Exploring the Boundaries of Formulaic Sequences: A Corpus-based Study of Lexical Substitution and Insertion in Contemporary British English

By

Kelly Louise Pycroft

Submitted for the Degree of Doctor of Philosophy

Department of English Language and Linguistics

University of Sheffield

May 2006

Abstract

This thesis presents an investigation into formulaic sequences; namely multi-word prefabricated phrases of either literal (e.g. *good morning*) or non-literal (e.g. *kick the bucket*) reading. A property of such sequences is variation. Formulaic sequences can be subject to varying degrees of lexical substitutions, grammatical variations, and insertions. This thesis investigates the boundaries of variation: the limits of lexical substitution and insertions for formulaic sequences, i.e. how much variation can occur before the sequence stops being fixed and becomes context-dependent. The boundaries between one formulaic sequence and another and the boundaries between a literal and non-literal reading are also explored.

The formulaic sequences for investigation were chosen from *The Longman Idioms Dictionary* (1998) and were explored using the British National Corpus (BNC). To investigate the limits of variation, I developed and used a technique that I term the *chaining process*. This is a systematic method of searching for sequences to find the maximum lexical substitutes and insertions. The frequencies of variant forms found during the study were recorded and analysed to highlight both so that both common and rare lexical substitutions and insertions could be examined, and their limits explored.

A result of using the chaining process was that sequences could be seen to "link" together. Formulaic sequences with the same underlying meaning and similar lexical set were found to form groups. Use of the chaining process showed how different formulaic sequences with similar meanings could link together in networks via common lexical substitutes, e.g. *flip your lid* and *blow your top* link via *flip your wig* → *lose your wig* → *lose your temper* and *blow your temper*. The use of the chaining process shows that formulaic sequences are more similar than different in terms of semantics as well as construction. Sequences are not autonomous; networks show that the boundaries of sequences are not as fixed as idiom dictionaries may lead us to believe. These phrasal networks formed via the chaining process provide a regular method of grouping formulaic sequences. This technique and analysis contribute to lexicography and inform cognitive models of storing and organizing language.

Contents

List of Tables

List of Figures

List of Abbreviations

BNC	-	British National Corpus
CCDI	-	*Collins Cobuild Dictionary of Idioms*
CIC	-	Cambridge International Corpus
CIDI	-	*Cambridge International Dictionary of Idioms*
COLT	-	Bergen Corpus of London Teenage Language
FLOB	-	Freiburg Lancaster/Oslo Bergen Corpus of British English
LID	-	*Longman Idioms Dictionary*
ODCIE1	-	*Oxford Dictionary of Current Idiomatic English Volume One: Verbs with Prepositions and Particles* ed. by Cowie, A. P. and Mackin, R.
ODCIE2	-	*Oxford Dictionary of Current Idiomatic English Volume Two: Phrase, Clause and Sentence Idioms* ed. by Cowie, A. P., Mackin, R. and McCraig, I. R
OED	-	Oxford English Dictionary (Online)

Acknowledgements

I would firstly like to acknowledge the University of Sheffield and the Department of English Language and Linguistics for the studentship which offered me the opportunity to carry out this research project. I would also like to thank the Department of Information Studies for the work space and equipment with which to fulfil the research.

I would like to give huge, huge thanks to Susan Fitzmaurice for stepping in as a last minute supervisor, for setting me deadlines and seeing me through the final stages. I would also like to thank Claire Cowie and Mark Sanderson for their support and encouragement, advice and assistance. I am especially grateful to Paul Clough for all his help with computational issues, and for his motivational encouragement and reassurance. I would also like to give recognition to Anita Naciscione, Henk Barkema, Rosamund Moon and Geoffrey Williams for their feedback and interest. Thanks also go to those at the Longman Idioms Dictionary for answering lexicographical queries.

I am indebted to my friends Becky, Justyna, Lia, Catherine and Anna for their encouragement, the celebrity gossip, and for reminding me to keep a balanced perspective. I would especially like to thank my mentee (or mentor?!) Angie for the coffee and lunch breaks, and for listening to all my whinges. Special thanks must also go to Dr. Michael Jakubovic and all at the University Health Service.

I would like to thank my parents Christine and Michael for their encouragement and not letting me quit during the difficult times. I'd especially like to thank mum for all the proof-reading, and being the only non-academic to have read the whole thing. Thank-you too to Ben and Deborah – yes I will leave university and get a job now! I am also grateful to Sue for all the support. At last I can use those champagne flutes! Many thanks also to Stirling, and the Dutton and Baines families.

Finally, this just would not have been possible without Blake – thank-you for everything.

1. Introduction

1.1. Formulaic Language – The Atomistic View

The field of phraseology, or the study of fixed phrases, is an area of increasing interest and study in linguistics (Cowie, 1998c: 209). As Fernando shows (1996: 26), the stereotypical attitude towards vocabulary is that it comprises a list of words: individual units which are typically contained in a dictionary. This is an *atomistic* view (Fillmore *et al.*, 1994: 503) extending to categories and combinations of grammar. The limited grammatical atomistic base comprises noun, verb, adverb, adjective, and preposition, and combinations of these generate sentences. Such an analysis leaves idioms, or fixed phrases, in a peripheral *phrasicon* (Gläser, 1998: 124). This analysis persists, despite the fact that there is evidence to show that fixed, or prefabricated, phrases contribute to a large proportion of everyday language (Jackendoff, 1995: 136; Wray, 2002: 5).[1] The phrasicon can be seen in the *idiom list hypothesis*, in which sequences are stored and retrieved in the same way as words but are stored separately from the lexicon (Gibbs, 1994: 92; Schweigert and Moates, 1988: 282).[2]

Not all of the phrases within the phrasicon have the same status and act in the same way. This general area covers for example: proverbs (e.g. *a bird in the hand is worth two in the bush*), similes (e.g. *as bold as brass*), sayings (e.g. *take the bull by the horns*), idioms (e.g. *fly off the handle*), binomials (e.g. *spick and span*), and phrasal verbs (e.g. *to give in*). Conventionalized discourse formulae, e.g. *good morning!* are also members of the phrasicon (Strässler, 1982: 19). In some literature even compound words are labelled as being idioms, e.g. *blackbird*, or names, e.g. *Jack Daniels* (Jackendoff, 1995: 134). Some sequences may be typically spoken, such as *good morning,* whilst others may be primarily written. The sequence *it's goodbye* ____, *hello* ____ is often found in newspapers (Longman, 1998: 140). The term *idiom* is often used to refer to any lexically and

[1] Jackendoff (1995: 133) showed that multiword units have kept the television programme "Wheel of Fortune" in business for many years - five shows a day, six days a week for ten years means at least 15000 different formulaic sequences, all of which should be familiar to the viewers!

[2] Figurative language is common in texts, however individual sequences appear rarely. Rundell found that the sequence *raining cats and dogs* (including verbal inflections) only appeared once in the 90million word written section of the British National Corpus (BNC), and not at all in the 10million word spoken section (1995: 29).

syntactically fixed expression which comprises two or more words, and has a figurative reading, or a reading other than that of the sum of its parts (e.g. Sinclair, 1991; Fernando, 1996). Thus the label idiom becomes "fuzzy", subsuming all items within the phrasicon.

Prototypical idioms such as *fly off the handle, kick the bucket,* and *over the moon* are considered to be "pure" (Moon, 1998a: 4) or "full" (Newmeyer, 1973: 327), and are thought to comprise one or more lexical items whose meaning as a whole differs from the compositional meanings of the constituent units; they are non-decomposable (Gibbs and Nayak, 1989a: 104). They also have a fixed syntax and lexical set, and have few variant forms (Fernando, 1996: 31). Howarth suggests that all idioms originate with a literal reading (1996: 22). Over time this becomes specialized either in the figurative reading of the whole, or of one element of the phrase. The phrase becomes associated with the same figurative meaning repeatedly, until it functions like a single word, and this leads to restrictions of flexibility and variation (Schraw *et al.*, 1988: 144). As Strässler says (1982: 179), the overall idiom then constitutes a lexeme in its own right, and can be entered as such in the lexicon, in a similar way to individual words (Keysar and Bly, 1995: 90). It is interesting to note that there are many phrases which demonstrate idiomaticity in terms of reading and relative fixedness, however, the number of phrases which actually fit the definition of (pure) idiom is very small.

Thus there appears to be a scale or continuum between the atomistic view of language and pure idioms. *Free expressions*; phrases which have lexical choices adhering to grammatical rules, demonstrate an *open choice principle* (Sinclair, 1991: 109), whereas idioms obey the *idiom principle* (Sinclair, 1991: 110), in which semi-preconstructed phrases constitute single choices, even though they might appear to be analyzable into segments. Between these two extremes are a wide range of sequences, often known as *collocations*. These are words that tend to be found next to, or near each other in a text, e.g. *hard* and *work, curry* and *favour.* Moon (1998a: 27) suggests that the simplest kind of collocation arises through semantics: members of the same semantic field co-occurring, e.g. *jam* with *butty, tart, doughnut, strawberry* etc.

What distinguishes a collocation from an idiom is the fact that in an idiom, the semantics are other than the compositional meaning of the phrase. There tends to be more flexibility

within a collocation, as opposed to in a fixed expression. Collocations are only tendencies and preferences of words to occur together, or near to each other. They can become fixed or restricted over time, through institutionalization. In this case, the substitutability of the lexical items involved becomes limited. The collocation becomes a fixed expression, or even a pure idiom. Many researchers in the field call the spectrum of formulaic phrases between pure idioms and open choice constructions *collocations*, (Fontenelle, 1998), or *restricted collocations* (*Oxford Dictionary of Current Idiomatic English Volume Two,* hereafter ODCIE2), and the scale from free expressions through collocations to idioms is known as the *collocation continuum*. The term used within this study for items between the poles of the continuum is *formulaic sequence* (see section 1.5).

1.2. Formulaic Sequences and Variation

1.2.1. Variation

This thesis reports the findings of an investigation of variation of formulaic sequences. Variation is a key property of items along the collocation continuum. Variation is a property which can determine whether a phrase is more open or more idiomatic. Moon agrees: "as far as phrasal lexemes are concerned, corpus evidence shows that their forms are by no means as fixed as some dictionaries appear to suggest" (1998a: 810). Many formulaic sequences demonstrate this feature of variation, despite the traditional belief that formulaic sequences are lexically, semantically, and syntactically fixed. Types of variation can be divided into lexical substitution, grammatical variation, and insertions/deletions, with an additional "level" of variation being style or register. Most formulaic sequences, even the most fixed ones, can undergo small grammatical changes, such as tense modifications e.g. *kick the bucket → kicked the bucket* or noun pluralization, *put hair on your chest → put hairs on your chest.* Lexical substitution involves changes of a main lexical component of the phrase, such as the noun, verb, adjective, etc. for example: *chance your arm/luck, quake/shake in your boots,* or *as tough/hard as nails.* Lexical substitution also incorporates examples where there is a context-dependent "open-slot", for example ____ *makes my blood boil! A* ____*-free zone,* or *roll on* ____. Grammatical variation involves substitutions of grammatical elements of the phrase, such as the determiner, negations, and syntactical rearrangements, for example: *take a lot of/some beating,*

not/never in a million years or *ring hollow → have a hollow ring.* Insertions/deletions involve adding elements to, or removing them from a phrase, for example: *sb/sth is (way) out of sb's league* or *fools rush in (where angels fear to tread).*[3] Many deletions occur in proverbs, where familiarity and shared knowledge amongst language users allows for ellipsis, e.g. *(speech is silver but) silence is golden.* Insertions can be lexical, such as adjectives or adverbs, e.g. *stem the **rising** tide,* or grammatical, such as *pack **quite** a punch.* Register variation involves, for example, one form being accepted as being more formal than another. The phrase *take the piss* is an informal version of saying *take the mickey/mick.* Another register variation may be dialectal, such as one form occurring in British English (BrE), and one being American English (AmE), as is the case with *be two/ten a penny* in BrE, and *be a dime a dozen* in AmE. Register variation can be exploited alongside other forms of variation.

1.2.2. Limits of Variation

Gibbs *et al.* (1989b: 58) note that "idiomatic phrases in English differ in the degree to which their lexical items can be changed and still maintain their figurative meanings". Pure idioms, such as *shoot the breeze* cannot support variation and maintain their non-decomposable meaning, however, more decomposable sequences such as *eat one's words/swallow one's words* can allow for some lexical flexibility. This is also the case with other variations, such as grammatical changes and insertions. Many phrases allow more than one type of variation, such as *come/follow hot/hard on the heels of sb/sth,* which has lexical substitution of the verb, the adverb, and also has context-dependent substitution. Another example is *kick/whip sb's ASS/butt,* which has verb, noun, and context-dependent substitution. Although many phrases demonstrate variation, there are few reports as to how much variation a sequence can undergo, i.e. what the limits of variation are.

---

[3] The notation practices used by the publishers of the Longman Idioms Dictionary are used here: /divides variant forms or substitutes, *sb* means somebody, and *sth* means something. ____ represents an open-slot. Sb/sth are also open-slot markers. Bracketing indicates material which can be added to or deleted from the sequence.

1.3. Statement of the Problem

In this study I aim to investigate the limits of variation of formulaic sequences. I aim to investigate how much variation a formulaic sequence can undergo; how much variation can occur before the sequence stops being a "fixed" expression and becomes context-dependent, or open choice, or whether there is a limit to the amount of variation a sequence can undergo.

The problem addressed in this project is:

- What/where are the boundaries/limits of variation in formulaic sequences?

The research problem can be further specified as three research questions:

- What/where are the boundaries of lexical substitution and insertion?
- What/where are the boundaries between one formulaic sequence and another?
- What/where are the boundaries between the literal and non-literal interpretations of a formulaic sequence?

Much previous work regarding formulaic sequences is descriptive and concerned with aspects of terminology and classification (e.g. Moon, 1998a). There are also many works concerned with cognitive aspects; how people process such phrases (e.g. Gibbs *et al.*, 1989b). There is less work regarding how sequences actually behave. My research aims to investigate the behaviour of sequences with regard to their variation.

Barkema asks "when are two forms one and the same lexicalised expressions and when are they two or more expressions which are synonymous?" or "where does one phrase end and another begin?" (1996b: 147). Another such question is "where does a phrase stop having a non-literal reading, and develop a literal reading?" These questions, alongside the question of the limits of lexical substitution and insertions, are addressed in this research project.

Previous researchers have tended to rely on intuition when searching for a sequence within a corpus. Earlier work has relied on intuition as to what substitutes or variants are possible.

This means that former studies have been illustrative rather than comprehensive. Researchers may have under-collected substitutes or variant forms. The full set of variant forms may not have been collected; some may be missed. To investigate the full variations a formulaic sequence may undergo, I developed a procedure which I call the *chaining process.*

The chaining process is a systematic set of rules for use when entering queries into a corpus. The corpus I have used is the British National Corpus (BNC). I selected ten formulaic sequences from *The Longman Idioms Dictionary* (hereafter the LID, (1998)), and these were taken as case-studies for the research project. Five case-studies were selected for the study of lexical substitutions, and five case-studies were chosen for a focus on insertions. The application of the process involves lexical searches of the corpus, looking first for all occurrences of the LID canonical form of the sequence. These are all recorded. The variant form (if any) given in the LID is then searched for in the same way. To find all noun substitutes, the verb and determiner are searched for (e.g. V det ___), including all verbal inflections. To search for verb substitutes, ___ det N is searched for. Pluralization and determiner changes are accounted for. Wider searches used the noun or the verb (e.g. V ___/___ N). The frequencies of all variant forms are also recorded. This process allows for maximum substitutes, insertions, and syntactic rearrangements to be found.

Exploring the boundaries of variation of sequences has useful implications outside the field of phraseology. Illustrating what actual variants are possible in use, which are literal or not, and where the boundaries are between sequences are of importance to students of the English language. The boundaries of variation of sequences are also useful for lexicography.

1.4. Outline of the Dissertation

Chapter 2 presents the literature review. The broader field of properties of formulaic sequences is discussed, narrowing the focus down to the area of variation, before pinpointing the specific research area of the boundaries of variation. Chapter 3 and 4 detail the methodology used, and the selection and preparation of the data used in the research. Chapter 3 outlines the tools used, the BNC and the LID, as well as other idiom dictionaries,

and this chapter also gives an in-depth description of the chaining process. Chapter 4 is concerned with the collection and preparation of the data; the creation of a database used to hold and display all of the sequences from the LID which showed any variation, selection of a dataset from the database, and selection from the dataset of the cases for investigation of lexical substitution and insertion.

Chapters 5 and 6 present the main findings of the research. Chapter 5 presents the analysis of lexical substitutions, emphasizing cases of networks formed involving the substitutes. Separate formulaic sequences can be joined in networks, demonstrating the boundaries between one sequence and another. Common and rare substitutes are demonstrated, showing the boundaries of lexical substitution. Chapter 6 analyses insertions, both lexical and grammatical, and provides examples of the boundaries of literal and non-literal readings, accenting the values of the verb status, functional or content, and the compositional status of the sequence.

Chapter 7 draws the thesis to a conclusion, discussing in further detail the findings of the research, and discussing ways to improve or build on this work in the future. In particular, it emphasizes the methodology as a useful technique for investigating the various boundaries of formulaic sequences.

