second “last time” sequence in the sample is collocation. (for *the last time*) Whereas the same sequence in the second sample is a sentence stem since it has a slot requires a simple past sentence. (*the last time* + *Sentence in simple past*)

3.6. Reliability Checks

The target sequences were classified by the researcher as collocation, idiom, fixed phrase and sentence stem. In order to ensure the reliability, some check instruments were applied to them. However, some checks have to be created since there are not certain check instruments for two categories: fixed phrase and sentence stem. Furthermore, final results of verb “do” are given in appendix.

3.6.1. Dictionaries

Some advanced dictionaries were used for each category. All classified formulaic sequences were, also, verified by using a hard copy advanced dictionary or a soft advanced dictionary program. Oxford Collocation Dictionary Software and Oxford Online Collocation Dictionary website (<http://llohe-ocd.appspot.com/>) were used to verify the collocations. For Idiom verification, two idiom dictionaries were used. These are Collins Co-built Idiom Dictionary (2007, 2e) and Oxford Dictionary of Idioms (2004, 2e).

For fixed phrases and sentence stems different check instruments were used since there weren't any certain materials. As it is mentioned above, the words forming fixed phrase can't be replaced by a synonym word. Therefore, the words forming the expression thought as fixed phrase were checked in a thesaurus dictionary to determine whether one of the words could be replaced by a synonym or not. While this reliability check is applied for one of the words forming collocation, in fixed phrase check, it was applied for all the components of it. That's because while collocation permits flexibility, any component of the fixed phrases cannot be substituted by a synonym word. In other words, if one word of the collocation is substituted with a new word, it is still possible to be a collocation, any word of the fixed phrase cannot be changed in no way. If it is possible to replace by a synonym word and the new expression is widely used, it means that it is not a fixed phrase. Nevertheless, if any word

can't be replaced by a synonym word and there is no possible change in the expression, it meant that it is a fixed phrase. Besides being a bit individual, fixed phrase classification is heavily depended on widely and strict usage of the phrase.

As it is stated above, there is no certain check instrument to verify sentence stem, either. However, the criterion detected by the researcher is used as the main criteria for sentence stems. According to that, the phrases which have a sentence structure linked to it is classified as sentence stem. These criterion, also, have been clarified for the other reliability checks which will be explained next.

3.6.2. Inter-rater reliability

In order to verify compatibility of criteria with category results, a sample categorization process applied to three inter-raters. The inter-raters are English instructors at a university. They are not native speakers. Firstly, a criteria sheet including necessary information for the rater was prepared and given them. Having been sure that they read and understood the issue, they were given a sample unanalyzed sheet. The sheet included 90 uncategorized target words. Additionally, the sheet also included non-formulaic items. The words were “come”, “good”, and “time”. All target words included 30 samples to analyze and the raters were supposed to categorize them according to criteria if they think the item was formulaic.

The results gathered from the raters were compared to researcher's categorization as shown in table 3.1. However, since the reliability results occurred below 80%, the items and formulaic sequence criterion were discussed with the raters and some additions were done. First of all, for collocation categorization, thesaurus assistance was offered and the researcher rechecked his choices. Then, the raters were asked to do analysis again. After the second analysis, it was observed that the results were still below the desired mean (82,2%). Also, the ratio of rater 3 was far below the expectations. That was because she had some objection to some target words. Besides other analysts had a few objections. There were 6 controversial items in total. All the items were discussed in group and some changes have been decided which are given below.

Item 1.

7238. was just lunch and you know she doesnt eat wheat. We had a **GREAT** time catching up. We even laughed about what jerks we

The rater 3 claimed that “great doesn’t belong to the collocation in the sentence given above. That sequence is a collocation because of “have time”. However, the other members including researcher objected to that for that reason: “great time” is also a collocation beside “have a great time” since “great” can be replaced by “good” which has a close meaning but not same while time can’t be replaced any word which has a close meaning to “time” such as “hour”. As a result, rater 3 was persuaded.

Item 2.

7244. u pull this crap. Okay youre right. Im great with books. Im **GREAT** with art. Im great with identifying French wines from

The rater 3 noticed that grammar collocation isn’t a criterion for collocation according to criteria worksheet. Therefore, this sequence wasn’t accepted as a collocation by the group members and was extracted from the score.

Item 3.

9910. e a really nice guy... Oh no. It has been so **GREAT** *getting* to know you. I am gonna kill Marshal

Rater 2 claimed that “great” forms a sentence stem in such a phrase:

...be great Ving

Although the researcher accepted the claim by considering it is a sentence stem so:

...be adjective Ving

the other two raters rejected the objection since it has so many slots to consider a sentence stem. They thought there was nothing related to FS. As a conclusion, only the researcher changed his mind.

Item 4.

9913. m not sure yet. I... lost my password. Thats **GREAT** Cara. I dont know yet. I lost my password. I

The researcher accepted “that’s great” sequence as a fixed phrase. However the other raters objected since it is a sentence. However, the researcher still thinks it is a fixed phrase by taking into consideration “you know” sample. Only rater 3 changed her mind.

Item 5.

7436. s. Maybe if you agreed upon that beforehandyeah. Little guy, **HAVE** a big day. Well talk about it tomorrow. Sweet dreamslug

Rater 3 objected to this item with the same reasons as item 1. Therefore she accepted that it as a collocation. Furthermore, it should be noted that this item is a collocation not because of big but because of “*have a big day*” altogether.

Item 6.

7484. ave sex. Save it. Lily saw you. Or did she? You see Ted you **HAD** so little faith in Heather and me... ... we decided to g

The rater 3 claimed that “*faith in*” in this sequence is the base word so it is collocation because of “*little faith in*”. The researcher and rater 2 accepted that it is a collocation but “*have little faith*” and “*have faith*” sequences were also collocation but rejected to include in as it is a preposition. And also, “*little faith*” doesn’t have frequency itself separately apart from “*have*” in the available corpus.

