TO BE OR NOT TO BE:

A Case Study of Formulaic Sequences in Finnish EFL Textbooks for Upper Secondary School

Master's thesis

Matti Ylisirniö

University of Jyväskylä

Department of Languages

English

October 2012

JYVÄSKYLÄN YLIOPISTO

Tiivistelmä – Abstract

Kivettyneet ilmaisut voidaan määritellä useamman sanan pituisiksi ilmaisuiksi, jotka käyttäytyvät yksittäisten sanojen tavoin. Näillä fraaseilla on aina osasanoista riippumaton merkityksensä, ja on hyvin todennäköistä, että mieli myös varastoi kaavafraasit kokonaisuuksina yksittäisten sanojen sijasta. Kaavafraasit ovat tärkeitä kielen sujuvalle käytölle ja oppimiselle, sekä kielelle järjestelmänä, sillä ne nopeuttavat kahta ensimmäistä prosessia ja ovat kulttuurisidonnaisuudessaan kriittisiä viimeiselle. Täten on perusteltua olettaa myös opetusmateriaalien sisältävän kaavafraaseja. Aiheitta on kuitenkin tutkittu äärimmäisen vähän, ja lähes ainoat tutkimuksen koskevat äärimmäisen rajoittuneita englantilaisia yliopistokirjoja. Tarkempaa tutkimusta etenkään suomalaisista oppikirjoista ei yksinkertaisesti ole olemassa, ja tutkielmani keskittyykin tähän aukkoon tutkimusperinteessä.

Tutkielmani tarkoitus oli selvittää, missä määrin kaavafraaseja esiintyy lukion englannin pakollisen viidennen kurssin tekstikirjoissa. Lisäksi tavoitteena oli saada selville, minkälaisia kaavafraaseja kirjat sisälsivät, sekä miten fraasit tuotiin esille ja opetettiin. Tutkimuksessa analysoitiin tekstianalyysin keinoin kahta suomalaista mainitulle kurssille tarkoitettua kirjaa kahdelta eri kustantajalta.

Tutkimukseni osoitti, että tekstikirjoissa oli hyvin erilaisia kaavafraaseja kuin luonnollisessa kielenkäytössä. Eritoten idiomit ja muut perinteiset sanonnat saivat osakseen suhteetonta huomiota. Lisäksi tutkimuksessa selvisi, että lähestymistavat fraasien opettamiseen olivat äärimmäisen erilaisia, vaikka kirjat pohjautuivat samaan sitovaan opetussuunnitelmaan. Ensimmäisessä kirjassa kaavafraasit olivat korkeintaan sivuroolissa, mutta toinen kirjoista oli osin suunniteltu niiden ympärille. Molemmissa tapauksissa on mitä suurimmalla todennäköisyydelle ollut kyse tietoisesta suunnitteluratkaisusta.

TABLE OF CONTENTS

1 INTRO

The current study focuses on area of lexicon called formulaic sequences. They are entities more than one orthographic word in length, which still behave as if they were single words. First of all, each sequence has a unified meaning that is at least partly independent of the component words. (Wray 2005:9) Secondly, the mind seems to store the formulaic sequences as unified entities and not as individual words. (Underwood et al. 2004:161) What is more, the importance of formulaic sequences flows directly from this behaviour. The sequences can be seen as a crucial part of language, because their existence in the mind seems to reveal something fairly profound on how the mental language storage organises itself. It has also been proven that formulaic sequences facilitate easier and faster language use, because they are available as ready-made block of meaning that do not have to constructed word-by-word under the time constraints of real-time communication. (Kuiper 2004:42) Thus formulaic sequences can give a learner an early if somewhat inflexible access to communicative competence that has not quite yet been reached. (Wray 2004:255) The blocks can be used even if one does not yet know how they are constructed. Formulaic sequences seem to be important also for the language as a system, as the majority of language can be seen as slightly formulaic. (Altenberg 1998:102) Moreover, certain forms of formulaic language belong to deep cultural layers of language, and are thus vital to understand. (Teliya et al. 1998:75) Formulaic sequences are clearly important for both the average language user and to the body of linguistic research.

As mentioned above, formulaic sequences play a vital role in how language is used and learned. Thus it is also natural to assume that they should also have a role in any learning material that wants to presents effective and accurate picture of the target language. It is particularly important for textbooks to do this because studies have shown the contents of the textbooks to dominate the language input received by the learners in the classroom. (Opettajakyselyn taulukot n.d) Simply put, the qualities of the textbook can be a deciding factor in what kind of instruction the student receives in formulaic sequences. On that account, it seems to be of paramount importance to determine how large the presence of formulaic sequences in textbooks is.

Although both formulaic sequences and textbooks have been extensively studied in the past, it is surprising that very little research has been done on the combined effect of the two issues. This gap in knowledge is particularly curious given the sequences overwhelming value to the language system, and the effect of their potential presence in any learning material. About the only studies that have been carried out on the subject have concentrated on university level EFL (Koprowski 2005) or EAP (Wood 2010) textbooks authored by native speakers. Even though these studies have provided valuable information on how formulaic sequences in the textbooks tend to be different than those in natural discourse, it is not entirely certain if this insight can be generalised. It is, after all, far from self-evident, for instance, that the same patterns would be followed by textbooks designed by non-natives for non-natives. The textbooks used by the Finnish school system are predominantly designed under such circumstances, and it is imperative that some light is shed on the situation. This is particularly important because there have been no previous studies whatsoever on the role of formulaic sequences in Finnish textbooks.

The purpose of the present study was, therefore, to explore the role of formulaic sequences in two relatively new Finnish EFL textbooks. The actual research question was threefold. How large was the presence? What kind of formulaic sequences were present? Finally, how were the sequences presented? The last question was added because previous research has shown the learning of formulaic sequences to be particularly vulnerable to changes in methodology. (Mäntylä 2004:180) The current study chose to focus on upper secondary school textbooks, because the target language of that level is already fairly advanced and can be expected to contain formulaic sequences. More specifically, the compulsory English culture course of the Finnish national curriculum was chosen, because formulaic sequences are highly culturally sensitive. The course in question has an explicit focus on cultural knowledge (Lukion opetussuunnitelman perusteet 2003), and one could thus assume that textbooks designed for the course would include formulaic language.

The present paper is structured in a simple fashion. Section two provides a detailed account of how previous studies see the properties and subtypes of formulaic language. Section three, on the other hand, focuses on the role and importance of formulaic sequences from various linguistic perspectives. Previous research on formulaic sequences in textbooks will also be explored. Section four goes into more

detail on how the current study was conducted. This account includes the full research questions along with description of the methods used in the data collection and analysis. Sections five and six, present the results of the current study. The first of these section focuses on the actual numbers of how many and what kind of formulaic sequences were found. The second result section provides a description of how the textbooks presented their formulaic sequences. The seventh section discusses the main implications of these results. The last section the will present the conclusions that can be drawn from the results. The limitations of the current study will also be discussed.

2 WHAT IS FORMULAIC LANGUAGE

In a study concerning formulaic sequences, the explanation of the theoretical background must begin with the definition of a **word**. This is done to avoid leaving the reader in an impassable terminological tangle. Aside from colloquial statements, one of the first attempts towards a scientific approach was made by Bloomfield (1933:178, as quoted by Carter 1998:5) when he defined a word as the minimal free unit of a language. This means that a true word could function all by itself as carrier of meaning and it would lose this capacity if reduced any further. For example, saying *Yes* is a sufficient answer for a query but *Y...* would make the message unintelligible. Yet this definition soon runs into issues with, for instance, the topic of the current study. Idioms such as *a turkey shoot* are independent and cannot lose any parts without losing meaning. (Carter 1998:6) Despite this, most people would say there are two or three separate words in the idiom depending on what one thinks of articles. To avoid this problem the definition could also focus on meaning instead of form by stating that a word is "the minimum meaningful unit of language" (Carter 1998:5). This too, however, is problematic as the definition neither accounts for sequences that seem to have too much meaning nor those that seem possess to none. That is to say, a *police state* can be counted as one or two words, as both parts carry some independent meaning. The conjunction *if*, on the other hand, carries almost no meaning. One could also adopt a phonological definition by saying that in a word there is but one stressed syllable. (Carter 1998:6) Yet problems again arise with compounds and their like. Finally there is also the old orthographic definition that simply states that a word is a string of letters that is limited on both sides by either a

space or a punctuation mark (Carter 1998:4), and this is the definition employed here. Despite its obvious problems with compounds and such, it has the benefit of being eminently suitable for the purposes of the current study. Firstly, it is of no consequence that the definition has innate bias in favour of written text, as the current study focuses exclusively on written material. Secondly, even the fairly glaring flaws in handling the meaning of words do not really matter as the concept is simply meant as a tool and not an accurate analysis of a lexical phenomenon. For the current study the definition offers a fairly unambiguous method of discussing the inner structure of formulaic sequences. More is not needed.

The key concept of this study is that of formulaic language and its basic unit, the formulaic sequence. Yet before these can be explained, it is crucial to understand how fundamentally arbitrary the terminology within the field is. Even while the scientific community is fairly unanimous on that the phenomenon exists, there are several different ways of referring to it. Terms that have been used range from **idiomaticity** (Fernando 1996:30) and formulaic sequence (Wray 2005:9) to **idiom principle** (Sinclair 1991:110) with even **the lexical chunk** (Lewis 1993, as quoted by Ellis 2006:128) being an acceptable scientific denomination in this context. This dilemma is further exasperated by the fact that the parallel labels are only partly interchangeable, or simply have wildly different implications. The terms above, for instance, include and exclude different parts of the phenomenon and do so using different criteria. While it is, for instance, justifiable to use the term idiomaticity as a synonym for formulaic sequence (Fernando 1996:30), the approach carries an inherent risk of implying idiom-like properties on formulaic sequences that do not have them. That is to say that the metaphorical meaning so essential for a traditional idiom is actually fairly non-existent in the wider context of formulaic sequences, and it would be sorely inaccurate to imply otherwise. Another good example of the problem is the aforementioned idiom principle which simply states that "a language user has available to him or her large number of semi-preconstucted expressions that constitute single choices, even though they might appear to be analysable into segments" (Sinclair 1991:110). While this is most certainly true and nearly identical to the definition employed in the current study, it has the same problem of unintended implications as the previous one. Moreover, as the label is never truly elaborated on it remains a vague superordinate classification, and thus offers little in

the way of details that could serve as the basis for an analysis. Any particulars beyond the level of pure axiom would have to be investigated with supplemental definitions which would make the original concept of idiom principle essentially irrelevant. The terms chunk (Willis 1990:39) or multiword-item (Lewis 1993:92) suffer from a similar problem of extreme vagueness on all levels. Even though the first of these has come to be used as the colloquial by-word for formulaic language, it has never really developed beyond its origins as a quick and easy catchall for taking holistic meaning into account in dictionary design. Thus the word chunk is critically lacking in scientific quality. For the multiword-item, on the other hand, the terminological problem is almost entirely the opposite as it is essentially too scientific. In itself it is quite commendable to develop a label that avoids all allusions and unintended meanings, but this unfortunately results in a term that tells next to nothing of the phenomenon it claims to describe. Technically multiword-item could cover anything up to and including ordinary compounds and spontaneous compositions, because the name carries no traces or implications of the critical formulaic qualities of a sequence being stored and used as a unified meaning. The label is essentially overdeveloped in its objectivity. In conclusion, it should be clear that terminology that could be brought to bear is extremely wide and varied. All of the labels presented here have their own merits and imperfections, and many of them could have very easily been chosen to serve as the basis of the current study. Thus it is crucial to understand that the labels I have chosen to use in this study represent only my personal view of what is the best way to discuss the phenomenon. Others are equally possible, but will not be used here.

Clearly there are many equally valid explanations for the phenomenon of formulaic language. The current study, however, adopts the term **formulaic sequence** for the following reasons. First of all, it has become an established term that is the most widely used one of all the alternatives. (Schmitt 2004) Secondly, the use of the word formulaic acknowledges and accounts for the fact these sequences are indeed formulas. That is to say that a formula has a distinct function and a form that is relatively fixed but allows for some variation. Thirdly, the label sequence carries an implication entity with a holistic meaning, and is wide enough to cover a vast majority of formulaic language, while still being exclusive enough to be analytically useful. These two component words essentially combine to create a definition that is

eminently suitable for analysis. Yet is must be admitted that the term suffers from some of the same problems of vagueness and the need of supplementary definitions as, for example, the idiom principle, but I consider the aforementioned qualities of the term formulaic sequence to offset the flaws. As for the definition itself, at its most basic level a formulaic sequence is a chain of separate orthographical words that nevertheless functions like a single entity or a word. In other words, the mind stores the sequence as a single unified entity and then uses it only in one fairly fixed form. Thus the term formulaic sequence denotes only and exactly what the name implies: a sequence of words stored and used as a formula (Wray 2005:9). Yet despite its accuracy the definition has one major limitation. The definition might cover a large number of instances, but it lacks details needed for case to case accuracy. Therefore, the current study incorporates Wood's (2006:21-23) five formulaic characteristics as an expansion of Wray's definition. Not only do they allow for a more detailed theoretical analysis, but they have the added benefit of having already served as a classification tool in Wood's empirical study. Due to their proven capacity it seems reasonable to assume that the criteria can be reliably used to identify formulaic sequences also in the data of the current study, and thus supplement Wray' more general description.

As was said before, Wood (2006:21-23) introduces five characteristics that indicate whether or not a chain of words is a formulaic sequence. The first of these factors is called **phonological coherence and reduction**. The criterion essentially states that a formulaic sequence is not only thought of as a single word but also pronounced as such. In other words, a formulaic sequence has a coherent intonation pattern of an individual word. Elements of reduction, such as reduction of syllables are also prevalent. (Wood 2006:21) If we take, for instance, the interjection *whatchamacallit*, it is easy to see that the sequence is pronounced with a single intonation pattern. What is more, the constituent words are clearly meshed together by the reduction of syllables, such as *should* becoming *cha*. The component words are, in fact, so integrated that the example could also have been classified as an orthographical word, but this only serves to highlight that it is difficult to draw a line between words and formulaic sequences. It should be noted, however, that this criterion might be of limited use to the current study whose data is exclusively in written form. Despite this, I have decided to include it as supporting measure that uses the pronunciation of

the sequence to draw a demarcation line in problem cases. The second criterion is called **greater length/complexity than other output** (Wood 2006:22) and it simply states that a sequence can be considered formulaic if it uses structures that are markedly more complex than those found elsewhere in the text. A learner could, for instance, use a simple *no* to express all other negatives except for *I don't understand* in requests for clarification. In such a case, the more complex form can be assumed be an undifferentiated chunk which the learner is unable to break into its constituent parts. In other words, it is used as a formulaic sequence. The third criterion is called **semantic irregularity** which states that a formulaic sequence may have a meaning that is something else than the combined meaning of the component words. (Wood 2006:22) It is, for instance, difficult to deduce that the combination of *to kick* and *the bucket* is an expression referring to death. The fourth criterion, **syntactic irregularity**, actually refers to two separate but connected criteria. On one hand, the principle means that formulaic sequences are syntactically irregular because they can be manipulated in a relatively limited fashion. Kicking anything else than a bucket would no longer refer to death. One also cannot substitute *you* in *How do you do* and still have the typical greeting. On the other hand, the criterion refers to the phenomenon that formulaic sequences do not have to follow all the standard rules of language. One can, for instance, say *to jump the gun* even though *to jump* is technically an intransitive verb. (Wood 2006:22) Wood's fifth criterion is what he himself calls **the taxonomy used by Nattinger and DeCarrico** (Wood 2006:21) and as the name implies, it essentially is a summary of the definitions offered by Nattinger and DeCarrico. As it is also the case that these definitions are covered in the **collocations** and **lexical phrases** chapters of the current study, I will not adopt Wood's fifth criterion into the definition of the umbrella term **formulaic sequence**. The contents of the criterion are already a part of my classification of the types of formulaic language and therefore they function just as Wood intended, that is, as a "guide to possible formulaicity" (Wood 2006:21), even if the current study uses them in a slightly different part of definition process.

In general, the current study defines a formulaic sequence as a chain of seemingly separate words that is nonetheless evidently used as an undivided entity (Wray 2005:9), such as the habitual greeting *How do you do?* This superordinate definition is supported by four mutually complementary criteria of phonological coherence and

reduction, greater length/ complexity than other output, semantic irregularity and syntactic irregularity. It is also important to understand that while all formulaic sequences conform to the superordinate outline, no piece of formulaic language has to follow all the four subordinate principles. The greeting above, for instance shows none of the semantic irregularity of the idiom *to kick the bucket*. Yet both of them are formulaic sequences. In other words, the general nature of any given formulaic sequence can easily be defined, but the exact characteristics are always more nebulous.

2.1 Types of formulaic sequences

2.1.1 Idioms

The following chapter deals with the subcategory of formulaic sequences called **idioms**. The term is unfortunately somewhat ambiguous, because it and the derivative **idiomaticity** are used to cover a wide range of issues. In itself there is nothing wrong with this approach as one label is just as good as any other, but problems of clarity and accuracy soon arise. It becomes difficult to tell what exactly is meant if the term idiom is used as a label for certain proverb-like expressions, idiomaticity as a catch-all term for formulaic language (Fernando 1996:30), and names such as pure idioms, semi-literal idioms and literal idioms (Fernando 1996:32) denote expressions that could just as well be subcategories of formulaic language all by themselves. While the terms themselves are accurate and well-motivated, they easily create an impression of connections that do not actually exist. In other words, not all formulaic sequences share the properties of idioms. Therefore the current study uses the terms **formulaic language** or **formulaic sequences** instead of idiomaticity, and idiom refers only to a specific subcategory of formulaic language which is presented in the following paragraphs.

In defining the phenomenon called the **idiom** the current study incorporates elements from several sources. First of all, an idiom is a formulaic sequence and as such it conforms to the superordinate definition presented earlier. For the most part the current study adheres to the definition that the two key characteristics of an idiom are the opaqueness of its meaning and the relative fixedness of all its attributes. (Moon

2006:46-47) The first one of these is fairly simple as it essentially means that the overall meaning of an idiom is not the combined meaning of its constituent words. The sequence *to kick the bucket*, for instance, does indeed have the surface meaning of foot moving towards and impacting a water container, but this has very little to do with the idiom's actual meaning as a euphemism of death. The example used here also highlights that the meaning of an idiom is not only opaque but often also metaphorical. Fixedness, the second part of the definition, means that there can be relatively little variance in the form or the meaning of an idiom. (Moon 2006:46, Fernando 1996:30) The aforementioned sequence *to kick the bucket*, for instance, could technically be modified with any number of nouns or verbs resulting in combinations such as **to kick the spoon* or **to carry the bucket*, but neither of these would retain the original allusion to death. There are, however, some exceptions to this, as will be shown later. In other words, an idiom is what an idiom is, and it can only be modified in a limited fashion. In conclusion, for the purposes of the current study an idiom is a discrete word sequence with holistic and often metaphorical meaning and form that can be varied only in a limited fashion.

Idioms are by definition fixed in both form and meaning, but it is crucial to understand that this is only relative, and that there is some leeway in both criteria. The image of a fixed and opaque idiom is a stereotype that holds true for the vast majority of cases, but it is not the entire truth. A good description of variance in meaning is provided by Fernando (1996:71-72), but before moving on with the portrayal, two caveats must be given. First of all, the model will be used only as it applies to what Fernando calls **pure idioms** (1996:32) as this is her closest equivalent to how idiom is defined in the current study. Using her entire representation would create unnecessary overlap as she uses term idiom to cover formulaic sequences the current study has chosen to place under different labels. Secondly, as the model itself is not an exhaustive one, neither shall this description be. It is merely intended to give example and show the basic principle along with some of its implications. The first category, the variance of form, is fairly simple and brings no contradictions to the primary definition of an idiom, because the issue here is simply that some idioms allow the change of some internal elements without any resulting shift in meaning. (Fernando 1996:71) It is, for instance, entirely possible to make an allusion to torrential rain with both *it's raining cats and dogs* and *it's*
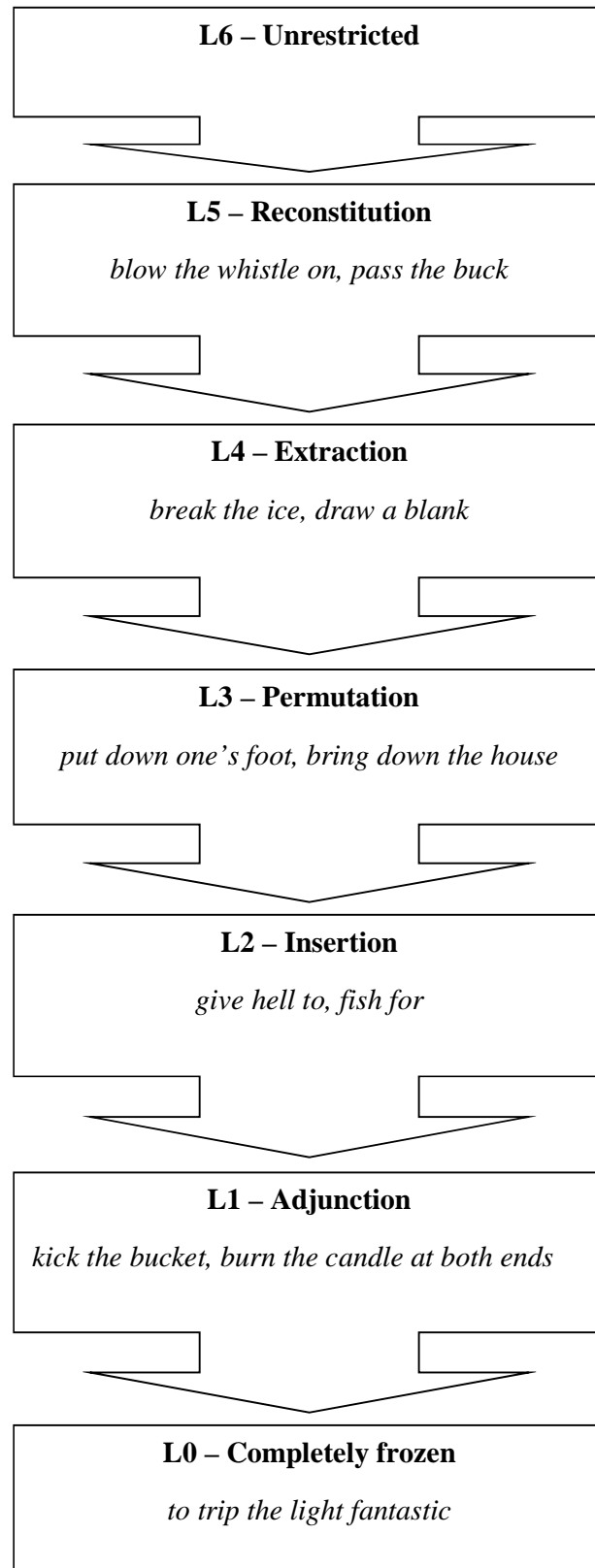
pouring cats and dogs. This and other similar examples simply represent parallel forms for the same metaphorical meaning. Fluctuation in the degree of metaphoricality, however, is more difficult to incorporate into the primary definition. Sometimes it is simply the case of an idiom having both a literal and non-literal interpretation, as for instance in *to roll out the red carpet* (Fernando 1996:71), where the relative literalness of the idiom is likely the results of its novelty (Wray 2008:28-30). In other words, people still remember the original context and the actual physical carpet. It might also be that only one of the words in the idiom has a non-literal meaning, such as in the expression *a thumbnail portrait*, where only the first word is metaphorical, and even it can be deciphered with the help of the literalness of the second one. (Fernando 1996:71) These examples should highlight that it is not a question whether or not idioms are metaphorical, but rather that there are multiple ways and degrees of being metaphorical. In essence, some kernel of metaphor must remain, or a sequence cannot be classified as an idiom.