1.5. Definitions

Section 1.1 showed that the collocation continuum encompasses a wide variety of multi-word phrases. Wray emphasizes that there is a high number of different general terms covering all of the "middle ground" e.g. amalgams, chunks, multi-word units, phrasemes, fixed expressions, formulae, collocations, restricted collocations, received expressions, phraseological units etc. (2002: 9).[4] In this thesis I adopt Wray's term *formulaic sequence* as a general term for any item between the end-points of the continuum; a term meaning any combination of words that is prefabricated, rather than created by grammatical rules:[5]

---

[4] In fact, there is so much terminology and classification debate in the field, that the term "idiomphobia" (Irujo, 1986: 300) may seem relevant!

[5] Another such general blanket term is Moon's Fixed Expressions and Idioms (FEIs). This has not been chosen for use here due to its emphasis on "fixedness" (Moon, 1998a: 2).

> A sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar
> (2002: 9)

This research will expand the reference meaning. This study looks at formulaic or prefabricated phrases, which are not necessarily pure idioms, and which may have either a literal or non-literal reading.[6] Formulaic sequences are often (although not always) non-literal (Fernando, 1996: 1). Figurativeness is often taken as a property of items under study in phraseology, although its presence is not a necessary factor. Howarth says that there must be a primary distinction made of word combinations into functional expressions (sentences; discourse/set formulae) and composites (syntactic/semantic idioms, sentence constituents), realizing that there can be both idiomatic and non-idiomatic expressions in each category (1996: 11). The term formulaic sequence also encompasses different syntactic structures found within the field, for example phrases such as *ulterior motive,* clauses such as *kick the bucket* and whole sentences such as *please tender exact fare and state destination.* The term does not emphasize "fixedness", which will be seen to be a concept central to the field.

In this dissertation, the term *idiom* will be used in its formal "pure" sense, and *formulaic sequence* used in a more general sense, subsuming fixed and variable figurative and non-figurative expressions. Following Naciscione's terminology, (2003: 24), the form of the formulaic sequence as listed in a dictionary will be known as the *base* or *canonical form,* i.e. having no inflections on the noun or verb. Any *variant form* is a variation of the canonical form, as opposed to being seen as an expression in its own right which happens to share the same vocabulary set, syntactic structure, and same meaning. Similarly, often phrases have an American and a British pairing within idiom dictionaries, e.g. *arse about face* is the British version of the American *ass backwards.* These are also known as being variants of the same formulaic sequence, rather than being separate expressions. The broad terms *literal* and *non-literal* are used, as opposed to *metaphoric*, or *idiomatic*, as these can be thought to be subjective, and may differ depending on each reader's interpretation.

---

[6] Wray's term looks at phrases which have a non-decompositional reading, a meaning other than that of the sum of the meanings of the component parts.

1.6. Delimitations of Scope

The delimitations of the research are restrictions I imposed on the project. The first delimitations of this research project were the tools used. The BNC and the LID were the corpus and idiom dictionary used. These tools are discussed and justified in section 3.2, however it is possible that the results of the investigation might have differed if different data sources had been used.

Further delimitations concern the selection of the dataset. All of the sequences in the LID that showed any kind of variation were taken and entered into a database, showing what type of variation they had, whether they demonstrated more than one kind of variation, etc. Lexical substitution was the most prevalent type of variation. Phrases which had only one type of variation, lexical substitution of the noun, were selected. The selection choice is discussed in section 4.3.

The data was chosen according to the sequences with VP NP structure and showing lexical substitution of the noun only. The syntactic structure was maintained when searching for the formulaic sequences in the BNC. This is a delimitation of the research project; it is possible to use the chaining process to investigate other syntactic variations of sequences, however it was decided to maintain the V det N structure.

The sequences were investigated for variation. This research project focussed only on lexical substitution and insertion, not grammatical variations or register changes. Finally, five sequences showing the most lexical substitutions were taken as case-studies, as were five sequences showing the most insertions. I decided to investigate sequences in more detail, rather than to focus generally on the whole dataset. This is discussed in section 4.4.

2. Literature Review

2.1. Introduction

The thesis Introduction (Chapter 1) outlines the research problem as being "what are the limits of variation in formulaic sequences?" The research problem was further detailed as three research questions:

- What/where are the boundaries of lexical substitution and insertion?
- What/where are the boundaries between one formulaic sequence and another?
- What/where are the boundaries between the literal and non-literal interpretations of a formulaic sequence?

This chapter details the previous work done in relation to the research problem; the parental backgrounds of properties of formulaic sequences and variation, and the immediate background of network theories. Section 2.2 discusses the properties of formulaic sequences with respect to the general issue of variation in phraseology, section 2.3 details types of variation, and section 2.4 examines the research problem; the limits and boundaries of variation. Section 2.5 concludes the chapter, summarising the state of the theory and how the research questions aim to add to previous works regarding the boundaries of variation.

2.2. Properties of Formulaic Sequences

The three main defining features of a prototypical idiom are:

1 It must be made up of two or more words[1] (Kövesces and Szabó, 1996; Mel'čuk, 1995)

2 The meaning of the phrase as a whole is different to the sum meaning of the words it is made up of (Sinclair, 1991; Fernando, 1996)

[1] A few authors allow for single words to be prototypical idioms, although this is an area of some dispute. Filmore *et al.* (1994) and other construction grammarians suggest that language is made up of single words, and constructions which hold meanings other than those of the individual word meanings added up, and thus act in the same way as single units. They suggest that single lexical items can also have meanings other than their literal expected ones.

3 Fixedness (Moon, 1998a; Gläser, 1998)

Principal properties of general formulaic sequences are compositionality, variation, institutionalization and familiarity, and figurativity. This section looks at previous accounts of properties and then focuses in turn on the four main features.

2.2.1. Previous Descriptions of Formulaic Sequence Properties

Gläser takes the view that there are some necessary properties of formulaic sequences and some features that may occur to lesser or greater extents in different phrases. She suggests important features of syntactic and semantic stability, and lexicalization, as well as idiomaticity (where meaning cannot be derived from the meaning of its constituents), and reproducibility. She also notes features such as carrying connotations and having expressive, emphatic or intensifying functions (1998: 127). This section details "lesser" features of formulaic sequences.

Syntax

Substitutability and syntactic variation are noted by d'Arcais (1993: 80), who said that the more frozen a sequence is the fewer lexical and syntactic variations it can undergo; phrases vary from being completely frozen to highly fixed. Moon too recognizes lexicogrammar as being a feature of formulaic sequences, implying that sequences have preferred vocabularies and restrictions on mood, aspect , and voice (1998a: 2). Syntactic stability is a commonly discussed feature of formulaic sequences; free combinations are more open to syntactic changes such as passivization than "pure" idioms. Nunberg *et al.* (1994: 492) suggest that prototypical idioms have an inflexible syntactic structure; "pure" idioms have fixed vocabularies and syntactic structures. Howarth (1996: 35) points out that common syntactic patterns found in the field are:

V + N
V + P + N
Adj + N
Phrasal Verb
Adv + Adj

This is supported by the OADCIE2 (1985: 2), which says that the most common clause patterns are e.g.:

V + Complement	=	*go berserk*
V + Direct Object	=	*ease sb's conscience/mind*
V + Indirect Object + Direct Object	=	*do sb credit*
V + Direct Object + Adjunct	=	*take sth amiss*

Kjellmer (1990) also found this, suggesting that nouns and verbs form the most collocates. Howarth (1996: 82) and Altenberg (1992: 228) suggest that the verb and its complementation is an important structure in language for providing information. The determiners in these phrases aid the relative flexibility, but they do not affect the cohesion between the principal lexical items under examination. As a result, attention is directed to the two lexical components; the verb and the noun (Benson, 1985: 64). Such phrases are often referred to as *verbal idioms* (Nicolas, 1995), and as they are the most productive of formulaic sequence syntactic patterns in terms of lexical substitutions, insertions, transformations etc., they are the ones most focussed on in research. Fellbaum (1993: 272) calls V det N idioms *verb phrase idioms* (VP). She studies VP idioms of the V NP PP type. Given that formulaic sequences can also take the form of clauses, single words, and whole sentences, actual syntactic structure is not a determining property of sequences, although syntactic flexibility is important within the property of variation.

Orthography

Orthography seems an obvious factor and is one of the defining features of an idiom: most formulaic sequences consist of, and are written as, two or more words. The usefulness of orthography as a distinguishing factor of idioms comes into play when considering that some linguists think of compound words as idioms, e.g. *blackbird* (Moon, 1998a: 2). Moon suggests that the boundaries between single-word and multi-word items can be seen when some formulaic sequences have single-word, hyphenated, cognates. She gives the example *break the ice, ice-breaker*, and *ice-breaking*. Although *ice-breaker* and *ice-breaking* are written as individual words, they consist of two words combined (1998a: 8).

Phonology

Phonology is a particularly important factor when considering the use of formulaic sequences as discourse functions. Wray (2002: 35) reports on tests done to determine whether listeners could hear a phrase removed from context and determine whether it was meant with its literal or figurative reading. It was found that literal phrases contained more pauses, and thus lasted longer than the phrase in its figurative sense. The literal reading was also pronounced with more articulatory precision. Prefabricated chunks are thus thought to be produced more fluently than newly created constructions. Similar findings were reported by Van Lancker (1981), who found that intonation was important when distinguishing literal and non-literal forms.

Zgusta (1967) suggested that other features of formulaic sequences include: a phrase must be a synonym of a single word, and the phrase must have a counterpart in a foreign language. Nunberg *et al.* (1994: 492) suggested three other features for a prototypical idiom: proverbiality (Many idioms have a plain "homey" meaning), informality (Idioms are typically related to colloquial registers, and to popular genres), and affect (Idioms are used to imply a certain attitude towards their referent. They are not used with regard to neutral topics). Other "lesser" features Gläser (1988: 269) notes include that comparative and superlative forms of nominalized idioms cannot occur e.g. *a wet blanket* but * *a wetter blanket.* Another such feature is that blocking of nominalization transformations occurs, e.g. *to play a waiting game* but * *the playing of a waiting game.* The permutation test may also apply, whereby idioms such as *man and boy* cannot be rearranged to *boy and man* without a change in meaning (1988: 269).

It is thought that these features may be optional (Moon, 1998a: 7). The more "compulsory" features of compositionality, variation, institutionalization and familiarity, and figurativity are discussed in the following sections.

2.2.2. Compositionality

A compositional, or decomposable, expression is one which has a meaning that is the total sum of the component words' meanings; summed from left to right along the phrase, e.g. *have a nice day.* For a phrase to be compositional, it can have either a literal or a non-literal

reading. For example, *good morning* has a literal compositional reading. The formulaic sequence *spill the beans* also has a compositional reading, although it has a non-literal meaning:

Spill	+	the	+	beans
(reveal)		(a)		(secret)

A non-decomposable expression is one which has a reading other than the sum meaning of its component parts, for example the ubiquitous *kick the bucket* has a non-decomposable meaning of "to die", and not "to kick a pail with your foot".[2] The standard view in phraseology is that "pure" idioms are non-decomposable.

The *lexical decomposition hypothesis* theory was proposed by Gibbs and Nayak (1989a: 104), and says that formulaic sequences vary in degree of compositionality. People try to analyse sequences as they would a literal phrase. They make assumptions as to how parts of the phrase contribute to the overall figurative meaning. They try to assign independent idiomatic meanings to the individual parts of the idiom, which are then combined to compositionally create the figurative interpretation of the phrase. There are three types of idiom:

Normally decomposable	–	each part of the phrase has an idiomatic referent for example in *spill the beans:* spill = to reveal, the beans = secret
Abnormally decomposable	-	there is indirect reference between a word and its referent, for example *carry a torch for someone,* torch = a metaphor for warm feelings
Non-decomposable	-	the phrase must be taken as a whole, e.g. *kick the bucket*

Barkema (1994a: 21 - 22) divides the area between the two extremes (abnormally decomposable idioms) into:

---

[2] Schraw *et al.* (1988: 413) suggest that many formulaic sequences, such as *kick the bucket*, take their figurative reading more often than their literal one.

Pseudo-compositional - these have meanings derived partly from the lexical items and partly from the syntactic structures, e.g. *bed and breakfast* has a greater meaning than the lexemes and syntax would predict (i.e. a private house/hotel that provides a place to sleep for the night, and breakfast the next morning for a fee)

Partly-compositional - only parts of the meaning are derived from the lexical items, e.g. *broad hint* = full and clear hint

Compositionality theory says that decomposable phrases can be modified and varied; however, non-decomposable phrases cannot be without losing their idiomatic meaning. When normally or abnormally decomposable phrases are interpreted, each part of the phrase has a meaning. As each part has a meaning, the individual parts can be modified and moved like any part of a literal expression (Voort and Vonk, 1995: 285). However, even very fixed expressions can be externally grammatically modified, or open to tense inflections, for example, *kick the bucket → kicked the bucket.* Nicolas (1995: 233) says that such modification applies to the phrase as a whole, rather than to a specific part of it. Normally and abnormally decomposable idioms will be processed faster than literal expressions due to familiarity – compositional literal phrases are new constructions, and thus take longer to interpret. Variant forms seem to be processed compositionally in the same way as literal expressions (Howarth, 1996). Glucksberg (1995: 20) says that there are no such things as non-decomposable formulaic sequences – in no matter how small a way, all phrases are analysable to some extent. Cserép suggests that "full analyzability does not require the analyst to attach all of the meaning exhaustively to the idiom constituents". He gives the examples of *show your true colours* having a compositional reading of "show + det + your + real + character", although he also points out that the meaning also implies that this character is unpleasant (2002: 68). This is reiterated by Barkema (1993: 262), who points out that there is more to the meaning of *bed and breakfast* than the basic sense of *bed* ("an article of furniture to sleep on") added to that of *breakfast* ("the first meal of the day").

Compositional idioms can be open to modifications such as lexical substitutions, for example *button/shut/zip your lips,* whereas non-decompositional idioms cannot: *kick the bucket* but **kick the pail.* Gibbs and Nayak (1989a: 104) suggest that syntactic flexibility

of sequences cannot be explained if phrases are taken as singular complete blocks. If idioms were simply long words whose constituents had no meaning on their own, then they would not be syntactically flexible, nor would you be able to substitute parts (Glucksberg, 2001: 69).

Gibbs and Nayak (1989a: 104) suggest that there are a number of ways to support the notion of compositionality. Examples they give are internal modification, such as *leave no stone unturned/leave no legal stone unturned;* the modification acts on the noun only, not on the whole phrase; pluralization in the phrase *touch a nerve/touch a couple of nerves* – note the meaning of the latter is not *touch a nerve a couple of times:* the pluralization acts on only part of the phrase. Another example is that of topicalization; Gibbs suggests that it is not possible to move parts of the phrase around a sentence if the phrase acts as a whole and the individual parts lack meaning. Compositional parsing is useful, as it means that individual parts of decomposable idioms can be rearranged into other syntactic structures, or lexically altered (and thus open to insertions) without disruption of the figurative meaning (Gibbs *et al.*, 1989b: 66). It follows therefore that compositional sequences are open to insertions. Barkema notes that compositionality concerns the relationship between the meanings of the individual words of the sequence and the sequence meaning as a whole. Compositionality is thus not restricted solely to non-literal sequences, but may also be true of any lexicalized expression, such as *pen and pencil* (1996b: 129).

Fillmore *et al.* (1994: 504) suggest that formulaic sequences can be divided into encoding and decoding idioms.[3] The former is an expression which can be translated without prior knowledge; it is compositional – *wide awake.* A decoding idiom is one which requires previous knowledge; its true meaning cannot be interpreted, e.g. *kick the bucket.* They also say that some formulaic sequences can be encoding and decoding idioms. These expressions are decoding in their figurative sense, (*kick the bucket* meaning "to die"), and encoding in the literal reading. This differs from Makkai's (1972, cf. Akimoto, 1983) use of the terms encoding and decoding, where he uses the former to refer to problems for the sender and the latter to refer to idioms which pose problems for the receiver.

[3] Fillmore uses the term "idiom" as a cover term.

A suggested test for compositionality is to replace a word in sequence with its dictionary definition (Grant and Bauer, 2004: 48). If this gives the same meaning as the formulaic sequence, then the sequence is compositional. If the meaning changes, the sequence is non-decomposable.