Also, the researcher discussed the raters’ preference in categorization and did the necessary changes in original corpus as well as explained above. Lastly, they were asked to do analysis again. The final analysis answered the expectations and the results are given below

	ITEM 1			ITEM 2			ITEM 3			TOTAL		
	1st Analyst	2nd Analyst	3rd Analyst	1st Analyst	2nd Analyst	3rd Analyst	1st Analyst	2nd Analyst	3rd Analyst	1st Analyst	2nd Analyst	3rd Analyst
Rater 1	90%	96.6%	96.6%	86.6%	93.3%	93.3%	60%	60%	83.3%	78.8%	83.3%	91.1%
Rater 2	100%	100%	100%	83.3%	86.6%	86.6%	73.3%	76.6%	80%	85.5%	87.7%	88.8%
Rater 3	90%	100%	100%	73.3%	86.6%	86.6%	40%	40%	73.3%	67.7%	75.5	86.6%
TOTAL										77.4%	82.2%	88.8%

Table 3.1. Inter-rater analysts’ and researcher’s taxonomy means

3.6.3. Native-speaker consultation

Some items, especially the ones which don’t have a special dictionary or software, were consulted to native-speakers who are academic and literate people. There were some informal phrases in the corpus since it is a spoken corpus. Some of the phrases which don’t

appear in sources were asked to two native-speakers. One of them is lawyer in San Francisco, USA. The other is an Assistant Professor in English Literature Department at Karabük University. They were asked, firstly, the meaning of the unknown phrase and secondly, they were asked to examine the phrase syntactically. Furthermore, the phrase were asked again to native speakers after replacing the related word in the phrase with a synonym of it in order to detect which one , former or latter, is widely used and naturalistic.

CHAPTER IV

RESULTS AND DISCUSSIONS

4.1. Introduction

This chapter intends to present the results and discussion of the findings in this research in the following order:

- I. Formulaic sequence frequency in English TV series (4.2)
- II. Common formulaic sequence types in English TV series (4.3)

4.2. English TV series as a formulaic sequence source

Research Question 1. *Are English TV series rich in formulaic sequence?*

The corpus can be regarded as a spoken corpus since the theme of the series bases on routine activities and an unremarkable plot. It can, however, be regarded as a written corpus as the conversations were written before. However, the conversations show us that they belong to spoken language. There are slangs, contractions and even swearwords which are not definitely used in a formal conversation. So, it is clear that the corpus should be regarded as a spoken corpus.

The thing whether the corpus is written or spoken will lead us to determine the FS level in written or spoken language. Many researchers claim that FS is used in spoken language more frequently than written language. According to a research conducted by Biber and et al. (1999), the FS rate in spoken corpus is 28% while it is 20% in written corpus. Additionally, the research by Erman and Warren (2000) shows that 58.6% of spoken language is formulaic while it is 52.3% in written language. According to Hill (2000), collocations (formulaic sequences) make up approximately 70% of everything we read , write, say or hear.

This study doesn't have a goal to compare spoken and written corpus FS rate. It is aimed to determine whether the target spoken corpus is rich in FS. The formulaic sequence density of target words in the corpus is given below

Table 4.1. Formulaicity in target word patterns

	Formulaic		Non-Formulaic		Total	
Come	335	%43	441	%57	776	%100
Do	131	%10	1204	%90	1335	%100
Have	450	%32	930	%68	1380	%100
Good	251	%54	215	%46	466	%100
Great	117	%33	235	%67	352	%100
New	85	%26	238	%74	323	%100
Night	148	%45	179	%55	327	%100
Time	475	%90	55	%10	528	%100
Girl	50	%21	188	%79	238	%100
Total	2042	%37	3685	%63	5785	%100

The percents were acquired by comparing each target word's formulaic and non formulaic patterns. The percentages show how extent the target words form formulaic sequence.

37% percentage result of formulaic sequence in the corpus proves that the target corpus is rich in FS since, at least, one out of three patterns is formulaic. This result verifies Biber and et al (1999) study which is mentioned above. Thereby, it can be asserted that English TV series can be a source for FSs.

Authentic language sometimes refers to words even never found in written language but spoken language and inconsistent with grammar rules. Moon (1998) states that by following the evidence of corpora ..., it is possible to discover multiword expressions that are not yet listed in classical handbooks of phraseology and to detect new usages and forms of known multiword expressions. Similarly Widdowson (1978) claims with regard to especially reading and listening skills frequency remains a significant factor, given various studies on

large corpora that have revealed striking discrepancies between rules and generalizations provided in syllabuses and actual language use. Gavioli and Aston (2001) suggest that while corpora in themselves are unable to advise what to teach, based on the frequency of use information, teachers and material designers can make better-informed decisions and justifications, or learners subjectively interpret instances of language to create models of their own; this being a potential which should not be overshadowed by discussions of the extent to which ELT syllabuses and materials should be shaped by corpus information for a better reflection of linguistic reality. This percentage is crucial to language acquisition since such a corpus means a wonderful exposition to the authentic language as Lewis (2000) states, a learner should be exposed a large number of words and chunks so that s/he can make generalizations about lexis and its collocational restrictions. Sinclair (1991) and Hausmann (2004) argue that conventionalized chunks (formulaic sequences) play an important role in the everyday language and greatly shape its structure. Partington (1998) emphasizes that the frequency of a structure or word sense is indeed an important component of potential language. Kennedy (1998) also maintains that frequency of occurrence of language items should be only one of the criteria used in determining the content of teaching since even less frequent items may deserve more attention depending on the goals of learners. As stated before in fluency, formulaic sequences play a vital role for being native-like. Or, "If you do not choose the right collocation, you will probably be understood but you will not sound natural (Van der Wouden 1997). Similarly, Kjellmer (1991) and Nattinger and De Carrico (1989) claim that if the learners acquire more chunks and become capable of producing them, it will enable them to process and produce language at a far faster rate without any hesitations or pauses and motivate them to participate in more social interaction. Moreover, their reading and listening skills will develop better as a result of instant recognition of these prescribed patterns and they will be more competent in the foreign language.