It would seem to be the case that idioms are not as fixed as the overall definition implies. To assume otherwise would render some clearly idiom-natured examples of data invisible to the current study and create unacceptable gaps in its coverage. Therefore it is fortunate that the syntactic variation among idioms can be shown to operate according to certain specific principles, and that idioms can indeed be classified into a hierarchical according to their relationship to these rules. (Fraser 1970:41) In essence, Fraser states that there are the five principles of **adjunction**, **insertion**, **permutation**, **extraction** and **reconstitution** that can be seen in all syntactic variation among idioms. (Fraser 1970:36-37) The first one of these, adjunction, simply means that some non-idiomatic unit is added to the idiom itself, such as when *John kicked the bucket* is nominalised into *John's kicking the bucket was sad* by the adjoining of the possessive marker and the –ing ending. (Fraser 1970:37) In essence one can treat adjunction as grammatical inflection of the idioms component words. The principle of insertion, on the other hand, works by adding entire constituent words into the idiom sequence. This happens, for example, when the indirect object that comes after the idiom itself in the *John read the riot act to the class* is moved to within the idiom in *John read the class the riot act*. (Fraser 1970:37) The third principle, permutation, is highly similar as it operates when two components of the idiom can switch places. This is the case when the particle and the

noun phrase in the idiom *lay down the law* permutate into *lay the law down*. The crucial thing to understand is that the permutating units must be internal components of the idiom, or else the process in question is that of insertion. (Fraser 1970:37-38) The principle of extraction also works with the internal components of an idiom, but does that by making them external. This happens, for instance, when the passivisation of *Her father laid down the law* removes the object from within the idiom and places it first in the sentence resulting in *The law was laid down by her father*. (Fraser 1970:38) The final process, reconstitution, is the most radical one because it changes the idiom's entire syntactic structure. A good example of this is when the idiom sentence *He laid down the law to his daughter* becomes a noun phrase in *His laying down the law to his daughter*. (Fraser 1970:38) Yet the key thing about Fraser's classification are not the classifications themselves but how they combine into what he calls the "Frozenness Hierarchy" (Fraser 1970:39), a seven stage ladder beginning with the category completely frozen and ascending through adjunction, insertion, permutation, extraction and reconstitution with the category unrestricted resting on the uppermost rung. The idiom's level on the model tells the upper limit of its flexibility because an entity on one rung of the ladder can also be subjected all the processes on the lower steps but not to those on the upper levels. (Fraser 1970:39) The expression *to fish for something*, for instance, is located on the level of insertion and is thus also capable of adjunction, but not permutation which is the immediate upper level. It must also be said that levels completely frozen and unrestricted are not actual processes, and thus they represent the theoretical opposite borders of idiom flexibility. Very few idioms are completely frozen, and a syntactically unrestricted sequence is not an idiom.

Figure 1. The frozenness hierarchy. Adapted from Fraser (1970:39)

L6 – Unrestricted

L5 – Reconstitution

blow the whistle on, pass the buck

L4 – Extraction

break the ice, draw a blank

L3 – Permutation

put down one's foot, bring down the house

L2 – Insertion

give hell to, fish for

L1 – Adjunction

kick the bucket, burn the candle at both ends

L0 – Completely frozen

to trip the light fantastic

Idioms can also be classified according to their function. This division is included in the current study in order to allow not only the analysis of the presence of idioms but also of what they are used for. Fernando (1996:72-74) constructs a functional classification by first dividing idioms into three broad categories: ideational, interpersonal and relational idioms. The current study, however, will only focus on the first one of these, because the last two are lacking in the essential metaphorical quality and therefore are not idioms in the sense meant by the current study. What is more, they denote more or less the exact same set of formulaic sequences as lexical phrases (Nattinger and DeCarrico 1992), and as such are discussed elsewhere in this study. **Ideational idioms**, however, will be discussed because they have qualities that could make them not only a subset of idioms but also a rough synonym for the entire class. Ergo a functional categorization of ideational idioms would be a categorization of all idioms. One caveat, however, must be given right away. Ideational idioms also include sequences such *to be exact*, which characterize the nature of the message (Fernando 1996:98), and thus fall under the category of lexical phrases. The contents of this chapter do not apply to them. The overall similarity is apparent, first of all, in the form of the majority of ideational idioms. Ideational idioms are usually sequences shorter than a clause, such as a nominal form of *backseat driver* or the verbal nature of *to smell a rat*. They can, however, be full clauses, which often is the case with proverbs such as *barking dogs seldom bite*. (Fernando 1996:98) Proverbs such as the example are commonly held to be almost stereotypical idioms, and the fact that Fernando has classified them as ideational idioms gives further credence to the claim that terms idiom and ideational idiom could be treated as synonyms. The second point of convergence is the overall function or meaning of ideational idioms, which is to describe "the speaker's or writer's experience of the world" (Fernando 1996:97). It does not require a grand leap of faith to say this is nearly the same thing as the idiom definition used by the current study. The sequence *to kick the bucket*, for instance was used as an example of a stereotypical idiom, and it could equally well be described according to the functional definition of an ideational idiom. Even the metaphoricality is just a method of describing such an experience, and as long as one makes sure that metaphor requirement is fulfilled by the potential idiom, it is justifiable to use ideational idioms both as synonym and a functional categorization for idioms. The actual classifications range from **actions** such as *wave out an olive branch* to

emotions such as *for one's blood to boil*, and it is important to understand that they are not meant to be an exhaustive and utterly exact categorization of idioms. They are supposed to be a tool for understanding what idioms are commonly used for. (Fernando 1996:72) These classifications are provided in Table 1.

Table 1. The functional classification of the message content of ideational idioms. Adapted from Fernando (1996:72-73)

Function	Examples
Actions	*give somebody and inch and he'll take a mile, wave out an olive branch*
Events	*the straw that breaks the camel's back, out of the mouth of babes*
Situations	*to be in Queer street, to be in a pickle*
People and things	*a back-seat driver, a fat cat*
Attributes	*lily-white, as green as grass*
Evaluations	*beauty lies in the eyes of the beholder, a watched pot never boils*
Emotions	*for one's blood to boil, lose one's heart*

2.1.2 Lexical phrases

The transparent end of the opaqueness spectrum holds the type of formulaic sequence called **lexical phrases**, which were first introduced by Nattinger and DeCarrico (1992:36). Though they are markedly different than the other categories presented in the current study, a lexical phrase is a sequence of words ranging in length between two words and nearly an entire sentence. Like all other formulaic sequences it is stored, retrieved and used as a whole. Yet it differs from the traditional formulaic sequences such as idioms in that its meaning is fairly transparent and the internal structure often follows the standard grammatical rules. Yet the most important defining characteristic of a lexical phrase is that it has a well-

defined syntactic and pragmatic function. These rules are best explained by using an example. The expression *How do you do?* fills the aforementioned requirements in the following way. First of all, the sequence is self-evidently longer than a single word, yet it is still only ever used in this unaltered form. In other words, the chain is like all other formulaic sequences in that it expresses a single unified meaning relatively frozen in form while being comprised of multiple words. Secondly, the expression is a lexical phrase because its meaning is still relatively easily deduced from its component words, and also because it has a grammatical structure that usually follows the standard grammatical rules of English. This is markedly different than idiom, which follow neither of these rules and can thus be utterly opaque to the uninitiated and have an internal structure completely at odds with the rules of grammar. Lastly and arguably most importantly, *How do you do?* is a lexical phrase because the expression holds a very specific function of being the standard culturally accepted greeting in the English language. When all is said and done, the key issue about lexical phrases is not the form of the form but how and why the expression is used. If there is a standardized multiword expression relatively locked in form, which without any additions or subtractions is necessary and sufficient for performing a function in discourse, then it is a lexical phrase. (Nattinger and DeCarrico 1992:36)

In addition to the attributes that define **lexical phrases** in relation to other *formulaic* sequences, there is a set of parameters that draw distinctions between different kinds of lexical phrases. As presented by Nattinger and DeCarrico (1992:38) these four constants are 1) length and grammatical status, 2) canonical or non-canonical shape, 3) variability or fixedness and 4) continuity or discontinuity. The first of these categories is also the simplest one, as it describes the outward appearance and the basic function of the phrase. In other words, the phrase might be a short pair of words such as *so long* or nearly an entire sentence such as *once a upon a time*. What is more, both of these have very different syntactic functions, as the first one "behaves very much like an individual lexical item" (Nattinger and DeCarrico 1992:38), whereas the other one is an utterance in its own right. The second category explains whether or not the form of a phrase follows the standard rules of grammar. The phrase *waste not, want not*, for instance, is non-canonical because it does not use the conditional marker *if* even though a conditional phrase would technically require it (Nattinger and DeCarrico 1992:33). The third category defines how flexible or

inflexible the form of the lexical phrase is. For example, the phrase *a good time was had by all* allows multiple permutations along the lines of *a good time was had by none, a good times was had by them* (Nattinger and DeCarrico 1992:34), while the aforementioned *how do you do* only allows for this particular form and no other. Thus one cannot say *how do they do* and not have the greeting sound slightly off. Lastly there is the parameter of continuity, which describes whether or not the components of the lexical phrase must be used in an unbroken progression (Nattinger and DeCarrico 1992:38). The structure of *how do you do*, for instance, is instantly broken by any additions while the phrase pair *once a upon a time* and *they lived happily ever after* (Nattinger and DeCarrico 1992:40) can have nearly anything between these two parts. Moreover, it is important to understand that all the separate categories represent continuums rather than binary oppositions. Having characteristic from one extreme end of a parameter does not necessarily rule out some contrary properties, as was shown by the aforementioned examples (Nattinger and DeCarrico 1992:38). All in all, the four criteria given here are a robust method of handling an otherwise nebulous category, and thus the current study will use them as the primary method of structurally categorizing lexical phrases.

The third set of demarcation lines within the category comes from the functions of the lexical phrases. In other words, **lexical phrases** can be classified according to what they are used for and thus there are the following categories: 1) social interactions, 2) necessary topics and 3) discourse devices. (Nattinger and DeCarrico 1992:60-64) While it might seem that this division is in conflict with the system presented in the next paragraph, this is not the case. It is more of a case of looking at lexical phrases from the perspective of their discourse functions. The two views are complementary. The first of these functional categories, **social interactions**, does exactly what the name implies – they construct and reflect social relations inherent in communication, and as such they can be further divided into two categories. On one hand, we have items of **conversational maintenance** which govern the flow of conversation. One could, for instance, start with *excuse me* (Nattinger and DeCarrico 1992:61) or use similar phrases for maintaining or ending the conversation. One the other hand, there are the tools of **conversational purpose** which echo the objective of the communication, such as when one wants to be polite and uses the phrase *if you don't mind*. (Nattinger and DeCarrico 1992:62) The second main category,

necessary topics, is somewhat dissimilar because it states there are not only common daily topics such as weather and shopping but also conventional ways of talking about them (Nattinger and DeCarrico 1992:63). It is not, for instance, uncommon to address a waiter and say *Check, please*. Lastly, we have the category of **discourse devices** which are "lexical phrases that connect the meaning and the structure of the discourse" (Nattinger and DeCarrico 1992:64). These artefacts are textual tools used for making sure that text flows and conveys the intended meaning logically. A good example of this is the exemplifier *in other words* that is used extensively in this text. In conclusion, the functional division of lexical phrases does exactly and only what the name implies. It does not describe what the phrases are; it only tells what can be done with them, and considering that lexical phrases are defined by their functions, this is an important division indeed.

2.1.2.1 Polywords

In the next four sub-chapters I will present the way in which Nattinger and DeCarrico have used their own structural and functional criteria to categorize lexical phrases. The resulting four categories are **polywords, institutionalised expressions, phrasal constraints** and **sentence builders**. It must again be stressed that the term lexical phrase or its subcategories are not the only method that could be used for classifying the phenomenon. They are, however, the most established and tested one, and therefore they compare favourably to newer alternatives that can be somewhat vague and ill-defined. The purely frequency based approach of lexical bundles, for instance, is very similar to the classification used here, but it does not make sufficient differentiation between pieces of formulaic language and other naturally occurring sequences. (Biber et al. 2004:373) While both lexical phrases and lexical bundles cover the same core of expressions, the latter classification covers many lexical entities that can be seen as ad hoc results of the language's generative rules. Thus to avoid the needlessly confusing situation of endless second-guessing whether or not sequences like "is going to be" and "have a lot of" (Biber et al. 2004:381) are formulaic language, the current study will use categories of Nattinger and DeCarrico as the best available option while still acknowledging their somewhat rigid and arbitrary nature.

The first form based subcategory of lexical phrases is called **polywords** and they adhere to the four aforementioned structural criteria in the following fashion. First of all, they are only a few words in length, and thus have a syntactic role similar to compounds or short interjections. Tools of **social interaction** such as *by the way* or *all right* are a good example of this category (Nattinger and DeCarrico 1992:65). One should also pay special attention to the fact that some polywords have become so conventional they are no longer even written as sequences of different words. Despite their unconventional form expressions such as *notwithstanding* are classified as polywords due to their etymology and function. (Nattinger and DeCarrico 1992:39) Polywords' relation to the second structural criteria is that depending on the case, they may or may not follow the standard grammar of a language. There are both canonical expressions such as *I'll say* and non-canonical ones such as *by and large*. (Nattinger and DeCarrico 1992:38) The third and the fourth criteria of variability and continuity are closely linked with regards to polywords, as the representatives of the category are highly fixed in form. In other words, expressions like *by and large* can only ever be used in one form and nothing can be inserted between the component words. (Nattinger and DeCarrico 1992:38)

2.1.2.2 Institutionalised expressions

The second subcategory is called **institutionalized expressions** and it is somewhat different than the previous one. First of all, institutionalized expressions can be as long as whole sentences and therefore they carry the function of independent utterances. A stereotypical sample of this category is the example used several times in the current study. The socially accepted greeting *how do you do* can be classified as part of the category due to its sentence like form and function. The last three structural criteria have a somewhat more nebulous relationship with the subcategory as institutionalized expressions can be both canonical and non-canonical, though there is a considerable preference on the former. Moreover, even though there is no variance of form, institutionalized expression allow for some discontinuity, which makes their form slightly freer than that of polywords. (Nattinger and DeCarrico 1992:39) In other words, while *how did they do* is not a lexical phrase, it is entirely acceptable to insert quite a lot between the components of the phrase pair *once upon a time...and they lived happily ever after* (Nattinger and DeCarrico 1992:40).

2.1.2.3 Phrasal constraints

The key issue to understand about **phrasal constraints** is that one cannot think of them as phrases in the typical sense of the word. It could be said that they are more like patterns or constraints that provide a readymade blueprint for constructing a wide variety of similar expressions. (Nattinger and DeCarrico 1992:41) If we take, for instance, the phrase *a year ago*, we quickly realise that the English language has a great number of similar time related expressions. We could even say that there exists a pattern underlying this all, and this is indeed the root of the issue. In other words, phrasal constraints are relatively short lexical phrases that support near limitless variability of some grammatical category at one or more points in their structure. The example presented here, for instance, can accept almost any time related noun to be inserted between the indefinite article and the word *ago*. That is, *a month ago* and *a nanosecond ago* both conform to the pattern just as well as the aforementioned example. In addition to the canonical expressions such as our example, the category includes sequences like *the sooner, the better* which do not follow the standard syntax. Despite their highly flexible nature, the *phrasal constraints* support discontinuity, the fourth criterion of a lexical phrase, only on rare occasions. (Nattinger and DeCarrico 1992:42)

2.1.2.4 Sentence builders

The last category of lexical phrases is formed by the **sentence builders** and they do exactly what the name implies. They are conventional expressions or formulas that provide a framework or a starter for expressing an idea, and they are almost of sentence length themselves. The basic structure of *a sentence builder* is that is has an expression such as *my point is that* which is followed by an empty slot that holds almost any idea or opinion. (Nattinger and DeCarrico 1992:41) Consider, for example, all the possible permutations of sentences beginning with the example above. In this regard, sentence builders are highly similar to *phrasal constraints* in that they too are less a type of expression than a pattern that allows great variability at one or more points in the structure. In this case, however, the slot is not filled with single words but whole sentence or entire ideas. As this category is essentially a collection of framing devices, it should come as no surprise that they are highly flexible in other ways too. That is to say, that sentence builders can be both canonical

and non-canonical in nature, and what is more, allow for highly discontinuous structures. (Nattinger and DeCarrico 1992:43)

2.1.3 Collocations

For the purposes of the current study, idioms and lexical phrases are the two main categories of formulaic language. Yet they are by no means all that there is. In particular, the current the current study has to consider lexical collocations, as they are arguably the most common variety of formulaic language (Moon 1998). Unfortunately the sheer size of the category and certain other properties discussed later make it impossible for the current study to take collocations into account except for in some highly specific circumstances. These circumstances will also be discussed later. Thus the following discussion is intended to serve as a brief introductory acknowledgement of the importance of collocations. The discussion should also highlight why the category is for the most part beyond the scope of the current study.

In principle, collocation is not a particularly complex phenomenon. In fact, it can be described as "group of words which occur repeatedly in a language" (Carter 1998:51). This definition is, however, somewhat vague and there are several complementary definitions, of which three are presented here. They have been chosen both because they seem to be among the most common approach and because they serve to highlight the somewhat problematic nature of the phenomenon. These three methods are, as presented by Herbst (1996:380), the text oriented approach, the statistically oriented approach and the significance oriented approach. The first of these is brought forward, for instance, by Halliday (2004:11) for whom a collocation is a way of saying that certain words tend to occur together because they share context. That is to say that one can expect to find both *fork and knife* in a text discussing kitchen, and thus one possible way of seeing collocations is to imagine them as a textual property. This is not all too dissimilar to the statistically oriented approach according to which "collocation is the occurrence of two or more words within a short space of each other in a text" (Sinclair 1991:170). The only major difference here is that of perspective, as approach only aims to define collocations from a point of view that is independent of all other linguistic influences. Thus the approach sees collocations as sequences of words that occur more often than could

be predicted by pure statistical chance. (Herbst 1996:382) The third definition, the significance oriented approach, offers a slightly different point of view as it defines collocations as "semi-finished products of language" (Hausmann 1984:398, as quoted by Herbst 1996:382). This definition is supplementary to the other approaches, as its stated aim is to distinguish between those frequently occurring sequences that are purely a result of regular language rules and those that have significance in themselves. (Nesselhauf 2005:14) In other words, the approach helps to separate purely grammatical but highly common sequences such as *I'm going to* from more meaningful entities such as *to foot the bill*. Based on these three definitions it is possible to say that collocations are statistically recurring sequences of words that may be context sensitive and may have independent meaning. By applying simple logic it is easy to deduce collocations could be a very common language phenomenon. Unfortunately, from these three definitions also arises the reason why collocations are beyond the scope of the current study. Firstly, if collocations are taken as a statistical phenomenon, they become impossible to include because a handful of textbooks simply is not a large enough corpus to draw any meaningful statistical deductions on. It would be impossible to say what a significant sequence is and what is not. It might be possible to use the textual approach to pick out collocations from the textbooks, as one could assume that material intended for learners is loaded with sequences relevant to the context. The data of the current study is, however, once again too small for that to work. Even the significance based approach is problematical, as it is all too similar to definitions of other formulaic sequences. In fact, some writers (Cowie 1994, as quoted by Nesselhauf 2005:15) include even idioms as sub-category of collocation. Thus the use of the significance approach would lead to an unavoidable confusion of terminology and classification. For these reasons the current study must exclude collocations even though they seem to be one of the most important parts of formulaic language. This rule has one exception. The current study focuses on learning material, and it is entirely possible that the textbooks explicitly label and bring to fore some collocations. These potential cases represent intentional focus on formulaic language, and it would create unacceptable gaps in the coverage of the study to not take them into account. This, however, is the only exception to the categorical rejection of collocations.

3 THE IMPORTANCE OF FORMULAIC LANGUAGE

This chapter focuses on a very simple goal: explaining why formulaic language is important. An entire chapter is devoted for this purpose to offset the fact that the definitions and subdivision of the phenomenon are somewhat arbitrary, and without careful explanation the reader might be left with a faulty impression that the phenomenon would be somehow questionable or lacking in empirical support. While the study of formulaic language faces the same issues as any field that touches of psycholinguistics, namely being unable to directly observe the inner workings of the mind, then within those limits the evidence for the role of formulaic sequences is fairly conclusive. Moreover, statistical data derived from corpuses also gives strong support for existence and the importance of formulaic sequences as a language phenomenon. Finally certain pedagogical studies have shown that formulaic language has a crucial function in the process of language learning.

3.1 The statistical frequency of formulaic language

Formulaic sequences are a phenomenon of language, and thus any statistical representation of language also includes a statistical representation of formulaic sequences. This seemingly self-evident conclusion is supported, for example, by statistical corpus studies that clearly indicate that people do not use the full extent of creativity allowed by the language. A study on the London-Lund corpus of spoken English, for instance, came to the conclusion that up to 80 % of the words in the corpus belonged to recurrent word combinations. (Altenberg 1998:102) Even though the number might seem fairly conclusive, it is in fact far from unambiguous. First of all, the corpus in question is based solely on spoken language, and one should thus be careful in generalising its findings to hold for all uses and mediums of language use. Secondly, the length of the corpus is only about 500 000 words long, which is somewhat less than is usual. Thirdly and most importantly, it is crucial to understand that the study discovered a wealth of **recurrent word combinations** and not formulaic sequences. While the two terms have some considerable theoretical and practical overlap, there are some critical differences. There were in fact very few recurrent word combinations that conformed to the classical qualities of formulaic language. Idioms, for instance, were practically non-existent in the data. (Altenberg

1998:120) What is more, few of the combinations in evidence were either semantically opaque or had a completely fixed form. In fact, many of the recurring combinations such as "a number of" (Altenberg 1998:120) could be considered to have been generated entirely through regular syntactic rules. (Altenberg 1998:121) Thus it is highly interesting to note what formulaic qualities the recurrent combinations did show. The pattern that kept repeating throughout the corpus was that nearly all recurrent clusters were tied to specific pragmatic functions such as agreement or greeting. (Altenberg 1998:104) For each function there seemed to be one, or at least a very limited set of somewhat fixed expressions that were being used over and over again even though other options would have been equally valid syntactically. Even though all of these sequences did not have a full set of stereotypical formulaic properties it is clear that there were "routinised ways of unfolding and presenting information in continuous discourse" (Altenberg 1998:121). The concept is very similar to what Nattinger and DeCarrico (1992) mean when they discuss conventionalised ways expression. Considering that this is their overall definition for a lexical phrase, it is possible to draw some tentative conclusions on Altenberg's (1998) study. On a general level, it can be argued that even though the fixed end of formulaic spectrum is not in evidence, spoken discourse is riddled with or even ruled by an assortment of prefabricated phrases that are tools for realising particular pragmatic functions. The statistical evidence would seem to tentatively support this conclusion, and it is indeed what Altenberg (1998:120) himself concludes. On the more specific level of the current study it can be argued that spoken discourse seems to be the domain of what Nattinger and DeCarrico (1992) call lexical phrases. Thus it seem that while the classical forms of formulaic language are exceedingly rare, the more flexible end of the spectrum is one of the governing factors in communication.