A factor relating to the area of compositionality is the notion of delexicality, in particular with relation to the verb of a verbal idiom, such as a V det N sequence. Akimoto details past accounts of idioms and says that Ruhl (1975: 440, cf. Akimoto, 1983: 11) lists what he calls *primary verbs* such as *be, do, have, get, take, give, come, go, put, make, hit, break,* and *cut.* Of these, it was suggested that *be, do,* and *have* are the primary verbs and can have little or no meaning. Altenberg (1992) and Howarth (1996: 28) also support this, the latter suggesting that verbs such as *have, take,* and *make* are often used in delexicalized senses. Fontenelle (1998: 191) suggests that the area in between free constructions and idioms is known as *collocations.* This area can be divided into three sections. *Free collocations* occur when a word is open to a range of tendencies, e.g. *hire staff/clerk/secretary/worker. Restricted collocations* occur where one word in the phrase has a specific or figurative meaning, e.g. *champion a cause.* Thirdly *support verb constructions* are such as *have a drink/take a bath/make a decision.* These phrases have a grammaticalized verb, e.g. *take/make/have/give/do.*

Howarth says that delexicalized verbs can be seen as having a large number of discrete meanings, or very little meaning, and thus are open to occurrence in a large number of formulaic sequences. On the other hand, verbs with a smaller number of senses are less common within sequences, and an example given is that *shrug* is often found within the same context as *shoulders.* (1996: 41). Howarth also says (1996: 95) that the delexicality applies only to certain senses of the verbs, not to the actual verbs themselves. Some verbs may have a tendency to occur more delexically than lexically, e.g. *have, make, take.* Note however, that Stein gives evidence for literal compositional sequences to show that the "empty" verb does actually carry some meaning. She gives examples such as *she sipped the tea* vs. *she had a sip of the tea* to demonstrate that *to sip* does not have the same meaning as *to have a sip.* The former indicates possible repeated action, whereas in the latter, only one sip is indicated (1991: 16). That the empty verb carries some meaning is also found in the imperatives *rest!* and *have a rest!* Whilst the meanings may be the same, the former is more

of an order, whilst the latter acts more as an offer (1991: 26). Howarth also questions the empty verb, asking why we say *take a look* but not *make a look*, and suggests that *make* might be "collocationally blocked" in the same way as one can *assume* or *adopt a form* (regarded as synonymous), but one can only *assume* but not **adopt importance* (1996: 40). Wierzbicka also says that activities in the *take a V* frame, where *V* is a deverbal noun, are seen as unitary, having a natural beginning and end. Sequences of a *have a V* frame are more arbitrary in duration (1982: 795).

Howarth (1996: 40) suggests that formulaic sequences containing a delexicalized verb are restricted collocations which occur towards the free end of the scale. The verb contributes little meaning to the formulaic sequence. He outlines an example by Aisenstadt (1981, cf. Howarth, 1996: 40), who suggests that, because of their "weakened grammaticalized meanings" they are in some expressions interchangeable, for example *have a look, take a look, give a look* etc. They may also be easily replaced with a synonymous expression, such as *to look*. There are four tests Howarth suggests for detecting delexicality:

The semantic equivalence of verb + noun to lexical verb

e.g.	*do harm to*	means the same as	*harm*
	give an answer		*answer*
	take action		*act*

The semantic equivalence of verb + noun to copula + adjective construction

e.g.	*take care*	*be careful*
	have significance	*be significant*

The noun being abstract and mass

e.g. *get inspiration from*
take account of
make contact with

The noun being used in a figurative sense

e.g. *give notice*
have a following
make strides
take a stance

(cf. 1996: 95 – 97)

In many cases, more than one of these features applies. Note that if the verb is delexical, or grammaticalized, then compositional meaning rests on the noun of the sequence. The sequence is then open to lexical substitutions and syntactical transformations such as passivization in the same way as fully compositional sequences. Some phrases are open to verbal substitution of delexical verbs, for example, *have/get/give sb support* (Howarth, 1996: 97).

In theories of grammaticalization, it has been suggested that verbs such as *take* have become more grammatical; functional rather than lexical. Vannebo (2003) looks at the "take and V" construction in Nordic languages, in examples such as:

Ta og kast'n overbord!
Take and throw-him overboard
"Throw him overboard!" (Bojer, *Den siste Viking*, 1921: 74)

Ta og ro deg ned noen hack!
Take and cool yourself down a bit
"Just cool it a bit!" (Lindell, *Drømmefangeren*, 1999: 216)
(2003: 169 – 170)

Although his work concerns *take* in the construction "take and V", rather than used as a main verb, Vannebo's findings seem to support what appears to be the case in the phrase *take a BEATING/hammering* (see section 5.3); the verb seems to have lost its original lexical content, "grasp", or "seize", and has developed a more general meaning. This process is known as *bleaching*, and is characteristic of the grammaticalization process.

Vannebo also points out (2003: 178) that *take* is one of a group of verbs which seem to be open to grammaticalization. Such verbs, in both European and non-European languages, are frequent verbs. Frequently occurring verbs tend to include those which may also act as auxiliaries, such as *be, have* and *do.* Verbs such as *take* and *give* are often thought to be *vector verbs*, that is, verbs which represent some stage between a full lexical verb, and a grammatical auxiliary. They are also thought to be core, basic verbs in many languages, and it is this characteristic which is thought to fit them for the grammaticalization process. As Vannebo puts it (2003: 179) a verb such as *take* is more likely to be open to grammaticalization than words of similar meaning, such as *grasp* or *catch.*

2.2.3. Fixedness and Variation

One of the commonly occurring themes in regard to formulaic sequences is that of fixedness. Take, for example, the phrase *it's raining cats and dogs.* This has a fixed vocabulary and syntax (it is not possible to have a substitute such as *it's raining horses and chickens*) and certainly has a meaning other than that of its literal sense. These phrases cannot also have a pronoun replace the noun and the meaning remain. Despite fixedness being one of the defining properties of idioms, many formulaic sequences, and indeed proper idioms to some extent, show variation. Fixed phrases are not in fact fixed (Sinclair, 1991: 83). Zgusta (1967), who gave one of the first descriptions of formulaic sequences, listed substitutability (whether one lexical item can be substituted for another), and additions (whether a word can be inserted into the expression) as two of her nine defining properties.[4] There are different types of variation which formulaic sequences may be open to, such as lexical substitution, (*hit the deck/hit the dirt*), or context-dependent lexical substitution, (____ *went down like a lead balloon*), grammatical substitution, such as changes of the determiner, (*take a lot of beating, take some beating*), syntactical transformations, a form of grammatical variation, such as passivization (*pin your hopes on/hopes were pinned on*), and insertions or deletions into, or from the sequence, (*sb/sth is (way) out of sb's league, fools rush in (where angels fear to tread)*). Barkema separates grammatical variation and lexical substitution into his characteristics of transformational

[4] Note that Zgusta separates lexical substitutions and additions. Many authors simply band these into one category of fixedness, and separate syntactic flexibility from these. Variation will be discussed in section 2.3.

deficiency or flexibility, and collocability respectively (1996b: 131).[5] Many works within the field discuss the property of variation to some extent. Variation of formulaic sequences and the boundaries of variation are investigated in this thesis. The property of variation thus requires an in-depth discussion and this will be carried out in section 2.3. It will be seen that as with properties of compositionality, figurativeness and familiarity, the degree of fixedness a sequence has may also vary along a continuum.

2.2.4. Institutionalization and Familiarity

The property of institution and familiarity refers to the knowledge native speakers have about a sequence. As Howarth says (1996: 36), it is this feature which makes a phrase noticeable before others. Institutionalization distinguishes between nonce forms (one-off combinations) and lexicalized phrases. As Wray suggests (2002: 59), virtually all formulaic sequences begin as novel constructions. Over time, they become known and accepted by other speakers, and only one or some of the possible meanings are used. Finally, lexicalization occurs when the phrase is used in the same way as a lexeme, in that it has a non-compositional reading. The phrase is accepted by other speakers as a known item (Howarth, 1996: 36).

Familiarity with a phrase helps it to become lexicalized and institutional; the string of words becomes fixed with a specific meaning. Collocations can become fixed or restricted over time, through institutionalization. In this case, the substitutability of the lexical items involved becomes limited. The ODCIE2 (1985: 3) says that historically, idioms are formed through constant literal use, then develop a non-literal extension, and then become fixed in the language with that meaning. Some people may have never known the original literal meaning. As Howarth says (1996: 24) in some cases, an expression contains a word which has otherwise died out in the language, for example *dudgeon* in *in high dudgeon.* Whether a sequence is institutionalized or not is largely intuitional.

Conversely, there is evidence to suggest that lexical phrases are learnt in first language acquisition: phrases are learnt in familiar chunks, alongside their function in context (Nattinger and DeCarrico, 1992: 11)

---

[5] Here they have been grouped together as variation and will be discussed in more detail in section 2.3.

> For example, children may initially use *wannago* holophrastically as a memorized prefabricated routine in certain set situations, and then after they become aware of similar phrases like *wannaplay* and *wannaget* in other contexts, they begin to analyze this phrase as a pattern with a moveable component, *"wanna* + VP". As children hear such moveable components in prefabricated patterns, they begin to analyze chunks into their separate pieces, and work their way to the actual rules of syntax.
> (1992: 25)

Pragmatically, figurative language is thought to be separate from typical open choice language. In order to interpret a figurative reading correctly, there must be some real-world knowledge. This is particularly the case with metaphoric language (Gibbs, 1994: 1). Idiomatic language is often thought of as being conversational, "slangy" or transitory, although conversely sequences may be found in formal, written (and specialized) registers. Many idiom dictionaries mark idioms which have tendencies to occur in formal or informal registers, with the ODCIE2 (1985: 27) noting that most formulaic sequences are neutral for formality or informality, although stylistic values are constantly changing.

Prefabricated language is seen as being a social feature. It creates an economy of effort between sender and receiver, and is a uniting factor of communities. It is also possible to have idiolectal idioms, and idioms within special communities (Fernando, 1996: 66). Shared knowledge occurs when the speaker and audience recognize the same idioms, thus idioms can be shortened or "punned" (Wray, 2002: 24). Familiarity of an idiom also maintains frozenness, as does frequency of the idiom within the community (Gibbs and Nayak, 1989a: 129).

There are many formulaic sequences which have particular contexts of use; they have distinct discourse functions. For example, phrases such as *good luck, how do you do*, and *once upon a time* have become familiarized discourse-markers for well-wishing, greetings, and story-telling respectively. As Wray notes (2002: 53), there are often lots of possibilities for what could be said in any given communicative situation, but not all are as accepted as discourse functions we actually use.

Nattinger and DeCarrio (1992: 59) comprehensively list types of discourse markers for both speech and writing, for example spoken social interactions, in greetings such as *how are*

you, *Good morning*, and closings such as *I must be going* and *Nice talking to you*, as well as conventions for checking comprehension, e.g. *all right?* A necessary topic included is that of autobiography, with examples such as *my name is*____ and *I'm from* ____. Discourse devices are another kind of discourse marker, with connectors such as *nevertheless* and *Because of*____. Wray (2002: 88) suggests that formulaic sequences can be used within situations as, for example overall membership markers; group chants and "in" phrases, rituals, such as prayers, forms of address and incantations, and as markers to show someone's place in a hierarchy, such as threats and quotations.

Moon (1998a: 217) suggests that the use of fixed expressions and idioms can be divided into five categories:

Informational	=	used to give information
		for sale, catch sight of sth
Evaluational	=	used to convey a speaker's evaluation and attitude
		kid's stuff, near the knuckle
Situational	=	used to describe or explain a situation
		talk of the devil, long time no see
Propositional	=	used to convey truth, advice, requests etc.
		you know what I mean, I kid you not
Organisational	=	used for organizational purposes of text or discourse
		By the way, for instance

Of these types, she found that evaluative sequences tended to be metaphorical. Situational sequences were found in discourse. The use of formulaic sequences is not only to convey meaning; they are used to add to the discourse in some way too. Fernando (1996: 42) says that most formulaic sequences can be manipulated according to the communicative intent of the language user. Features of institutionalization and familiarity, and use such as a discourse marker may have an effect on the boundaries of variation. Familiarity is often called upon in creative language. The creative use of formulaic language is shown by Philip (2000: 222 - 223):

Like a *red rag to a bull*
As a *red rag to Diane's bull*

The effect of creativity occurs because something else is expected; the familiar, original construction.

Partington (1996: 139) believes that fluency of language in use is due to the Idiom Principle; speakers rely on the predictability of prefabricated language. Using manipulations of familiar phrases causes the receiver to reflect back to the original, and brings about what Partington calls a *smugness effect*; readers are flattered to appreciate the relation of the variant form to the original phrase, particularly in puns.

Gläser (1998: 125) says that the use of formulaic chunks of language can have an emphatic or intensifying effect in a text. She also points out that as well as systematic variations of phraseological units, there are also creative modifications of such expressions, made to achieve a stylistic effect. Gläser is well known in the field for her work on the use of formulaic language stylistically within different genres:

> In general, idioms, being metaphors or metonymies, may add to the imagery of any text, ranging from advertising and instruction to academic discussion and prose fiction. Authors of scientific writing are prone to modify idioms, proverbs and quotations for intellectual punning and sophisticated allusions. Journalists will exploit idioms and phrases in headlines and commentaries. The authors of textbooks intended for pupils and students may use phraseological units to enhance the intelligibility and memorability of a text.
> (1998: 144)

Some genres are dense in multi-word expressions, others are not. There are more prefabricated units in general newspapers than in news reports and journalism, where creativity appeals to readers and makes language vivid and interesting (Collins, 2002: vi). Fictional dialogues report a high frequency of formulaic sequences. Horoscopes also show a high number of sequences. Formulaic expressions are less common in spoken dialogue than might be predicted, although as Wray points out (2002: 76), the connection between slang and colloquialism and idioms is close, and many people believe slang forms are idioms (Fernando and Flavell, 1981: 2). There are certain performance genres such as

sports commentary, where speakers save time by using formulaic sequences. Discourse functions and conventions occur more in speech. This is supported by Erman and Warren (2000), who found that multi-word constructions, termed *prefabs*, were slightly more common in speech than writing, but this was mainly due to the inclusion of reducibles (e.g. *don't, wouldn't, I've)* in the study.

2.2.5. Figurativeness and Metaphor; Semantic Meaning and Transparency

The mind is not literal (Gibbs, 1994: 16). Figurative, or non-literal, language pervades much of any text, as demonstrated by Wray (2002: 5), and Erman and Warren (2000: 29). One important function of figurative language is that it presents a way of talking about abstract concepts (Cacciari and Glucksberg, 1995: 43). Gibbs (1994: 27) suggests that words have a linguistic definition (that which is found in a dictionary), and an encyclopaedic definition, based on real-world knowledge. Many words have referents to real-world objects or concepts. So when people interpret expressions compositionally based on their literal meanings, there can often be great differences as to what the actual meaning is. The sense of a word may change according to context. Thus it is not so clear-cut as analysing the literal reading first, and then the figurative one. Figurative readings may be easier to process, as they are not as reliant on context for their meaning as literal readings are. This view of importance of context and use is also taken by Glucksberg (1995: 3), who views formulaic sequences as having three meanings: firstly, the literal interpretation, accessible to anyone who knows the language. Secondly, there is the figurative reading, accessible to those who know the phrase. Finally, there is the meaning intended by the speaker in using that particular sequence. Grant and Bauer say that to understand non-literal language one must first take a compositional untruth and then pragmatically reinterpret it to gain the truth (2004: 50).

Often authors refer to figurative language as being idiomatic (Fernando, 1996). *Idiomaticity* is a narrower term than *figurative*, as it has elements of pure idioms only; fixedness and a meaning other than the sum meaning of the component parts. *Figurative* incorporates metaphoricity and idiomaticity, metaphoricity involving real-world knowledge and imagery, as well as the meaning of the lexemes involved. This is supported by Gibbs *et al.*

(cf. Cacciari and Glucksberg, 1995: 45), who divides formulaic sequences as having literal readings, figurative readings (incorporating idioms), and metaphoric readings.

Metaphoricity is a special form of figurative language in which one concept is used to describe another concept, e.g. ANGER IS HEATED FLUID IN A CONTAINER – *he flipped his lid.* Idiomaticity is when the words in a phrase mean something other than their literal reading. All idioms are idiomatic, but not all idiomatic expressions are idioms. Idioms are a special subgroup of idiomatic expressions. Gibbs claims that images formed from conceptual metaphors help us to bypass the literal reading in favour of the figurative one (1993: 74).

As there seems to be a scale of compositionality, there also seems to be a continuum of figurativeness applying to formulaic sequences. At one end of the spectrum, there are set *transparent*, literal expressions, such as *good morning,* and at the other end are proper *opaque*, non-literal idioms such as *pass the buck.* Opaque sequences do not have a viable literal interpretation. Barkema (1993: 262) suggests that formulaic sequences contain lexical items with one or more senses which are not basic (i.e. the literal, usual dictionary sense). If they contain no items with basic senses, they are *entirely idiomatic*; if they contain some lexical items with, and some without, basic senses they are *partly idiomatic*, e.g. ***rain*** *cats and dogs* is partly idiomatic, and *paint the town red* is entirely idiomatic, the un-emboldened italicized words in these examples being ones without the basic sense.

Many idioms have both a literal and a figurative reading. However, some phrases such as *bark up the wrong tree*, do not have "good" literal interpretations (Popiel and McRae, 1988: 476). This is not to say that the literal interpretation is not ever valid, more that people simply never use the literal version. In fact, Popiel and McRae's 1988 research shows that the figurative interpretation of formulaic sequences is the reading most frequently used, and is more familiar to speakers of the language than the corresponding literal reading. Preference for the figurative reading over the literal one depends on lexicalization and the degree of familiarity with the phrase as an idiom (Schraw *et al.*, 1988: 413). Lexicalization is the process by which a string of words acts in the same way as an individual word; the string has a non-decompositional reading. This is a result of familiarity. Barkema (1996a) says that the link between a phrase's literal and figurative reading is known as its

motivation, and this is also demonstrated by Cacciari and Glucksberg (1995: 43). Sequences which are known as being opaque have no clear motivation, for example, it is not clear why the phrase *kick the bucket* has the idiomatic reading it does. However, in a phrase such as *spill the beans,* the motivation is clear, as it is easy to relate "spill" to "reveal", and "beans", as discrete concrete units to "secrets" again, discrete units. Such expressions as *spill the beans* and *pop the question,* are transparent in motivation, and have a clear link between the literal and figurative readings. For most formulaic sequences, it is possible to establish some relation between meaning and form (Vega-Moreno, 2003: 84).