The target words' average rate is over 30%. There are three words below 30%. These are "Do", "New", "Girl". The word which has the lowest rate is "Do". The formulaic rate rises to 43% when the verb "Do" is extracted from the corpus. The rate rises to 47% when "girl" is extracted, too. This result means that almost one out of two word is formulaic. On the other hand the highest formulaicity rate belongs to "time" with 90%. This is because the word "time" is rich in both collocation and sentence stem. The FS types of each target word will be

given below. The formulaicity of each word will be explained while explaining the second research question.

A second result to be inferred from the table 4.1 is that except from two words non formulaicity rate of the words is higher than formulaicity. The target words, that is, consist of more non-formulaic patterns than formulaic patterns. The exception words are “time” and “good”.

The result above has aroused a question. Which word type has more formulaic sequence? Which is the most abundant source for FSs? The table 4.2 below shows that nouns form formulaic sequence structures more frequently than others.

Table 4.2. Formulaicity rate of target word types

	Formulaic	Non formulaic	Total
Verb	916 %26	2575 %74	3491 %100
Adjective	453 %40	688 %60	1141 %100
Noun	673 %61	422 %39	1095 %100

According to the table 4.2, nouns have more FS than adjectives and verbs. The table shows that formulaicity rate in target words in descending order is noun, adjectives and verbs respectively. However, according to Laufer’s study (1991) verbs have the highest frequency while nouns are in the second line.

Considering that three of the target words decreases the formulaicity rate, it is wondered whether the order changes when extracted the words below 30% formulaicity rate. The table 4.3 below shows the formulaicity rates of target word types without the words whose formulaicity rate is below 30%. (do,new,girl)

Table 4.3 Formulaicity rate of target word types without the words whose rate below 30% (do, new and girl)

	Formulaic	Non formulaic	Total
Verb	750 %35	1371 %65	2121 %100
Adjective	368 %45	450 %55	818 %100
Noun	623 %73	234 %27	857 %100

As it can be seen in the table 4.3 the order doesn't change even if the words whose formulaicity rate is below 30% are extracted. However the most critical rise is seen in nouns with 12%. Adjective formulaicity rate, which is in the second most formulaic row, increase only 5% while verb formulaicity increases 9%. However, the final order is still same; nouns, adjectives and verbs respectively.

Nevertheless, it will be a generalization to claim that nouns are the words which have the most frequent FS. It is generalization because the words having below 30% formulaicity rate consist of a verb, an adjective and a noun. That is, every word type might have a low formulaicity rate. Besides, a noun which is in the highest rate group might fall behind with a verb which is in the lowest rate group (girl-come). Therefore, it will be right to state that every word has exclusive formulaic properties. Furthermore, the table 4.2 and 4.3 both reveal nouns have the highest formulaicity rate.

4.3. The most frequent FS types in English TV series

Research question 2. *What type of FS is frequently used in English TV series?*

As stated by Moon (1998), terminology in this field has always been problematic, and there is no generally agreed common vocabulary. Different terms sometimes used to describe identical or very similar kinds of unit; at the same time, a single term may be used to denote very different phenomena. Moon (1998); for instance, define idioms as a fixed phrase while these two terms are studied separately in this study. In addition to this, Liu (2010) entitles almost all types of sequences as collocation. It is therefore there has been a

requirement to clarify the FS types. For that reason, the researcher analyzed the formulaic sequences in 4 categories: Collocation, fixed phrase, idiom and sentence stem.

All the process to determine the types of FSs were run meticulously. There are 2042 FSs forming by using target words. Which FS type is most frequently used in these FSs. The results have been shown below. According to this table, collocations are the most frequently used FS type in target corpus. In other words, collocation is almost the unique FS type in the corpus.

4.4. FS type frequency

	Collocation	Fixed Phrase	Sentence stem	Idiom
Come	50 / 2,4%	263 / 13%	20 / 1%	2 / 0,09%
Do	124 / 6%	7 / 0,3%	0 / 0%	0 / 0%
Have	444 / 22%	0 / 0%	0 / 0%	6 / 0,3%
Good	201 / 9,8%	36 / 1,9%	12 / 0,5%	2 / 0,09%
Great	117 / 5,7%	0 / 0%	0 / 0%	0 / 0%
New	78 / 4%	7 / 0,3%	0 / 0%	0 / 0%
Night	148 / 7,2%	0 / 0%	0 / 0%	0 / 0%
Time	438 / 21,4%	0 / 0%	36 / 1,8%	1 0/ 0,04%
Girl	50 / 2,4%	0 / 0%	0 / 0%	0 / 0%
Total	1650 / 80,7%	313 / 15,5%	68 / 3,3%	11 / 0,5%

These results were attained by counting the FS types in the corpus comparing the score to the total FS number and the percentages were acquired. According to these results, it has been found out that the highest frequency in FS types belongs to collocations with dramatic percent (80,7%). The second highest frequency belongs to fixed phrases with 15,5%. Idioms and sentence stems have low rates (0,5% and 3,3% respectively).

The zero (0) digits mean that there is no example of that FS type in that word. “Do”, for instance, doesn’t have any sentence stem form in the corpus. As it can be inferred from table 4.4, idiom including target words is used very rarely in the corpus. There are only eleven idioms.

When the table is studied in respect of word type and FS type relationship, the results are given below in table 4.5.