It was established in the previous paragraph that formulaic language does indeed have a role in spoken discourse. Yet the written language must also be discussed and it is interesting that the role seems to be markedly different than its equivalent in spoken discourse. This conclusion is supported by, for instance, a study of the Oxford Hector Pilot corpus; a database of written British English consisting of approximately 18 million words with the focus of the corpus on journalistic language. The corpus also had small sample of transcribed spoken communication,

but as this accounted for only 3 percent of the data its effect can be considered negligible. (Moon 1998:80) The study compared the data of the corpus against a set of 6700 mostly classical formulaic sequences such as idioms derived Collins Cobuild English language dictionary (Moon 1998:79-81). The overall results of the study were fairly conclusive as 70 % percent of the sequences occurred less frequently than once in a million words with 40 % having no presence at all outside random occurrence (Moon 1998:82). With this caveat in mind it is interesting to note what type of expressions did have a notable presence. That is to say that 46 % of the expressions found in the corpus were collocations of some sort (Moon 1998:85) with all the sequences that occurred more than 100 times per million words belonging to this category (Moon 1998:83) The more fixed formulaic sequences such as lexical phrases, idioms and proverbs occurred almost at random (Moon 1998:85) with the only truly common ones being those with exceedingly transparent meaning, such as *play part in* or *take steps to* (Moon 1998:87). Yet despite their rarity these expressions covered 54 % of the occurrences within the corpus. (Moon 1998:85) Thus some tentative conclusions can be drawn. First of all, written language would seem to be much less formulaic than spoken language. This might be because the written medium places much lighter time constraints on communication, and thus allows for the use of the full flexibility of the language system. One should also take note that although idioms and their like are very infrequent there seems to be a considerable number of them in written discourse. Moon's (1998) study itself found several thousands of separate classical formulaic sequences, even if any given one occurs more or less randomly. One could thus argue the role of the more fixed sequences is to be recognized rather than actively used. The results of the study and their implications must be, however, taken with some caution. The study was after all conducted by comparing a predetermined list against a corpus and not by actually checking which clusters recur naturally within the corpus. While these two things can reasonably be assumed to be more or less the same thing, certainty will not be reached until the latter method is used.

3.2 The psycholinguistic perspective

Based on the statistical evidence it would seem that formulaic sequences are a cornerstone of language use. Yet it still remains a question whether or not the

sequences derived from a corpus area actually what the mind uses. The psycholinguistic perspective is not directly connected with the study of learning materials, but it still plays a crucial role. It is the opinion of the current study that it is impossible to achieve any meaningful understanding of the role of formulaic sequences in learning material, if one does not take into account how the mind processes them. To do otherwise would discount an important body of scientific proof for the existence and importance of formulaic language.

One of the most convincing attempts to shed light on the situation was done by Underwood et al. (2004) in a study where test subjects were presented with a reading task during which their eye-movements were measured. The reasoning behind the procedure relied on the psychological phenomenon that the longer it takes for the mind to understand and process a word the longer the eye fixates on the word in question. (Just and Carpenter 1980:330, as quoted by Underwood et al. 2004:154) Consequently, it was hypothesized that if the mind indeed stores formulaic sequences as holistic units of meaning then the words within them would receive fewer and shorter fixations than the same words in non-formulaic contexts. (Underwood et al. 2004:123) In other words, formulaic sequences would require less processing time from the mind and would not be particularly fixated on as they would essentially be large well-known words. (Sinclair 1991, Pawley and Syder 1983) To test this hypothesis the reading text was embedded with some fairly typical formulaic sequences such as the idiom *by the skin of his teeth*. (Underwood et al. 2004:169) Subsequently the results seemed confirm the hypothesis almost completely as formulaic sequences did undeniably receive significantly fewer and shorter fixations. (Underwood et al. 2004:161) In the light of these results it seems fairly evident that there is at least some truth in the maxim that formulaic sequences are holistically stored, and that they are an actual psycholinguistic phenomenon.

It has now been established that formulaic sequences have at least some real psychological presence, and that this presence manifests in way that seems to be concurrent with the theories regarding formulaic sequences. Yet the issue is not this simple. There are, for example, the unvoiced assumptions that all corpus-derived sequences are actual formulaic sequences and that all formulaic sequences would offer the same kind of advantage in processing load. Both of these conjectures seem to have some serious weaknesses as was proven by Schmitt et al. (2004) in their

dictation test regarding formulaic sequences. The aim of the study was highly similar to that of Underwood et al. (2004) as this study too was designed to gauge whether formulaic sequences are stored as holistic entities. The study also had an explicit aim of discovering if seemingly formulaic sequences drawn from corpus data were actually formulaic sequences. (Schmitt et al. 2004:128) Thus the study constructed a dictation test in which 25 sequences were embedded in text which the participants heard and had to repeat to a tape recorder under a time-constraint. This was done in order to insure that the participants would not be able to accurately rely on either their short time memory or generative capabilities given by knowing the syntactic rules of the language. (Schmitt et al 2004:131) The working hypothesis was that any true formulaic sequence would stand out as a more fluent stretch of language because in principle they should have been readily available as holistic entities in the participants' minds. (Schmitt et al. 2004:131) The results were, however, somewhat more ambiguous than in Underwood et al. (2004). There was certainly an overall trend that the 25 sequences were produced more accurately than other dictation material. (Schmitt et al. 2004:142) Yet all the participants also showed hesitation and tendency to transmute the target sequences into semantically similar forms such as *as a consequence of* becoming *as a result*. (Schmitt et al. 2004:146) These results led the study to conclude that some sequences are indeed readily available in the mental lexicon, but it is very difficult to say whether they are available as holistic meanings or rules that allow a rapid reconstruction of grammatical items. In other words, the conclusion of the study was that "corpus data on its own is a poor indicator of whether those clusters are in the mind as wholes" (Schmitt et al. 2004:147).

Yet before one uses Schmitt et al. (2004) to discount the entire concept of formulaic sequences as holistic entities some crucial details need to be explored. First of all, the key word in the quote presented in the previous paragraph is *alone*. Schmitt et al. (2004) does not deny the existence of holistic meaning in formulaic sequences, but simply states that one cannot rely solely on corpus data in its discovery, as many of the language's recurring clusters may be simple grammatical constructions. Secondly, the truly fixed end of the formulaic spectrum was notably absent in the target cluster of the study, with the potentially lexical phrase *you know* being one of most formulaic expressions included. (Schmitt et al. 2004:130) The ambiguous results could just be the result of the exclusion of idioms and other clearly holistic

entities. All in all, the study by Schmitt el al. (2004) serves to highlight that formulaic sequences are a phenomenon that is increasingly difficult to pin down. On one hand, statistical data does not seem to be entirely accurate. On the other hand, any true verification of the said data runs into the traditional problem of psycholinguistics that the inner workings of the mind cannot be directly observed. Despite these caveats the studies Underwood et al. (2004) and Schmitt et al. (2004) combine to create relatively reliable conclusion that formulaic sequences actually do have a psycholinguistic presence. One merely needs to be cautious in applying this conclusion of holistic meaning on seemingly formulaic elements.

Formulaic language does indeed have a real, albeit somewhat complicated, psycholinguistic presence. The basic implications of this presence are fairly simple as most theories about formulaic sequences hinge on the assumption that these entities make communication more fluent by easing the processing load of the mind. (Wray 2005:101) This in turn hinges on some basic assumptions on how language is processed. The traditional Chomskyan view has been that the mind has a limited memory capacity but has access to abundant online processing, and thus it would be easier and faster to use comprehensive syntactic rules to construct ad hoc meanings than to recover existing meanings from memory. (Skehan 1998:31) Not only does this approach present a neatly defined and systematic account of language processing but it also explains the language's natural flexibility. On the other hand, if it is assumed that it is the processing capacity that is severely limited and the memory that is fairly abundant, then formulaic language begins to offer the aforementioned benefits. The situation with both of these approaches is akin to being given a pile of Lego-blocks and being instructed to build a house. With the traditional view the blocks are the small ones (words) that can be used to build practically anything, and the builder is given detailed instructions (grammar system) on how to combine the pieces into a meaningful whole. This method indeed results in a house that can even be quite creative but during the construction one has to constantly look at instruction sheet which slows down the process, but if time is abundant, then this is a non-issue. Yet if the blocks are some of the larger units such as ready-made roofs and walls (formulaic sequences) with labels indication the proper use clued on them, it should be fairly self-evident how a house is build, and thus time is saved as one does not need to constantly look at the instructions. This tendency to abscond completely free

use of language is a phenomenon much discussed in the field and Sinclair (1991:110), for instance calls it the idiom principle, and Pawley and Syder (1983:209) consider it be a crucial part of how native speakers achieve fluency under the time stress of true communication.

Yet, these conclusions about the benefits of formulaic language need not be based merely on deduction, as has been proven for instance by Kuiper (2004) in his study of the language used by auctioneers and sports commentators. Both of these professions present situations where the speaker has to produce detailed language under extreme time constraints, and it was found out that the language used was almost entirely formulaic, which was argued to be a consequence of the mind compensating limited on-line processing with a reliance on memory. (Kuiper 2004:42) In essence, if a speaker had to narrate a quick horse race of ten or so contestants, the only way to retain the accuracy of all the details was to employ formulaic language. (Kuiper 2004:40-41) One could argue that this is simply the case of a cultural convention with no direct psychological presence, but it interesting to note that although non-formulaic speech was indeed used in less constrained parts of the discourse, the formulas always returned and did so in adverse proportion to the time available. (Kuiper 2004:40) Had the formulaic sequences of this context been a mere cultural convention, they would have probably been ascendant in all parts of the discourse and not just in the parts that seemed to have a clear psychological need for them. While it is clear that not all contexts function under the draconian time-constraints presented above, it is equally clear that limited time is an issue for all spoken communication. Thus also formulaic sequences are an issue for all spoken communication.

3.3 The role of formulaic sequences in language learning

This chapter has two intimately connected aims. First of all, the text here discusses how formulaic language is important for the process of language learning. Secondly, the methods of teaching formulaic language will also be examined.

3.3.1 Learning

If the issue is reduced to its essentials, one could say that there are two sets of factors that make formulaic sequences important for language learning. First of all, there are

the factors that make formulaic sequences extremely useful for the learning process itself. That is to say they can be seen to significantly accelerate the learning process and thus facilitate communication to a large degree. Secondly, it can be argued that the correct use of formulaic sequences is one of the clearest indicators of how well the learner has integrated himself to the target language's culture, or to put it in simpler way, how advanced the learner's communicative competence is. All in all, the following sub-chapter will argue that formulaic sequences are both the most basic aid of language learning and also its ultimate test.

As it was stated above, there are essentially two ways of looking at the relationship between formulaic sequences and language learning: the early stages and the advanced stages of the process. The next two paragraphs focus on the former. It seems to be a well-documented principle that children learn their first language in chunks. (Bolinger 1975:100) What this means is that a child takes an expression such as *What's that* and memorises and uses it as a single entity more akin to *whatsthat* than the grammatical combination of the words *what*, *is* and *that*. Further segmentation and grammatical analysis comes only later. (Bolinger 1975:100) While it is clear that all the combinations chunked by the learner are not formulaic sequences in the statistical and cultural sense, the process itself strongly implies that in the early stages of learning the mind views everything as formulaic sequences until contrary evidence is received. It is generally held that reliance on formulaic language becomes a foundation for learning, because it gives the child a communicative competence that is technically above his or her grasp of the language system.

So the real question is whether or not this principle applies also for second language learners, which cannot be directly assumed due to other fundamental differences in the nature of the processes. What is more, the dilemma is actually two separate questions, as there are two distinct groups of second language learners, those still in the critical period of language learning and those already beyond it. It seems that in the case of the first group the process of chunking is very much in play, as certain studies have shown that young children successful in second language learning proceed almost exactly like first language learners: by taking more or less all input as formulaic. (Wong Fillmore 1976, Hakuta 1976) Yet in the name of accuracy it must be stated that formulaic learning is a tool and not the end-point of the process as the

eventual segmentation and systemic analysis were the marks of true learning, whereas is some less advanced cases the language is never analysed into its components and chunks becomes fossilised (Wong Fillmore 1976:549, as quoted by Wray 2002:171) If one moves to examine adults and other learners past the critical period of language acquisition, the relationship between formulaicity and learning becomes somewhat different. The basic principle with older learners seems to be that they employ formulaic learning to a much lesser degree than children. It seems that after the critical period the analytical processes of the mind become so dominant that adults rarely use an expression before they are able to point out and analyse the its individual constituent parts. (Wray 2002:212) In addition the irrevocable changes in the brain structure there might also be an alternate explanation for the adult's avoidance of formulaic learning. There might also be some social pressure that creates the expectation that adults must be able to exactly understand all the nuances of an expression before they can use it (Wray 2002:170) which is essentially in polar opposition to children who seem to have such and urge to communicate that they are willing to take the risk of inaccuracy inherent in formulaic learning (Hakuta 1976:333). Thus it would seem that psychological differences between adults and children also manifest in their relationship to formulaic learning.

Yet the case of chunking and post-critical period learners is not nearly as clear cut as one might assume based on the discussion above. In other words, it seems that even though chunking is not the natural primary choice for adults, they can be essentially made to benefit of it in some limited scenarios. This was shown in a study which followed an individual with rudimentary passive command of Welsh attempting to learn the language in a very limited timeframe of one week. (Wray 2004:252) Due to the time constraints the teaching was done by almost totally eschewing the attempt to present the language as a system of grammatical rules, and instead gave the learner undifferentiated expressions meant for a purpose. They used entities such as *Ychwaneg wch y madarch*, which is Welch for *Add the mushrooms* (Wray 2004:256). The learner was essentially forced to take all the language input as formulaic, and the interesting point is that this method gave the learner good enough command of the Welsh language to allow her to perform in a cooking show. (Wray 2004:255) What is more, this was not a case of complete and instantly forgettable rote learning, as the individual not only retained bulk of the material for a considerable while but also

began some limited experimentation with it. (Wray 2004:258) All in all, the study in question seems to show that formulaic learning can and does have a clear and beneficial role even with adult language learners. Therefore, the case of formulaic sequences seems to be fairly clear. During the critical period the most natural way of learning for both natives and non-natives seems to be formulaic. For adults the case is somewhat less clear. They do not naturally chunk language input, but can benefit from it under certain circumstances. Yet one could argue that the main benefit of formulaicity for adults is more of a cultural one, which is an issue handled in the next paragraph.

One of the objectives of this chapter was to show the role of formulaic sequences in sociocultural integration and thus also in learning. This second part in the argument flows from a very simple chain of deduction. It begins with the seemingly self-evident dual statement that language is an expression of culture (Teliya et al. 1998:57) and that integration into this culture and its practices is the true test of language learning (Dörnyei et al. 2004:105). When this preface is combined with terminology that defines formulaic sequences as culturally sensitive expressions, it does not take much of a leap of faith to say that formulaic sequences are important to cultural integration and to language learning. What else could be the conclusion when lexical phrases are defined as culturally conventionalized functional expressions (Nattinger and DeCarrico 1992) and idioms as being highly opaque without the knowledge of their proper cultural context (Wray 2008:28-30). Yet this conclusion need not be based on logic alone. Certain studies have shown, for instance, that the incorrect use of formulaic sequences is one of the most important telltale signs in separating the performance of even the most advanced learner form that of a native speaker. (Herbst 1996:390) This cultural difficulty could be a result of formulaic sequences requiring much more than simple communicative competence from the user. If one considers, for instance, a prototypical lexical phrase, the mealtime greeting, it's easy to understand where these problems arise. Both German and Japanese have this greeting, *Guten Appetit* and *itadakimasu* respectively, and so the lexical phrase should be relatively easy to master. This is, however, not the case, as the social customs surrounding the expressions are entirely different and result in the German version being used by the host, whereas the Japanese *itadakimasu* and carries the meaning *I gratefully receive* and is always used

36

by the guest. (Coulmas 1979:259) The problem is similar, and one could even argue, greater with less functionally oriented metaphorical formulaic sequences such as idioms that are often so tied up with specific cultural context that a learner without cultural immersion might entirely miss the meaning of the idiom or its parts. The Russian idiom *chuvstvo zakonnoy gordosti* translates literally into *a feeling of justified pride*, and it derives the metaphor from the Soviet times when the expression was used as official political jargon. Nowadays it has become a metaphor used to critique any totalitarian regime (Teliya et al. 1998:75). Here the difference alluded to by Herbst (1996) comes into existence simply because the non-native speaker more than likely does not understand or use the full cultural import of a piece of officialise turned into a metaphor of defiance against the system. To put it simply, the non-native speaker does not understand when something is mocked, because he has not lived through the Soviet times, and thus he is marked as a non-native speaker. It seems relatively clear that formulaic sequences of all kinds are intimately tied with the culture of its users. One could even argue that mastery of these expressions is a true test of full communicative competence. Thus, despite the rarity of some of the sequences, the phenomenon as a whole has a clear and well-motivated role in language learning.

3.3.2 Teaching

It should now be clear that formulaic sequences have a role to play in the language learning process. One could even state that they offer certain clear benefits to the learner. Yet this assessment would be somewhat incomplete without discussion on how formulaic sequences could actually be taught, and to put it bluntly, all the answers to this question tend to work on the axis of explicitness and implicitness. First of all, it is important to understand that using the extreme ends of the spectrum is generally considered to be a bad idea. The main problem with explicit methods, such as learning list of idioms by heart, is the sheer number of formulaic sequences. How is one to choose the ones that are actually useful? (Mäntylä 2004:180) How does one make the choice between *to ride the gravy train* and *to kick the bucket*, or why the lexical phrase *my condolences* should be more useful than *How do you do*? What is more, it is easy to see that explicitness is quite risky, when one combines the sheer number of formulaic sequences with the rarity of any given sequence. Any

overt attention might lead the learners to see formulaic sequences as much more common than they actually are, and thus use them too often and even out of context. (Moon 2006:60) Due to these complications it has even been suggested that formulaic sequences in general, and specifically idioms, should be left almost entirely for contextual learning. (Mäntylä 2004:180-181) The reasoning behind this policy is that aforementioned problems could be sidestepped, if formulaic sequences are given in proper context straight from the beginning. The learner should be unlikely to misuse or overuse a sequence if its natural context is used to show how, why and when the sequence is actually employed. If one, for instance, learns to associate the idiom *to ride the gravy train* with its proper old fashioned context, then the learner is less likely to make himself seem old fashioned by using the idiom in modern conversation. Thus implicitness would seem to be the key to success.

Unfortunately complete implicitness is just as problematic as explicitness, even if the basic argument in favour of implicit contextual learning is entirely solid. The underlying assumption with implicit learning seems to be that learners should pick out formulaic sequences from context by themselves with as little explicit aid as possible. Yet it seems that left to themselves learners rarely manage notice formulaic sequences. At least it seems that very little actual learning leading to active use actually happens. (Bishop 2004a:18) Given the problems associated with the either end of the spectrum, it is no wonder that the preferred method lies somewhere in the middle. Even this concession holds on to the basic premise that implicitness and contextual learning are paramount, but the main modification is that the learners should be carefully aided to notice the formulaic sequences. Some good results in comprehension and recall of formulaic sequences has been achieved, for instance, in a study where a text was made into a hypertext document where the formulaic sequences were highlighted and made into links to a database on the accurate meanings of the sequences. This method had the dual benefit of making the formulaic sequences stand out without separating them from their context and additionally clarifying their meanings (Bishop 2004b:241) Not all text can be modified in this fashion, but it is fairly evident that making formulaic sequences easy to notice is one of the key issues in their teaching. It is, however, clear that recognition and recall are a somewhat different issue than the active use of formulaic sequences. While there is no panacea for facilitating active use, it seems that the

passive knowledge achieved by the aforementioned means can be leveraged to aid the process. The key thing seems to be a kind of an endless loop of formulaic learning. First the awareness is raised by making by making the sequences stand out by using, for instance, by highlighting them in context (Bishop 2004b). After this some carefully formed lightly explicit instruction, such as discussion on the nature of the highlighted sequences (Jones and Haywood 2004). Optimally, the next to follow is a production rehearsal that can be anything from close exercise and essay writing (Jones and Haywood 2004) to recording a monologue (Wood 2009). This in turn is followed by examination of the product and its formulaic sequences. In other words, what went wrong and what went right. The process is repeated in ad infinitum and in every cycle some of the inaccuracy should be shaved off. It really does not seem to matter whether the learner is dealing with monologues as a preparation for a discussion (Wood 2009:48) or the practise of academic writing and its stock expressions (Jones and Haywood 2004:274). The key thing in changing passive knowledge into active use seems to be that the learner is himself empowered in the process (Jones and Haywood 2004:273). If the learner gets to try out and rehearse the use of formulaic sequences in a monitored environment using his personal data so to speak, there is a fair chance that at least some of the sequences are transferred into active long term use. In other words, the learner not only encounters the sequence in a proper context but also gets to use it in a proper context.

All in all, it seems that formulaic sequences are not so much pedagogically atypical after all. One simply needs to take into account that they are somewhat more sensitive for a proper context and methodology than some other language material. This is best taken into account by ensuring that the sequence is encountered in a proper context and is easily noticeable. After this it is merely a question of trying to smuggle the sequences into the learner's active use. In other words, teaching formulaic sequences is a question of balancing explicitness and implicitness.

3.4 Previous research on formulaic sequences in textbooks

The body of research on formulaic language and textbook is not particularly extensive, which creates a niche for the current study. Yet, the research that has been conducted has some interesting implications. One of more far-reaching conclusions that is that textbooks and corpuses do not contain the same formulaic sequences.

While one could expect the vast majority to be collocations followed by some lexical phrases and few idioms (Moon 1998) a recent study of four British-made EFL textbooks made using formulaic definitions nearly identical to those of the current study came to the conclusion that the distribution of formulaic sequences is often wholly arbitrary (Koprowski 2005). Collocations were indeed the largest group in all the textbooks studied, but idioms for example received attention that was out of all proportion considering their rarity in corpuses with figures ranging up to 14 percent of all the formulaic expressions. (Koprowski 2005:327) In addition to the overall distribution pattern, the study came to the paradoxical conclusion that the textbooks focused on individual sequences that were quite rare. It was in fact the case that the more sequences there were, the rarer any individual one was. (Koprowski 2005:329) The possible explanation offered for all these phenomena was a relatively simple one. They might, for example, be a result of the fact that textbooks were designed around certain lexical fields and structural units with formulaic sequences having to adjust to that fact. (Koprowski 2005:329) If a textbook chapter focuses exclusively on lexical phrases containing the verbs *take* or *put*, it is quite inevitable that the selection of sequences becomes quite skewed and includes expression that much less common than others (Koprowski 2005:329) What is more, the arbitrariness of the choices was reinforced by the utter lack of unity between the four textbooks. In fact, the agreement rate on chosen formulaic sequences was only about 1 percent, which is curious considering that all the books had highly similar stated pedagogical aims. (Koprowski 2008:330) All in all, it seems that the relationship between textbooks and formulaic sequences is a problematic one. On one hand their importance is recognized. On the other hand, there seems to be definite lack of objective criteria in how the sequences are chosen. While textbooks do indeed teach formulaic language, they do so based on a set of criteria that is not derived from formulaic language itself.