Cacciari and Glucksberg (1995: 43) also provide an example of motivation in the following example:

Carrying coals to Newcastle

Literal reading - take some coals to Newcastle-Upon-Tyne

Non-Literal Reading - Newcastle produces coal itself, so this action would be unnecessary, thus the figurative reading is "an unnecessary act"

To understand the motivation behind this phrase, some cultural knowledge of Newcastle is required. Such knowledge is not required for more transparent formulaic sequences. Howarth (1996: 24) argues that the phrase *move the goalposts* has a clear link between its literal sporting origin and its figurative use. Pure idioms such as *shoot the breeze* are thought to be unmotivated. The motivation of a formulaic sequence depends on people's age, knowledge and background. As Vega-Moreno says (2003: 306), phrases may be transparent if they have a word within which carries a literal reading, such as *miss* in *miss the boat,* or they may have a word which acts metaphorically (*blow* in *blow your top*) or the sequence may act non-literally in an exaggerated manner, such as *eat one's heart out.* Some sequences may also carry a "support" or functional verb, such as *have, make, give,* which may have many different senses, or carry no meaning (see section 2.2.2). Non-decomposable sequences have a meaning which has no clear motivation.[6]

[6] Although there is often some historical motivation which has been lost.

A metaphor occurs when two conceptual domains are referred to, and one is talked about in reference to the other. Metaphors do not contain the same link between the word's literal and figurative meaning as formulaic sequences do. Metaphorical language underlies much of our everyday language, and helps us to organize our thoughts and experiences. Note that an idiom is an expression which is fixed and has a non-decomposable meaning. A metaphor is a way of representing one concept in terms of another. Metaphor is a common base for sequence formation, and is a form of motivation, but not all formulaic sequences are metaphoric, in particular, discourse functions such as *how do you do?* Metaphors are a specific type of figurative language. They also need not be multi-word units, for example Glucksberg (2001: 4) gives this illustration of metaphorical language:

> *Crime in our city has become an epidemic that will soon infect even our finest neighbourhoods*
> "crime" is representative of "disease". The concept "disease" is used as a metaphor for "crime". Crime can have properties of disease, e.g. infectious, an epidemic, a virus, we might try to cure it etc. If it can have the same properties as a disease, we can use the same vocabulary to talk about it.

Metaphors are often spoken about in terms of conceptual metaphors and linguistic metaphors. Conceptual metaphors are capitalized, such as KNOWING IS SEEING. They represent underlying concepts which unite linguistic metaphors. The linguistic metaphor is a surface representation of a conceptual metaphor, e.g. a conceptual metaphor may be LOVE IS A JOURNEY, and the linguistic metaphorical expression *our relationship is at a crossroads* (Cameron and Low, 1999: 79). Other examples of conceptual metaphors include ARGUMENT IS WAR (your claims are *indefensible*), LOVE IS A JOURNEY (Our marriage is *on the rocks,* we're *at a crossroads*), and TIME IS MONEY (you're *wasting* my time, he's living on *borrowed* time) (Lakoff and Johnson, 1980: 7).[7] Conceptual metaphors often occur when trying to understand difficult abstract concepts, such as love, in more familiar delineated terms (Gibbs, 1993: 60).

It is interesting to note that there are often several formulaic sequences with the same or similar meaning; *blow your stack, hit the ceiling, flip your lid* = to get very angry, or *spill the beans, let the cat out of the bag* = reveal a secret. This may be due to an underlying

[7] Lakoff and Johnson created their own examples of each metaphor rather than relying on real-life data.

conceptual metaphor. Whilst conceptual metaphors do not determine the words or structures of a formulaic sequence, they do explain why certain phrases are used for certain situations. Gibbs (1995: 106) carried out an experiment in which subjects were asked to describe the mental image they had for idioms. There was found to be a high degree of consistency in the images for idioms with similar meanings. The images they described were highly specific. This suggests that conceptual metaphors underlie figurative expressions:

> Many idiomatic phrases refer to a single concept e.g. anger, but may have different underlying conceptual metaphors, e.g. *blow your stack* vs. *bite your head off*: ANGER IS HEATED FLUID IN A CONTAINER vs. ANGRY BEHAVIOUR IS ANIMAL BEHAVIOUR. Because our ordinary concepts are often understood via multiple and sometimes contradictory metaphors, it is often no wonder that we have so many different kinds of idioms to reflect the sometimes subtly different aspects of our everyday experience.
> (Gibbs, 1995: 108)

A debate surrounding metaphors and figurative language is whether they are "dead". Some people believe that idioms are dead metaphors (Gibbs, 1994: 91). For example, the phrase *kick the bucket* has two suggested origins. The first is that it arose to refer to the slaughtering of pigs. The pigs would be hanged from a beam and then their throats cut, the blood draining into a bucket below. Any post-mortem spasms the animal would have caused the pig to kick against the bucket. The second origin of this phrase comes from a method of suicide whereby a person would hang themselves by attaching a noose around their neck, and standing on an upturned bucket. To commit the act, they would "kick the bucket" away (Flavell and Flavell, 2000: 117 – 119). These origins of the phrase have long been forgotten, and the phrase is often used now as a frozen fixed expression meaning "to die", with no motivation behind its meaning. It is known as a dead metaphor. However, conceptual metaphors cannot be dead, otherwise we would not be able to understand new creative language:

> Most scholars traditionally assume that idioms like these [*blow your stack, spill the beans, kick the bucket, pop the question*] may have once been metaphorical in their origins, but have lost their metaphoricity over time and now just exist in the speaker's mental lexicons as stock formulas or as "dead" metaphors. Just as speakers no longer view *face of the clock,* or *arm of a chair* as metaphoric, few contemporary people recognise phrases such as *have the munchies* or *to get pissed off* as particularly creative or metaphoric.
> (Gibbs, 1995: 98)

Kövesces and Szabó (1996) support the view that conceptual metaphors are behind figurative language, and thus formulaic sequences are not dead. Their experiment shows that using conceptual metaphors, second language learners can learn new idioms better than simply memorizing them from a list. Students were asked to complete phrasal verbs in sentences with the prepositions *up* or *down.* The students were English language learners and had no prior knowledge of the phrasal verbs being tested. Students who were previously exposed to conceptual metaphors such as MORE IS UP and WRITTEN OR RECORDED IS DOWN performed much better than those who had not.

2.2.6. Conclusion: Uniting the Properties

Barkema (1996a: 69) proposes that instead of trying to define each type of sequence in-between the two poles according to a single feature, it is better to think of formulaic sequences in terms of a set of features. Each feature forms a scale, and each feature is open to variation. Formulaic sequences score along the scales for each feature.

> Because of their scalar nature, an infinite number of classes could be distinguished by each time choosing slightly different positions on each of the...scales. Each class would then be only slightly different from another.
> (Barkema, 1996b: 133)

Barkema says that due to the impracticability of making a sequence class at each possible point on each of the feature scales, it is better to divide the scales into three parts – the extremes, and the middle section.[8]

As Barkema indicates (1996b: 133), formulaic sequences have many properties, and sequences vary individually in the degree to which each property is present. To label them

[8] This would lead to an infinite number of categories and terminology.

according to one defining property only generates many exceptions to the rule, and to try to group them results in generalisations. Perhaps Barkema's multi-dimensional descriptive model represents a comprehensive system of defining formulaic sequences sufficiently, as it treats formulaic sequences in terms of bundles of these properties working together, each along a scale, the scales being divided into the extremes and middle section for ease.

The properties Barkema (1996b: 133) focuses on are the familiarity, compositionality, fixedness, and collocability. The scale for familiarity (institutionalisation) can be divided into *free expressions*, which are open lexical choices following grammatical rules, *restricted expressions*, which have some degree of fixedness, either semantically, syntactically or lexically, and *lexicalized expressions*, which are fixed, and accepted as familiar lexicalized expressions. Compositionality depends on sense: *basic* (the literal sense of a word), *extended* (an extension of the literal sense), and *derived* (an established non-literal reading of the word). The three parts of the compositionality scale are thus *fully compositional* (phrases which are sum meanings of basic or derived senses), *pseudo-compositional* (phrases which have at least one item with an extended sense), and *non-decompositional* (phrases which have no basic meanings in its comprising words). Fixedness, or flexibility, can be subdivided into *fully flexible*, *semi-flexible*, and *inflexible*, the former being open to variant forms, for example, lexical substitutions, insertions etc., semi-flexible phrases being open to variant forms, and *pseudo-variant* forms, and the latter being open only to *pseudo-variant* forms. Pseudo-variant forms (1996b: 144) are variations which cause the expression to lose its formulaic meaning and use, for example *man and boy* is an accepted sequence, but the plural form *men and boys,* or the reversal *boy and man* causes the expression to lose its formulaic meaning, as would simply adding determiners *the man and the boy.* Finally, collocability (lexical substitution) is the degree to which certain lexical items are fixed and non-substitutable, for example, *the tip of the iceberg* has a certain figurative meaning, whereas *the summit of the iceberg* holds a literal reading. Collocability ranges from *free* or *collocationally open*, where any item from the same lexical class can be substituted, to *collocationally closed*, where none of the component parts can be substituted. In between these extremes, there are *collocationally limited* phrases, in which a component part can be substituted but only from a small set of variant choices. This also extends to other variation types, such as passivization, or insertions etc.

A collocationally closed expression withstands any type of variation. I have summarized Barkema's properties in Table 2-1:

Property	"Open" Extreme	Mid-point	"Upper" Extreme
Collocability	Free	Collocationally Limited	Collocationally Closed
Fixedness	Fully Flexible	Semi-Flexible	Inflexible
Compositionality	Fully Compositional	Pseudo-Compositional	Non-decompositional
Familiarity	Free Expressions	Restricted Expressions	Lexicalized Expressions

Table 2-1– Properties of Formulaic Sequences

All of the properties act in combination to mould the formulaic sequences. Barkema (1996b: 150 – 154) offers the following examples:

The bottom/first/highest rung of the ladder (an expression meaning "the ___ level in an organisation")

This expression has in the past been classified as an idiom, a figurative idiom and a non-decomposable metaphor. Referring to Barkema's multi-dimensional scheme, this sequence is a lexicalized expression, entirely idiomatic (non-decomposable), semi-flexible and with some pseudo-compositionality (*bottom/first* etc. can be substituted, but *rung* and *ladder* are collocationally closed).

Bed and Breakfast

Multi-dimensionally, this is a lexicalized expression, it is pseudo-compositional, having more meaning than the words it contains, but still retaining some of the compositional reading, it is inflexible and it is collocationally partly limited (*breakfast* can be substituted with *board),* partly closed (*bed* is fixed). Previously, it has been labelled a binomial expression, or simply a collocation.

The scheme also allows for discourse functions, e.g.:

If I may say so

This is fully compositional, inflexible, and collocationally closed.

Ultimately, Barkema's system may not provide neat category labels for each sequence. However, it takes into account the fact that different properties are available in formulaic sequences to differing extents, and it seems to be a more comprehensive manner of describing sequences, than trying to pin them down to one defining feature. Whilst dividing each property into a scale of only three steps may be limiting, and there may be arguments for more properties relevant to formulaic sequences, it seems that this approach is an inclusive one for talking about all types of formulaic language.

2.3. Fixedness and Variation

The collocation continuum has open choice sequences at one end of the spectrum, and idioms at the other, with most formulaic sequences appearing between the two poles, depending on their properties (see section 2.2). Formulaic sequences have less flexibility than free grammatical constructions. However, they are by no means fixed. There is a range of variations that formulaic sequences can undergo, whilst retaining their figurative or formulaic meaning. Moon says that although there is always some explicit form of fixedness present, it is not always lexical fixedness (1998a: 122). The older the idiom, the more frozen it will be (Cutler, 1982: 317 – 320).

Barkema (1996a) suggests that factors influencing the amount of variation a sequence may show are inherent, such as the linguistic characteristics of the phrase, the context the phrase is used in, and the frequency of occurrence of the canonical form. So a familiar formulaic sequence may be open to more variations than a lesser known phrase. Other factors may be contextual, such as the genre and medium of use, and also usage of the sequence; the frequency of the phrase may affect its variability. Carter (1987, cf. Partington, 1996: 26) concurs with Barkema, suggesting that if a phrase has an irregular syntactic pattern, for example, *the more the merrier*, then it is likely to be more fixed than one with a regular pattern, such as *to break someone's heart.*

Before detailing types of variation, it is important to recognize that all formulaic sequences have specific grammatical variations. Such variations include tense and inflection agreements. Fernando (1996) suggests that such variations arise due to discoursal needs, without having a significant effect on the sequence or its meaning. This is also suggested by

Glucksberg (1995: 16) in reference to the canonical phrase *speak your mind.* This can be quantified, for example, *speak your minds,* or it can be altered according to tense (*spoke your mind*). Fernando also suggests (1996: 42) that proverbs are often resistant to any changes.

Some changes in sequences may be due to unintentional mistakes. Moon suggests (1998a: 135) that there are also cases of spelling differences, and differences in homophones. She demonstrates this with examples such as *flying colours/colors,* where the spelling difference is due to register difference between British English and American English; the two varieties being different registers. Another example Moon gives is the formulaic sequence *the spit and image* becoming *the spitting image* over time, reflecting historical or etymological changes. Sound changes arise due to misunderstandings in processing the phrase, for example, *dull as dishwater/ditchwater,* or *damp squib/squid.* Changes in homophony include *toe/tow the line,* and *strike a chord/cord.* In some cases, the deviant form becomes the institutionalized one. Such pairings are thought to exist as the hearer tries to make sense of the formulaic sequence.

Perhaps the most comprehensive work on variation in formulaic language has been carried out by Rosamund Moon in her 1998 book *Fixed Expressions and Idioms in English.* Her work is descriptive, using a database of 6776 British and common American FEIs, gathered using the *Collins Cobuild English Language Dictionary.* She then investigated them using the 18million word Oxford Hector Pilot Corpus. The main types of variation she outlines can be divided into three broad categories, lexical substitution, grammatical variation and deletion/insertion. Partington suggests (1996: 143) that there are four basic mechanisms for altering a phrase; substitution, abbreviation, insertion/expansion, and reformulation. I group abbreviation and insertion together to indicate three broad categories:

Lexical Substitution:	Such as substitution of a main lexical component of the phrase; the verb/noun etc.
Grammatical Variation:	Such as structural changes, substitutions of grammatical elements of the phrase, and negation

Deletion/Insertions:	Additions of elements within the phrase, or attached to it. The opposite of this is deletion; an element is removed from the sequence

These variation types are ones often discussed by authors in the field. It is also important to acknowledge that these variations can be exploited stylistically in, for example, register change, such as the differences between British and American English, or an informal and formal version of a phrase etc. The next sections discuss the types of variation, and also register variation in more detail.

2.3.1. Lexical Substitution

Lexical substitution is perhaps the most common form of variation within a formulaic sequence. Terminology is somewhat consistent within the field; Moon (1998a) uses *lexical variation* and then subcategories such as *noun variation, verb variation* etc., Philip (2000) uses the same terminology as that used here: *variation* for change to the formulaic sequence, and *substitution* for the type of variation. Čermák (2001) uses lexical *variability* and Gibbs *et al.* (1989b) use *lexical flexibility*. Glucksberg (2001) uses *lexical substitution.* Moon in particular found that verb substitution was the most common form of variation in her research (1998a: 124), although no she does not report actual frequencies. Partington (1996: 126) says that lexical substitution, alongside rephrasing, is the most common type of variation.[9] This encompasses replacements of the lexical elements of the phrase, such as the noun (*quiet as a lamb/mouse*), verb (*upset/overturn the applecart*), and adjective (*don't get cute/smart with me*), as well filling gaps in sequences with context-dependent open-slots.

Phrases allow for lexical substitution in differing degrees. Some phrases allow for no lexical substitution at all, such as *kick the bucket,* or *see red.*[10] These would be collocationally closed, according to Barkema's scheme as shown in section 2.2.6. The amount of substitution, or insertion, a sequence can withstand depends on its meaning. For

---

[9] Barkema discovered that 1664 out of 2174 phrases he studied (Barkema, 1994) were flexible, the majority taking additions. The next most popular variant type was lexical substitutions, then syntactical rearrangements, e.g. passivizations, then interruptions.

[10] In Moon's study, she also found that even the most frozen, *kick the bucket,* had variant forms in the Hector Corpus: *kick the pail,* and *kick the can* were also found meaning "to die".

example (Glucksberg, 1995: 7) shows that it is possible to say *he silently kicked the bucket,* as it is possible to die quietly. However, *he sharply kicked the bucket* is not possible; although it is possible to kick something sharply, we do not tend to die sharply. Similarly, the action of kicking can take place over a long time – *he was kicking the bucket all week* – but one cannot die over a long time.