Table 4.5 Word type-FS type relationship

Word type	Collocation	Fixed Phrase	Sentence Stem	Idiom
Verb	618 / 37,5%	270 / 86,2%	20 / 29,5%	8 / 72,7%
Adjective	396 / 24%	43 / 13,8%	12 / 17,6%	2 / 18,2%
Noun	636 / 38,5%	0 / 0 %	36 / 52,9%	1 / 9,1 %
Total	1650 / 100%	313 / 100%	68 / 100%	11 / 100%

The scores above in table 4.5 were attained by accumulating each FS type score in consideration of word types. Then, each word type’s score was compared to related FS type and the percentages were attained.

According to the table 4.5., it is clear that collocation has a pervasive percentage. Therefore it can be concluded that collocations should be assessed more intensely than others. Hill (2000) confirms the importance of collocation and claims that the awkward stretches of language that learners produce are due to lack of collocational competence. He also puts forward that “lack of competence in this area forces students into grammatical mistakes because they create longer utterances because they don’t know the collocations which express precisely what they want to say.” Lewis (2000) supports Hill, “...the fewer collocations students are able to use, the more they have to use longer expressions with much more grammaticalisation to communicate something which a native speaker would express with a precise lexical phrase and correspondingly little grammar.” The result of FS type analysis leads to that conclusion: “...collocation deserves special treatment in FL learning and

teaching as it is the most important factor in the creation and comprehension of natural languages” (Eryıldırım 2002).

According to these results, in collocation verbs and nouns are very near. However, adjectives are not much far from them. So, as it can be understood from table 4.5. collocation is very frequent in three word types. When fixed phrases are studied, it is clearly understood that fixed phrases are formed by verbs more frequently than the other word types. However, it should be stated that most of the fixed phrases are comprised of one: “come on”. Only 26 of 263 fixed phrases including “come” is formed by different “come” fixed phrases. That is, 237 out of 263 fixed phrases in verb “come” is “come on” which constitutes 87,7% of all verbal fixed phrase. Each word type has sentence stem structures. the word type which has the highest sentence stem rate is noun group (52,9%). Furthermore, when table 4.4 is studied, it will be understood that 36 out of 68 sentence stem is formed by using “time”. Besides, it should be noticed that each word type has only one word to form a sentence stem. These are “come” in verbs, “good” in adjectives and “time” in nouns. Thus, it can be alleged that the exclusiveness of the word is effective in creating a sentence stem. When idioms are studied in table 4.4 and 4.5, it will be seen that it is not frequently used in target corpus. The percentage of idiom is 72,7% in verbs, 18,2% in adjectives and 9,1% in nouns. There are only 11 idioms in the corpus; 8 idioms in verbs, 2 idioms in adjectives and 1 idiom in nouns.

CHAPTER V

CONCLUSIONS

5.1. Introduction

There were two main objectives of this study. One of them was to find out whether the English TV series are rich in FS. As mentioned before, to get an answer for this research question, several processes have been applied. Secondly, it was aimed to figure out which formulaic sequence type is most frequently used in the target corpus.

This chapter summarizes the findings of the study. This chapter also points to the limitations of the study that might help future researchers and provides further research suggestions.

5.2. Conclusions

This study clearly indicates that TV series are rich in formulaic sequences. Many researches announce FS frequency rate between 32% and 58,6% as rich in FS. Depending on these results, it can be asserted that English TV series are rich in FS with 37%. This percentage is really high in view of the fact that FSs cannot be learned but acquired. This result also means that English TV series might be a source for educational materials.

The second research was to determine that which FS type is most frequently used in English TV series. According to research results, collocations are the most frequently FS type used in the target corpus. The second FS type most frequently used in the target corpus is fixed phrase. The other types, idiom and sentence stem, have low rates. So, it can be stated that collocation teaching should be one of the main objectives of educational program.

5.3. Limitations of the Study

The corpus which is used as a source for the study was formed by using scripts of a TV series which is considered as compatible with daily spoken language. However, it is still likely that the scripts might not match with spoken language.

Some resource books which were used as a checking instrument involved a wide range of phrase as FS. On the other hand, the types of FS in this, study were identified by a joint

criterion of the researcher and the other checking instruments. In other words, a phrase which is described as a collocation, e.g. in a collocation dictionary, might not be considered as a collocation by the researcher or other inter-raters. Therefore, it should be noted that the criterion to decide the category of FS types are subjective. This is partly because many researchers in their study use same terms in different meanings or same meanings in different terms. For that reason, the criterions of FS types in this study were formed by looking into literature and trying to form a joint terminology.

5.4. Suggestions for Further Research

This study indicates that a significant portion of spoken language composes formulaic sequence. It may be further researched how the word types form FS by studying them in detail. According the research results, collocation is prevalent FS type in the corpus. The structures of collocations might be studied.

Other oral resources like song, news, movie scripts might also be included to the corpus and they might be contrasted with each other in the view of FS frequency rate.

To be able to integrate available FS sources to teaching is another aspect of the field. In formulaic sequences, noticing FSs has always been problematic. Schmidt (1990) claims that noticing is essential for learning and for vocabulary build-up means of awareness raising techniques needs to be researched.