The distribution of formulaic sequences in textbooks tends to be somewhat arbitrary. Despite this it also seem to be clear that there are some patterns to be seen, and this in turn has some implications on how the data of the current study is to be analysed. A study concentrating on six English for academic purposes (EAP) textbooks concluded that the actual designed learning material such as texts or exercises had a surprisingly small role in teaching formulaic sequences. The learners received a very large proportion of the formulaic sequences from instructional language such as

headlines and exercise directions. (Wood 2010:101) Moreover, the formulaic sequences in these contexts were also much more useful, i.e. more common in corpus data, than their counterparts in the actual learning material. The flipside was that formulaic sequences in actual texts were much more relevant for academic context. (Wood 2010:103) In other words, the instructions contained high frequency lexical clusters like the pair *on one hand/on the other hand*, whereas the text had sequences such as *in the case of* which were more specific to academic discourse. (Wood 2010:107) It can even be argued that many of the sequences found in the actual texts were formulaic sequences only in a sense that they are conventionalised ways of saying in the academic discourse, but would be considered purely grammatical compositions in any other contexts. It is not the intention of the current study to suggest that this would be the case in its own data, as senior secondary school textbooks have markedly different aims from academic textbooks. Namely one could expect material concentrating on overall cultural learning to contain a much more even-handed distribution of formulaic sequences than a purely academic text. It is, however, reasonable to assume that the underlying principle highlighted by Wood's (2010) study holds true. The formulaic sequences in the instructions should be different to that found in the actual texts. What is more, it can and should be included in a credible textbook analysis because it seems to be one of the most important sources of formulaic input.

4 DATA AND METHODS

4.1 Research questions

The aim of the current study is to examine the role of formulaic sequences in Finnish EFL textbooks for upper secondary school. In other words, the objective is to examine what kinds of formulaic sequences are present the material and how large the presence is. Moreover, the current study also examines how the formulaic language is introduced, as previous studies (Bishop 2004a, Mäntylä 2004) have shown this type of material to be especially sensitive to methodology. These topics are important because of three interconnected reasons. First of all, the significance of formulaic sequences themselves derives directly from the theory presented earlier. If one assumes, as this study does, that formulaic language assists learning, facilitates communication and serves as the final arbiter of cultural integration, then the role of

these sequences becomes of crucial interest and importance in any learning material. Secondly, certain studies have shown that textbooks are the single most important governing factor in the contents of formal classroom learning (Opettajakyselyn taulukot n.d.) and thus function as the deciding factor in what kind of language input the learner does and does not receive at school. To put it bluntly, it could be argued that the learner encounters formulaic language in a classroom only if the textbook is designed to enable it, and thus it is essential to examine how the textbooks perform in this role.

The relationship of formulaic sequences and textbooks is a somewhat neglected subject. Yet the little data there is has some weighty implications on the research questions of this study. First of all, there is the overall statistical information suggesting that the distribution of the formulaic sequences should follow a particular pattern. (Moon 1998) Various types of collocations should be the largest group found, if that was within the bounds of the current study to properly examine. Apart from that, one could expect to see mostly lexical phrases and few pure idioms deemed culturally important for the learner to understand. This assumption cannot, however, be taken entirely at face value because certain studies on textbooks imply that the distribution of formulaic sequences often seems to be based on the designer's subjective notions of usefulness rather statistical information derived from corpuses. (Koprowski 2005:331) At this point it is still impossible take a firm stand on which of these possibilities holds true for the current study, but it is fairly certain that it will be one or the other. What is more, it could be tentatively hypothesised that the textbooks lean more towards the latter, because Finnish textbook designers do not use corpuses that would make immune to this kind of formulaic bias.

Thus the research questions of the current study are the following:

1) What kinds of formulaic sequences are present in the textbooks and how large is this presence?

2) How are the formulaic sequences presented?

4.2 Data collection

The data of the current study comes from two textbooks designed for the fifth compulsory English course of the Finnish upper secondary school. The two books werethe *ProFiles Course 5* (Ikonen et al. 2009) and the *Culture Café Book 5* (Benmergui et al. 2004). As there was only one book from each series, the sources will henceforth be only referred to as the *Culture Café* and the *ProFiles*. The two books were chosen because they both are relatively new textbooks in active use in the school system. As such, they could be reasonably expected to reflect both recent scientific opinions on formulaic sequences and the most current actual practices on teaching formulaic language. What is more, according to the fifth national curriculum (Lukion opetussuunnitelman perusteet 2003) the aim of the learning on this level is to familiarise the learners with effective and versatile communication strategies and even rehearse the most demanding ones. Because formulaic sequences arguably fill all these requirements, one should be able to find them in the specified textbooks. Lastly, the fifth compulsory course was chosen because it specifically focuses on communication and cultural learning. (Lukion opetussuunnitelman perusteet 2003) If formulaic sequences are to be found anywhere at all in the upper secondary school textbooks, then these books meant for the fifth course are the place.

It must also be made clear how the data within these books was chosen and limited. The basic principle was to include all material in the textbooks themselves. This contains all texts, exercises and authentic texts. Even instructional subtext such as headlines and exercise directions were included because they not only contain formulaic sequences but the ones present in the instructions are markedly different than the ones in normal discourse. What is more, the instructional metalanguage is often so repetitive that the learners meet its formulaic sequences considerably more often than those contained in the actual learning material.(Wood 2010:111)

4.3 Methods of analysis

The current study uses content analysis as its primary method of investigation. First and foremost content analysis is a subtype of qualitative research, as its aim is to build a theoretically solid overall understanding of the phenomenon being studied with relatively little emphasis being placed on raw statistics (Tuomi and Sarajärvi

2009:85). What is more, this method can be defined as textual analysis because it examines documents in a systematic and objective way in order to construct a concise and generalised description of a phenomenon. (Tuomi and Sarajärvi 2009:103) In the current study the documents are the textbooks and the phenomenon formulaic sequences. Content analysis can be done in three ways, and the first one is the traditional data grounded content analysis, in which the document is studied with no presuppositions. The aim is simply to see whether patterns exist in the text, and not to try proving or disapproving the existence of some pre-given pattern or category. (Tuomi ja Sarajärvi 2009:108) Given that the current study has such presuppositions, the method cannot be employed here, but if it were, it would result in attempt to ascertain what kind of word combinations naturally occur in the textbooks. Then there is also the theory grounded content analysis, which is what the current study uses. This method entails taking a predetermined theory or a concept and comparing it against the data. (Tuomi ja Sarajärvi 2009:113) The current study, for instance, takes the presupposition that formulaic sequences exist and do so in a specific form and then tries to determine whether they are present in the data. In addition there is also the theory guided content analysis, but it is merely a variation of the theory grounded method with the only difference being the phase in which the theory is incorporated into analysis. Thus the distinction is not relevant (Tuomi ja Sarajärvi 2009) and will not be explored in any detail.

In this paragraph I will go into more detail on how theory grounded content analysis is conducted and how this translates into the methodology of the current study. First of all, it must be stated that Tuomi and Sarajärvi originally wrote in Finnish, and consequently both the term theory grounded content analysis and its acronym TGCA are my own translations. According to Tuomi and Sarajärvi (2009) there are four main stages to TGCA: building the framework of analysis, reducing the data to its essentials, the search for patterns and similarities and finally the composition of the overall view. In the first of these the researcher takes a theory and uses it to decide what is and is not searched for in the data. (Tuomi and Sarajärvi 2009:113) In the context of the current study this meant synthesising definitions for formulaic sequences and building a view of how they might be taught. It was also decided in this phase that collocations apart from the singular exception of explicit pedagogical contexts would be excluded from the analysis. This was because that collocations in

free text would have been impossible to reliably identify without extensive statistical analysis. In the second part of the TGCA the data is reduced to its essentials by searching for the issues specified by the framework (Tuomi and Sarajärvi 2009), which in the case of the current study meant two things. On one hand, the textbooks were scoured for combinations that filled the requirements set for formulaic sequences in the theoretical background of this thesis. On the other hand, it was also noted whether the sequences were presented explicitly or implicitly. The results of the individual books were then combined into tables. The process was repeated separately for each of the books after which a common summary table of all the results was constructed. The pattern looked for in the third stage of the TGCA was simply whether the overall distribution and the pedagogical context of the formulaic sequences conformed to the hypotheses of my study, and if it did not, what kinds of patterns were to be found. In the final phase an overall picture and a tentative explanation of the results was constructed based on the theoretical background and previous research on the field.

5 DISTRIBUTION OF FORMULAIC SEQUENCES IN THE EXAMINED TEXTBOOKS

The following section will present the result of the analysis on the two textbooks, *Culture Café. Book 5* and *ProFiles. Course 5*. Even though the current study is a comparative one, the presentation will be structured around themes rather than individual books, and the comparisons between the two sources will be handled within these themed sections. The two main sections correspond to the research questions of the current study: what kind of formulaic sequences were found in the material and how these sequences were presented. Within these sections relevant subdivisions will be made, which in the case of the first question means that idioms and lexical phrases will be discussed separately. For the second theme a similar partition will be made between explicit and implicit contexts. Both of these sections will include the necessary tables and numbers, but as the current study is not a statistical analysis, these will serve as illustrations and examples rather than the actual data.

It is worth noting that the current study counted each occurrence of a formulaic sequence as a separate event even if it was the same phrase being repeated continuously, as was the case when the sequence *Up close and personal* (Ikonen et al 2009:78-88) was repeated a dozen or so times in a single chapter due it being printed in the upper margin of each page in the unit. The current study chose to count these and other similar occurrences individually because it can be generally assumed that repeated exposure aids learning. It is reasonable to expect that the authors of the books now this fact, and thus the repetition of an expression can be taken as intentional and meaningful. Yet in some cases it was necessary to count separate individual phrases instead of occurrences. This was done to reveal the actual individuals behind the repetition patterns.

The current study uses the following terms and abbreviations to refer to the textbooks. As there is only one book from each series, and thus no change of a misunderstanding, the current study has opted to use the shorted versions *Culture Café* and *ProFiles* instead of the full names of the textbooks. Alternatively the abbreviation *CC* may be used for the former and *PF* for the latter.

5.1 The general situation

The purpose of the following paragraph is to provide a short introductory glance at formulaic sequences in the textbooks. This is mainly done with Table 2 below. To begin with, the diagram lists occurrences of formulaic language rather than how many different individual sequences the textbooks contained. Thus each duplicate and possible variant of each individual phrase was counted separately. The idiom *up close and personal*, for instance, is registered 11 times, because that is how often it was found in the *ProFiles* textbook (Ikonen et al. 2009:78-88). This was done to illustrate the overall amount and types of formulaic language found in the textbooks. A more detailed account of individual expressions, types and repeating patterns will be provided in later paragraphs of the current study. Secondly, it is the presence of collocations and phrasal verbs in the table does not represent a deviation from the principles laid down in the theoretical background of this study. No collocation or phrasal verb was actively looked for in the data. Members of these groups were only taken into account if the textbooks paid explicit attention to them and used labels that made it apparent that the phrases were seen as formulaic sequences. Essentially, they

are not present because of what they are, but because of what is done with them. Again, this will be discussed in greater detail later in this thesis.

Table 2. The overall distribution of formulaic sequences within the textbooks

	ProFiles	Culture Café
Formulaic sequences	802	388
Idioms	443	152
Lexical phrases	252	167
Collocations	87	69
Phrasal verbs	20	---

These numbers reveal several things. To start with, it is quite evident that there was a remarkable difference between the two textbooks. Overall, formulaic sequences were 116 % more numerous in *ProFiles* than in *Culturé Café*. Moreover, this held true for all categories of formulaic sequences, as ProFiles had a 207 % advantage in idioms, 58% in lexical phrases and 26% in collocations. It is also noteworthy that while *ProFiles* had almost twice as many occurrences of idioms than occurrences of lexical phrase, the other textbook roughly similar amounts of both expression with there being slightly more lexical phrases than idioms. While both books introduced formulaic language, the numbers show that *ProFiles* did this in much greater detail and brevity than *Culturé Café*. This tendency is especially evident in the fact that while the former handled an entire class of expressions, the phrasal verbs, as at least partly formulaic, the latter did no such thing. In essence, even a preliminary look into the data provided an answer to one of the research questions of the current study. That is to say that yes, formulaic sequences did indeed have a presence in the two textbooks. What is more, this presence proved not to be a constant across the two sources. Otherwise, the volumes would have been more equal.

The sheer numerical presence of formulaic sequences is, of course, important, but the figures above also reveal something important about internal distribution of the formulaic language. In essence, corpus data from previous studies (Moon 1998) suggests that language naturally uses enormous numbers of borderline collocations, some lexical phrases and very few idioms. This, however, was not the case for either the *Culture Café* or the *ProFiles*. On the contrary, idioms were, without a doubt, the most common type of formulaic sequence in *ProFiles* with lexical phrases having a somewhat smaller presence. In the *Culture Café* there were very nearly equal amounts of lexical phrases and idioms with the former group having a slightly more substantial presence. The collocations of either of the textbooks were more difficult to factor in as the numbers only represent those sequences that were explicitly identified. Nevertheless, the textbooks seemed to essentially invert the natural pyramid-like distribution (Moon 1998) mentioned above. This result, however, corresponds with another previous study (Koprowski 2005:327), which suggested that textbooks often do not follow corpus data in their choice of formulaic sequences. Koprowski's conclusion that especially idioms tend to receive attention out of all proportion with their actual frequency seemed to fit particularly well to the data of the current study. Thus one could preliminarily argue that the internal distribution of formulaic sequences *ProFiles* and *Culture Café* is for the most part in line with how textbooks are generally designed.

5.2. Lexical phrases

The focus of the following chapter is to report what kind of lexical phrases were found within the two textbooks, *Culture Café* and *ProFiles*. These results are presented from the two complementary perspectives of structural and functional classification of lexical phrases. A more detailed discussion of these perspectives can be found in theoretical background section of this study. It is also crucial to understand that all the figures, except for those in Table 4, refer to the amount of individual expressions unless otherwise stated. The choice was made because pure occurrence numbers would have obscured any individual expressions. What is more, the exact nature of recurring lexical phrases is, in itself, something that will be discussed. Thus all the duplicates were filtered out of results. Naturally this also

means that lexical phrases from the vocabulary lists are not included, as they are, by definition, duplicates of material found elsewhere in the textbooks.

Table 3. Lexical phrases – a comparative table

	ProFiles	Culture Café
Occurrences	252	167
Individual lexical phrases	118	116

The table above takes the absolute number of occurrences of lexical phrases and juxtaposes it with the number of individual expressions that were found in the textbooks. The first point that rises from this data is that *ProFiles* had a much less distinct numerical advantage in lexical phrases than it seemed based on pure occurrence counts presented in table 3. In fact, a comparison between the numbers of individual expressions proved the difference to be a rather insignificant amount of two more lexical phrases in *ProFiles* than in *Culture Café*. In essence, both books had nearly identical amounts of phrases that were repeated throughout the material. This, however, calls into attention the second point that can be seen from the data presented in the table. That is to say that both the textbooks had a set of lexical phrases that was used repeatedly, and in this regard the difference between the sources is somewhat more significant. While *Culture Café* has approximately 44 % more occurrences of lexical phrases than actual individual phrases, the corresponding situation for the *ProFiles* is that there are over two times more occurrences than individuals. This then begs the question of which were the sequences that kept on repeating, and how did these repetitions occur. Therefore, a list of five most common lexical phrases from both books will be provided. The numbers of occurrences come in parenthesis after the phrase in question.

Table 4. The list of the some of the most common lexical phrases in the textbooks.

ProFiles	Culture Café
there's more to (3)	*whereas* (5)
how come (6)	*try as sb may* (3)
however (4)	*no matter* (3)
all right (3)	*for example* (7)
on the contrary (6)	*let me tell you* (2)

It is clear that there was a certain pattern that kept on repeating. The most common formulaic sequences were usually very short polywords such as *all right* (PF:70), while other types of lexical phrases repeated more or less randomly. What is more, the phrases listed here are in no numerical order and do not represent an exhaustive list, as each all recurring phrases were only repeated between 1 and 7 times. A hierarchical structure would simply serve no purpose. The list is a source of possible examples, and not as an exhaustive account. Thus the purposes of the current study are better served by an analysis of the patters these examples highlighted. There were actually two separate kinds of repetition to be found in the textbooks. On one hand, there were the phrases occurred in unconnected parts of the books. A good example of this is how the lexical phrase *however* appeared independently in different texts and units or within the same unit. On the other hand, there was the type where the books contained a phrase such as *there's more to* (PF:14) in a text and then later called explicit attention to it within the same unit. This type of repetition was mostly found in the textbook *ProFiles*, and what is more, it was so intimately connected with methods of teaching formulaic sequences that further discussion about it is concentrated in the chapter 6, which is the part of the current study concentrating on the how the formulaic sequences were presented.

5.2.1 Lexical phrase by structural classification

The following chapter focuses on the four different structural subcategories of lexical phrases (Nattinger and DeCarrico 1992). More specifically, the focus is on how the overall number of individual lexical phrases was divided amongst the different structural categories. These results will be first provided in Table 5 located below. It will be followed by an analysis of these numbers along examples of the most salient issues brought forth by the figures.

Table 5. Individual lexical phrases by structural classification

	ProFiles	Culture Cafe
Lexical phrases	118	116
Polyword	57	60
Phrasal constraint	6	12
Institutionalised expression	31	27
Sentence builder	24	17

The first thing that is immediately apparent from the figures above is that both the textbooks had remarkably similar patterns of distribution with regards to different structural categories. The pattern is made even more evident by the fact that the overall numbers of individual lexical phrases were nearly identical. Thus the pattern of distribution also resulted in the textbooks having, not just similar ratios, but also highly similar actual numbers of lexical phrases in different functional categories. The greatest dissimilarity was in the categories of phrasal constraints and institutionalised expressions. The details of the divergence are discussed later along with other aspect of the said categories.

5.2.1.1 Polywords

The first structural subcategory of lexical phrases, polywords, was the most common one both in *ProFiles* and in *Cultural Café*. The latter had two more polywords, but the difference is fairly minor considering the overall number of expression, and thus

was not considered meaningful by the current study. There were also noteworthy similarities in what kind of polywords were presented, which the current study will try to illustrate by providing examples from both the books. The specimens are provided in their immediate context with the actual phrase being underlined. The formulaic sequence in each example is underlined. If, however, the phrase comes from a gap filling exercise the notation is somewhat different. The gap is represented by a short empty line, and the target phrase is given in parenthesis and underlined. These two conventions are used in all the examples of the current study, and any deviations or additions will be explained separately.

1) *After all, the book, Gibbon's Decline and Fall, was by one of her favourite authors, science fiction novelist Sheri S. Tepper.* (CC:14)

2) *However, negative reactions can be off-putting.* (CC:81)

3) *In Hamlet, for example, he appeared in a few scenes as the Ghost.* (PF:115)

4) *No wonder Gemma _____ in an advertising agency.* (PF:168)

5) *The wind and the rain were, nevertheless, merciless.* (CC:132)

All the examples, and a considerable part of the category they represent, easily conformed to the model presented by Nattinger and DeCarrico. In other words, they were a couple of orthographical words in length and filled the syntactic role of individual words. Interestingly, this function was that of a pure discourse device for almost all the polywords in both the books, but this will discussed in greater detail later in this study. It is also noteworthy that a considerable number of the polywords were of the type that had completely assumed the unified written forms of single words (Nattinger and DeCarrico 1992:39). Both *however* in example 2 and *nevertheless* in example 5 show this tendency. In the latter one it is particularly clear that the polyword is a direct amalgamation of the words *never*, *the* and *less*. Even though the previous five examples represent the typical polywords found in the data, there were some notable exceptions. Almost all of these seemed to be situations the expressions had trouble fitting the aforementioned syntactical functions of single words. The next two examples will highlight this tendency.

6) *Let's go; there's no time to losing/lost/loose.* (CC:127)

7) *In the meantime, the beat goes on.* (PF:30)

On the surface it might indeed seem that these phrases should not have been classified as polywords, but there is a case to be made for why the current study chose to do so. Essentially, a process of elimination was used. They are clearly lexical phrases as they are short formulaic sequences that have a specific communicative function. Yet they are neither long nor traditional enough to be institutionalised expression or flexible enough to be phrasal constraints. They also do not serve to begin and frame discourse in a way that would make them sentence builders. This is then combined with the fact that both the expressions were short enough to be polywords, and had functions that are very close to those of single words. The first one serves as a short phonetically unified (Nattinger and DeCarrico 1992:38) exclamation of a desire to move, while the second is single expression of time. Thus the two phrases are not perfect polywords, but they fit the category well enough. The presence of the exceptions reveals two things. The two textbooks contained more than one type of polywords. This is because the category is flexible enough to allow for and inevitably lead to borderline cases.

5.2.1.2 Phrasal constraints

Even though *Culture Café* and *ProFiles* contained polywords in nearly equal amounts, this was not so for all the subcategories of lexical phrases. Phrasal constraints, in particular, revealed an interesting difference between the two textbooks. Both sources of data had relatively few phrasal constraints, with *Culture Café* having 12 and *ProFiles* having only 6. In other words, the former had almost two times many individual expressions than the latter. One cannot even say that the contrast would be a result of *Culture Café* having introduced a type of phrasal constraint absent from *ProFiles*. Both books had similar expressions, but the former simply contained more of them. The following examples should serve as representative sample to highlight this resemblance.

8) *To cut a long story short, the prince turned into a frog and ended up on a French menu.* (PF:51)

9) *Monica and I have no plans whatsoever to get married.* (PF:38)

10) *Some graffiti is art, but most of it is rubbish, to put it bluntly.* (PF:26)

11) *What saved Lick, though, was the music, the cornet <u>to be precise</u>.* (CC:56)

12) *Just <u>once a week</u> at first.* (CC:57)

13) *<u>No matter what</u>, a work of art should give credit to _____ (ammattitaito), not discredit it.* (CC:81)

First of all, these examples tell us that most of the phrasal constraints came from less flexible end of the spectrum. As it has been said earlier, phrasal constraints are more formulas than fixed expressions (Nattinger and DeCarrico 1992), but textbooks seemed to contain expressions from the less flexible end of the spectrum. Most contained only one slot for changeable elements. The expression in the example 11, for instance, only seems to allow the changing of the noun between the words *no* and *whatsoever*. The phrasal constraint in the example 8, on the other hand, is even more rigid, as it allows for no actual changes in structure, but only the adding of qualifiers such as *extremely* before the adjectives of the phrase (Nattinger and DeCarrico 1992:41). The example number 13 was chosen to emphasize that both the textbooks contained phrasal constraints that were on verge of being polywords. The expression *no matter what* is, after all, very close to polyword *no wonder* in the example 4. The difference between these two superficially similar formulaic sequences is that the latter does not need and cannot have any extra elements fill its function of mock-disbelief. The former, however, has no meaning unless first two words are followed by a choosable third one that specifies the condition that can be ignored. Whether this results in a phrase such as *no matter what* or *no matter the odds* is irrelevant for the formula. Thus the example 4 and 13 are fundamentally different despite their extreme initial resemblance.

5.2.1.3 Institutionalised expressions

Institutionalised expressions were perhaps the most contradictory group of all the structural categories of lexical phrases. Parts of the results seem almost self-evidently clear. That is to say that there is no clear difference in the amount of individual phrases between *ProFiles* and *Culture Café*, as the former has 31 and the latter 27 of them. A closer look also revealed little noticeable difference in what kind of institutionalised expressions these were, or to put it more accurately, there was no

definitive type or pattern to be found. The examples below should serve to highlight this apparent lack of patterning.