It is thought that the more compositional a phrase is, the more variations it will be open to (see section 2.2.2). Non-decomposable sequences tend to be fixed and collocationally closed, and pseudo-compositional sequences are open to limited collocations. Free expressions are open to a free amount of variations. Lexical substitution is often studied in conjunction with fixedness and compositionality of a phrase. Gibbs (1998: 65) found that if a formulaic sequence is normally decomposable, then it is more likely to accept lexical substitutions than a non-decomposable sequence. Individual components of a normally decomposable sequence contribute to the overall meaning of the phrase, and thus it is possible to substitute them as long as the substituted lexeme makes the same contribution to the phrase as the original did. This is not the case for non-decomposable idioms, where the individual components do not contribute to the overall phrase meaning, and so it is less possible to substitute them. Gibbs also found that abnormally decomposable idioms (see section 2.2.2) were as lexically flexible as normally decomposable ones, although noun substitutions were more disruptive.[11]

Lexical substitution is substitution of a main lexical component of the phrase, such as the noun, verb, or adjective. There may also be context-dependent or open-slot substitution (termed *discontinuous idioms* (Fraser, 1970)), where the sequence has place for any lexical item to fill the gap, for example *bang goes* ____, or *somebody's had more* ____ *than you've had hot dinners,* where the open-slot can be filled by any noun or noun phrase (in the examples) according to context, or for example *be in a* ____ *FRAME of mind,* ____ *for ENGLAND* etc. There may also be *frames* such as *a few bricks shy of a full load, a few pickles short of a jar, a few sandwiches short of a picnic etc.* The frame is: QUANTIFIER + NOUN PHRASE + *short/shy of* + NOUN PHRASE (Moon, 1998a: 159). Such frames will be discussed in section 2.4.3.

[11] Abnormally decomposable sequences have an indirect reference between a word and its referent.

Moon (1998a) found that verb variation was the most prevalent type of lexical substitution, and this was usually the main verb in the phrase. Substitution of the noun was only slightly less common than verb variation. Often, the noun tends to be the head word of the sequence, and is substituted by synonyms, or items from the same semantic field. Of course, there may be exceptions, for example, *be left holding the bag/baby,* where the nouns are from differing areas. Moon says:

> In metaphorical FEIs, the nouns often appear to be the locus or focus of the metaphor. Variations do not have changed meanings, but mental images of the metaphor may differ considerably: for example, the images generated by *burn one's boats* and *burn one's bridges.* The distinctions are therefore greater than those between many verb variations, and there would be more reason to regard such pairs as discrete, but cognate, lexical items.
> (1998a: 127)

The noun is important within formulaic sequences, and so lexical substitution of the noun is discussed in this research project. Less common lexical substitutions involve changing the adjective or modifier of the phrase, for example, *take the high/low ROAD.* Moon proposes that this is because formulaic sequences comprise mainly nouns, verbs, and grammatical words which act together to form a phrase. She proposes that formulaic sequences may be used to convey abstract notions. Adjectives are more concerned with describing a referent, and thus do not collocate with abstractions. In context-dependent sequences where the open-slot can be filled by any adjective, the lexical choice makes a significant contribution to the overall sequence meaning. The vocabulary used in formulaic sequences is restricted – Moon notes that there are only 4000 items in a word frequency list based on her database of FEIs, compared with 20 – 30,000 in normal dictionaries.

Variant forms can be exploited in terms of style and register. Gibbs (1995: 102) gives the example of *break the ice* leading to *shatter the ice.* The verb is substituted by *shatter* to give a stylistic effect. There is a new meaning of "breaking down an uncomfortable social situation dramatically", which is an exaggeration of the original meaning. A similar example is *not spill a single bean,* again a semantic extension of *spill the beans.* Semantic productions are easier to interpret than lexical substitutions from different semantic fields. The more familiar the original form, the easier the variant form will be to understand. As well as systematic lexical substitutions, which tend to be synonym substitutions, there are

also stylistic substitutions, whereby a user familiar with the phrase can manipulate it for language play according to context. Fernando (1996: 45) gives the example of *fools rush in where angels fear to tread* becoming *Marie stepped in where Jeanette feared to tread.* Partington (1996: 126) also gives an example:

> ...comparing these two *give us this day our daily starch* and *a tale of two teams,* where in both cases the final word is substituted for another, the former seems to be somewhat more successful in engaging the reader than the latter. This is because the substituting word – *starch* – has a connection with the substituted word *bread,* they belong to similar semantic fields, and the humorous effect is achieved by the contrast in register between The Lord's Prayer and the semi-technical item *starch.* In the latter, the substituting item *teams* seems to have no semantic relation to the original item *cities*, and so the alteration seems pointless, it is an empty pun. In *the naked and the well-read* the substitution of *well-read* for the *dead* of the original book title may also appear rather empty, but there is a platonic link between the rhyming old and new items...the majority of such word-phrase plays are opaque in sense if no context is apparent.
> (1996: 126)[12]

In terms of the lexical substitutions studied in previous research, it seems that there can be distinctions made between *stylistic* substitutions, and *systematic* substitutions (synonyms) with semantic substitutions occurring in each. Moon (1998a: 124) suggests that individual substitute forms of a formulaic sequence may not have the same focus, intensity, or distribution, although it is useful to group together phrases which vary in a similar way. She outlines *quasi-systematic* substitutes, for example noun synonyms (*a skeleton in the closet/cupboard),* verbal hyponyms (*upset/overturn the applecart),* male/female noun equivalents (*you can't keep a good man/woman down),* or singular/plural noun forms (*not give a hoot/not give two hoots).* As Moon says (1998b: 92) such variations are thought to be systematic as they show some regularity, or predictability, although Church *et al.* (1994: 156) point out that even the most synonymous of words are not identical in meaning. Other such variations include changes of possession (*get/give a raw deal*) or causative/resultative structures (*one's hair stands on end/make someone's hair stand on end).* Ullmann (1967: 240) says that words are stored alongside the words they associate with. Some of the associations are based on senses, others on form and meaning (the similarity in form and meaning). Central associations are made by, for example, synonymy, antonymy, and

[12] There is more context present in the example *give us this day our daily starch* than in *a tale of two teams.* The extra context sets up a much stronger expectation as to what words make up the sequence, and thus the substitution of the final term is more engaging than in the latter.

collocations. The number of associations per word will be very changeable, and for some very common terms it will be very high. Aitchison says that meaning and sound are "interwoven" (1997: 75). Words are possibly grouped in clumps with similar rhythmic pattern, and with similar beginnings or ending (1994: 143). Gläser (1998: 130) also refers to systematic variants, suggesting that substitutes may be *contextual synonyms*, as they share a common semantic field. Note that in Moon and Gläser's accounts there are no suggestions as to which is the most commonly occurring type of the phrase, for example, *upset the applecart* or *overturn the applecart.* This is common throughout studies within the field; there is no indication as to frequency of variant forms.

A phrase such as *drag your FEET/heels* demonstrates a systematic lexical substitution; *feet* for *heels.* Both possible noun substitutes are from the same semantic class; body parts, and are also hyponymic, *feet* being super-ordinate over *heels.* Another such example is *blow your own HORN/trumpet,* where there is noun substitution from the same semantic field, although *horn* is the AmE version, and *trumpet* the British form. The noun substitutes involved in *take a BEATING/hammering* represent systematic synonymic substitution. In Gibbs *et al.* (1989b), speakers' assumptions about the lexical flexibility of idioms were tested, investigating idiom compositionality and degree of flexibility - which types of idiom were more open to lexical substitutions. For the experiment, Gibbs *et al.* restricted lexical substitutions to close synonyms, for example *burst the request* for *pop the question,* where the lexical substitute refers to the same activity as the original lexical item. They note that more radical substitutions may have more effect on people's intuitions regarding idioms. However, he also notes that some substitutes which are not synonyms are actually less disruptive than a synonymic substitution, for example *hit the hay* and *hit the sack* (as was also found in the dataset). Some of the figurative meaning is lost if the substituted form is the synonymic *hit the straw.* Gibbs *et al.* (1989b: 66) then suggest that lexical substitutes are acceptable if they come from the same semantic field, *hay* and *sack* having similar features with regard to sleep and beds. Gläser (1998) also uses synonymous words in her substitution test. She investigated the modification of formulaic sequences in different contexts and genres, applying tests of lexical substitution, paraphrasing, and deletion of items from the phrase. Her results showed that variant forms may have a similar meaning, but a different stylistic effect, depending on the genre in which they occur. Čermák (2001: 6) says that stylistic usage, or communicative usage creates a constant degree of variations,

and some of these may become more stable over time. In literal language, substitution of a component part leads to paraphrasing. In formulaic sequences, there is a restricted amount of lexical substitution that can occur whilst retaining the figurative reading of the phrase, or the inherent fixedness of the sequence. Often sequences can be exploited, and even given a literal reading, such that the reader has to reflect on the actual formulaic sequence being referred to, for example, *extract the Michael* from *take the mickey* (Moon, 1998a: 173).

2.3.2. Grammatical Variation

Grammatical variation consists of changes of the functional words of a phrase, such as the determiner (*no/small/little wonder*), coordinator/subordinator (*hit and/or miss*), degree modifier (*not too/very tightly wrapped*), auxiliary verb (*can't/don't/shouldn't judge a book by its cover*), negation (*not/never in my wildest dreams*), or verb particle (*come up/out smelling of roses*).[13] It may also involve structural changes such as passivization, or compounding. Barkema (1996b) called syntactic flexibility *transformational deficiency.* Again, some sequences are open to more grammatical changes than others. For example, *bite the bullet* cannot have a determiner change to *bite a bullet* and retain the same figurative meaning (Fellbaum, 1993: 274).

Fellbaum (1993: 272) also suggests that grammatical substitution of the determiner is similar to that of lexical items; some sequences allow for it, and others do not. Whether it occurs or not depends on whether the phrase is compositional or not; the more compositional a sequence, the more variable it will be. Determiners found in formulaic sequences tend to be indefinite, definite, or null. Quantifiers or negative determiners are found less commonly, although they can occur as variant forms. The possible candidates of determiners for substitution within formulaic sequences are fewer than those possible within free combinations. Non-decomposable sequences do not allow for substitution of determiner. If a sequence retains its non-literal reading when the determiner only is substituted, then it cannot be stored holistically. Pawley suggests that the use of the definite article is related to prior knowledge of the sequence. He gives examples such as *call the doctor/the priest/the fire brigade* and *he met the girl of his dreams* and suggests that "in

[13] Examples taken from the LID (1998).

such cases, the speaker is not referring to a definite individual thing, but to a representative of an institutionalized category" (1985: 110).

Moon (1998a) shows that variation of quantifiers, or degree modifiers, is systematic and predictable, for example, *no/little love lost.* There are a few cases where the change is one of attitude, for example, *a fifth/third wheel, a nine-day/one-day wonder.* Changes in the preposition or verb particle are also grammatical changes, usually not involving a change in meaning. There are sometimes marked frequency distinctions, for example, *in keeping* is more common than *out of keeping.* Reversal of binomials is another such change. There are often no meaning distinctions between the pairs, for example *day and night/night and day,* and between *on and off/off and on.* Sometimes, sequences are reversed as a stylistic exploitation. In such cases, the meaning does change.

> *Elway does it his way* is an extreme case of grammatical substitution. Of the 5 lexical items which make up the original quotation (the Sinatra song – *I did it my way),* as many as 3 have been replaced, the only ones which remain unaltered are *it* and *way.* This begs the question of how the text receiver is expected to recognize the original. The solution must be along the following lines. Each of the words in the new version is related to the one in the corresponding position in the original – thus *I* and *Elway* are both personal phrase subjects, *does* and *did* are parts of the same verb and *his* stands in the same relation to *Elway* as *my* to *I.* There seem to be two important deductions to be made. Firstly, it is clearly very much the phrase pattern, the phraseology, which is being recognized. Secondly, in order to recognize the original phrase, the receiver is expected to think on the level of word lemma rather than word form – particularly in order to make the link between the verbs *does* and *did.*
> (Partington, 1996: 126)

As with lexical exploitations that require the recipient of the sequence to refer back to the canonical form of the phrase, Partington demonstrates a similar effect with grammatical substitutions. He shows how inflectional changes, due to context and word play, can alter a phrase such that at first glance the original may be unrecognizable. The second deduction Partington makes; that the lemma is important rather than the specific word form, is an essential point for this research project, see section 4.3.3.

Passivization is another form of grammatical variation. This is often used as a test for compositionality; if the formulaic sequence can be passivized and retain the same

figurative reading or not, for example, *to lay down the law/the law was laid down* (Gibbs and Nayak, 1989a: 102). The figurativeness does not remain in *kick the bucket – it was the bucket that Pete kicked* (Schenk, 1995: 254). However, it could be suggested that such transformations would be acceptable stylistically in the correct context. Pullman (1993: 252 – 254) also shows that in some appropriate contexts, idioms can be syntactically varied, ,for example, *the beans were spilled.* Note that it is the components of the phrase that are open to variations, not the block as a whole. This provides support for theories of compositionality. Other types of syntactic transformations formulaic sequences can undergo are outlined in Schenk:

Clefting:	*it was Mary's leg that Pete pulled*
WH-movement:	*whose leg did Pete say that Mary pulled?*
Topicalization:	*Mary's heart Pete broke*
Raising:	*The roof seems to have caved in on John's dreams*

(cf. 1995: 254)

Fraser (1970: 37 - 39) groups formulaic sequences in a six-step hierarchy, whereby the higher up the hierarchy a phrase is, the more syntactic operations it is open to. The six levels (plus completely frozen) are:

L6	Unrestricted	any operations allowed
L5	Reconstitution	syntactic rearrangement (*he laid down the law/his laying down of the law*)
L4	Extraction	removal of part of the sequence to a position elsewhere in the sentence (*the law was laid down by her father*[14])
L3	Permutation	the particle moves in a verb-particle-noun phrase (*put on some weight/put some weight on*)
L2	Insertion	placing some non-idiomatic constituent into the sequence (*read the riot act/read the class the riot act*)
L1	Adjunction	operations such as possessive marker, nominalization etc. (*John hit the ball/John's hitting the ball*)

[14] The direct object *the law* in the canonical *laying down the law* becomes the subject noun phrase in this extraction.

L0 Completely Frozen no operations allowed

In this, the notion of a continuum of fixedness remains. Such variations as represented by each level are grouped and classed here as being simply structural rearrangements.

Horn (2003) studies Jackendoff's notion of *metaphorical semantic composition.* Jackendoff (1997, cf. Horn, 2003: 248) says that if a formulaic sequence can be broken down compositionally, e.g. *[break] [the ice] [break down] [a social barrier],* then movement will be allowed – *the ice was broken.* Jackendoff does allow for exceptions to this. *[raise] [hell]* and *[cause] [a serious disturbance]* are acceptable, but * *hell was raised.* Horn takes this theory a step further by suggesting that movement can only occur if the verb in its literal reading and in its figurative reading acts in the same way; it takes the same thematic arguments. For example, in *break the ice* in the literal sense, an object is broken. This is similarly the case in the figurative sense; the social barrier is broken (2003: 249).

Moon (1998a: 113 - 115) also suggests that processes of nominalization are examples of manipulation. This involves the formulaic sequence becoming a noun, for example, *cry wolf* (verb) becoming *cries of wolf* (noun). By means of nominalization, compounds of sequences can be formed, for example *break the ice* becomes *ice-breaker,* or *blaze a trail* becomes *trail-blazer,* the compounds becoming institutionalized in their own rights.

Philip (2000) looks at variation with reference to the formulaic sequence *like a red rag to a bull.* Possible variations she found were substitution of the preposition *like,* e.g. *as/just as/than,* additions/subtractions e.g. *red rag to a bull* (omission of *like* changing the expression from a simile to a metaphor) or *waving a red rag to a bull,* and lexical substitution of the content words, e.g. *red flag to a bull/red flag to a dragon.* She suggests that variations such as these do not change the figurative reading of the expression, but that they do interrupt the reading of the canonical expression – the reader is more likely to be made aware of the wording than simply accepting the phrase as a "chunk" of language (2000: 223).

2.3.3. Insertions/Deletions

Another variation type which may provide some support for theories of compositionality (see section 2.2.2) is insertions, whereby lexical material can be inserted into the phrase. Such material is often adjectival (Moon, 1998a: 130), although it can be adverbial, prepositional, or interjectional. If material can be added to, or inserted into a phrase, then the phrase cannot act as a holistic chunk. Authors such as Nicolas (1995) and Pulman (1993) refer to insertions as *internal modifications*. Akimoto (1983) uses *modifier additions*, Zgusta (1967) uses simply *additions*, and Abeillé (1995) uses *adnominal modifiers*. Glucksberg (1993) refers to insertions as *adjectival* or *adverbial modifications*. The terminology used here, *insertions*, demonstrates that the type of variation being discussed is material inserted into the formulaic sequence. A term such as *internal modification* could apply to lexical or grammatical substitutions, or syntactical change as well as insertions, and thus may seem a confusing label.