REFERENCES

- Adolphs S, N.& Durow, V.
 2004 “Socio-cultural integration and the development of formulaic sequences”, In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.107-127), Benjamins, Amsterdam.
- Aston, G .
 2001 “Corpora and language pedagogy: Matching theory and practice”. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: studies in honour of H.G. Widdowson* (pp.257-270) Oxford: Oxford University Press.
- Bahns, J.
 1993 “Lexical collocations: a contrastive view.” *ELT Journal* 47, 1:56-63
- Bateson, M.C.
 1975 “Linguistic models in the study of joint performances.” In M.D. Kinkade, K.L. Hale & O. Werner (eds.) *Linguistics & anthropology, in honor of C.F. Voeglin*. Lisse: Peter de Ridder, 53–66.
- Becker, J.
 1975 “The phrasal lexicon”. In B. Nash-Webber and R. Scha *Theoretical issues in natural language processing 1*. Cambridge, M Bolt, Beranek, and Newman.
- Beneke, J.
 1981 “Cultural monsters, mimicry and English as an international language.” In R. Freudenstein, J. Beneke & H. Ponisch (eds.) *Language incorporated: teaching foreign languages in industry*. Oxford: Pergamon & Munich: Max Heber Verlag, 73–94.
- Biber, D et al
 1999 “*Longman grammar of spoken and written English*”. Harlow, UK: Pearson Education Limited.
- Bishop, H.
 2004 “The effect of typographic salience on the look up and comprehension of unknown formulaic sequences” In N. Schmitt (ed.), *FS Acquisition, Processing and Use*, (pp.227-249), Benjamins, Amsterdam.
- Bolinger, D.
 1976 “Meaning and memory.” *Forum Linguisticum* 1: 1–14.
- Bower, et al.
 1969 “Hierarchical retrieval schemes in recall of categorized word lists.” *Journal of Verbal Learning and Verbal Behavior*, 8: 323-343.
- Brown, P. & Levinson, S. C.
 1987 “*Politeness: Some Universals of Language Usage*.” Cambridge: CUP.
- Butler, C.S.
 1997 “Repeated word combinations in spoken and written text: s implications for Functional Grammar.” In C.S. Butler, J.H. Connolly, R.A. Gatward & R.M. Vismans (eds.) *A fund of ideas: recent developments in Functional Grammar*. Amsterdam: IFOTT, 60-77.

- Carter, R. & McCarthy, M.
1988 "Vocabulary and language teaching." New York: Longman.
- Carter, R.
1988 "Vocabulary, cloze and discourse: an applied linguistic view." In R. Carter & M. McCarthy (eds.) *Vocabulary and language teaching*. Harlow: Longman, 161–180.
- Conklin, K. & Schmitt, N.
2007 "Formulaic sequences: are they processed more quickly than non-formulaic language by native and nonnative speakers?", *Applied Linguistics*, 1-18
- Connell, B., J. Bridgley, & M. Edwards.
1996/1999 "Examining the Media." London: Hodder and Stoughton.
- Conrad, S.
2000 "Will corpus linguistics revolutionize grammar teaching in the 21st century?" *TESOL Quarterly*, 34/3, 548-559.
- Cowie, A. P.
1998 "Phraseology: Theory, analysis, and applications." Oxford: Clarendon Press.
- Crowther, J., Dignen, S., & Lea, D. (Eds.).
2002 Oxford collocations dictionary for students of English. Oxford, England: Oxford University Press.
- Dagut, M. and Laufer, B.
1985) "Avoidance of phrasal verbs — a case for contrastive analysis." *Studies in Second Language Acquisition* 7: 73–80.
- DeCarrico, J. & Larsen-Freeman, D.
2002 "Grammar." In N. Schmitt (ed.), *An Introduction to Applied Linguistics*, 19–34. London: Arnold.
- De Cook, S.
2000 "Repetitive phrasal chunkiness and advanced EFL speech and writing." In C. Mair and M. Hundt(eds.) *Corpus Linguistic and Linguistic Theory*, (pp 51-68), Amsterdam: Rodopi.
- Dörnyei, Z., Durow, V., & Zahran, K.
2004 "Individual differences and their effects of formulaic sequence acquisition", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.87-107), Benjamins, Amsterdam.
- Edwards, V. & Sienkewicz, T. J.
1990 "Oral Cultures Past and Present: Rappin and Homer." Oxford: Basil Blackwell.
- Ellis, Nick C.
1996 "Sequencing in SLA: Phonological memory, chunking, and points of order", *Studies in Second Language Acquisition* 18:91-126

- Ellis, N.C. & Sinclair, S.G.
 1996. "Working memory in the acquisition of vocabulary and syntax: putting language in good order." *Quarterly Journal of Experimental Psychology* 49A(1):234–250.
- Erman, B. & Warren, B.
 2000 "The idiom principle and open-choice principle.", *Text*, 20: 29-62.
- Eryıldırım, S.
 2002 "Lexical collocations in FL learning and teaching: The influence of input and transparency." In *Expanding Horizons* proceedings of the 6th METU international ELT convention.
- Ferguson, C.
 1976 "The structure and use of politeness formulas." *Language and Society* 5: 137–151.
- Fillmore, C.J.
 1979 "On fluency." In C.J. Fillmore, D. Kempler & S.-Y. W. Wang (eds.) *Individual differences in language ability & language behavior*. New York: Academic Press, 85–101.
- Firth, J. R.
 1957 "Modes of meaning". In J. R. Firth, *Papers in linguistics 1934 -1951* (pp. 190-215). Oxford, UK: Oxford University Press.
- Flavell, L. & Flavell, R.
 1992 "Dictionary of idioms and their origins." London: Kyle Cathie.
- Foley, J. M.
 1990 "Oral-formulaic Theory: A Folklore Casebook." New York: Garland.
- Foley, J. M.
 1995 "The Singer of Tales in Performance." Bloomington IN: University of Indiana Press.
- Foster, P.
 2001 "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, Peter & M. Swain (Eds.), *Researching pedagogic tasks: second language learning, teaching and testing* (Chapter 4, pp.75-93). Essex: Pearson.
- Francis, G. & Sinclair, J.
 1994 . 'I bet he drinks carling black label': a riposte to Owen on corpus grammar. *Applied Linguistics*, 15/2, 190-200.
- Frankenberg-Garcia, A.
 2005 "Pedagogical Uses Of Monolingual And Parallel Concordances." [Electronic version] *ELT Journal*. 59/3, 189-198.
- Gavioli, L. & Aston, G.
 2001 "Enriching Reality: Language Corpora In Language Pedagogy." [Electronic Version] *ELT Journal*. 55/3, 238-246.
- Granger, S (Ed)
 1998 "Learner English on computer". London and New York: Longman.