14) *To be, or not to be, that is the question.* (CC:44)

15) *Say 'Merry Christmas!' Jim, and let's be happy.* (CC:99)

16) *A God damn whale.* (PF:69)

17) *What on Earth happened?* (PF:129)

The main problem in applying the label of institutionalised expressions was that the category itself has only ever been loosely defined. Roughly put, an institutionalised expression is simply a conventionalised phrase filling a very specific function in communication (Nattinger and DeCarrico 1992), such as the traditional Christmas greeting in the example 15, or the formulaic curse in the example 16. Perhaps the only true common factor the textbooks had for their institutionalised expressions was that most of the expressions were usable independently of any immediate syntactic context. Both the previous examples, for instance, could have been used alone. One could argue that this was and is a natural corollary of the definition of the phenomenon. Conventionalised greeting, for instance, must be sufficient alone, or it does not fill its function. The loose definition of institutionalised expression also had a second highly interesting and somewhat unforeseen implication on the current study. It was at times increasingly difficult to draw a demarcation line between institutionalised expressions and idioms. The phenomenon was, in fact, so prevalent and pervasive that some formulaic sequences had to be classified as belonging to both groups, which unavoidably led to some overlap in the expression counts for the both groups. The next examples were chosen to illustrate the nature of these borderline cases.

18) *I took it like a man.* (PF:15)

19) *All the world's a stage.* (PF:115)

20) *If you prick us do we not bleed? If you tickle us, do we not laugh? If you poison us, do we not die? And if you wrong us, shall we not revenge?* (CC:44)

21) *Friends, Romans, countrymen, lend me your ears. I come to bury Caesar, not to praise him.* (CC:44)

The first three of the examples above were pure borderline cases, as it was impossible to definitively label them as idioms or institutionalised expressions. Thus they were placed in both categories. The last one, on the other hand, remained on the lexical phrase side of the fence, but not by much. It could equally well have been an idiom. The key issue here is the level of metaphoricality the expressions showed, and in principle it should have been easy separate idioms from all other expression using this factor alone. In practice, however, it proved to be nearly impossible to determine when exactly does is phrase opaque enough to start being an idiom and not institutionalised expression. The example *If you prick us do we not bleed*, for instance, could be understood completely literally, but doing so would leave out the deeper cultural allusions and the full meaning. Underlying this and all other examples is also the fact that they all have the hallmark of institutionalised expressions, a specific and independent communicative function. Thus the most accurate choice was to label the problem cases as belonging to both groups. It is also noteworthy many of these cases were Shakespeare quotes, because both the textbooks had an explicit chapter on the said playwright. One could argue that this led to a situation where the data contained an inordinate amount fixed expressions that had a specific function but were not metaphorical enough to qualify as idioms. On the other hand, it could also be argued that this simply highlighted the nature of the borderline cases by providing stereotypical examples.

5.2.1.4 Sentence builders

Sentence builders, the last structural subcategory of lexical phrases, had similar average presences in both *ProFiles* and *Culture Café*. The figures of 25 and 17 individual expression respectively led to a situation where sentence builders made up approximately one fifth of the individual lexical phrases in both textbooks. Apart from these pure statistics, however, the nature of the presence was somewhat more contradictory. In essence, there seemed to be two separate types of presence. The first one was formed by almost stereotypical sentence builders, while second group consisted of phrases that had slightly more trouble in fitting the definition of the category. The examples below should shed light on the first group.

22) *I wonder if I'd meet the director's requirements.* (CC:30)

23) *I must say that* I think there are lots of other much nicer names. (CC:49

24) *I would like to conclude by saying*... (PF:132)

25) *As you all know by now*... (PF:132)

As it was stated above, the examples represent phrases that were nearly stereotypical sentence builders. They are "…lexical phrases that provide a framework for whole sentences. They contain slots for parameters or arguments for expression of an entire idea." (Nattinger and DeCarrico 1992:42). Examples 22 and 23 provide the best illustration of the framework, as the phrases clearly serve to introduce and establish the opinions that follow. The context of 24 and 25 went as far as to explicitly label the frameworks under headings of "Concluding" (PF:132) and "Making the main points" (ibid.) respectively. The existence of the free slots was very strongly implied in 22 and 23, but it was almost explicitly stated for 24 and 25. The triple use of full dots can, after all, be considered an almost universal symbol in a continuing expression, and apart from the use of underlying and italics, the examples were printed here in their original form.

Yet as it was stated earlier, not all sentence builders were as stereotypical as the previous four examples. The problem was not that they would have been incompatible with the general definition but that they had characteristics of other lexical phrases or even wholly different formulaic sequences. This tendency is highlighted by the examples below.

26) *Legend has it* that he caught the pneumonia walking home in the rain after a night of heavy drinking with a couple of mates. (PF:115)

27) It' hard to believe she's over 80. *You can say that again.* (PF:38)

28) I can remember *like it was only yesterday*. (CC:15)

29) She is a born leader and, _____ (try as I may), and I can't explain it any better.

The problem could be something as subtle as the hint of idiom-like metaphor in the examples 26 and 28. The example 27, on the other hand, was nearly independent enough to qualify as an institutionalised expression. Particularly problematic was the

type exemplified by 29, because the phrase is very much like a phrasal constraint. Nattinger and DeCarrico (1992:40) even go as far as to classify a similar expression *be that as it may* as an institutionalised expression. Yet the current study overrode this categorization with a simple core fact present in all four of the above examples. Despite their secondary characteristics, all the phrases still seem to contain an empty slot for a sentence. The formulaic sequences serve as a framing device for the larger structure they are a part of. The current study chose to include and highlight this dilemma, because it sheds light on the fact that the sentence builders present in the two textbooks were often far from self-evident or unambiguous.

5.2.2 Lexical phrases by functional classification

The aim of the following chapter is to provide an account of how the lexical phrases were divided amongst the different functional categories. In other words, it is an account of what the phrases were used for. It should again be stressed that the figures used in Table 6 are counts of individual lexical phrases and not of occurrences. Even though the chapter is about the functional classifications, there is some slight overlap with the contents of the previous chapter. This is simply because the points of convergence between the two types of categories will be discussed in this context.

The figures in Table 6 show that the textbooks were highly similar in their distribution of lexical phrases among the different functional categories. While these were not the same distribution patterns as with the structure classes, it is interesting that the two textbooks showed converging tendencies even in this regard. The overall pattern is clearly that both the books had pure discourse devices as the largest group. In both cases the most numerous category was followed in descending order by phrases of necessary topics, phrases of conversational maintenance and phrases of conversational purpose. Yet it is also clear that *ProFiles* had much sharper differences in the amount of phrases in the different categories, while the last three categories were nearly equal in *Culture Café* with even discourse devices being less clearly separate.

Table 6. Individual lexical phrases by functional classification

	ProFiles	Culture Café
Lexical phrases	118	116
Discourse device	61	48
Necessary topic	30	26
Conversational maintenance	19	22
Conversational purpose	8	20

5.2.2.1 Discourse devices

The largest functional group in both textbooks were the discourse devices. What is more, this point is impossible to discuss accurately without touching on the fact there was remarkable convergence between the structural category of polywords and functional category of discourse devices. Not all discourse devices were polywords, as will be discussed later, but much would have not been needed to tip the scales.

30) *Whereas the faster music became almost frantic.* (CC:60)

31) *For example, the audience knew that if a character delivered his lines while standing under the picture of the moon, he was not thinking with his head but with his heart.* (CC:41)

32) *That critic, however, isn't an expert on modern art.* (PF:149)

33) *On the other hand, I think reading books improves my language skills.* (PF:149)

The examples should clearly illustrate the most common type of discourse device, and moreover that the point of convergence between polywords and discourse devices fell on a very specific type of phrase. To being with, discourse devices are defined as "lexical phrases that connect the meaning and structure of the discourse." (Nattinger and DeCarrico 1992:64), and every one of the above examples fitted this definition quite well. They are, in fact almost stereotypical in this regard. The last of the examples, for instance, is a tool used to bind to differing points of view into the same discourse structure. Essentially, all the examples and the vast majority of

discourse devices in the data were logical connectors of various kinds. It is interesting, that all the above examples and the majority they represent were also polywords due to their short word-like forms and specific discourse functions. In hindsight, this probably should not have come as a surprise, as both definitions seem to point out to the same type of expression. Even so, the two groups were nigh identical in the data of the current study, and there is fairly strong evidence to suggest that this would also be the case in general.

Yet the key term in understanding the discourse devices of the two textbooks is indeed "nigh identical", because there were some exceptions to the pattern of overlap between discourse devices and polywords.

34) *What saved Lick, though, was the music, the cornet <u>to be precise</u>.* (CC:56)

35) *We all have the same hardware, our brain, but it's programmed to expect and appreciate different things, depending on where we grew up and the country we live in. <u>Or words to that effect</u>.* (PF:14)

Both these examples show a situation where the textbooks introduced phrases that were fairly clearly discourse devices. Both are employed to modify or set the tone of the discourse they belong to. Yet neither these examples nor the type they represent, were polywords. The first one is a clear phrasal constraint while the second has the qualities of an institutionalised expression. In other words, they were unlike the stereotypical discourse devices found in the data of the current study. In addition to being a feature of the two texbooks, these exceptions can be taken as a natural corollary of how discourse devices are defined. That is to say that definition clearly points to and neatly covers the type of overlap between polywords and discourse devices, but the definition is loose enough to allow for host of borderline cases that bear less resemblance to the type of logical connector that Nattinger and DeCarrico (1992:64) have clearly intended as a stereotypical example of a discourse device.

5.2.2.2 Necessary topics

Necessary topics had a relatively large presence in the textbooks. In terms of individual phrases the group was the second largest one in both *ProFiles* and *Culture Café*. Even the absolute numbers of individual expressions were highly similar. Yet

their presence was not entirely unproblematic, as necessary topics are perhaps the most loosely defined of all the functional categories. Thus they are also the most comprehensive one. It is, after all defined as "…topics about which the learners are often asked, or ones that are necessary in daily conversations." (Nattinger and DeCarrico 1992:63) It is a direct result of this definition that necessary topics had a large and relatively complicated presence in the data of the current study. The examples below have been chosen to illustrate this fact.

36) *The first thing we do, <u>let's kill all the lawyers</u>.* (CC:44)

37) *Grand as the watch was, he <u>sometimes</u> looked at it on the sly on account of the old leather strap that he used in place of a chain.* (CC:99)

38) *<u>Mind the gap</u>* (PF:8)

39) *<u>Thirty or so years</u> after he arrived in London, Chanu decided it was time to see the sights.* (PF:54)

It is apparent that there is little that could be called a pattern among the necessary topics. Perhaps the only one was that there was some slight overlap between institutionalized expressions and necessary topics. This is, for instance, the case with examples 36 and 38, which are both expressions of common topics. The first one articulates a common sentiment towards the legal profession, while the second one is a formulaic warning often found in the subway. Both are also established enough to be considered institutionalized expression, and the first one is even on the verge of becoming an outright idiom. Yet the pattern demonstrated by the examples 36 and 38 was far from universal, and it would be most accurate to merely state that necessary topics were institutionalized expressions more often than not, but that is all. The remaining examples, 37 and 39 are somewhat dissimilar phrases of necessary topics. Both of them are formulaic sequences that handle the common topic of time and its passage in a conventionalized form, and thus they were classified as necessary topics. The first of these has such a short and word-like form that it easily fitted the definition of a polyword. Yet the other one clearly has an empty slot in it structure for different amounts of years, and thus was classified as a phrasal constraint. In conclusion, it should be clear that necessary topics of the two textbooks were a diverse lot. They had a relatively substantial presence, but that was only to be expected considering that they by definition handle common topics. Yet the diversity

of individual expressions made it impossible to detect any large patterns in their presence.

5.2.2.3 Phrases of conversational maintenance

Phrases of conversational maintenance were again one of those groups that had nearly equal amounts of presence in both textbooks. As *Culture Café* had 22 individual phrases and *ProFiles* had 19, the numerical difference between the two is negligible, thought to the advantage of the former. These numbers made the grouping the third largest one in both textbooks. Even though many kinds of lexical phrases could potentially share function of conversational maintenance, there were certain patterns to be found in the data.

40) *I'm sorry, sir. <u>It's just</u> - I took a deep breath - she is wearing a blue skirt.* (CC:73)

41) *<u>Ladies and gentlemen</u>...* (PF:132)

42) *Decorative souls may add a little lace, <u>and welcome</u> – my tablecloth is your tablecloth, knock yourself out.* (CC:9)

43) *Five of her co-stars have met their deaths in gangland violence since the film was released in 1998. <u>Enough said</u>.* (PF:105)

The first two examples illustrate that phrases of conversational maintenance were often sentence builders. Yet this was not really a surprise considering that phrases of this type are defined as "…regularities of conversational interaction that describe how conversations begin, continue and end." (Nattinger and DeCarrico 1992:60). Example 40 is essentially a tool of nominating the topic of the blue skirt and giving the speaker time to continue, while 41 functions as a highly formulaic and formal method of beginning a conversation. What makes both these phrases also sentence builders is that they have trouble functioning without a sentence following them. Essentially, they have an empty slot in their structure. Yet both textbooks showed that this was hardly the only possible form for a phrase of conversational maintenance. This is shown by examples 42 and 43. The first of these carries the function of greeting someone and beginning a conversation. The second one quite evidently ends conversations. The first one is a short formulaic greeting, and thus could be classified as either a polyword or an institutionalized expression. The current study leaned towards the latter definition. For much of the same reasons the

second one was labeled a polyword. The formal difference between the two expressions is minimal as they both combine a verb and an adverbial into a formulaic sequence. It was mainly a question of degree of formulaicity that gives the two different labels. For *welcome* the development has gone far enough to completely fuse the component words, whereas *enough said* still the parts as orthographical individuals. In conclusion, all the examples should highlight that there was a similar prominent pattern to be found in both the textbooks, but that this pattern was far from universal.

5.2.2.4 Phrases of conversational purpose

Phrases of conversational purpose were noteworthy due to the fact they were the only functional subcategory of lexical phrases where *Culture Café* had a numerical advantage over *ProFiles*. What is more, the advantage was also fairly major as the former had 20 individual phrases of conversational purpose while the latter only had 8. Yet there were only very faint discernible patterns of phrases to be found in either of the books, and these were similar in both. Thus it is only possible to say both books contained similar expression, but *Culture Café* simply had more of them.

44) *Personally, darling, <u>to speak quite candidly</u>, I don't much care about the name Ernest… I don't think it suits me at all.* (CC:49)

45) *<u>The odds are</u> greatly in my favour.* (CC:36)

46) *<u>In your opinion</u>, what might have been the source of inspiration for each piece?* (PF:89)

47) *Some graffiti is art, but most of it is rubbish, <u>to put it bluntly</u>.* (PF:26)

48) *<u>What was</u> the title of the assignment <u>again</u>.* (PF:14)

49) *I'm a philistine <u>through and through</u>* (PF:14)

Phrases of conversational purpose are defined as "types of speech act, i.e. functions that describe the purposes for which conversations take place" (Nattinger and DeCarrico 1992:62). Thus the greater than normal amount of examples is warranted as it highlights the fact that the phenomenon and its definition can cover a wide range of formulaic sequences. That was at least the case for the textbooks examined by the current study. Perhaps the strongest one of the faint patterns to be found was that both books included more than one phrase of the type highlighted by the examples

44 and 47. These phrases explicitly flag the purpose and set the tone of their sentences. Coincidentally, this type of phrase was often, though not always, a phrasal constraint. Apart from this particular pattern, it was also fairly common for a phrase of conversational purpose to be a question. Both 46 and 48 explicitly label their sentences as queries. Structurally the phrases are a phrasal constraint and a sentence builder respectively. Then there were also phrases that simply functioned as statements such as *the odds are* in 45. Lastly there was the type exemplified by 49, although it might wrong to call this a type, because the common factor here was that phrases such as this were borderline cases in whether they expressed conversational purpose at all. It would have been, for instance, entirely possible to classify *through and through* as a discourse device. The current study made the choice call the expression a phrase of conversational purpose, because the sequence still is more an expression of message reinforcement than a logical connector. This example and all the previous ones should highlight the fact phrases of conversational purpose were a complicated and even a contradictory group. This was the case for both textbooks, and it is the opinion of the current study, that the variety is a natural corollary of how the category is defined.

5.3 Idioms

The following chapter focuses on the presence of idioms in the two textbooks, *ProFiles* and *Culture Café*. The issue will be discussed from three interconnected points of view. The first of these are the pure numbers of how many idioms the books contained. Secondly, the chapter will also delve into the varying formal aspects of idioms. Lastly, the chapter will include an analysis of the functions the idioms exhibited.

Table 7. Idioms – a comparative table of occurrences and individual phrases

	ProFiles	**Culture Café**
Formulaic sequences – occurrences	802	388
Idioms – occurrences	443	152

Idioms – individuals	157	100

The figures in Table 7 show that there were some clear patterns in the amounts of idioms. First of all, corpus studies (Moon 1998) suggest that natural discourse gives idioms a very small role and an almost random change of occurrence, but this was not the case for either of the textbooks. For both the books idioms comprised of roughly half of all occurrences of formulaic language, which is considerably more than results of the aforementioned corpus data. It is, in fact, even more than the 14 % Koprowski's (2005) study found to be the relative quantity of idioms. It is, however, important to understand that the similarity between the two textbooks was only superficial. The best sign of this is that *ProFiles* not only had over twice the number of idiom occurrences, it also had more individual idioms than *Culture Café* had occurrences. There was also a marked difference in ratio of occurrences versus individuals. While *Culture Café* had roughly one and half times as many occurrences as individuals, *ProFiles* had nearly three times as many. The difference in the ratios was almost entirely due to the fact different patterns of repetition. Both the textbooks had idioms that recurred in text, but *ProFiles* had an overwhelming numerical superiority in exercises that focused on and thus also repeated idioms. Thus the difference was mostly of textbook design rather than the quality of texts. This difference, however, will be discussed in more detail in later chapters dealing with ways of presenting formulaic sequences.

The opacity of meaning and the relative rigidity of form are the two main structural characteristics of an idiom, and this paragraph focuses on how the first of these manifested itself in the idioms found in the data. It must be mentioned that the example 50 is not given in its original form, as the sentence is from a punctuation exercise, and the commas missing from the original form have been added here for the sake of clarity.

50) *But on the other hand <u>a man's gotta do what a man's gotta do,</u> hasn't, he added.* (CC:133)

51) <u>*Break a leg!*</u> (CC:36)

52) *I will <u>wear my heart upon my sleeve</u> for daws to peck at.* (CC:44)

53) *Many people were surprised when Madonna <u>turned her hand to writing</u> children's books.* (PF:23)

54) *But hop over the border into Italy and it's <u>a whole new ballgame</u>.* (PF:15)

55) *<u>I'm a philistine</u> through and through.* (PF:14)

Opacity of meaning is difficult to scale. One can easily say that and idiom is highly metaphorical, lightly metaphorical or something in between, but is exceedingly difficult to create an absolute scale out of something that is inherently subjective. Despite this it was fairly easy to see that the vast majority of idioms in the textbooks had only a slight touch of metaphorical meaning, and examples 50 and 53 serve to illustrate this point. Both undeniably have some metaphorical content. The first one can be taken as a reference to honour bound obligation, while the second doesn't have to allude to an actual hand being moved. On the other hand, the core of both expressions can be understood from their literal meanings alone. The prevalence of fairly transparent idioms had been confirmed by previous corpus studies (Moon 1998), and thus the textbooks were very much like natural discourse. Yet the textbooks also contained much more opaque idioms, though to a lesser degree. The examples 51 and 54, for instance, are anything but transparent, with *Break a leg* having a meaning that is utterly impossible to deduce from the component words. It is also worth noting that the books contained some idioms that derived their entire opacity from culturally sensitive knowledge. All idioms are almost by definition culturally sensitive (Teliya et al. 1998), but some fill this definition better than others. The example 52, for instance, can only be fully understood by someone who knows British culture well enough to recognize the idiom as a Shakespeare quote. *I'm a philistine* in 55, on the other hand, is a Bible reference, and could thus be opaque for learners with non-Christian cultural backgrounds. All in all, it could be argued that the internal distribution of idioms was in line with natural discourse, even if the actual number of idioms in the books was not.

A relatively fixed form is the second main structural feature of all idioms. While the criterion has been studied and codified (Fraser 1970), and should thus have been open for further study, the two textbooks did not make the process easy, or to put it more accurately, the books presented idioms in way that made it difficult to evaluate their fixedness. This was the case especially for *Culture Café*. It should be noted that

66

more than one occurrence is listed in the examples below when necessary, because variation between different forms is the only reliable way to spot and evaluate the fixedness of form. The occurrences are separated by a slash (/).

56) *...special effects that would later become his <u>stock-in-trade</u>. / Jackson's <u>stock-in-trade</u>. (CC:26/CC:28)*

57) <u>*BRIDGING THE*</u> *CULTURE* <u>*GAP*</u> (PF:21)

58) Jason sure thinks a lot of himself and his talents, doesn't he? Yeah, I wish he wouldn't (<u>blow his own trumpet</u>) all the time (PF:38)

59) <u>Try your hand</u> at this ancient craft by following the step-by-step instructions below to create a simple everyday object. / Will <u>tried his hand</u> at his father's trade of glove-making to help support the family. (PF:93/PF:172)

The vast majority of all the idioms in the two textbooks were like the one in 56, where it was impossible to pass judgement on how fixed the form was. In other words, the idiom either occurred only once or was always repeated in the exact same form. In the first case there was no way to spot the variation that could potentially exist, and in the second one, there was no variation to be found. These types were particularly prevalent in *Culture Café,* which had no idioms whatsoever in any kind of inflected form. Thus the last three examples concern *ProFiles* exclusively. They reveal that if there was any variation to be found, then it was done in one of two possible ways. It could be that the variation was made self-evident by the context, even if there was only single occurrence. It is, for instance, clear that the word *culture* has been added to the expression *bridging the gap*. It is also equally clear that the pronoun in *blow his own trumpet* can be inflected and correlates with the preceding pronoun. Then there were also some instances, such as the phrases in 59, which provided explicit evidence on the variation. The pronoun in the idiom can, after all, self-evidently be both *your* and *his*. Given that there rarely was any variation of idioms' forms even in *ProFiles*, the current study does not try to construct statistically reliable figures on the types of variation. Nevertheless, it was possible to see some tentative patterns. Most of the variation was comprised of simple grammatical inflection of the idiom's internal components, which was the case in both 58 and 59. In other words, the variable idioms mostly stayed on adjunction, lowest rung of Fraser's (1970) ladder. In addition to this there were some

isolated incidents of wider variability. Such is the case in the example 57, where a non-idiomatic element has been added to the idiom. Thus the expression is placed on insertion, the second level of Fraser's (1970) model. All in all, it was relatively clear that both the textbooks presented idioms in very fixed forms. *Culture Café* naturally exhibited a much stronger tendency towards this, due to its complete rejection of variable forms.

It is possible to classify idioms according to what they are used for. The current study chose not to present these results in a statistical form for two separate reasons. Firstly, the framework of functions proposed by Fernando (1996) is intended to serve as tentative list of possible functions, and not as watertight model of all the meanings an idiom can take. Secondly, the data of the current study proved to be rather difficult to unequivocally classify using Fernando's (1996) model. In other words, a considerable number of idioms in both books had two or more possible functions, which led to enormous overlap between different categories. Given these two details, a statistical presentation would have been a misuse of the original model, and would not have given meaningful results. Thus the examples below and the analysis that follows them are intended shed light on the confusion that seemed to be inherent for the idioms

60) *She'd shaken us up a little, but surely that was just an act designed to <u>weed out</u> <u>the deadweight</u>.* (CC:93)

61) *What this example suggests is that we are <u>barking up the wrong tree</u> by trying to define installations.* (CC:79)

62) *It'll <u>break your heart</u> to dump a good idea but professional writers do it all the time.* (CC:33)

63) *I must say this young painter is completely different from the rest - she really <u>breaks the mould</u>.* (PF:23)

64) *Yes, <u>an ugly customer</u>, one you wouldn't forget in a hurry...* (PF:63)

The model mainly deals with idioms that reflect their user's experience of the world (Fernando 1996:97), and idioms in the data filled this definition quite handily. Apart from this superordinate definition, however, the functions of the idioms proved to be difficult to pin down. Naturally there were some expressions that clearly had only one function. This was, for instance, the case for *to weed out* in the example 60, which clearly is a metaphorical expression describing an action and nothing more.