There are several types of insertion possible in formulaic sequences. Barkema (1996a) differentiates between the terms *interruption* and *addition*, calling the former material which interrupts the syntax of the phrase, for example, *a hard nut, as always, to crack* and the latter being material which fits into the syntactic structure (*an appallingly hard nut to crack*). Additions are found more commonly than interruptions and this will be seen in the data for this research project; 71% of the 35 sequences in the dataset showed additions (See table 4-8, section 4.3.3). Naciscione (2001) refers to *interruptions* as *clefts*, and *additions* as *insertions*. Barkema (1996a: 71) also introduces a third class of insertion under the label *permutation*.[15] This occurs when, instead of there being a pre-modifying adjective, the syntactic pattern is rearranged, resulting in a post-modifying adjective: *a tough nut to crack →a nut too tough to crack.*

For the purposes of this project, *additions* will be the term given to material added outside the syntax of the formulaic sequence, and *insertions* will be the term given to material added within a phrase. In previous studies, insertions have been mainly restricted to material such as adjectives or quantifiers. Nicolas (1995) suggests that internal modification is modification of the NP of a V NP structure, Pulman (1993) proposes that internal

[15] Barkema's *term selection* also corresponds here to *lexical substitution*, and *permutation* to *structural rearrangement* (see section 2.3.2).

modifications can be adjectives or quantifiers, such as pre-determiners of adjective modifiers, and Akimoto (1983) suggests insertions to be pre-modifications of the noun, such as the definite/indefinite article, deictics, adjectives, and quantifiers.

A discussion regarding insertions or additions to the phrase involves the effect of the extra material to the phrase. Ernst (1981) classifies three types of modification effect caused by additional adjectival material:

External modification	-	material which gives context to the phrase *Carter doesn't have a political leg to stand on*
Internal modification	-	material which specifies the literal lexis, and which can be interpreted figuratively *we were reduced to scraping the bottom of every single barrel*
Conjunction modification	-	material which specifies the literal lexis, and which cannot be interpreted figuratively *Had such fun pulling his cross-gartered leg for so long*

External modifiers are known as *domain delimiters*, as they specify what domain the idiom is to apply to politics, economics, and so on. "Taken alone with the noun it appears next to, [such a modifier] doesn't make quite as much sense as we would expect" (Ernst, 1981: 51).

Nicolas (1995: 236) suggests that as insertions modify a component part of a complex expression, the modification must thus act over the whole phrase; for example, if an adjective modifies the noun of a V NP formulaic sequence, then it follows that the NP of the phrase is modified, and thus the phrase as a whole. This is in agreement with external modification. Abeillé (1995: 21) clarifies this as saying that an adjectival insertion is "external" if it acts in a similar way to an adverb, for example in: *politically, Carter doesn't have a leg to stand on.* It is "internal" if it only modifies part of the phrase, for example *scraping the bottom of the barrel → scraping the bottom of every single barrel;* only "barrel" is quantified by the insertion. It is coordinate if it involves both the literal and figurative meaning of the noun, for example, *pulling his leg* = "to tell someone something

that is not true, as a joke" (Longman, 1998: 206). The insertion *cross-gartered* affects the literal noun, and also the formulaic sequence as a whole. All of these types of modifications are syntactically insertions, however semantically they have differing effects on the reading of the phrase.

Moon (1998a), Naciscione (2002), and Gläser (1998) discuss reasons for such variations as insertions. One of the most important reasons is cohesion. Cohesion is the quality of making texts flow; it is where the interpretation of one element within a text depends on the interpretation of another element. Cohesion also plays a part in creating genre and registers. Examples of cohesion are metaphors, puns or allusions to formulaic sequences; these are specific types of wordplay, and help to further the interaction between the speaker/writer and the reader/listener. Formulaic sequences are often evaluative, and are used to convey attitude towards something. Gläser (1998) suggests that there are also *systematic modifications* possible, alongside *creative modifications* for creating a particular stylistic effect (see section 2.3.1).

Truncation, or deletion, is the "flip side" to insertions or additions. In such cases, an element is removed from the phrase and the meaning retained. Many sequences open to reduction are proverbs or sayings which are reduced from the longer canonical form. In many cases, as Moon suggests (1998a: 131), the reduced form "feels" elided, carrying an allusion to the original form, for example, *don't count your chickens (before they're hatched).* Some formulaic sequences become so familiar in their deleted form that these forms themselves become the institutionalized form, for example, *silence is golden* comes from the longer *speech is silver but silence is golden.*

Deletion is useful with regard to phraseological allusion (Naciscione, 2001: 99). This is where an implicit mental reference to the image of the formulaic sequence is made by using one or more of the sequence components, hinting at the overall formulaic image. Naciscione gives the example of *to choose the lesser of two evils* in D.H. Lawrence's *Mr. Noon*: "It is a choice of evils. Which do you choose?" (2001: 101). In many cases, the discourse is incomprehensible without awareness of allusion to the canonical sequence. As Naciscione suggests, a sequence may be retrieved by allusion using only very little

information, giving the example *a cat has nine lives* → "Well, you were drowned *four* times over. You are not a *cat* you know" (2001: 103).[16]

2.3.4. Register Variations

Register variation is a different type of variation. It occurs when one of the major types of variation, such as lexical substitution or grammatical variation is exploited, and the different variation types found are distributed differently according to register. Variant forms with register variations are often synonymous, but one form is of a higher register than the other. *Beat one's breast* is more formal than *beat one's chest* (Moon, 1998a: 132). Usually, it is the colloquial form that is the canonical one. In contrasts between British and American English, it is usually the noun that carries the variation. A few cases reflect cultural variations, as in: *turn on a sixpence/dime.* There are some formulaic sequences which are semantically the same, but have completely different lexis: *in inverted commas/quote unquote.* Moon also suggests two other types of variation: calques and false variations. Calques are foreign phrases which exist in the English phrasicon alongside their translations: *carpe diem* and *seize the day.* The two versions have differing frequencies of occurrence, and are often restricted via register. False variations are phrases which look as though they may be variations of the same phrase but are in fact separate sequences in their own rights, for example, *give and take* and *give or take.* There are few cases where a constituent keeps its meaning in different phrases, such as *beans* in *spill the beans* and *I don't know beans.* (Fellbaum, 1993: 279)

Whilst fixedness is a key factor to formulaic sequences, absolute frozenness is not a necessary property. Of the 6776 FEIs studied by Moon, 40% showed one or more types of variation. Not all variations are stylistic. The size of corpus is important. The BNC contains no occurrences of the phrase *kick the bucket*, so is even less likely to show variations on this sequence. The larger the corpus, the more occurrences and variant forms there will be per phrase. It is not true that the more frequently a sequence occurs, the more fixed it is (Barkema, 1996a: 70). Variations occur more in predicate structures than in adjective groups, for example. Variations occur systematically and stylistically:

[16] Example used by Naciscione taken from B. Shaw *Back to Methuselah.*

> Genre clearly plays a role too: while variations occur across the range of text types, it is often associated with journalism. Variations found in journalism cannot be dismissed out of hand as mannerism and journalese. In fact, journalism represents the cutting edge of language change, or the popularization of language change: variations fossilizing here may foreshadow what later becomes institutionalized more widely.
> (Moon, 1998a: 121)

Variations that occur stylistically are prevalent and unpredictable. Gläser (1998: 142) says that there is more stylistic potential for formulaic sequences in literary and everyday genres than can be found in a textual analysis. Journalists in particular exploit phrases in headlines and commentaries. Naciscione (2001: 225) also notes how manipulation of formulaic language is used to great effect in advertising and marketing, suggesting that "figurative language makes the advertisement easier to remember, therefore figurative use of language is one of the most striking features of the style of advertising".

2.4. The Study of Variation in Formulaic Sequences

2.4.1. Introduction

So far, this chapter has offered the background to the field of variation within phraseology. It has been shown that variation and fixedness is a key property of formulaic sequences. The different types of variation were also outlined. The amount of variation sequences are open to is a much debated issue. Barkema (1996c: 51) points out that handbooks of grammar do not describe the distribution of structures and their frequencies; they do not point to the actual variations permitted, only theoretical ones. In particular, Barkema asks, how does one discover to what extent variability is limited (1996c: 50)?

Barkema refers to grammatical variation in his work, however, the question may also apply to lexical substitutions; as has been demonstrated throughout this chapter, many sequences are open to lexical substitutions, but to what degree? This is also queried by Howarth: "...how limited is a limited set of substitutes?" (1996: 43). A similar question relating to insertions is posed by Čermák (2001: 4) in regard to the phrase *to pull strings.* He asks how many words, and what kinds of words can be inserted, and maintain the figurative meaning. It is possible to have *pull long/short/interesting/political strings,* but not *pull fish/nylon strings.*

There is little previous work done on the limits of variation: how much variation can occur without the sequence becoming a free construction, or how much variation can occur restrictively, without the phrase becoming variable according to context. Similarly, there is no definitive account of how much variation occurs before the formulaic sequence loses its figurativeness. Barkema, Moon, and Howarth have attempted to answer the question as to how limited variation is. Barkema investigated full flexibility of received expressions (formulaic sequences) (as opposed to "potential flexibility"; the rules of grammar allow for unlimited embedding and recursion). He says:

> [full flexibility] is the flexibility of non-received expressions to which morphological and syntactic rules can be applied without any restriction. Therefore, to be able to determine how limited the flexibility of any received expression is, one should know what the flexibility characteristics are of non-received expressions that have the same "base pattern" [...] For example the non-received expressions *the roof of the house* and *the end of the chapter* have the base pattern of the received ones *the tip of the iceberg* and *the villain of the piece.*
> (1993: 274)

Kytö and Rissanen suggest that the way to collect full relevant instances of a sequence is by a combination of intuition and systematicity (1993: 253). Barkema's method was to first take a list of the structures under investigation. This is a list of the base, or canonical, forms taken from the Longman Dictionary of Contemporary English. Then, using corpora, he made a list of all of the variant forms of the canonical expressions, and noted the frequencies. Barkema used two corpora; the 20million word Birmingham Collection of English Texts, and a subcorpus of the Nijmegen corpus and the British component of the International Corpus of English (subtotalling 111,713 words) for the free constructions.[17] For ambiguous constructions, a list is also made of "counterfeit" forms (1996c: 52). These are the forms of the canonical formulaic sequences which have a literal reading, for example *cold feet* = lack of courage, but the counterfeit form is *cold feet* = feet lacking in warmth. For constructions which do not have "counterfeit" forms, such as *blind alley,* all free constructions having some similarity to the canonical form are included, for example, *blind man/woman, long/short alley* etc. He looked at grammatical variations of noun phrases. Barkema compared the results for the lexicalized expressions against those of the free constructions. He compared the free constructions in a corpus with received

[17] This smaller subcorpus was used for free constructions as they are more frequently occurring than lexicalized expressions (Barkema, 1996c: 55).

expressions of the same base pattern in a corpus, and compared the results to find flexibility in use, rather than flexibility as prescribed by grammar (1994b: 45). His method seems to point to some use of intuition, particularly when finding the similar lexical forms to the canonical form. This means that his method may not be a comprehensive way of collecting all possible substitutes. Note that Gibbs *et al.* (1989b) say that not all lexical substitutes are synonyms, nor necessarily from the same semantic field.

Moon used a collocation technique:

> A specific query such as "show all matches of the lemma **spill**, used as a verb, with the word **beans** occurring within a window of between 2 and 5 words of **spill**, and preceded immediately by **the**" yielded 7 matches
> (1998a: 50)

As Moon herself notes, intuition is a large part of this process, as searches are based on what the researcher enters, not what should or could have been looked for (1998a: 49). This also means that due to the unpredictability of formulaic sequences, transitory or stylistic variant forms might be missed by using this technique. Intuition is unavoidable when studying formulaic language. The collocation technique adopted by Moon means that some variant forms and substitutes may be missed. Howarth also says that collocations that span over sentences or paragraphs, and cases where the noun is substituted by a pronoun may also be missed (1996: 74). Moon used significant occurrences for her collocations, setting her threshold at five; any sequence having an occurrence frequency of four or less was disregarded as being a random occurrence within the corpus (1998a: 57). Again, this means that some substitutes may have been excluded, and thus full flexibility was not recorded.

Howarth also used collocation searches in corpora, although he only noted those collocations which had collocational significance, and thus had an occurrence of more than one.[18] This distinguishes significant from "casual" collocations (1996: 69). Significant collocations are collocations which occur more often than their respective frequencies would predict. Again, this results in the method not collecting the comprehensive set of variant forms a sequence may have. Here, intuition is again relied on by the researcher as to

[18] Howarth used a subcorpus of social science texts created from the Lancaster Oslo Bergen (LOB) social science texts, and social science texts provided by the University of Leeds – the corpus totalled 240,000 words. (1996: 75).

whether two forms are of the same collocation. Note however, that Howarth did not restrict the distance between collocating items, and thus can study collocates over sentences or paragraphs.

Barkema, in his exploration of collocability in order to provide a multi-dimensional model of formulaic sequences found that:

> In a small number of cases it is possible to replace a whole expression by a synonymous one in a number of steps. An example is from *near miss* to *narrow shave: Near miss – near thing – close thing – close shave – narrow shave.* The same goes for *old stager* and *good hand: old stager – old hand – good hand.* The insoluble question in relation to dominoes like this is: when are two forms one and the same lexicalized expressions, and when are they two or more expressions which are synonymous?
> (1996b: 147)

These *chains* or *steps* link variation; or lexical substitution in this case, to grouping formulaic sequences. Thus exploring the boundaries of variation of sequences can be situated within theories of formulaic sequence groups.

2.4.2. Lexicographical Approaches

Jackson (2002: 147) says that the tradition of storing words in alphabetized lists dates back to Old English Latin – English glossaries. However, around this time, glossaries of vocabulary organized by topic also started to appear, with areas including parts of the body, family relationships, religion, war etc. These continue today in thesauri. A developing tradition within phrasal dictionaries is similar; grouping sequences together with similar themes or topics (see section 3.2.1.2). An alternate theory of grouping sequences which dominates Western phraseology and lexicography is the idea of *Lexical Functions* (LFs), as suggested by Igor Mel'čuk (1998). Lexical Functions are general abstract meanings which are related to deep syntactic roles. They group together lexical phenomena which have always been thought of as being separate, but in fact have the same logical nature.

Mel'čuk outlined an original set of about 60 basic LFs, known as *Simple Standard Lexical Functions*. These have since been added to by authors such as Fontenelle (1998). These make up the foundations for the description of collocations. LFs can occur in isolation or in

combination. So for the function *the one who undergoes,* if this function is put onto the verb *to shoot*, the outcome is *target* (the one who undergoes the shooting is the target). If the object is a *hotel,* the outcome is the *guest* (the one who undergoes a stay in a hotel is a guest) (Mel'čuk, 1998: 32). LFs are a conceptual grouping, grouping together words and collocations that can undergo the same functions within phrasal dictionaries.

Howarth found that for foreign language learners, it is not pure idioms which are problematic. These can be rote learnt together with their meanings. The problematic area concerns items within the collocation continuum, which are often overlapping in terms of form and meaning, and it is these sequences which are often neglected in lexicography (1996: 162). He also notes that the difference between collocations and pure idioms is also insufficient in idiom dictionaries (1996: 168). Another problem in the field is that idiom dictionaries do not specify the boundaries of variations; they show only a limited selection of the substitutes available in actual use. Likewise, they do not cater for the boundaries between literal and non-literal cases, in particular in decomposable sequences they do not show which, if any, component carries a literal meaning (1996: 174). Native speakers themselves may be unable to suggest where such boundaries lie.

Kövesces and Szabó suggest that the traditional atomistic view of formulaic sequences treats them as a matter of the lexicon; each sequence needs to be treated alone in terms of form and meaning. They then can relate to each other in similar ways as words, via synonymy, and so on. They say that this traditional view does not take into account the conceptual system and encyclopaedic knowledge of speakers (1996: 329).

A way in which dictionaries of English formulaic sequences tend to represent lexical groups of any kind is via *thematic panels*. The LID has ten basic concept words, each divided into groups of idioms with similar meanings. The idea of these thematic panels (an idea common to most English idioms dictionaries) is that it helps learners of the language see which idioms are related/similar or different. This is supported by Kövecses and Szabó (1996: 329) who give an example as to how dictionaries systematize idioms:

Spitting fire
The *fire* between them eventually *went out*
The painting *set fire* to the composer's imagination
Fire away!

Each of these examples contains the word *fire.* However, as Kövecses and Szabó say, this way of grouping formulaic sequences demonstrates only those sequences containing the word *fire;* it does not show any conceptual basis. They suggest a better way of grouping sequences is by concept, in a manner akin to a thesaurus, for example (1996: 329):

Sparked off
Burning the candle at both ends
Snuffed out
Fanned the flames

These sequences are linked by "aspects" of fire, not the orthographical form word *fire.* Kövecses and Szabó support the conceptual view of sequences, saying:

> ...idioms are products of our conceptual system, and not simply a matter of language (i.e. a matter of the lexicon). An idiom is not just an expression that has meaning that is somehow special in relation to the meaning of its constituent parts, but it arises from our more general knowledge of the world (embodied in our conceptual system). In other words, idioms (or at least the majority of them) are conceptual, and not linguistic, in nature.
> (1996: 330)

The systems of lexical functions and thematic panels seem to group sequences in a more comprehensible manner than simply alphabetically, although they still demonstrate the problems emphasized by Howarth (1996: 174) regarding the boundaries of sequences, and the boundaries of literal/non-literal meaning.