- Hakuta, K.**
 1976 "A case study of a Japanese child learning ESL. *Language Learning* 26:321– 352.
- Hausmann, F. J.
 2004 "Was sind eigentlich Kollokationen?" In K. Steyer (ed.) (2004a), pp. 309–334.
- Hill, J.
 2000 "Revising priorities: from grammatical failure to collocational success." In Lewis, M. (ed.) *Teaching Collocation*, Hove: Language Teaching Publications.
- Hunston, S.
 2002 "*Corpora in Applied Linguistics*". Cambridge: Cambridge University Press.
- Howarth, P.
 1998 "Phraseology and second language proficiency." *Applied Linguistics* 19: 24–44.
- Irujo, S.
 1986 "Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language.", *TESOL Quarterly*, 20: 287-304.
- Irujo, S.
 1993 "Steering clear: Avoidance in the production of idioms.", *International Review of Applied Linguistics*, 21: 205-219.
- Jackson, H.
 1988 *Words and their Meaning*. London: Longman Group UK Limited
- Jespersen, O.
 1924/1976 "Living grammar." In *The philosophy of grammar*. London: George Allen & Unwin, 17–29. Reprinted in D.D. Bornstein (ed.) 1976. *Readings in the theory of grammar*. Cambridge, MA: Winthrop Publishers, 82–93.
- Kennedy, G.
 1998 "*An introduction to corpus linguistics*." New York: Longman.
- Kerbel, D. & Grunwell, P.
 1997 "Idioms in the classroom: an investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching and Therapy*" 13:113–123.
- Kjellmer, G.
 1991 "A mint of phrases." In K. Aijmer and B. Altenberg (Eds.), *English corpus linguistics* (pp. 111-127). Longman, Harlow.
- Kuiper, K.
 2004 "Formulaic performance in conventionalized varieties of speech" In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.37-55), Benjamins, Amsterdam.

- Kuiper, K.
1996 “*Smooth Takers*”. *Vol. 1: Studies in Everyday Communication*, Hillside NJ, Lawrence Elbaum associates.
- Kuiper, K. and Haggio, D. C.
1984 “Livestock auctions, oral poetry and ordinary language.” *Language and Society* 13: 205–234.
- Laufer, B.
1991 “‘How much lexis is necessary for reading comprehension?’”, In P. J. L. Arnaud & H. Bejoint(eds.), *Vocabulary and Applied Linguistics*, (pp 126-132), Basingstoke: Macmillan.
- Leech, G.
1982. “*Semantics*”. Harmondsworth: Penguin Books.
- Lewis, M.
2000 “Learning in the Lexical Approach” In Lewis, M. (ed.) *Teaching Collocation*, Hove: Language Teaching Publications.
- Lewis, M.
1997 “Implementing the lexical approach”, Hove: Language Teaching Publications.
- Liu, D.
2010 “Going beyond patterns: Involving cognitive analysis in the learning of collocations”, *TESOL Quarterly*, 44.4-30
- Lightfoot, David W.
1979 “*Principles of diachronic syntax.*” *Cambridge: Cambridge University Press.*
- Lord, A. B.
1960 ” *The Singer of Tales.*” *Cambridge MA: Harvard University Press.*
- Miller, G.A.
1956 “*The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information.*” *Psychological Review*, 63, 81-97.
- Moon, R.
1998 “Fixed Expressions and Idioms in English. A Corpus-Based Approach”. Oxford: Clarendon Press.
- Moder, C.L. & Martinovic, A.
2004 “Discorse Across Languages and Cultures.” (pp.117), Benjamins, Amsterdam. 9027230781
- Nattinger, J. R. & DeCarrico, J. S.
1992 “*Lexical phrases and language teaching*”, Oxford: Oxford University Press.
- O’Dell, F., & McCarthy, M.
2008 “English collocations in use: Advanced”. Cambridge, England: Cambridge University Press.

- Ong, W. J.
1982 *"Orality and Literacy: The Technologizing of the Word: New Accents."* London: Methuen.
- Partington, A.
1998 *"Patterns and meaning"*. Amsterdam : John Benjamins Publishing Company.
- Pawley, A. & Syder, F.H.
1983 "Two puzzles for linguistic theory: nativelike selection and nativelike fluency.", In J.C, Richards & R.w. Schmidt (eds) *Language and Communication*. Longman: London 191-225
- Pawley, A. & Syder, F.H.
2000 "The one-clause-at-a-time hypothesis." In H. Riggenbach (ed.) *Perspectives on fluency*. Ann Arbor: University of Michigan Press, 163–199.
- Read, J. & Nation, P.
2004 "Measurement of formulaic sequences", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.23-37), Benjamins, Amsterdam.
- Schmitt, N. & Carter, R.
2004 "Formulaic sequences in action: An introduction", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.1-23), Benjamins, Amsterdam.
- Schmitt, N. & 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*, (pp.173-190), Benjamins, Amsterdam.
- Schmitt, N. *et al*
2004 "Knowledge and acquisition of formulaic sequences: A longitudinal study", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.55-71), Benjamins, Amsterdam.
- Schmitt, N., Grandage, S. & Adolphs, S.
2004 "Are corpus-driven recurrent clusters psycholinguistically valid?", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.127-151), Benjamins, Amsterdam.
- Signes. C.G.
2001 "Language Teaching and Media Literacy". In Ferrer H. *et al* (ed), *Teaching English in a Spanish Setting Vol. III* (pp.123-136) València: Universitat de València
- Simon, H.
1974 How big is a chunk? *Science*, 183, 482-8.
- Sinclair, J.
1991 "Corpus, Concordance, Collocation". Oxford: University Press.