Idiom could and did, however, become much less ambiguous, as is shown by *the deadweight*, the second idiom in the example 60. It could equally well be label from the category of people and things or a straightforward attribute. It is equally unclear whether the example 61 was a description of an action or an evaluation of the said action. What is more, two was in no way the upper limit of potential functions. Example 62, for instance, has four with event, situation, evaluation and emotion. The same goes for 63 that could have any of the labels action, event, evaluation or attribute. The multitude of functions can be partly traced to Fernando's (1996) classification as it is a slightly subjective. It is, however, also possible that idioms themselves have multiple parallel functions. This was at least the case for the ones found *Culture Café* and *ProFiles*, and it is the opinion of the current study that there is no reason to assume that this would not also be true in other contexts.

5.4 Collocations and phrasal verbs

The purpose of the following chapter is to discuss the presence of collocations and phrasal verbs in the two textbooks. As it has been said before, these two categories technically fall outside the bounds of the research questions. This was the especially true for phrasal verbs, which the current study does not see as formulaic sequences, and did not expect to encounter in formulaic contexts. They are only included because *ProFiles* seemed to think otherwise. The current study chose not to include any theoretical background on phrasal verbs even after it became apparent that they had a minor formulaic role. The choice was made, because such addition would have meant an unacceptably large expansion of the theoretical background for a very small group. Suffice to say, that they can be seen to have some formulaic properties. All in all, it must be admitted that even the limited inclusion of collocations and phrasal verbs stretches the limits of the study questions. Yet the stretch is not unacceptable. Its effect is very limited due to the restrictions discussed in the next paragraph. Moreover, the two groups could not be completely excluded with a clear conscience, when they were so clearly seen as formulaic by the textbooks.

It is very important to understand that neither collocations nor phrasal verbs were actively looked for in the data. All the figures and examples only list collocations and phrasal verbs that were explicitly labelled as formulaic sequences in the textbooks. This leads to three things. Firstly, the figures below are not entirely compatible with

the other figures in the current study, because they do not list all the collocations and phrasal verbs that could potentially and very likely have existed in the data. Secondly, the current chapter will not go into detail about the possible characteristics or subdivisions, but only state the existence of the two categories and sizes of their presence. Thirdly, as the two categories are only interesting because of how they were presented, their main discussion will be in chapter 6.1, which focuses on contexts and methodology of formulaic sequences.

Table 8. Collocations and phrasal verbs – a comparative table

	ProFiles	**Culture Café**
Collocations – occurrences	87	69
Collocations – individuals	33	37
Phrasal verbs – occurrences	20	n/a
Phrasal verbs - individuals	7	n/a

First of all, it is fairly easy to see that both the books contained collocations in similar proportions. *Culture Café* had slightly more individual expressions but *ProFiles* repeated what it had more often. The main difference between the two books was their handling of phrasal verbs. Both books of course had phrasal verbs in some capacity as it is impossible to avoid them in natural text, but only *ProFiles* handled some of them as formulaic sequences. Even then it is important to understand that 7 individual expressions give phrasal verbs a very minor role compared to the overall number of formulaic sequences found in the books.

5.5 Summary

The first research question of the current study concentrated on two fairly clear-cut topics. This paragraph focuses on whether there were formulaic sequences to be found. However one may decide to look at the issue, the results were quite clear. To begin with, *ProFiles* had more of nearly everything. This was particularly evident in

the absolute number of occurrences, of which *ProFiles* had over two times as many as *Culture Café* did. The same difference in the number of occurrences held true for all the subcategories and even led to *ProFiles* handling some phrasal verbs as formulaic sequences. There were no occasions of this in *Culture Café*. The differences in the amounts of individual sequences were, however, much less straightforward. *ProFiles* had only slightly more idioms than *Culture Café* and with lexical phrases the difference was practically non-existent. Some structural categories of lexical phrases even had the situation reversed. To put it bluntly, the results of occasions and individual sequences combine into a simple fact. *ProFiles* repeated what it had, while *Culture Café* did not.

The second part of the first research question was simply this. What kinds of formulaic sequences did the two textbooks contain? The textbooks had remarkably similar profiles when one looks at individual sequences. To begin with, the relative numbers of idioms were out of all proportion with what previous corpus data has suggested. The absolute numbers were, of course, different but the same principle applies. As it comes to lexical phrases, the largest category for both books was formed by short polywords, and nearly all of them also belonged to the functional category of discourse devices. It is also noteworthy that both textbooks had surprisingly large contingents of collocations that were explicitly labelled as formulaic sequences and thus had to be taken into account.

6 PRESENTATION OF FORMULAIC SEQUENCES IN THE EXAMINED TEXTBOOKS

This chapter focuses on the second aspect of the current study, that is to say, how the formulaic sequences were presented in the two textbooks *Culture Café* and *ProFiles*. There is, however, some slight overlap with chapter 5 as it is impossible to discuss the patterns of teaching formulaic sequences without touching on the subject of what sequences were being taught. This was especially true for collocations and phrasal verbs, which in the context of the current study derive their entire position and meaning from being explicitly taught. It is also important to understand that while the actual method of analysis remains the same, this chapter introduces a change in how the actual results are presented. They are given separately for each of the books

and not organized according to themes, as was the case in chapter 5. The choice was made because there was a vast difference in how the two textbooks presented formulaic sequences, and these profiles are easier to present and make apparent when the analysis of the books is kept separate. Presentation of both these profiles is of course followed by short sub-chapter that explicitly summarizes and compares the results.

6.1 Culture Café

6.1.1 Implicit contexts

Implicitness is one of the key concepts involved in the teaching of formulaic sequences, and the current study defines it in the following fashion. A formulaic sequence is implicitly presented whenever it is a natural part of its co-text and no special attention is paid to it. In other words, it is up to the reader whether or not he notices the sequence. A good example of implicitness is the beginning of the previous sentence where the discourse device *in other words* stands. The concept of implicitness was particularly crucial for the textbook *Culture Café*, because it was the primary means of introducing formulaic sequences in the book. The following examples have been chosen to shed light nature of implicitness in *Culture Café*..

> 65) *A few years <u>down the line</u>, you are somewhere else, reading this. (CC:9)*
>
> 66) *Jackson's next films marked <u>a sea change</u> in his work, both in terms of budget and subject matter.* (CC:27)
>
> 67) *Oh, <u>by the way</u>, your careers counsellor _____ after your future plans the other day.* (CC:30)
>
> 68) *<u>Reading between the lines</u>* (CC:51)

These four examples illustrate the nature of implicitness in *Culture Café*, and thus also its typical way of presenting formulaic sequences. The textbook introduced formulaic sequences mostly as undifferentiated parts of whatever text they belonged to. They were embedded in context, and no attempt was made to acknowledge their existence. There was, however, more than one type of context that contained formulaic sequences. The vast majority of implicitly presented sequences were found in the reading texts of the book, and therefore the examples 65 and 66 come from such contexts. It is only natural that most of the sequences came from reading texts,

because reading text themselves formed the bulk of the material in the textbook. Yet there were also other contexts that contained implicit formulaic sequences. The example 67 is from a gap filling exercise focusing on vocabulary of the related reading text. The sequence *by the way* was neither the target expression in the exercise, nor was it found anywhere else in the exercise or the reading text. The situation is an example of a recurring situation where a sequence was both implicitly presented and totally secondary to the purposes of its context. One could even argue that cases where the sequence was embedded in an exercise text actually made the sequence more difficult to notice, as the reader´s attention is explicitly directed towards other issues. The last of the implicit contexts found in *Culture Café* were headlines of various kinds, which constituted a borderline case between implicitness and explicitness. The sequence *reading between the lines* was printed boldface and in larger font than rest of the text, and it could thus be argued that the book uses these textual means to explicitly raise the sequence to the reader's attention. Yet the current study argues that this example and the likes of it were, in fact, implicit because the same means were also used for all the other headlines. In this light it seems that the highlighting of formulaic sequences was completely coincidental.

6.1.2 Explicit contexts

The textbook *Culture Café* contained very few occasions where a formulaic sequence was presented explicitly. Broadly defined, there were only four types of such behaviour and even then the explicitness was fairly weak. Given the relatively low number of types, the current study reports and discussed them all. The decision is also influenced by the fact that all the types were somewhat different from each other, and thus a fruitful comparison between them can be made. The results are placed in a tentative descending order or explicit, but the reader should take note that this order is not absolute, as the lines between anything but the extreme cases can be murky.

The most explicit type of presentation in *Culture Café* was found in two separate exercises, and interestingly, these two were also the two most controversial cases, because they contained all the collocations acknowledged by the current study in *Culture Café*. While there were some differences in the actual methodology of the

two tasks, the basic principles were similar enough to warrant both of them being covered by this single entry.

69) *curtain _____ (closing) _____* (CC.44)

70) *comic _____ (relief)_____* (CC:44)

71) *we are <u>barking up the wrong tree</u>* (CC:80)

72) *it <u>compels engagement</u> from the viewer* (CC:80)

73) *deep down he was a loner (<u>at heart</u> he remained an outsider)* (CC:66)

The first two examples, 69 and 70, come from a gap filling exercise where the reader was given 10 sets of word pairs that had occurred in the preceding reading text. The first part of these pairs was presented directly while the second part was located in a hint box. The reader was prompted to write the correct second part on the line following the first part and then translate the whole word pair. Examples 71 and 72, on the other hand, come from an exercise where 7 phrases from the preceding reading text were printed in a short context with the actual phrase highlighted by the use of italics. The student was then prompted to think what the phrases meant in the given context. The example 73 comes from an exercise where the student was given short expressions and told to find the alternate expression that had been used to express the same thought in the reading text. The last example gives the text's version in the parentheses with, while the hint precedes it. The actual formulaic sequence is underlined. Despite their superficial differences the three exercises and their five examples exhibited profound similarities. First of all, each of the three tasks explicitly used labels that can be taken as admissions that the expressions were considered formulaic. The first task called the phrases words that "often appear together" (CC:44) while the second one explicitly labels the italicized expressions as "phrases" (CC:80). The third one uses the weakest label "useful expressions", which still recognizes the sequences of words as coherent entities. The second similarity was in how the tasks both isolate the sequences and presents them in context at the same time. All three did this by focusing the reader's attention on specific expression, but also referring back to the reading text and thus making sure that the proper context was understood. The final major similarity was that all the exercises contained a high proportion of expression that were not particularly formulaic. All of the expressions in the first exercise were collocations or expressions that stretched

even that definition. The situation was similar for the second exercise with the difference that two out of seven expressions were idioms with the rest being collocations. The pattern was even more marked for the last exercise which contained only one expression, the example above, which the current study considered even slightly formulaic. The rest could have, at most, been called collocations but the current study chose not to do even that. The prevalence of collocations can be taken as correlation with the results of corpus studies on natural discourse (Moon 1998), but further speculations about this are beyond the scope of the current study. Suffice to say that the most explicitly presented formulaic sequences in *Culture Café* were also the least formulaic ones.

The second strongest case of explicitness in *Culture Café* was of a particularly strong type. In fact, the only reason it did not rank higher than this, was that there was no exercise connected with the otherwise highly explicit presentation.

74) *To be or not to be: that is the question.* (CC:44)

75) *Why, then the world's mine oyster.* (CC:44)

76) *To be or not to be* (CC:44)

77) *Greek to me* (CC:44)

All of the examples above are from an exercise that familiarized the reader with famous quotes from the plays of William Shakespeare. Though technically, the context was not an exercise, because no task was involved. The book merely told that the quotes are important to recognize. Secondly, this case of explicitness is actually two connected contexts as there were 10 long passages explicitly labeled as Shakespeare's famous quotes along with 6 shorter ones given in the related instructions. There was one quote present in both context, and thus it is also printed in both examples 74 and 7. Additionally, the idiom in the example 77 was reprinted and translated as *Se on ihan hepreaa* (CC:44) in the tiny related vocabulary infobox. Both of these contexts can be called explicit for two reasons. To start with the quotes in both contexts are textually highlighted so that it is not solely up to the reader to notice them. Secondly, the relatively short forms of even the longer quotes make citations easy to notice and absorb. Thirdly and most importantly the instructions of the text explicitly state that the given quotes have become "catch phrases in modern day speech" (CC:44) or "widely spoken pearls of wisdom" (CC:44). What is more,

the reader is prompted to become conversant with these expressions as he is "sure to meet them in the future." (CC:48). All these factors combine to create a highly explicit presentation that not only presents actual sequences, but also presses home the fact that formulaic sequences are culturally sensitive conventionalized expressions. The presentation also indirectly seems to acknowledge that individual formulaic sequences are rare, when it encourages to the reader to familiarize himself with the expressions, but does not provide a task to train their usage.

The third strongest case explicitness *Culture Café* was found in an exercise near the beginning of the book. All the following examples come from this context, and they have been slightly simplified for the sake of intelligibility by omission of the quotation marks and the serial numbers found in the original. The underlines were also not part of the original context. Both the quotation and the translation are given in the same entry and separated by a slash (/).

> 78) *The odds are greatly in my favour. / Pidän parittomista.* (CC:36)
>
> 79) *I had two slices of bad pizza, went to bed and grew a conscience. / Söin pizzaa, menin sänkyyn ja oksensin omantuntoni.* (CC:36)
>
> 80) *You are not as thick as to open your mouth. / Ole hiljaa, läski.* (CC:36)
>
> 81) *Have you ever cut classes? / Oletko koskaan leikannut lasia?* (CC:36)
>
> 82) *Break a leg! / Katkennutta jalkaa sinullekin!* (CC:36)

The exercise in question contained 10 quotes from English speaking movies along with their incorrect translations. Out of these 10 excerpts there were 5 that were considered formulaic sequences. All of these are presented in the examples above. The other 5 quotes were non-formulaic expressions that had been mistranslated, such as *I'm invincible* becoming *Olen näkymätön* (CC:36). The instructions of the exercise explicitly state that the translations are wrong, and that the reader should try to figure out what why that is. There are compelling reasons to call this exercise explicit *Culture Café* even though half of the mistranslations were not formulaic sequences and no terminological label is given for the expressions. First of all, half of the expressions were clear formulaic sequences, and many of those that were not could have been called borderline cases. As such the exercise contained too many

formulaic sequences for the occurrences to be either coincidental or unintentional. Moreover, whenever a quote contained a formulaic sequence, then that was the element suffering from mistranslation. Yet the most compelling of the justifications is that the exercise directly concentrates on three key aspects of formulaic language: opaqueness of meaning, rigidity of form and cultural sensitivity. The exercise essentially forces the reader to acknowledge that the sequences cannot be understood literally and cannot be translated directly, and that correct translations might require some completely different expressions in Finnish. Thus it can be said that the level of explicitness in the exercise is somewhat difficult to grade. On one hand, the exercise not only focused on familiarizing the reader with individual formulaic sequences, but also seemed to discuss the features of the entire phenomenon. On the other hand, no labels are given and no sequences were highlighted using textual means. The exercise essentially seemed to stop just short labeling the sequences but did just about everything else.

The second weakest type of explicitness could essentially be called the random items basket of the formulaic sequences in *Culture Café*, and it comprises of exercises found mostly in the revision section at the back of the book.

83) *But the shape - _____ (let alone) the material - is not necessarily important.* (CC:124

84) *To/In/On the contrary, I think he shouts.* (CC:134)

85) *"Perhaps I'm barking _____ (up) the wrong tree but I think 'sampo' is the object attached ___ the seat there."* (CC:131)

86) *Therefore, it would only be too easy to give up and panic.* (CC:132)

These examples represent a type of paradox. The exercises themselves make no mention of having something to do with phraseological entities. The first example is, in fact, the only one that was found under the heading of vocabulary exercises. The textbook considered the rest to be grammar exercises. Thus in the traditional Chomskyan sense it would be impossible for the last four examples to discuss formulaic sequences, which are by lexical items by definition. The contradiction becomes even more evident, when one considers that the vocabulary section contained only 10 occasions of formulaic language while the grammar section had nearly 30. What is more, all the target expressions in the exercise were not formulaic

sequences. Yet there is sufficient reason to call all these occasions explicit. Essentially, the explicitness of this type focused the reader's attention on formulaic sequences as discrete and relatively inflexible entities. The examples 83 and 86 used the formulaic sequences as necessary and sufficient answers for their exercises. No other phrase was either necessary or suitable to serve as the answer. The example 86 even went as far as to print the formulaic sequence in boldface in order to specify which phrase was to be translated. On the surface it might seem that the examples 84 and 85 did not see their target phrase as fixed entities given that they broke internal structure of their target phrases. The focus on the prepositions was, however done in a way that highlights that there is only one suitable preposition for the given expression. In other words, the sequence was given as inflexible. All in all, explicitness of this kind was fairly weak and somewhat difficult to pin down. Yet its existence cannot be doubted.

The weakest case of explicitness in *Culture Café* was found near the end of the book in a collection of poems and short stories intended as optional reading material. There were only 7 cases of explicitness, but because the cases were nearly identical with each other, only two examples are given here.

> 87) *The teacher <u>forged on</u>, and we learned that Carlos the Argentine bandonion player, loved wine, music, and, in his words "making sex with the womens of the world."* (CC:93)

> 88) *The "Dillingham" had been <u>flung to the breeze</u> during a former period of prosperity when its possessor was being paid $30 per week.* (CC:98)

All the cases such as these two were presented in the same simple fashion. The sequences seemed to occur naturally, and they were highlighted only with number in small font referring to the corresponding translation printed in the margin. The translation for the first example was *paahtaa eteenpäin* and *lisätä huolettomasti* for the second one. The attention given to these seven expressions can only barely be called explicit as presentation of formulaic sequences can by no means be called the primary purpose of these annotations. There were, after all, well over 200 other translations given. Yet one cannot deny the fact that formulaic sequences, all of them idioms, were deemed important or difficult enough to warrant their own annotations. Moreover, the annotation numbers always came right after the end of the expressions themselves and can thus be considered as effective textual means as any underline or

special font would have been. The annotation number does, after all, break the reader's attention from the text in the same way. Additionally, the translations themselves open the meanings of the idioms to the reader. Therefore, the current study took the stance that the annotations represent a form of explicit attention, even though this attention is fairly weak.

6.2 ProFiles

The textbook *ProFiles* handled formulaic sequences in a very different way than its counterpart *Culture Café*. This might not be readily apparent, because both books did, after all, have the same basic patterns of implicitness and explicitness. Yet *ProFiles* applied these patterns in a much more systematic fashion and thus contained many, many more formulaic sequences. The presentation is structured somewhat differently than with the *Culture Café*. The current study has chosen to analyse the archetypal cases of implicitness and explicitness at the end of this section, as the main form of presentation in *ProFiles* does not neatly fall into either of these categories. Thus the special case will be discussed first. What is more, the archetypal categories had similar presences in both textbooks, and thus the current organization of this chapter directs the reader's attention to differences instead of similarities.

6.2.1 The Phrase Bank task cycle – from implicit to explicit

This first category concentrates on a very specific kind of explicitness that from now on will be called the Phrase Bank task cycle. This type always began with a reading text being embedded with formulaic sequences, which were rehearsed in later exercises. As such, the type is a combination of the implicitness and explicitness. What is more, it corresponds with the strongest type in *Culture Café*. Therefore the type would not normally qualify for an entry of its own, but in the case of *ProFiles* it is justified. This is, first of all, because the type constituted the textbook's primary explicit means of presenting formulaic language. Secondly, it can be considered that that the pattern was essentially institutionalised within the book, because it was both explicitly labelled and methodically applied. Thus despite the terminological overlap it deserves an independent analysis outside the main entries of explicitness and implicitness. It also needs to be stated that the following analysis is organized to fit

its subject matter. Because the category itself was a multi-staged process, the following analysis will focus on those stages and how they were realised. The current study also uses the same phrases as examples for each of the stages in order to make it explicit, how the task cycle itself repeated those phrases. This also aids in the understanding of how the different stages functioned.

The first phase of primary method always began by introducing a formulaic sequence in a reading text. The following examples will highlight this particular part of the process.

89) *But hop over the border into Italy and <u>it's a whole new ballgame</u>.* (PF:15)

90) *After a while, they jumped out of a first-floor window, ran to a churchyard and tried <u>to commit suicide</u> with a shotgun.* (PF:29)

91) *I found myself <u>getting hooked on</u> the first-person narrative all over again.* (PF:42)

92) *It <u>won't cost you an arm and a leg</u> either as the event is included in the price of your Tower of London admission ticket.* (PF:94)

For all intents and purposes, the first phase was always almost stereotypically implicit. The sequences were embedded in their respective reading texts and no textual means were used to highlight them for the reader, and as such, everything that has been said about the implicit contexts also applies to this phase. Even though 3 of the four examples above are idioms, there was no such clear division for the actual occasions themselves. Apart from the general overrepresentation idioms evident also in all the others contexts, the distribution of formulaic sequences was fairly even. There was, however, one marked exception to this rule. A considerable number of collocations and all the phrasal verbs taken into account by the current study originated from this context. The collocation *to commit suicide* in the example 90 is a good illustration of this. It should be noted, however, that the implicit contexts themselves did nothing to focus attention on the collocations. In truth, the current study looked for the collocations in the reading texts only after they had been identified by the exercises. In other words, they were not taken into account because of what they were, but because of what was done with them. Thus their inclusion does not constitute a breach of the borders established in the study question. To not

include them in this limited context would have been a violation against the other part of research question to include all explicit presentations of formulaic language.

The second part of *ProFiles* main method is its most important part, because here it became apparent that formulaic sequences were an institutionalised part of the textbook. In other words, *ProFiles* had an exercise category called the "Phrase Bank". These exercises were defined as containing conventionalised expressions from the reading texts (PF:3), which is essentially the traditional definition for formulaic language. What is more, there was such an exercise connected with 7 out of 11 reading texts in the book. These facts make it clear that a relatively large part of the textbook explicitly and systematically concentrates on formulaic sequences. The following examples have been chosen to illustrate the actual nuts and bolts of the process.