2.4.3. Lexical Frames

One linguistic theory concerning groups of idioms is the *frame theory* (cf. Moon, 1998a). This says that formulaic sequences cluster according to a common structure, but have variation in one of the constituent parts. The meanings of the formulaic sequences are

similar or identical. They are known as lexicogrammatical frames, for example *down the tube, down the drain, down the chute, down the pan, down the plughole,* and *down the toilet.* They are semi–synonymous, although may differ in distribution. These frames also emphasize possible restrictions for example in the frame *in the buff, in the altogether, in the nude, in the nuddy,* and *in the raw,* you cannot have **in the naked* or *in the bare* (Moon, 1998a: 147).[19] This then goes some way towards describing the boundary of lexical substitution for the phrase IN + det + NOUN (wearing no clothing). The frames are often strictly syntactical, for example, ON + det + NOUN (physical activity) *on the boil, on the fly, on the march, on the hop, on the run* etc. (Moon, 1998a: 147).

A popular frame has the structure V + det + NP, where the verb is to some extent semantically depleted (*make/take/get* - see section 2.2.2). Such frames are productive – individual realisations are not individual formulaic sequences, but are part of a larger group, for example, *make a killing/fortune/mint/profit.* There are underlying semantic constraints. It seems that the frame is institutionalized, alongside the semantic meaning, thus the lexical entries are not simply arbitrary collocations.

Nattinger and DeCarrico (1992: 36) classify sequences according to structure: grammatical sequences have open lexical slots (e.g. V + det + N), collocations which are strings of specific lexical items that have a tendency to occur together, and lexical phrases. Lexical phrases can be non-productive strings of lexical items, which allow for no changes lexically or grammatically. They can also be generalized productive frames which have a generally specified semantic and syntactic frame, and a pragmatic function, such as *a + N [TIME] + ago – a long time ago, a month ago.* Other frames are recognized by Barkema (1994a) as being collocationally open, such as ____ *after* ____ *time after time, day after day, month after month, year after year* etc. Renouf and Sinclair (1991) also investigated grammatical frameworks such as *a* ___ *of, for* ___ *of,* where the intermediate word is lexical. Wray suggests that variants within these frames are not semantically interchangeable, the example she gives being *if it's good enough for my sister it's good enough for him,* which does not have the same meaning as *if they're good enough for a wedding reception they're good enough for her party* (2002: 32).

[19] Moon says that this is a restriction, but does not explain her methods for determining the limitations of the frames.

"Similes are essentially frames with fossilized lexis" (Moon, 1998: 150), e.g. *as dead as a doornail, as quiet as a mouse, as good as gold.* They function as emphasizers. The two frames associated with similes are:

As + ADJ + as + NP

V + like + NP

Individually, similes are rare. Within similes, there are fewer adjectives than variable nouns; it is the nouns which tend to be the variable part of simile frames. The noun variant may cause a different meaning to the adjective. Binomials may also be a frame. There are rules guiding the construction of binomials, for example, in gender pairings, the male counterpart tends to go first. The first item is considered the more dominant or positive one. Pairs of sequences form parallel frames, where one phrase is the opposite of the other, for example, *from the bottom up, from the top down* (Moon, 1998a: 156).

These examples are fairly rigid examples of frames. There may also be frames with much less lexical stability, where the lexical variation is seemingly without limits, whilst the syntax and pragmatic use remain fixed. Frames such as *is the Pope a Catholic,* or *one sandwich short of a picnic* can have seemingly endless lexical variants within their underlying structure. These are termed *empty frames* by Wray (2002). It seems that the underlying notion and one version require memorizing (the former frame being used to answer another's question emphatically and the latter used to suggest that someone is mentally lacking). New versions are memorized depending on quality of wit. The creativity and humour of these frames is important, but the underlying frame must be recognized. Most are culturally dependent.

2.4.4. Idiom Schema

Lexical Frames are the most rigid system of grouping formulaic sequences. They have a fixed syntactic structure and some of the lexemes within the sequence are specified. Moon suggests a further development to lexical frames, involving more variation. She suggests a notion of idiom schema, e.g.:

Shake in one's shoes
Quake in one's shoes
Quake in one's boots
Quiver in one's boots
Quake in one's Doc Martins

The schema of *shake in one's shoes* is an example of a formulaic sequence cluster, where verbs meaning "shake" are associated with nouns meaning "footwear" to demonstrate fear (Moon, 1998a: 161).

Similarly:

Fan the fire of something
Fan the fires of something
Fan the flames (of something)
Add fuel to the fire
Add fuel to the flames
Add fuel to the flame
Fuel the fire
Fuel the fires
Fuel the flame
Fuel the flames (of something)

While the first and the last items are discrete, they also represent ends of a continuum.

> I am terming these kinds of FEI cluster idiom schemas. They have some reference in common, a metaphor in common, and cognate lexis, but without (necessarily) any very fixed structure or fixed lexis. The notion of idiom schemas can be used to explain a number of things: in particular, (extreme) variability, evaluative content, apparent compositionality, and the ease with which allusions to FEIs or exploitations are decoded. Idiom schemas represent concepts embedded in the culture and associated with particular lexicalizations. They are characterized by an underlying conceit (the relationship between tenor and vehicle) and an overlying preferred lexical realization, usually with connoted evaluation. The exact form of words may vary or be exploited, but is still tied to the underlying conceit which provides the driving or motivating force in the FEI.
> (Moon, 1998a: 163)

The main idea behind idiom schema is that they have some basic common underlying concept (or conceit), and an "overlying preferred lexical realisation"; a common lexical "set".[20] The exact lexemes involved can vary, but the underlying meaning remains the same. Idiom schemas can be linked to frame semantics, a theory which suggests that every time we experience a new situation, we memorize it as a frame, and use this to develop prototypes of a situation. Conceptual metaphors may also support these schemas. Prototype theories support idiom schemas in that there are variant forms which may be closer to the canonical formulaic sequence (the prototype) than other versions. As Moon suggests (1998a: 168), there are rules behind idiom schema: the metaphor and meaning must remain the same, and variant lexical items must be recognized as belonging to the group.

Moon suggests that idiom schema demonstrate a diachronic and dynamic process, whereby a metaphor stabilizes, destabilizes, and restabilizes (1998a: 164). She says that it could be the case that all metaphorical formulaic sequences form schema, some of which have simple forms, for example *spill the beans* (fixed lexical set), and some allow for substitutions, transformational changes etc. She also suggests that schema can be related to frame semantics and theory, whereby a frame is created when one encounters a new situation. This is then the framework for similar situations, adapting to fit any changing details as necessary. Thus each framework develops into a network as different features from similar situations are connected to the same underlying framework. Further support for schema comes from Pulman (1993: 256) who says that in sequences such as:

---

[20] Moon suggests that her frames can be seen as rigid schema, e.g. *in the nude/buff/muddy* etc. for wearing no clothing.

I've got some loose ends to be tied up
I'm tying up a few loose ends
A few loose ends need tying up

it is the same idiom in each sentence rather than three separate sequences. However, this discredits canonical form theories, as there is no syntactic canonical form which can represent all of the variational properties discussed above (structural rearrangements, tense changes, determiner/quantifier substitutes, and so on). Only in rejecting the notion of a syntactic and semantic canonical form, is it possible to accept an idiom schema theory.

2.4.5. Networks

As idiom schema build on lexical frames, networks build on idiom schema. The *network theory* (Aitchison, 1994) suggests that words are stored in the mental lexicon with words as nodes, linked together to others by relations due to characteristics and use; they form an interconnected system. Words are organized into semantic fields, and within these fields they are related to each other by links. This network theory is also compatible with prototype theories. Strong links occur between, e.g. co-ordinates and collocations, for example *salt and pepper,* or opposites, and *bright red, strong tea* etc. (Aitchison, 1994: 97). Associations join words in a network by senses, or by form and meaning. The number of associations, or paths attached to each word (node) differs from one to the next, and is changeable at any period throughout time as words become more or less in vogue. For commonly used words, the number of paths attached will be very high. Wierzbicka says that the meaning of a word is composed of semantic primitives. The meaning of words can overlap (a*bc* overlaps with *bc*d) and these semantic primitives can offer a tool for forming semantic groupings (1996: 170).

If individual words can be joined in networks by form and meaning, then this research thesis suggests that phrases can be too. This idea builds on work done by Williams (1998), who works on collocations between terms within a semantic field of medical terminology. One example he gives is in plant molecular biology for DNA. Common links to DNA are *chloroplasts, plastid, plant cells* etc.

> This leads to the supposition that if the wordings are seen to reflect the underlying conceptual frameworks, when lexical items are found to co-occur with other lexical items, patterns of co-occurrence will form "collocational chains". Following these "collocational chains" through would allow us to go beyond the immediate "contextual framework" to isolate the full frame of reference of a given item within the lexis of a discourse community...The word "network" is used here to signify a web of interlocking conceptual clusters realised in the form of words linked through the process of collocation. It is hypothesised that the patterns of co-occurrence forming the collocational networks will be unique to any one sub-language and serve to define the forms of reference within that sublanguage. The networks are formed by following through the "collocational chains" which consist of initial look-up forms, referred to here as nodes in that they act as points of intersection in the graphical representation, and their collocates.
> (Williams, 1998: 156)

So in analogy to word networks, Williams suggests collocational networks, where lexical members of a collocate form a node, and the link between them is a collocation. Thus, as words collocate with several different words, a framework, or network develops, with similar concepts, making the framework conceptual and relevant to particular areas of use. Certain words (such as *DNA*) will collocate with more items, thus showing that they are central concepts to the subject framework. Support for networks comes from Hanks (2004: 246) who says that humans do not store lexemes in isolation in their brains in a list. They store sets of syntagmatic patterns associated with each lexeme. Another example comes from Järvi *et al.* (2004), who used Williams' networks to examine which concepts were emphasized by the Nokia phone company in their quarterly reports. The word *nokia* was found to be central to most reports, linked to industry nouns such as *mobile, networks*, and *phones*. In early reports from 2000, there had been a drop in financial performance. Collocates showed concepts such as *increased* and *new*. As Järvi *et al.* (2004: 2 - 3) suggest, perhaps the aim of Nokia is to increase the trust of shareholders with positive messages. A further example of collocational networks is given by Ferrer i Cancho and Solé (2001: 1), who suggest that two words may be linked together if they co-occur frequently, forming a network whereby common words are central and rare words are on the periphery.

The idea here of these networks is that some concepts will be more accepting of collocates than others, and will thus prove to be more central concepts to the discourse community. The collocates form links and networks, with the central concepts acting as *nodes*. This

notion also seems to work for idiom schema. Take, for example, Moon's examples relating *shake in one's shoes* to *quake in one's Doc Martins* around the central underlying notion (akin to a conceptual metaphor) of "showing fear". The forms *quake in one's shoes,* or *shake in one's boots* are more prototypical to the schema than *quake in one's Doc Martins.*

Similar work which unites networks and Barkema's ideas of *chains* (section 2.4) from e.g. *near miss* to *close shave* is that done by Howarth (1996). In his attempt to explore the boundaries of substitutes, he suggests *overlapping collocation clusters* (1996: 44), in which verb and noun collocations are seen as verbs with synonymous figurative senses collocating with lexical sets of nouns. He demonstrates the clusters thus:

> For example, while *introduce, table* and *bring forward* collocate with *bill* and *amendment,* only *introduce* and *table* of those three collocate with *motion.* By shortening the set of verbs to *introduce/ table,* the set of nouns can be extended to *bill/ amendment/ motion.* At a more restricted level it can similarly be seen in the case of *pay heed* that extending the possibilities at the noun position and including *attention* as a substitute of *heed* results in excluding *take* as a synonym of *pay.* (1996: 102)

This is further illustrated as:

Bill (1)→*introduce* (2)↓

+bring forward + table

+amendment ← (3)

So *bill* and *amendment* collocate with *introduce, bring forward,* and *table,* although the nouns *bill, amendment,* and *motion* collocate with the verbs *introduce* and *table.* Howarth found this pattern of "chaining" for phrases where both the noun and the verb are open to a degree of substitution (for example not open substitution, but not closed either). This overlapping is supported by Altenberg and Eeg-Oloffson (1990: 8), who suggest that recurrent word combinations have variable length and frequent overlapping.

2.5. Conclusion

This chapter has detailed the background to the research project. Formulaic sequences are characterized by a variety of features such as compositionality, transparency, institutionalisation and fixedness, with fixedness and variation being the largest area. Variation can divided into lexical substitution, insertion and grammatical variation, with register variations being exploitable as well. There is much literature describing formulaic sequences and how they can be varied, yet there is little research exploring the boundaries of the variation. Previous methods do not comprehensively collect variant forms of a sequence. Exploring lexical substitution and insertion, this research project uses the idea of sequence groups, in particular networks, to explore the limits of manipulation. The following chapters (3 and 4) describe the methodology used to provide data to investigate the research questions, outlining the chaining process and the selection of data respectively.

3. Data Sources and Method

3.1. Introduction

The current chapter describes the sources used for the research project, and the extent to which they provide suitable data for analysis. The analytical procedures are also specified. Section 3.2 details the data sources used; the LID and the BNC. Section 3.3 explains the data preparation technique, here known as the *chaining process*, which will use the data sources to investigate the limits of variation. Section 3.3 also details the normalization technique used to standardize the data according to different domain sizes of the BNC. This chapter thus gives the tools and techniques used in this project to obtain the data (Chapter 4) and investigate cases in the empirical chapters (Chapters 5 and 6).

3.2. Data Sources

3.2.1. The Longman Idioms Dictionary

The aim of the LID is to reflect the range of idioms in use in British and American English today (1998: vii). It also provides notes as to the distribution of the idioms, for example, whether a particular phrase is used mostly on television, by children, in British English or American English, or whether it is considered to be rude or old-fashioned, and so on. The LID claims to include a wide range of idioms in the language, as well as some of the newer idioms, such as *it's all gone pear-shaped* (1998: vii). The target audience of the dictionary is composed of teachers and students of English as a foreign language and grammar, and the dictionary can also be used for general interest by native speakers. The LID takes its data from a variety of corpora known collectively as the Longman Corpus Network, which means that the formulaic sequences are taken from naturally-occurring data. However, as individual occurrences of formulaic sequences within even large fixed corpora are rare, it is necessary to use a variety of sources. The LID states that the World Wide Web and "keeping our ears tuned to the media and language on the street", are used in order to provide a corroboration of sources with which to validate an entry (1998: vii).[1] Whilst the presence of idioms in language is verified amongst a variety of sources, the fact that entries are not based on frequency of occurrence alone means that the dictionary has an element of

[1] Personal communication: Stephen Bullon, Managing Editor, Longman Dictionaries (08/ 08/ 05).

subjectivity regarding its entries. This will be evident in cases such as *throw a WOBBLY/wobbler* or *take a BEATING/hammering* in Chapter 5, where the frequency data of lexical combinations differs from the entry given by the dictionary.

3.2.1.1. Definition of Idiom

The dictionary calls each of its entries *idioms*. The LID takes its entry requirement, and definition of idiom to be the following:

> An idiom is a sequence of words which has a different meaning as a group from the meaning it would have if you understood each word separately
> (1998: vii)

The definition given by the LID is more general than the definition of pure idioms as given in section 1.1. It does not take into account the differing levels of compositionality or analysability of meaning, or the degrees of collocability or fixedness (see section 2.3), but simply refers to all multi-word units occurring along the collocation continuum as being *idioms*. The LID contains formulae, such as *of all things*, which cannot be classed as being on the same level of idiomaticity as the prototypical idiom *kick the bucket*, and which is also more lexically fixed than such a phrase as *shut your FACE/gob/trap/cakehole/mouth*. The LID does not include: collocates (such as *as usual*), nouns with their operating verbs (such as *make a point*), or phrasal verbs (such as *break down*) (1998: vii). Each entry has two or more examples demonstrating its use, taken from the Longman Corpus Network; idioms with only one example are not used frequently. Idioms which are not used often occur infrequently in corpora.