- Sinclair, J., & Renouf, A.
 1988 "A lexical syllabus for language learning." In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). London: Longman.
- Spöttl, C. & McCarthy, M.
 2004 "Comparing knowledge of formulaic sequences across L1, L2, L3, and L4", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.191-225), Benjamins, Amsterdam.
- Tognini-Bonelli, E.
 2001 "Corpus Linguistics at Work.", *Studies in Corpus Linguistics* 6. Amsterdam/Philadelphia: Benjamins.
- Underwood, G., Schmitt, N. & Galpin, A.
 2004 "The eyes have it: An eye-movement study into the processing of formulaic sequences", In N. Schmitt (ed.), *Formulaic Sequences Acquisition, Processing and Use*, (pp.153-173) Benjamins, Amsterdam.
- Van der Wouden, T.
 1997 "*Negative Contexts : Collocation, Polarity and Multiple Negation.*" London, UK: Routledge.
- Weinert, R.
 1995 "The role of formulaic language in second language acquisition: A review". *Applied Linguistics*, 16(2), 180 – 205.
- Widdowson, H. G.
 1978 "*Teaching language as communication.*" Oxford: Oxford University Press.
- Wilkins, D.A.
 1972 *Linguistics and Language Teaching* London : Edward Arnold.
- Williams, E.
 1994 "Remarks on lexical knowledge." In L. Gleitman & B. Landau (eds.) *The acquisition of the lexicon*. Cambridge, MA: MIT Press, 7–34.
- Wood, M.M.
 1986 "*A definition of idiom.*" Bloomington, IN: Indiana University Linguistics Club.
- Wray, A.
 2002 *Formulaic Language and Lexicon*, Cambridge: Cambridge University Press.
- Wray, A.
 2000 Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21/4, 463-489.
- Yorio, C. A.
 "Idiomacity as an indicator of second language proficiency." In K. Hyltenstam and L. K. Obler (eds), *Bilingualism across the Lifespan*, , 1989 55–72. Cambridge: CUP.

APPENDIX

COLLOCATIONS

4783. mystery celebrity who collects a mystery item asked you to **DO** a mystery Canadian sex act. Yes. Tell us. OK Ill tell you
4821. babysitter I spent weeks with round like a ball. Mom! Just **DO** me a favor. If you know someone special not go as I did.
4844. . Bens parents are getting divorced. They are? Who wants to **DO** some coloring? Great did Lily send you? I dont have the b
4845. d s. iWhats happening? i Hey Barney want to go upstairs and **DO** stuff to me that I wont even let Marshall do? iDifferent
4846. ers and a mask of Chuck. But there came a day that I had to **DO** tests on cast. and can stay. The rest is not required. He
4857. ession If youre not prepared to do the time dont pretend to **DO** the crime and laugh and get free drinks and not stand up
4858. ctive. So whenever youve been dating somebody for a while I **DO** the Front Porch Test. You knowhow is this person gonna fi
4629. ney? Laurette? Anyway this is everything in the past now. I **DID** the best I could but sometimes I wonder Barney is as cam
5001. atch with an LCD. Ill take the CRT. Barney this is nuts. Im **DOING** just fine thank you. No youre not. You need to learn t
5005. nts anymore. When I was pitching to your "task force" I was **DOING** the most inspired work of my career. In a way you idio
6548. a baby. Get away from my computer. Okay just **DO** a quick Google search for caca spaniel. The res
6549. ght that was a figure of speech pushups like **DO** a bazillion pushups. No one can
6563. e night. Come on Lets drink green beer. Lets **DO** *green* Jello shots Wheres your St. Pattys Day sp
6567. thought you were trying to impress her. You **DO** magic. *How* is juggling any lamer than magic Mag
6600. ie. Its our first weekiversary. Hey Wendy uh **DO** *me* a favor and send a glass of champagne to tha
6614. . comi Baby are you still sore Trish made me **DO** *pushups* but she only gave me credit for ten. Th
6615. only gave me credit for ten. Then she had me **DO** *situps* and then she made me cry using only her
6618. Robin This is about... You know Ive seen you **DO** *some* bad stuff. I mean *some* really terrible stu
6636. em. And you guys. And Heloise. She helped me **DO** *the* decoupage. I dont know Barney. I dont recog
6637. Arthur in the future if you want someone to **DO** *their* best work try to motivate them not with f
6667. eed your take on this. What... what should I **DO** *Well* what I teach my kids in kindergarten... Oh
6390. e here. Give me the good news. Im sorry. You **DID** *too* good of a job writing this thing. Itsits i
8891. es inibriation style. So what you want me to **DO** a shot Oh no. I want you to
8894. Lets **DO** *all* the stuff we talked about doing and never g
8895. o much about. Wait wait wait wait wait. Lets **DO** *all* the single ladies in this joint a big favor
8896. ed. Claire is my age and she and her husband **DO** *all* this classy grownup stuff. Maybe we should

8899. es my purse Im okay. The problem is you cant **DO** any of that couple stuff unless you have someon

8904. we can sit around and cry or we can run and **DO** awesome air kicks before our leg gets chopped o

8905. etaphor. How could the leg be well enough to **DO** awesome air kicks one day and yet still so sick

8911. charity work I **DO** charity work all the tim e. Remember when I sai

8912. ins This is crazy When did it get so hard to **DO** charity work I

8922. a shot Oh no. I want you to **DO** five shots. More interesting. Barney I think yo

8925. ye Ted. Slow... Wow right. Thats why we dont **DO** shots. *Friends* dont let friends drink and dial.

. t. But it has to be alive. Wait no no Theyll **DO** lab experiments on it. Thats so mean. Shouldnt

8987. a dog person Is she open to a threeway Ted I **DO** more research than this before buying a cell ph

9008. ur time because I really like you. I want to **DO** right by you and I think the best way for me to

9014. ed I found a way for you to help someone. To **DO** some good. This is Walter. Walter is homeless.