93) *aivan eri juttu* _____ (*a whole new ballgame*) (PF:17)

94) *jäädä jnk lumoihin* _____ (*to get hooked on*) (PF:45)

95) *tehdä itsemurha* (*A to commit suicide*) (PF:32)

96) *to be very expensive* _____ (*to cost and arm and a leg*) (PF:98)

The first two examples represent how five out of seven Phrase Bank exercises functioned. These tasks always began with explicit instructions to find certain phrases in the reading text. The reader was then given the Finnish translations of the target phrases and an empty line beside each on which to write down the phrase. Yet there were two Phrase Bank exercises that were an exception to the rule, and both of them are represented in the examples above. The example 95 illustrates an exercise where the phrases were given in Finnish, and the reader was prompted to connect each phrase with its English equivalent. These, along with 4 extra alternatives, were given in a separate box beside the phrases. Each alternative was given an alphabetical label which the student was supposed to write before the correct Finnish translation. Thus the student did not actually have to produce the target phrase. The example 96, on the other hand, illustrates the last subtype of Phrase Bank exercises. This type was highly similar to the majority, and the only major difference was that both the hint and intended target phrase were in English. As can be seen from the example 96, the student essentially had to comb the text for and formulaic synonym for the non-formulaic hint, and then write it down. There was also the difference that

the instructions of the task were somewhat more expansive, as they stated the target phrases to be "colloquial and idiomatic expressions" (PF:98). This served to strengthen the explicitness of the focus on formulaic sequences. Lastly the exercise with the example 96 gave the first target phrase as a model on how the task was to be done. No other Phrase Bank exercise provided such instructions. Thus all but one of the exercises seemed to rely on the same two-step process. The reader was to find the target phrases in context with the aid of the hints that provided. Secondly, the reader was to write down the phrases, which can be considered a very limited form of production practise. It is notable, though, that apart from this production was never rehearsed in this context. The subtype illustrated by the example 96 is a small deviation from the two step rule, as it did not require the reader to either find or produce the target phrases. Yet it could be argued that even it followed a strongly modified version of the same pattern, because even it required the student to have read the reading text.

The last phase of method constructed around the Phrase Bank exercises was the one where the actual use of the target phrases was rehearsed. These drills were usually neither extensive nor particularly free-form, but they existed nonetheless. Interestingly, these exercises were not technically a part of the main task cycle. The Phrase Bank exercises and their implications were discussed in the previous paragraph, but this last phase fell under a completely different label called Word Power exercises. The textbook described them simply as vocabulary exercises (PF:3), and tasks connected with the formulaic sequence task cycle were far from the only ones to fall under this label. Yet the connection between these drills and Phrase Bank exercises cannot be denied. All Word Power exercises with such connections made it explicitly clear that the student was supposed to use the phrases from a preceding Phrase Bank exercise. It is noteworthy, however, that even the Word Power tasks directly connected with Phrase Bank contained non-formulaic vocabulary as target phrases and the instructions often only stated that "some of the expressions in the exercise 4D will help you." (PF:46) The 4D refers, of course, to the Phrase Bank exercise. The wording makes it explicit that formulaic sequences are not the sole focus of the exercise. In addition to the target phrases used as examples for all the stages, there are four extra ones presented below. they represent exceptional contexts that did not coincide with four standard examples.

97) *Circuses have been around since ancient times, but the latest approach is _____ (a whole new ballgame). (PF:18)*

98) *Many people believe that Kurt Cobain didn't _____ (commit suicide), but was murdered. (PF:32)*

99) *What's your theory? _____ (Why are people hooked on the characters) and are dying to know more about them. (PF:46)*

100) *This phone _____ (costs an arm and a leg), but it's worth every penny. (PF:99)*

101) *Sanoisitko, että se oli käännekohta urallasi? – So would you say that it marked a turning point in your career? (PF:84)*

102) *valtavat lipputulot (raked in millions at the box office) – miten vaikutti ohjaajan elämään? (PF:108)*

The first four examples represent the type of exercise that covered 7 out of 10 cases. All of these occasions were nearly identical gap filling exercises where the reader had to place the target phrases in context. The main difference between the tasks was whether or not the reader was given clues, and whether those clues were in English or Finnish. The examples 97 and 99 had Finnish translations as clues, while 98 and 100 had none whatsoever. There was, in fact, only one occasion where the clues were in English and this was in an exercise not represented in the examples above. It is also noteworthy that only the exercise of the example 99 encouraged any further production beyond the act of filling the gaps. The exercise in question did this by prompting the reader to find a pair to work with and to act out the exercise that was constructed in the form of a dialogue. The last two examples and their corresponding exercises were a marked deviation from the main pattern of gap filling. Both of them were communicative exercises where the target phrases were embedded. The one represented by the example 101 was structured so that both students were given their own lines in Finnish and English translations for their partner's lines. The example 101 lists both of these in the same entry. The students were to act out the dialogue, and it was implied that each partner should use the translations to check correctness of the other's output. The exercise containing the example 102 was of similar purpose but of slightly different methods. While the previous had the entire conversation laid out in both languages and whole sentences, the one with the example 102 had a more parsimonious approach. Students were only given short bullet point summaries of each line in Finnish. It was stated that the expressions from the previous Phrase Bank exercise would be of use, but the students had to infer

themselves where the phrases were to be used. The clues were not detailed enough to unequivocally point out the right place.

Technically, there was also a fourth phase to the Phrase Bank task cycle, but it was intended as a revision, and was nearly identical to the third phase. In other words, the textbook contained a section called the BackTrack which was explicitly said to "help you revise the words, word partnerships, phrases and structures that come up in each unit." (PF:161) All the formulaic sequences in this context were originally from the Phrase Bank exercises. Thus the purpose of the section can be indirectly but very clearly linked to formulaic sequences in general and specifically to the Phrase Bank task cycle, even though the term Phrase Bank was never mentioned. It is also interesting that while the aforementioned introduction mentioned formulaic language fairly explicitly, the instructions on the task themselves made no mention of the phenomenon. At most the reader was prompted to complete the task using the clues given. The hint given for the example 103, for instance, was the Finnish translation "lyhyesti sanottuna" (PF:170) Thus one could argue that these exercises were intended as a more advanced form of practice, as the target phrases were flagged less explicitly, and consequently more strain was placed on the reader's own recollection of the reading texts. On the other hand, the reader was also given the correct answers to the tasks in a section further back in the textbook. Thus it could also be argued that the strain was not particularly great.

103) *Here's the plot _____* (*in a nutshell*). (PF:170)

104) *Will* <u>*tried his hand at his father's trade*</u> *of glove-making   to help support the family.* (PF:172)

The exercises themselves were very similar to the ones in previous phase of the task cycle. All but one of them were gap filling exercises, and every single one of them used Finnish translations of the target phrases as clues. The task that contained the example 103 was as stereotypical specimen. The only exception to the rule was the exercise with the example 104. It gave the reader a sentence in which a formulaic sequence was embedded and underlined. The reader was then prompted to translate the underlined parts into Finnish on a line below each sentence. As such, the exercise tested the reader's understanding of the target phrase, whereas the gap exercises seemed to focus on the mechanical aspects of remembering and producing the exact form of the formulaic sequences. There is also one last issue that must be discussed

in relation with the subject. Even though the BackTrack section explicitly places some focus on formulaic sequences, it is also clear that this is not the main focus. The instructions themselves state this, and the section contained topics up to and including nationality words. The focus could and did change in the middle of an exercise. This was the case for the task with the example 104, which began as gap filling exercise and focused on expressions such as "tuotteliaimpia näyttelijäkirjailijoita" (PF:172), which self-evidently was not a formulaic sequence. Therefore, it could be argued that the last phase of the task cycle was constructed as a clear part of the revision section, but this was not the only purpose of the revision section.

This last paragraph is intended as short summary about the main properties of the Phrase Bank task cycle. All the sequences present in it were embedded in reading texts. The main phase of the task cycle was the Phrase Bank exercises which explicitly focused on formulaic language. It is noteworthy, however, that some of the target phrases were collocations and other borderline cases, that would not have been included in the current study, had they not been explicitly focused on. The Phrase Bank exercises mainly consisted of tasks that prompted the reader to look for the target phrases in the text and then write them down. Next the cycle further rehearsed the use of the sequences with mainly gap filling exercises. Lastly, the BackTrack, the revision section at end of the book, devoted some but not all of its attention to sequences introduced earlier in the cycle. Given all these facts, the current study considers it justified to use the label Phrase Bank task cycle, as the formulaic sequences and their use are institutionalized and systematic part of the *ProFiles* textbook.

6.2.2 Purely explicit contexts

It was previously stated that *ProFiles* used the Phrase Bank exercise cycle as the main form of explicitness. This is indeed the case, but it is not the whole truth. Therefore the following chapter reports and analyses those forms of explicitness that did not belong to the main method. They were 6 major patterns of explicitness and they are organized here in a descending order of strength. The last three were, however, very nearly equally explicit, and their ordering can thus be considered somewhat arbitrary.

It is interesting that the strongest case of explicitness in *ProFiles* focused on collocations, and thus technically fell outside the bounds of the research question. Yet the sheer explicitness of the presentation made it impossible to not take these collocations into account. First of all, the exercise in question explicitly used the term collocation to describe the target phrases. More importantly, however, the task contained a relatively extensive info box on collocations as a phenomenon. It described collocations as "words that 'co-locate' or go together regularly." (PF:109) Given this fairly clear definition, it is noteworthy that half of the target phrases in the exercise were seen as idioms by the current study. These expressions were *to rub shoulders with, to ring a bell, to fall in love, to go on record* and *to shed light on* (PF:109). Particularly the second one is metaphorical enough to qualify as an almost stereotypical idiom. Yet this confusion need not be taken as anything more than a simple sign of the overall difficulty in defining and categorizing the phenomenon of formulaic language. These dilemmas have been discussed in detail in the theoretical background of the current study. Whatever is the underlying definition, however, the metatext even mentioned that learning collocations "will help to make your English more natural and fluent." (PF:109) This could be taken as a direct reference to how formulaic language is not only culturally sensitive, but also potentially beneficial for the mental processing of language (Wray 2004).

105) *to achieve* _____ (*fame*)_____ (*saavuttaa mainetta*) (PF:109)

106) *to ring* ____ (*a bell*) _____ (*tuntua tutulta*) (PF:109)

107) *Although some actors* _____ (*achieve fame*) *and fortune, most stay relatively unknown throughout their career.* (PF:109)

108) *Did you say Julianne Moore? Her name* _____ (*rings a bell*), *but I can't remember what she looks like.* (PF:109)

The collocation exercise itself was simple. The student was given beginnings of 10 collocations and told to match them with the correct endings from the box above. After having done this, the reader was supposed to translate the resulting collocation on a line given for this purpose. The examples 105 and 106 illustrate this process, and for the sake of clarity the current study has chosen to add possible translations in the second parentheses beside the examples. The second stage of the process was a gap filling exercise, where the reader had to place each of the collocations in correct sentences. The student was also told to pay attention to the correct grammatical

forms of collocations, as the collocations were given in their basic forms, but the contexts necessitated inflection. The stage is illustrated by the examples 107 and 108. It is also noteworthy that the phrases were not randomly chosen, but each one originated in the preceding reading text. In essence, the collocation exercise was highly similar to the Phrase Bank task cycle. The collocations did, after all, follow the same pattern. They were from the text, were initially drilled with simple mechanical task, and whose proper context and meaning were introduced with a gap filling exercise. The similarity is also more than superficial, because this collocation exercise and the immediately preceding Phrase Bank exercise both contained the two expressions *to go on record* and *to shed light on*. Due to these similarities the task could have been classified as a part of the main Phrase Bank task cycle, but it contained enough independent expressions that the current study saw this as impossible. What is more, the explicitly stated focus of the exercise gave it a distinct identity and profile.

The second strongest case of explicitness is *ProFiles* was also a very strong one. It comprised of two exercises that fell under the general heading of Word Power. The exercises were labelled as MUSIC IDIOMS and BOOKISH IDIOMS respectively. Thus the task made their focus on formulaic sequences explicit straight from the beginning. This was similar to what was done with the collocations presented in the previous paragraph, but there was one notable difference. Neither of the idiom exercises offered any explanatory metatext on idioms as a phenomenon. Therefore the readers were not offered any explanation on why the idioms in the tasks were idioms. It is possible that the textbook considered the term idiom to be either self-explanatory or part of common knowledge, and thus did not need to be explained in the same way as collocations. The hypothesis of familiarity could be supported by the fact that all the idioms in the task were highly opaque proverbs, and thus almost stereotypical idioms.

107) - *Jason sure thinks a lot of himself and his talents, doesn't he? - Yeah, I wish he wouldn't* _____ (*blow his own trumpet*) *all the time.* (PF:38)

108) *The team is very successful because all the members are on the same page.* (A *have the same aims*) (PF:51)

The details of the task themselves were fairly simple. The first one handled the music idioms through a simple gap filling exercise, as is shown by the example 107. The context of each idiom was structured in the form of a short conversation, where the first line always acted as a clue for the meaning of the target idiom, though this was not told in the instructions. The reader was also given each idiom in separate box above the exercise, so that he only had to pick the correct alternative from the list. What is more, each idiom in the list was already in the grammatical form needed to fit the context. Thus the reader himself did not have to inflect or modify the idiom in any way. The reader was also prompted to figure out the meanings and the possible Finnish equivalents of each idiom. The exercise focusing on bookish idioms, however, was not a gap filling exercise, but something quite different, as is shown by the example 108. The reader was given 8 sentences embedded with a single idiom each. The idioms were highlighted with boldface font. The reader was supposed to match the idioms with their correct meanings, which were given in English in a box beside the sentences themselves. The student was to indicate his choice by writing the alphabetical designation of the meaning before the corresponding sentence. The instructions stated it outright that the context of each idiom would help in deciphering the correct meaning. The reader was not prompted to translate the idioms. Despite these technical differences, the two idiom exercises seemed to share the same design philosophy. Both exercises focused explicitly on idioms. Both of them were designed around using the context to decipher the meaning of these idioms, and neither of them forced the reader to produce the idioms independently.

The third strongest case of explicitness in *ProFiles* was somewhat atypical, as there were no actual exercises directly involved. The case is formed by the so called *How come?* info boxes which were defined as giving information on origins of words and phrases. (PF:3) The definition alone links the info boxes with formulaic language. There were ten of these boxes scatted throughout the textbook, and five of them could be considered to have concentrated on formulaic sequences. All of these instances were idioms. The current paragraph discusses their case through two examples that were found on the same page. The first one discussed the phrase to be a *philistine*, while the second one concentrated on the opposed pair *highbrow* and *lowbrow*. (PF:16) Both of them were phrases that were found in the preceding reading text, and this was indeed made plain when the boxes gave the exact page and

line numbers for each of the phrases. This was also done in all the other *How come?* boxes, as all the other occasions were also connected with specific texts. The boxes can be considered particularly explicit because they go into painstaking detail in explaining the meanings of the phrases. The box, for instance, not only explained that *philistine* is a biblical allusion, but also explained when it was first used to mean crude and vulgar behaviour. The level of detail can be taken as fairly direct admission that the target phrases contained a level of meaning that cannot be fully appreciated without specific background information. This is essentially the same thing as saying that idioms are culturally sensitive and thus opaque to outsiders. (Teliya et al. 1998) One could also argue that printing the phrases in boldface and placing them in bright yellow boxes is the equivalent of increasing their typographic salience (Bishop 2004), even if the these things were not done in the original reading text context of the idioms. The functions of these info boxes seemed to be similar to how the early parts of the Phrase Bank task cycle raised the formulaic sequences to the reader's attention. A direct reference to the phrase's exact position functions similarly to a prompt to go and find an expression the text. The main difference here is that the info boxes go no further. As a final note, the similarities between the info boxes and the task cycle were not completely coincidental, as certain few expressions occurred in both. The idiom *highbrow*, for instance, was handled in this way.

The fourth strongest case of explicitness in *ProFiles* was rather atypical in many ways. First of all, it had absolutely no connection whatsoever with the reading texts or the Phrase Bank task cycle. Secondly, it focused on very different kind of formulaic sequences than the rest of the book. There was a list of 13 expressions that were labelled as "useful phrases for making speeches" (PF:131) The list was a part of a larger whole focusing of how to write a speech.

> 109) *Ladies and gentlemen…* (*formal*) (PF:132)
>
> 110) *On behalf of all us, I'd like to…* (*neutral*) (PF:132)
>
> 111) *It just remains for me to say…* (*formal*) (PF:132)

There were 13 phrases on the on the list organized under the three subcategories of introduction, making the main points and concluding. The three examples are from these categories respectively. The overall tone of the expression was given in parentheses after the expression itself. The three possible registers were formal,

neutral and informal. The explicit registers mitigated the fact that the phrases were essentially given without context apart from a short example speech that demonstrated the use of similar expression. Nothing was actually done with the expression, as there was no exercise connected with material. Despite these details, the list can be taken as a highly explicit presentation of formulaic language. To begin with, the aforementioned definition is very nearly the same things as saying that the focus is on conventionalised expressions with specific discourse functions, that is to say, on lexical phrases. Indeed, all the phrases would have been thus classified even without their explicit context. What is more, even the functions were made explicit in the aforementioned fashion. Secondly, the list itself indirectly strengthens the phrases' identity as fixed expressions by giving them in isolation and providing no opportunity to modify them. All in all, the list represented a serious attempt to provide access to a very specific subcategory of formulaic sequences. Yet that was all the list was, as there were no tasks connected with the information.

The fifth strongest case of explicitness in *ProFiles* was difficult to classify. It was indeed explicit, but it was somewhat questionable whether it focused on formulaic sequences. The exercise focused on non-finite clauses, and its explicit Finnish headline can be roughly translated as conventionalised phrases resembling shortened clauses. The terminology used makes the connection between lexical phrases and the exercise somewhat explicit, but further features seemed to offset this.

112) *Generally speaking, Finns are really good at languages.* (PF:159)

113) *generally speaking yleisesti ottaen* (PF:159)

114) _____ (*Generally speaking*), *you should call boss Mr Phipps, but most people call him Bob.* (PF:159)

The exercise focused on 10 sequences, and it provided four potential contexts for each of the phrases. The first possible context was in a model sentence, such as the example 112, in which the actual phrase was printed in boldface. Three phrases were given such a sentence. The second phase listed all the 10 phrases and their Finnish translations. Some of the expression were given alternative preposition or spellings, such as "compared with/to" (PF:159) The example 113 is an illustration of this phase, and the English expressions were again in boldface. The third phase was an infobox list which again presented all of the phrases in English, though in normal

90

font and without translations. The fourth and final phase was a gap filling exercise which prompted the reader to choose the correct phrase for each context. Five phrases of the ten phrases were handled this way. In light of these details, it seem fairly clear that exercise and the infolists concentrate on what the exercise metatext seems to label as lexical phrases. The current study, however, had some trouble in classifying some of the expressions as lexical phrases. The most extreme case was the expression "including" (PF:159), which clearly has a discourse function but cannot by any stretch of imagination be called a multiword-item. The expressions "compared to" and "judging by" (PF:159) are technically both several words long and have specific discourse functions, but is very difficult to say whether they are fixed phrases or purely grammatical constructions. The current study would not have taken them into account without the explicit context. Despite the caveats, one should take not that the exercise did repeat the phrases some many times that the context must be called explicit based on quantity alone, even if some of the individual expressions stretched the limits of formulaic language.

The second weakest form of explicitness in *ProFiles* was found in a task that was a borderline case between explicitness and implicitness. To begin with, the exercise was one of the so called Text Wise exercises which were explicitly labelled as concentrating on reading and listening comprehension. Thus the main focus of the exercise seems unlikely to have been vocabulary items such as formulaic sequences. Yet the examples below show that this was not the case. The original contexts have been shortened due to their excessive length.

115) *The name probably doesn't ring a bell...* (PF:104)

116) *When the film was released critical acclaim...* (PF:105)

The exercise was connected with a reading text, but it actually began as a primer on the previous page. The student was told that some parts of the text would be highlighted. The reader was then prompted to decipher the meaning of these eight parts using the context. The parts were indeed well highlighted using textual means, as they were not only printed in boldface but also surrounded by large red circles. Five of the eight highlights were straight-forward idioms like the one in the example 115. Yet there were three, like the example 116, that could have, at best been called collocations. Nothing apart from the comprehension was required of the reader, but

the exercise had an interesting connection with some of the other tasks in the reading text. The sequences kept cropping up in the later exercises. They were not part of the Phrase Bank cycle itself, but the collocation exercise discussed earlier contained for instance the phrase in the example 116. This primer task was, in fact, remarkably similar to the Phrase bank cycle, but did not share enough phrases to be included in it. All in all, the exercise was a remarkably straight-forward example of presenting formulaic sequences in context by and bringing them to the reader's attention through textual means.

The weakest form of explicitness in *ProFiles* was an exercise on quotes from the plays of William Shakespeare. These quotes are idioms almost by definition, but the way they were handled was only lightly explicit. In contrast with the corresponding section in *Culture Café,* very little terminology was used. The exercise only stated that the phrases were "famous quotes and expressions from the Bard's plays, still in use today." (PF:112) As such it is only implied that the expressions have become culturally sensitive conventionalised expressions. The below example represents the standard workings of the exercise, and it is presented in its original form, except for the name of the play which has been omitted to save space.

> 117) _It was Greek to me._ (*C I couldn't understand a word*)
> (PF:112)

The exercise itself was a fairly standard one. The reader was given five quotes from five separate plays and the approximate meanings of each in English. The reader was then prompted to match each quote with its meaning, by writing the correct alphabetical designations on a line before the quote. Thus no production of the idioms or their meanings was required. The exercise actually walked a very fine line between implicitness and explicitness. Granted, it explicitly focused on certain idioms which it gave as entities isolated from any context of use, and thus could be classified as fairly explicit indeed. Yet it belonged to neither the Phrase Bank nor Word Power exercises, which were the textbook primary vocabulary tasks. It was, in fact, a Kick Start task, which the book explicitly defines as preparing the reader for the vocabulary and contents of the upcoming reading texts. One could thus argue that the point of the exercise was only to prepare the reader for the text to come, and the idioms were only means to this end. Interestingly though, none of these idioms could be found in the text. Therefore, the exercise seemed only prepare the reader for the

overall topic of William Shakespeare, but not the actual vocabulary of the reading text.

6.2.3 Purely implicit contexts

The following chapter analyses the textbook *ProFiles* and its implicit formulaic sequences. This includes all contexts regardless of the whether they were reading texts, instructional metatext or something else entirely. An exception is made, however, with the implicit contexts connected with Phrase Bank task cycle, as they have been already discussed in detail. It should also be noted that the following chapter will not be particularly long, because the patterns to be found were fairly self-evident and they can thus be discussed with a rather compact presentation.

The main issue about implicit contexts was fairly simple. Even after the Phrase Bank sequences were discounted, the reading texts still remained the most important source of implicit formulaic sequences in the textbook. Many of the texts actually seemed to contain just as many implicit sequences that had no connection with the task cycle. The examples below have been chosen to serve as an illustration of this.

> 119) *It was all over, you would have said, but the hanging.* (PF:63)
>
> 120) *It's woven into the fabric of our lifestyle, and national identity, if you like.* (PF:24)
>
> 121) *It sounds like the sequel is just crying out to be made.* (PF:104)

The first possible context was those reading texts whose exercise batteries did not include a Phrase Bank task in the first place. There were four such texts, and most of them were poems, book excerpts, quote collections or short stories. In addition there were two listening texts. Of these, only the printed textual element was analysed. In other words, all the contexts were authentic English texts that had not been specifically designed and crafted for the textbook. The example 119 and its idiom are from such a context. These texts also contained relatively few formulaic sequences compared their Phrase Bank counterparts. In fact, the last of these texts had no formulaic sequences whatsoever, even though it explicitly concentrated on famous last words, and could thus been conducive to formulaic language. Given the relatively low number of cases and the differences between the individual texts, it is fruitless to attempt a reliable analysis on the types of formulaic sequences these texts

contained. Yet there seemed to be some tentative evidence of the overrepresentation of idioms. There is, however, somewhat more evidence on implicit formulaic sequences that had no connection with the Phrase Bank task cycle, but were still found in texts connected with the cycle. There were not many of these non-embedded formulaic sequences, but both main categories of the current study, lexical phrases and idioms, were in evidence. The example 120 corresponds to the former and the example 121 to the latter. All this seems to indicate that the Phrase Bank texts were seeded with formulaic sequences that were, for one reason or another, deemed not to require explicit attention.