9015. cry about it or would you run a nd jump and **DO** some awesome air kicks while you still could Aw

9076. inist or something. I know what you mean You **DO** Well Im always putting my career ahead of my re

8894. Lets **DO** all the stuff we talked about doing and never g

8895. o much about. Wait wait wait wait wait. Lets **DO** all the single ladies in this joint a big favor

8896. ed. Claire is my age and she and her husband **DO** all this classy grownup stuff. Maybe we should

8899. es my purse Im okay. The problem is you cant **DO** any of that couple stuff unless you have someon

8987. a dog person Is she open to a threeway Ted I **DO** more research than this before buying a cell ph

9008. ur time because I really like you. I want to **DO** right by you and I think the best way for me to

9014. ed I found a way for you to help someone. To **DO** some good. This is Walter. Walter is homeless.

9302. ead. I was winning Fin e. Ill go next. I was **DOING** a report on live TV about a hansom cab drive

9336. aving to go to a farmers market. Okay. Ah... **DOING** some math here. Table six has got three brid

9337. h Id just spend the day giving back you know **DOING** some good. Canceling out Barney. Exactly. Wh

9339. gonna find that someone is by going out and **DOING** stupid single stuff with Barney. What But ma

9340. tly exactly likelike just now when I saw you **DOING** the chicken dance out there Im not gonna lie

6086. I am. I cant believe you recognized me. You **DID** that story on the giant pizzas Would you check

6324. thout... Oh hey look check it out. Ted Mosby **DID** an interview in Adult Video Weekly. Adult Vide

6368. get one of these Mom Man what a show huh She **DID** some disgusting stuff. Really sticks with you.

6378. some ketchup and film Guess Teds the one who **DID** the sh opping. Ooh look a microwave pizza Gues

6380. contact. iSo that night as Lily and Marshall **DID** *their* best to hate Gael... i Hey Robin. Gael.

6578. d cut me open and crawl inside me. Yeah dont **DO** *any* of that stuff to me it skeeves me out. But.

6592. in beauty school and Lily didnt want her to **DO** *her* hair so she said she could

6648. e. Apology accepted Barney. As you can see I **DO** *just* fine on my own. I dont need a babysitter.

6649. the only weak part of my game. Im sure youll **DO** *just* fine son. Did you *just*... Im ready Bob.

6656. to teach him how to live. I even got him to **DO** *my* laundry once. I thought it was a Mr. Miyagi

6828. eat Okay Hey so I think I figured out how to **DO** *your* toast in a completely appropriate way. Fin

6548. a baby. Get away from my computer. Okay just **DO** a quick Google search for caca spaniel. The res

6600. ie. Its our first weekiversary. Hey Wendy uh **DO** *me* a favor and send a glass of champagne to tha

6601. o get in shape. I dont have to put on makeup **DO** *my* hair wear some t rendy outfit. Yeah this pla

6618. Robin This is about... You know Ive seen you **DO** *some* bad stuff. I mean *some* really terrible stu

6831. will make his move. i Hey there how are you **DOING** *Fine* Barney. I mean um fine stranger. Wow yo

6832. en happening to your Uncle Barney. i iHed be **DOING** *great* with a woman... i So I throw the press

6843. iKids its sometimes possible to think youre **DOING** *just* fine in life... i iNext message Can you

6848. robin. Curt The Iron Man Irons. Im gonna be **DOING** sports. Oh. Welcome. You look really familia

6856. od. How are you Good thank you. Great. Youre **DOING** *really really* well. Now ask her what shes

6860. e years later I ran in to Cathyi and she was **DOING** well. *Ted* Oh my God I havent seen you in so

6390. e here. Give me the good news. Im sorry. You **DID** *too* good of a job writing this thing. Itsits i

4780. r a horse; thats impressive. Barney I dont get it. You dont **DO** a damn thing in any of these clips. Exactly. Because that

4781. break these bricks with your forehead. What? But you didnt **DO** a damn thing in your video. You just stood near a horse a

4782. d surgery. Vaginal rejuvenation surgery? You know who didnt **DO** a lot of operating on vaginas in college? Who? Dr. X. He

4791. .. people who seem like bold risk takers but never actually **DO** anything. Actually doing things gets you fired. In fact I

4792. se Im a man. Youre a woman. The assumption is that you cant **DO** anything. But you have to prove society wrong. I cant bre

4843. ut Lily was right. We werent ready. We both need to move on **DO** our own thing. Still doI guess. We should make a pact. If

4865. ing I remember was waking up. i Okay Im all warmed up. Lets **DO** this thing. Turns out getting in a fight was a terrible i

4866. se kids what really happened. i Okay Im all warmed up. Lets **DO** this thing. What happened? This happened. Holy crap! iTur

FIXED PHRAES

6862. doctors office theyd say Try again. How you **DOING** *there* Robin Im good. Oh so I made a decision

6963. h by myself. Now its like two bites and I am **DONE**. *Baby* can we grow sandwiches behind the garag

6605. en he stared at me until I la ughed. Okay Im **DONE**. Great. Ill put the paints away. You just go

6607. . Im the bride you cant look better than me. **DONE**. *Im* going to get the bouquet. You guys get th

6965. dy but that chick knows what I like. Okay Im **DONE**. *But* I really dont want you reading this unle

6966. e in the winter Ill grow it back out Okay Im **DONE**. *I* know it was weird showing up with Gael. Im

6566. be about Clint Hey buddy its Clint. How you **DOING** *champ* Hey Clint. Listen um Hey hey no I dont

ÖZGEÇMİŞ

Mustafa AKSAR, 1985 yılında Kırıkkale’de doğdu. Uludağ Üniversitesi İngiliz Dili Eğitimi Bölümü’nden 2007 yılında mezun oldu. Ekim 2008 yılından bu yana Karabük Üniversitesi Yabancı Diller Yüksekokulu’nda okutman olarak çalışmaktadır.

VITAE

Mustafa AKSAR was born in Kırıkkale in 1985. He graduated from English Language Teaching Department, Uludağ University in 2007. He has been working as an instructor in Foreign Languages Department at Karabük University since October 2008.