The reading text were, however not the only context that contained implicit formulaic sequences. This naturally included the seemingly coincidental instances where the phrase was located in the instructional metatext of an exercise or in the exercise itself. The example 122 is an illustration of the first situation while the 123 serves a similar role for the second. Both kinds of occasions were characterised by almost stereotypical implicitness, as no textual or other means were used to raise them to the reader's attention. It was completely up to the reader to notice them. The textbook also used apparent formulaic sequences as headlines for exercises. These were usually highlighted and made easier to notice using textual means such as fonts and colours. This was especially true for the lexical phrase how come that was used as headline for a series of infoboxes. Not only were the aforementioned textual means used, but the sequence was repeated ten times throughout the textbook in this function. This make *how come* the single most repeated formulaic sequence in *ProFiles*.

122) *Write a paragraph or two for the EveryDay Art blog, describing the object and giving grounds for your choice.* (PF:9)

123) *However, from what I've understood,_____ (you were on the verge of) a career change soon after your second book came out. Why was that?* (PF:46)

124) *HOW COME?* (PF:16)

6.3. Summary

The second research question of the current study focused on a simple question. How were the formulaic sequences present in the material given to the reader? Were they

94

perhaps taught in some explicit fashion, or was it completely up to the reader notice and learn the phrases? A detailed answer to these questions has already been provided, but the current chapter aims give a short summary that makes the two textbooks easier to compare.

Culture Café handled the formulaic language in a distinct fashion. Very nearly all of the sequences were found in the reading texts, and no special attention was paid to them. In other, words they were completely implicit. There were, however, certain flavours to the implicitness, as the texts were not the sole implicit context, and neither was there only on kind of implicitness. There were, for instance, discourse devices embedded in the context sentences of exercises, where it was readily apparent that the sequence could not have been the focus of the task. There were also certain few implicit phrases that were almost explicit due to strength of textual means used to highlight them. The explicit contexts in *Culture Café* were much more limited than the implicit ones. There were only six occasions and five types of this behaviour in the textbook, and none of these occasions were particularly strong in their explicitness, and the strongest one of them concentrated on collocations, which itself is very weak form of formulaic language. More importantly, however, the textbook did not have any systematic approach to formulaic language, and all explicitness seemed to be more or less random. This is reinforced by the fact, that no explicit terminology was used.

The textbook *ProFiles* was very systematic in its approach to formulaic sequences. There were formulaic sequences in nearly all the reading texts. Some of the phrases remained wholly implicit, and no further attention was paid to them, and some phrases were simply embedded in exercises and the like. Yet many of the ones in the texts were singled out to be parts of the task cycles that dedicated several interconnected exercises to rehearsing the formulaic sequences of its reading text. It is also important that the task cycle used fairly explicit terminology in reference to formulaic language. Yet the task cycle was not the strongest form of explicitness, even if it was the most standardized one. The textbook contained multiple occasions that were only tenuously connected with the reading texts. These were very or even completely explicit in their focus and the use of related terminology. It was particularly interesting that by far the strongest case of explicitness discussed collocations.

As it was stated before, the two textbooks had very different approaches to formulaic language. The difference is very easy to summarize into a single concept. The *ProFiles* seemed to be at least partly with formulaic sequences in mind, and they were systematically integrated with the textbook. *Culture Café*, on the other hand, relied almost entirely on implicit contexts, and the little explicit attentions there was seemed to be more or less random.

7 DISCUSSION

7.1 Content that does not follow corpus evidence

The basic result of the current study was exceedingly simple. The formulaic sequences found in the two textbooks did not conform to the patterns suggested by previous corpus studies (Moon 1998). To summarize, both textbooks contained a several hundreds of individual formulaic sequences. Idioms and lexical phrases were the two largest groups in both textbooks, while collocations and other similar formulaic expressions shared a relatively minor presence. Even though the current study only counted those collocations that were explicitly labelled as formulaic sequences, the situation is still at odds with what natural discourse usually has. One might have expected there to be more collocations that filled this requirement. Consequently there should have been structure resembling an inverted pyramid, with collocations as the uncontested winner. The presence of lexical phrases should have been considerably narrow and with idioms being practically non-existent at tip of the pyramid. The results of the current study seemed to be much more in line with Koprowski's (2005:327) conclusion that textbooks generally tend to overvalue idioms. The current study shares also the speculation (Koprowski 2005:329) that the situation exists simply because implications of corpus data might not be a primary concern in textbook design. To put it bluntly, if the topics of learning material are the first thing to be chosen, the formulaic subject matter is what comes with those topics, and they more than probably will not be the ones suggested by corpus data. A very good example of this is how both the books dedicated an entire reading text on William Shakespeare. Both of these contained some famous quotes from his plays, and these have undeniably become idioms in the English language. It is, however, debatable whether these idioms would be particularly common in any corpus data. Thus it is the opinion of the current study that corpus evidence can only have had a

96

secondary role in the design of the two textbooks. It is questionable whether true distribution would even be feasible, let alone desirable, as that would imply unfiltered language input, which in turn would be relatively opaque for a learner.

The previous paragraph dealt with overall patterns and their implications. This paragraph, however, goes into more detail regarding the imbalance of idioms and lexical phrases. It can probably be taken for granted distribution of the formulaic sequences is at least partly a result of conscious design process. Yet it is also entirely possible that the results of the current study were affected by the study itself. To begin with, the current study concentrated on an English course that explicitly focuses on cultural knowledge (Lukion opetussuunnitelman perusteet 2003), which might partly explain the imbalance of idioms and lexical phrases present in both books. It could be argued that idioms are much more culturally sensitive than any other formulaic sequences (Teliya et al. 1998), and thus the textbook authors might have seen this particular course as a suitable place to focus on idioms, even at the expense of natural distribution. The current study might have found a similar bias towards lexical phrases, had it chosen the second compulsory course, which focuses on communication (Lukion opetussuunnitelman perusteet 2003). Such a course might indeed want to concentrate on lexical phrases considering that they can facilitate communication (Kuiper 2004:42) and can aid learning in both children (Bolinger 1975:100) and adults (Wray 2004:252). The situation could of course be only verified with a detailed study, but it cannot be discounted that certain bias towards idioms is inherent and unavoidable for any textbook focusing on the fifth course. At least this seems to be the case for *Culture Café* and *ProFiles*.

The current study and collocations had a somewhat complicated relationship. The phenomenon is such an integral part of formulaic language that it was impossible to discount completely. Yet it would also have been impossible to reliably recognize collocations in the textbooks. Thus the current study decided on a policy to take into account only those collocations that were explicitly labelled as formulaic sequences. While the policy is a tolerable compromise, it created an institutional flaw in the results, as there is absolutely no way to be certain that these were all the collocations in the textbooks. It is, in fact, highly likely that the textbooks contained vastly more collocations than could be taken into account. Several studies do, after all, place collocations as the most common form of formulaic language (Moon 1998), and

there is no reason to assume that this would not hold true for the two textbooks also. The definition of a collocation as "group of words which occur repeatedly in a language" (Carter 1998:51) implies such a wide phenomenon that no amount of conscious textbook design could markedly diminish its presence in any discourse. It is unfortunate that this also makes the full extent of the presence impossible to uncover without extensive corpus study, but the current study considers it virtually guaranteed that the textbooks contained a vast hidden presence of collocations.

7.2 Two sets of methodology

At this point it has been well established that the two textbooks had very different approaches to presenting formulaic sequences. To put it bluntly, *Culture Café* was almost completely implicit and formulaic sequences had a very secondary role. *ProFiles*, on the other hand, seemed to be partly designed around formulaic sequences, and had a wide variety of explicit and implicit contexts. Yet there was one interesting similarity that was present in both books. That is to say that the vast majority of formulaic sequences were only present once in the textbooks, and most phrases did not thus receive the repetition that is usually seen as the necessary for learning vocabulary. It is particularly surprising that *ProFiles* left so many formulaic sequences to their own devices, when the book had otherwise been so clearly designed to accommodate formulaic language. The following chapter discusses the myriad implications, causes and background factors of the policies of the two textbooks.

Culture Café touches on formulaic language only indirectly, and the current study thinks that there are two possible reasons for this, and both of them are connected with textbook design. First of all, it is important to understand that *Culture Café* makes extensive use of authentic English texts such as newspaper articles, novellas and poems. In fact, nearly every reading text was a member of this category, and moreover, the modifications done to the texts seemed to be very light, if they existed at all. Even without explicit knowledge of the designer intent it seems fairly clear that the use of original texts was a core value of the design process. One could further hypothesize that this priority might have ended suborning certain other language features, such as formulaic sequences. Overt attention to formulaic language may even have been considered to be at odds with the intent of providing

language training that is as natural as possible. Many forms of formulaic language are, after all, quite rare. The conclusion is strengthened when one considers the overt attention that does exist. The strongest case was on collocations, which can be very common indeed. Idioms, such as the Shakespeare quotes, were effectively and quite accurately relegated to rare passive vocabulary that is unlikely to require active use. One could thus argue that the lack of explicit attention in *Culture Café* is a side-effect of the textbooks evident focus on authentic language. One could even argue that it is effectively impossible to have both in the same textbook.

There is also a second possible reason for the lack of explicit attention is also connected with textbook design. Although in this case the issue is directly connected with discussion on how formulaic sequences should be taught. Earlier in this study it was told how there is an influential opinion according to which one should be extremely careful when teaching formulaic sequences. The reasoning is that the subject matter is so rare that overt attention is likely to cause the students to mistake the formulaic sequences as being much more common than they actually are. Thus they would also mistakenly use them too often. (Moon 2006:60) It is also possible that this trail of thought has been the driving force in designing the textbook's approach to formulaic sequences. This conclusion is especially easy to draw when one considers that the school of thought considers the best solution to be that the learner has to notice the sequences from their natural context. Too much overt help might lead to the aforementioned problems. (Mäntylä 2004:180-181) The current study considers it highly interesting how well the theory and its implications coincide with the structure of *Culture Café*. One could even argue that lack of explicit attention is a conscious design decision that reflects the designers' knowledge and opinion on formulaic sequences.

The textbook *ProFiles* had a very distinct approach to formulaic sequences. Consequently a considerable number of formulaic sequences were presented only explicitly. More important than this, however, was the Phrase Bank task cycle that was the main method used by the textbook. The pattern revealed two major issues. First of all, the cycle was applied so systematically and regularly, that only one conclusion was possible. The textbook not only acknowledged the role of formulaic sequences, but it had also been specifically designed to accommodate them. Secondly, it the cycle was surprisingly similar to what is generally considered to be

the optimal method of teaching formulaic sequences. It always began with purely implicit sequences embedded in the reading texts. The sequences were very rarely highlighted in context, and such the method does not technically follow the suggestion that sequences should be made easier to notice (Bishop 2004b:241), but the same is achieved with the first stage of the cycle. It did occur only after the initial reading, but it forced the reader refer back to the text, and thus raised awareness in way similar to regular textual methods. It is even conceivable that the students would create their own textual methods by underlining the target sequences while looking for them. The task cycle contained little explicit instruction on the nature of the sequences (Jones and Haywood 2004), but one could argue that the overall label of the Phrase Bank exercises and the explicit terminology of each separate task filled the same function. Light instructions is optimally followed by some light production task such as a gap filling exercise (Jones and Haywood 2004), and it was clear that *ProFiles* followed this advice almost to the letter. The textbook did not, however, devote any time for more free-form production exercises, which are considered essential for releasing the formulaic sequences into actively used vocabulary (Jones and Haywood 2004:274). This omission is particularly interesting considering how much the previous phases of the optimal method and the task cycle overlapped. It is impossible to know for certain, but the reason might be that the textbook designers have delegated this role for the essays that are inevitably written on every English course. It might also be that directly connecting a true production task with formulaic sequences was considered too aggressive a move, as that might have led to gross overuse of formulaic sequences in the product. Whatever the exact details of the method might have been, two issues are fairly clear. Firstly, *ProFiles* was designed to accommodate formulaic sequences. Secondly, the designers have been aware of the current consensus on how to best teach the sequences. There are simply too many similarities for any other conclusion to be possible.

7.3 The combined effect of content and presentation

All the issues discussed in the previous paragraphs are all important themselves. It is, however, their combined effect that is particularly interesting. The main issue is of course that the reader can develop a slightly lopsided view of formulaic sequences using these two textbooks, as both of them devoted nearly all of their explicit

attention on idioms and collocations. Lexical phrases, on the other hand, were given relatively little attention, and even more importantly they are the only form of formulaic language that neither of the books explains. Both books, of course, contained a large number of lexical phrases, but it is questionable whether the books help the reader to understand their role conventionalised expressions with a specific discourse function. One could not really even expect this from *Culture Café*, as systematic approach to formulaic sequences clearly has not been a design concern. For *ProFiles*, on the other hand, the flaw is fairly glaring. The only truly explicit and focused presentation on lexical phrases presents the expressions in the isolated context of speech making. No reference is made to the existence of similar expressions in other contexts. Needless to say, no section or exercise was specifically labelled for lexical phrases. It has already been discussed that the situation exists because the books were designed for a culture course. Yet one could argue that lexical phrases are at least as culturally sensitive as any other form of formulaic language, and would thus have merited a more substantial role. This is especially true considering that collocations did receive attention in both books, and they are probably the least culturally sensitive of all formulaic sequences. Whatever the cause for the oversight might be, the consequences are easy to predict. The reader will not develop sufficient knowledge base on lexical phrases. Something quite different could have been expected, at least from *ProFiles*.

8 CONCLUSION

Formulaic sequences are a crucial part of the language. This holds true almost regardless of the perspective. The sequences have an important part to play in language processing, language learning and even in the cultural side of learning. Given these facts it is entirely justifiable to say that any textbook should be conversant with the phenomenon. Therefore, it is quite surprising that very little research has been done to shed light on their actual role in textbooks. For Finnish EFL material, the body of research is particularly non-existent. The purpose of the present study was therefore to correct the state of affairs and to find out how Finnish EFL textbooks for upper secondary school take formulaic sequences into account. This purpose boils down to three very simple questions. What kind of formulaic

sequences were present and how large the presence was, and also how the sequences were presented?

The current study discovered the relationship between formulaic sequences and textbooks to be a complicated issue. First of all, it was discovered that neither of the textbooks much resembled natural discourse in their choice of formulaic language. Simply put, there were far too many idioms and far too few lexical phrases. The result was, however, in line with previous studies that suggested that textbooks often choose their formulaic language more or less at random. Both the *Culture Café* and the *Profiles* showed a similar distribution pattern with light idioms and pure discourse device lexical phrases as the largest groups. Despite this the textbooks were anything but unanimous in their actual approach to presenting formulaic sequences. On one hand, there is *Culture Café* that placed practically the whole weight of learning process on the learner himself. Most of the sequences were just embedded in the reading texts, and no help was given to the reader in locating them. What is more, there were so few occasions of explicit focus on formulaic sequences that, the phrases can only have had a secondary role in the design process, if even that. *ProFiles* on the other hand seemed to be at least partly designed with formulaic language in mind. There were, of course, sequences embedded in the texts, but there were also numerous individual occasions of explicitness along with a full task cycle meant solely for rehearsing the formulaic sequences of each reading text. All in all, the current study made it clear that there are various possible approaches to formulaic language.

The implications of the current study seem to be fairly clear. First of all, there is no such thing as common approach to formulaic language, even though all the Finnish textbook designers are bound by the same national curriculum. Secondly, a common approach might not exist, but it is obvious that at least the authors of *ProFiles* have been aware of the body of research on formulaic sequences. The similarities in the textbook's methodology and the suggestions of the research are simply too great to be a coincidence. Thirdly, it became clear that corpuses cannot have had much of a role in the design of the textbooks. They were either not taken into account at all due to other design concerns, or they were not consulted in the first place. It is, however, impossible provide accurate answer without consulting the authors themselves. Lastly, it might be that a focus on formulaic sequences is quite simply incompatible

with certain design philosophies. It is, for instance, questionable whether a textbook could retain its focus on authentic material, if the texts were pedagogically but still artificially embedded with formulaic sequences.

The current study was not without its limitations. First of all, even though clear and predetermined criteria were used in identifying formulaic sequences, it was not possible to eliminate subjectivity from the selection process. It was particularly challenging to try separating purely grammatical constructions from lexical discourse devices and certain institutionalised expressions from lightly metaphorical idioms. All further studies should thus be conducted with the support of extensive corpus data in order to minimize the subjectivity of the textbook analysis. Additionally, a cadre of several judges could be used to evaluate whether any single phrase is a formulaic sequence. The second major limitation of the current study was its scope. The results of a case study based on two textbooks cannot be generalised very far. Better results would require considerably wider data. Thirdly, the current study was unfortunately unable to take the audio-visual parts of the textbook into account due to practical reasons. This is a clear flaw, and it should be remedied by further studies.

There are several different possibilities for further research that would benefit the field of formulaic studies. First of all, previous studies have tended to focus on the endpoint of the design process, the textbook itself. Without some research into the attitudes and priorities of actual authors, any speculation on the design of the textbooks is likely to remain just that, pure speculation. Secondly, knowledge on textbooks and formulaic sequences will remain somewhat insufficient until it is studied whether or not the designs are actually efficient. This would mean studying how well the students learn to use formulaic language using the textbooks. On the whole, there are many aspects of formulaic language and learning materials that remain virtual unknowns.

BIBLIOGRAPHY

Primary sources

Benmergui, R., A. Lampinen, S. Leinonen-Nuorgam, J. Mäki, T. Päkkilä and R. Silk. 2005. *Culture Café. Book 5.* Helsinki: Otava.

Ikonen, J., A-M. Mäkelä, L. Nikkanen, O-P Salo and T. Sutela. 2009. *Profiles. Course 5.* Helsinki: WSOY.

Secondary sources

Altenberg, B. 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word Combinations. In A.P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.

Biber, D., Conrad, S., & Cortes, V. 2004. If you look at … lexical bundles in academic lectures and textbooks. *Applied linguistics, 25,* 371 – 405. (16 September 2011) http://applij.oxfordjournals.org/content/25/3/371.full.pdf+html

Bishop, H. 2004a. Noticing formulaic sequences – A problem of measuring the subjective, *LSO Working Papers in Linguistics* 4, 15 – 19. (19 March 2012) http://ling.wisc.edu/lso/wpl/4.1/LSOWP4.1-04-Bishop.pdf

Bishop, H. 2004b. The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 227-248. (19 March 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formulaic+sequences

Bloomfield, L. 1933. *Language*. London: Allen & Unwing.

Bolinger, D. 1975. *Aspects of language* (2nd edition). New York: Harcourt Brace Jovanovich.

Carter, R. 1998. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.

Coulmas, F. 1979. On the sociolinguistic relevance of routine formulae, *Journal of Pragmatics* 3 (3/4), 239-266.

Cowie, A.P. 1994. Phraseology. In R.E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon. 3168-3171.

Dörnyei, Z., V. Durow and K. Zahran. 2004. Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 249-268. (13 March 2012)

http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formul
aic+sequences

Fernando, C. 1996. *Idioms and idiomaticity*. Oxford: Oxford University Press.

Fraser, B. 1970. Idioms within a transformational grammar. *Foundations of Language* 6, 22 – 42.

Hakuta, K. 1976. A case study of a Japanese child learning English as a second language. *Language Learning* 26 (2), 321-351.

Halliday, M.A.K. 2004. Lexicology. In M.A.K, Halliday, W. Teubert, C. Yallop and A, Čermáková. *Lexicology and Corpus Linguistics: An introduction*. London: Continuum International Publishing. 11-22. (17 February 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10250796

Hausmann, F.J. 1984. Wortschatzlernen ist Kollokationslernen, *Praxis der neusprachlichen Unterrichts* 31, 395-406.

Herbst, Thomas. 1996. What are collocations: Sandy beaches or false teeth, *English Studies* 77 (4), 379-393. (16 February 2012) http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=85d07010-89d6-4fa0-8e28-3e4848cd4135%40sessionmgr14&vid=4&hid=17

Jones, M and S. Haywood. 2004. Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 269-300. (19 March 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formul aic+sequences

Just, M.A & Carpenter, P.A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87, 328-354.

Koprowski, Mark. 2005. Investigating the usefulness of lexical phrases in contemporary coursebooks, *ELT Journal* 59 (4), 322-332.

Kuiper, K. 2004. Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 249-268. (2 March 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formul aic+sequences

Lewis, M. 1993. *The lexical approach*. Hove and London: Language Teaching Publications.

Lukion opetussuunnitelman perusteet 2003. Finnish National Board of Education [online]. (3 April 2012) http://www.oph.fi/download/47345_lukion_opetussuunnitelman_perusteet_2003.pdf

Moon, R. 1998. Frequencies and Forms of Phrasal Lexemes in English. In A.P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.

Moon, R. 2006. Vocabulary connections: multi-word items in English. In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, 40 – 63.

Muuttuvat teksti –ja mediamaisemat –hanke. Opettajakyselyn taulukot. 20XX. Soveltavan kielentutkimuksen keskus. (3 April 2012) https://www.jyu.fi/hum/laitokset/solki/en/research/projects/tolp/kyselyaineisto/opetta jakysely/opettajakyselyn-taulukot/view

Mäntylä, K. 2004. *Idioms and language user: the effect of the characteristics of idioms on their recognition and interpretation by native and non-native speakers of English*. Jyväskylä Studies in Humanities. University of Jyväskylä.

Nattinger J. R. and J. S DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Philadelphia: John Benjamins Publishing Company. (16 February 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10074867

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*. New York: Longman, 191 – 226.

Schmitt, N., Grandage, S. & Adolphs, S. 2004. Are corpus derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Philadelphia: John Benjamins, 249-268. (23 February 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formul aic+sequences

Skehan, P. 1998. A Cognitive Approach to Language Learning. Oxford: Oxford University Press.

Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Teliya, V., N. Bragina, E. Oparina and I. Sandomirskaya. 1998. Phraseology as a Language of Culture: Its Role in the Representation of a Cultural Mentality. In A.P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press, 55-75.

Tuomi, J. and A. Sarajärvi. 2009. *Laadullinen tutkimus ja sisällönanalyysi*. Tammi: Helsinki

Underwood, G., Schmitt, N. & Galpin A. 2004. The eyes have it: An eye-movement study into processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 249-268. (24 February 2012) http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formul aic+sequences

Willis, D. 1990. *The lexical syllabus*. London: Collins Cobuild.

Wong Fillmore, L. 1976. *The second time around: cognitive and social strategies in second language acquisition*. Unpublished PhD thesis, Stanford University.

Wood, D. 2006. Uses and Functions of Formulaic sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *Canadian Modern Language Review* 63 (1), 13-33. (12 October 2011)
http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=42cfc6ba-50c0-4509-a76a-6899eb293180%40sessionmgr10&vid=4&hid=24

Wood, D. 2009. Effect of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *The Canadian Journal of Applied Linguistics* 12 (1), 39-57. (19 March 2012)
http://search.proquest.com/docview/216969938/fulltextPDF/1325D7F4998213769E6/1?accountid=11774

Wood, D. 2010. Lexical Clusters in an EAP Textbook Corpus. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication*. London: Continuum International Publishing.

Wray, A. 2004. "Here's one I prepared earlier": Formulaic language learning of television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Philadelphia: John Benjamins, 249-268. (18 November 2011)
http://site.ebrary.com/lib/jyvaskyla/docDetail.action?docID=10052865&p00=formulaic+sequences

Wray, A. 2005. *Formulaic Language and the Lexicon.* Cambridge: Cambridge University
Press.

Wray, A. 2008. *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.