3.2.1.2. Variation

As well as listing the different idioms, the LID shows where variations may possibly occur, and says that "very few idioms are fixed in form" (1998: ix). Take, for example, the phrase *put sth on the back BURNER.*[2] The LID lists this expression as demonstrating several different possible types of variation, for example syntactic variation: *be on the back BURNER;* lexical substitution of the verb in the form *be/stay on the back BURNER*, or adjectival substitution in *be on the front BURNER.* There can also be context-dependent, or open-slot substitution in the canonical version (sth). An expression such as *be smiling/grinning from EAR to ear* is listed in the LID as having only one variation by contrast: lexical substitution of the verb. So whilst the LID does not take fixedness into account in its definition of an idiom, it acknowledges possible variant forms for individual entries. The LID uses the label *variant* to cover all types of variations. It uses two ways to illustrate lexical substitutes, firstly:

Sell like hot cakes
(also **go like hot cakes)**

The type above is used to show that the bracketed form is less frequent than the other. In cases where the phrase is laid out in the form:

Take a beating/hammering

the dictionary suggests that each form of the noun is equally frequent. Other types of variation as shown in the dictionary include:

[2] The stylistic conventions of the LID are followed here – headwords of the expression are capitalized. *Sth* refers to "something", and "somebody" is abbreviated to *sb* in expressions that are context-dependent. Other conventions include bracketing to show which parts of the expression can be omitted without changing the meaning, e.g. *a BIRD in the hand (is worth two in the bush),* lexical substitutions are shown with strokes, *leave/fly the NEST,* and context-dependent substitution may also be marked with ____, e.g. ____ *is a two way STREET* .Typically, phrases are listed according to the first noun in the phrase. The noun is often the headword of the sequence. In the absence of a noun, the phrase is entered at the first significant word: the verb or adjective. Determiners, pronouns and prepositions tend not to be used as keywords.

- Deletion: *Cut your (coat according to your) CLOTH*
- Dialectal variants: *not have two PENNIES to rub together* (British English) *not have two CENTS to rub together* (American English)
- Structural variants: *in the CLUTCHES of/in sb's CLUTCHES*
- Spelling variants: *catch/take/get (a lot of) FLACK/flak,*
- Acronyms: *same SHIT different day/SSDD*

Variants which occur less frequently than the main idiom are bracketed. Variants which have a similar meaning, but cannot be defined together are grouped with the main entry being defined, and the closely related idiom listed below with only an example, no definition, for example *give sb a clean BILL of health* (main entry), and *get a clean BILL of health.* The surface meanings of these differ due to the direction of the verb, however the underlying meaning is the same: that someone is completely healthy. The latter phrase is given after the definition and examples given for the former. This is also the case for *sb has been sold a BILL of goods* and *sell sb a BILL of goods:*

> **Sb has been sold a bill of goods**
> *Especially AmE*
> Used in order to say that someone has been given an untrue description of a situation, by someone who is trying to gain an advantage: *The weapon looks good, but it doesn't work – Congress has been sold an expensive bill of goods.* | *Independent politicians like Ross Perot have tried in the past to convince the American people that they have been sold a bill of goods by the government.*
> **Sell sb a bill of goods:** *He is not an objective witness – he's trying to sell the jury a bill of goods!*
> (1998: 26)

The layout illustrated above follows for nominalized or verbalized forms of idioms, for example *sit on the FENCE; FENCE-sitting* (1998: x). Idioms which are opposite in meaning are also listed under the main entry, and noted as being an opposite, for example *in/into the LIMELIGHT,* and *out of the LIMELIGHT.*

The LID refers to insertions in its guide for use only. It is noted that a bracket notation is used to show words that can be left out without changing the meaning of the idiom. This looks as though the dictionary is explaining deletions, rather than insertions. This is

confusing. Is the information in brackets found originally in the phrase, but then nowadays truncated, or is it information which has recently become a regular addition to the canonical phrase? There seem to be two main sorts of bracketed material in the dictionary; the first type is almost phrasal, and is typically found at the end of long proverbs, and the second type is one-word modifiers found inserted into the phrase, for example *beat a **(hasty)** RETREAT* (1998: 284). The dictionary does not have any indicator for phrases which may allow for context specifiers, for example, under the definition for *overstep the MARK/limits/bounds* there is no marker to say that this phrase regularly allows for insertions.

The LID also contains a section known as an *idiom activator*. This section of the LID takes what it calls *concept words* and groups idioms according to these concepts. The idioms in each concept have similar meanings. There are ten such concept words used: *angry, different, difficult, easy, not understand/know, problem, same, start, stop*, and *understand.* One example is the concept word *PROBLEM,* which has the following concepts and related formulaic sequences:

Cause problems for yourself	-	*the CHICKENS have come home to roost*
	-	*foul your own NEST*
	-	*be your own worst ENEMY*
In a situation with difficult choices	-	*between the DEVIL and the deep blue sea*
	-	*a CATCH 22 situation*
	-	*Hobson's CHOICE*
Having problems	-	*in/into deep WATER*
	-	*in a tight SPOT*
	-	*be up the CREEK (without a paddle)*

(Longman, 1998: 192)

Another concept word is *EASY,* which subdivides as follows:

Sth is easy	-	*as easy as falling off a LOG*
		as easy as PIE
		be a BREEZE/doddle
		be a DUCK shoot
		be a piece of cake/piss
Sth will be easy because the difficult part is done	-	*be DOWNHILL (all the way) from here*
		be HOME and dry
		be HOME free
Sb does sth easily	-	*sb can do sth in his/her SLEEP*
		no SWEAT
		sb takes to sth like a DUCK to water

(Longman, 1998: 198)

The members for each underlying concept do not lexically relate to each other, but they retain the similar underlying meaning. *The Cambridge International Dictionary of Idioms* (hereafter CIDI) and *The Collins Cobuild Dictionary of Idioms* (CCDI) also contain such sections, the former having 15 panels, subdivided in the same way as in the LID, whilst the latter has 32 non-subdivided categories. These will also be used for reference.

The LID contains over 6000 idioms, and 2210 show variation of some kind, such as lexical substitutions, syntactic arrangement, opposition, dialectal variation etc.[3] Of these, 1450 formulaic sequences show only one type of variation. There are 593 idioms in the dictionary showing two types of variation, making up 9.9% of the total number of idioms in the dictionary. There are 167 entrants in the dictionary which show three types of variation. The LID contained many phrases, such as *knee high to a GRASSHOPPER,* or *every MAN to himself* which showed no possible variations.

[3] 37% of the overall dictionary idioms display variation, which is comparable with the 40% Moon found (see section 2.3.4). The statistics for entries with one, two and three types of variation are also comparable to Moon; she found 14% of her total database to have two or more variations. Here it is 12.7%.

3.2.1.3. Register

The LID makes use of a number of labels that add relevant information about the idiom. These include, for example, AmE or BrE when a sequence is used in either American or British English respectively. The sequence *play/be a GOOSEBERRY* is typically British in use, whereas *take a RAIN check (on sb)* is American. Other such systematic labels include whether a sequence is spoken (for example *get a LIFE!*), old-fashioned (for example *sing for your SUPPER*), slang (for example *sb has lost the PLOT*), or taboo, such as *give sb a BOLLOCKING.*

As will be illustrated in section 4.2.1, these labels are adopted for the database created in this research project. It is suggested that idioms are used particularly in journalism "where writers frequently use them to bring their stories to life" (Longman, 1998: viii). Notes regarding the context and users of the idioms are also given by the LID, for example the expression *HOP/jump to it* is labelled as "often used by adults speaking to children". There is not a set list of such terms as employed by the dictionary – it is simply noted as and when an expression shows a clear tendency for use in a particular genre, or by a certain social group; they are not used systematically.

3.2.2. The Cambridge International Dictionary of Idioms and The Collins Cobuild Dictionary of Idioms

As mentioned in section 3.1, this research project took the data from the LID. However, the CIDI and the CCDI were also used for cross-referencing canonical forms and substitutes, definitions, reflecting the familiarity of variant combinations found and for the idiom activator sections.

The CIDI (1998) lists more than 7000 idioms, and again uses *idiom* as a blanket term. The entries each have examples taken from The Cambridge International Corpus (CIC), a 700million word monitor corpus of written and spoken British and American English, including written academic and business texts, and a 27million word Learner's written English section. The CIDI includes stereotypical idioms, idiomatic compounds, similes, exclamations and sayings, and clichés. The CIDI is similar to the LID in its variations included, and layout of entries. The CIDI does, however, include grammatical information

where relevant, such as for the phrase *a rich seam,* the additional note says "often + of" indicating that this phrase is often followed by a prepositional phrase (1998: xiii). The CIDI also notes where idioms share the same figurative keyword for example *have your knife into, put/stick the knife in, turn/twist the knife, a turn/twist of the knife* all use *knife* in connection with unpleasant behaviour (1998: xiv), and any commonly occurring idioms are highlighted. However, the dictionary does not mention different layouts for different frequencies of variant forms in the same way as the LID does.[4] For dialects of English, the dictionary uses British, American and Australian English. Register labels used are: informal, formal, very informal, old-fashioned, taboo, humorous and literary.

The CCDI uses the Bank of English monitor corpus, a monitor corpus of over 525million words of spoken and written British, American and Australian English, and includes over 8500 entries, comprising a range of formulaic sequence types under the general term *idiom.*[5] It does not include fixed expressions and formulae such as *at least* or *how do you do,* although it does include phrasal verbs. The CCDI includes notes as to where a phrase originated. For variations, it says that "[idioms] are difficult because they have unpredictable meanings of collocations and grammar" (2002: v), i.e. they are not fixed. In this dictionary, there is no _/_ demarcation used for one-word variations, separate lines are used for each variant form, the commonest one being listed first. Minor variations are mentioned in the sentence explanations for that idiom, e.g. "verbs such as "poke" and "shoot" can be used instead of "pick"" (for *pick holes in*). Variations are dealt with in a separate paragraph if they require special comment (2002: x).

The dialectal types used are again British, Australian and American English, and the register terms used are: journalism, literary, offensive, old-fashioned, rude, spoken, very rude, and written. These are more systematic than the other dictionaries, using four categories: Geographical identity (BrE/AmE etc.), Genre (for example journalism, novels etc.), Date and Currency (for example old-fashioned), and Formality (rude, offensive etc.).

---

[4] Although the notation of _/_ for one-word variations, and separate lines for variations of a larger effect are used, as well as grammatical variations being included as sub-entries.

[5] "The Bank of English Corpus is jointly owned by HarperCollins Publishers and the University of Birmingham. In 2005 the corpus [stood] at 450million words" (http://www.titania.bham.ac.uk).

The CCDI and the CIDI also have idiom activators. The CCDI calls its idiom activator a *thematic index*. It lists 32 themes, and lists idioms alphabetically within each theme. There is no subdivision of theme as is the case in the LID. The CCDI suggests that the index helps when looking for idioms on a particular theme, aids foreign students of English wishing to find an English idiom with meaning similar to one in their own language, and helps when an exact form of an idiom is not known, but the general meaning is. The CIDI has 15 *theme panels*, each subdivided in the same way as the Longman idioms activator. There is a very short explanation in the CIDI's introduction saying that idioms within the panels are grouped according to meaning or function. Each theme panel has a paragraph containing many of the idioms from within that theme panel, showing how they may be used in context, for example taken from "health":

> I'd been **feeling a bit off colour** for a while. I'd been more tired than usual and getting lots of headaches, and was generally a bit **below par**. It was worst in the morning. I'd get up **feeling like death warmed up**....
> (Cambridge, 1998: 444)

3.2.3. The British National Corpus

The formulaic sequences were selected from the LID and investigated using the BNC. It is important to note that the LID also relies on corpus data – it used the Longman Corpus Network to investigate the most frequently occurring formulaic sequences, alongside new phrases in the English language, and these were entered into the dictionary.

The Longman Corpus Network, owned by the Longman publishing company, is a compendium of corpora.[6] This corpus network provides the data for the LID:

Corpus	Mode	Number of Words (Million)	Comprising	Type
Longman Learner's Corpus	Written	10	Essays and exam scripts sent in by teachers, and students of the English language asked to donate their work	Monitor
Longman Written American Corpus	Written	100	Wide range of sources, e.g. newspapers, journals, novels, technical and scientific texts. Based on the written section of the BNC to allow for comparisons to be made between British and American written English	Monitor
Longman Spoken American Corpus	Spoken	5	Relatively new corpus. Over 1000 Americans of all ages groups, levels of education and ethnicity, from over 30 U.S. states were recorded	Static
Longman/ Lancaster Corpus	Written	30	Written texts from various locations, dates and text types within the English speaking world	Static
The British National Corpus: Written Component	Written	90	Written texts from various domains, representing a wide range of sources e.g. newspapers, journals, etc. from modern British English	Static
The British National Corpus: Spoken Component	Spoken	10	5 million words from natural spontaneous speech 5 million words from context-governed speech, e.g. pre-written scripts Also containing COLT – The Bergen Corpus of London Teenage Language	Static

Table 3-1 - Composition of the Longman Corpus Network

Whilst idiomatic language is highly frequent in texts, Degand and Bestgen (2003: 249) illustrate that searching for one individual idiomatic phrase does not yield substantial results; they suggest that searching for a stereotypical sequence such as *spill the beans* would occur only once per million words investigated. Using other means of text retrieval, such as corpora, enables more occurrences to be found. The BNC is one of the largest, most accessible corpora of modern British English. It also aims to be representative of British

[6] For more information about The Longman Corpus Network, see http://www.longman.com/dictionaries/corpus/index.html

English as a whole. Searching for the phrase *kick the bucket* in the Freiburg Lancaster/Oslo Bergen Corpus of British English (FLOB), a one million word collection of modern British English texts, provides no results.[7] Using the BNC, there are seven retrievable results. For the phrase *spill the beans,* again, there were no occurrences in the FLOB corpus, but 23 occurrences in the BNC. There are larger-still corpora than the BNC, such as the COBUILD Bank of English project, a 450million-plus word monitor corpus of British English, or the CIC, containing over 600million words, however, such collections are not as accessible for research as the BNC.

Another benefit in using the BNC for this research project, other than accessibility, size, and representivity, is that it forms part of the Longman Corpus Network, the data source for the LID. This enables comparisons to be made between the results the dictionary finds, such as possible variant forms, and the behaviour of such phrases, and the findings in this study. Ideally this research project would take the idioms from the LID and explore them in the Longman Corpus Network, however, this was inaccessible, and thus the BNC was used. At 100 million words, it is sufficient in size and notably larger than Moon's 18million word Oxford Hector Pilot Corpus (1998a), or those used by Barkema or Howarth (see section 2.4.1).

The BNC aims to represent contemporary British English, and attempts to do this in a balanced way. It is a mixed corpus, containing spoken and written texts. The LID sequences, and thus the idiomatic expressions under investigation here, are taken from both British and American English. Using the BNC as a corpus to study formulaic sequences illustrates how such phrases are used in British English only.

The 4124 texts which make up the BNC are of roughly equal length (between 40,000 and 50,000 words), and sampled to create a balanced view of English, as opposed to more domain-specific corpora. Almost 90 of the corpus' 100million words are taken from written sources, and the remainder from spoken domains. Each word in the BNC is tagged for "part of speech", a feature which was useful in the methodology (see section 3.3.1); for example,

[7] FLOB is based on the Brown corpus of standard written American English, a one million word amassment of texts taken from 15 different genres, aiming to be representative of the language. http://khnt.hit.uib.no/icame/manuals/flob/INDEX.HTM

distinguishing between *pin* as a noun and *pin* as a verb when investigating the case study *pin your HOPES/faith on* (see section 6.5).

3.2.3.1. The Written Component

In order to maximize variety of language styles available, the texts of the written component were selected according to three criteria: domain, time, and medium.

The domain is the genre in which the text was written. Less than 25% of the texts were taken from imaginative domains, whilst the rest could be labelled as "informative", such as texts from applied, social or pure sciences, or world affairs texts. As Aston and Burnard say (1998: 29), the target domains were based roughly on trends in book publishing in Britain. Table 3-2, taken from Aston and Burnard (1998: 29) shows the composition of the whole BNC (i.e. including the spoken part as a domain) by domain.

Domain	Texts	Percentage of Total Texts	Words	Percentage of Total Words
Imaginative	625	15.2	19664309	19.64
Natural and Pure Science	144	3.5	3752659	3.75
Applied Science	364	8.8	7369290	7.36
Social Science	510	12.4	13290441	13.28
World Affairs	453	11.0	16507399	16.49
Commerce and Finance	284	6.9	7118321	7.11
Arts	259	6.3	7253846	7.25
Belief and Thought	146	3.5	3053672	3.05
Leisure	374	9.1	9990080	9.98
Unclassified	50	1.2	1740527	1.74
Spoken	915	22.5	10365464	10.35

Table 3-2 - Domains of the BNC, the number and percentage of texts and words per domain

Time of production was also a criterion for texts. Over 80% of texts were dated between 1975 and 1993, almost 2% were taken from earlier than this range (the BNC also includes samples from well known classic texts), and around 17% were unclassified. The majority of the texts were contemporary at the time of the creation of the BNC.

The medium of the texts, the type of material in which the texts were published, is illustrated in Table 3-3 (Aston & Burnard, 1998: 30). *Miscellaneous published* includes brochures and pamphlets, and *miscellaneous unpublished* includes texts sampled from letters, memos, diaries, reports, and so on. Television scripts and such are classified under *to be spoken.*

Medium	Texts	Percentage of Total Texts	Words	Percentage of Total Words
Book	1488	36.1	52574506	52.52
Periodical	1167	28.3	27897931	27.87
Miscellaneous Published	181	4.4	3936637	3.93
Miscellaneous Unpublished	245	5.9	3595620	3.59
To be Spoken	49	1.2	1370870	1.37
Unclassified	79	1.9	364980	0.36
Spoken	915	22.2	10365464	10.35

Table 3-3 - Media of the BNC, the number and percentage of texts and words per medium

Information, such as the domain and medium of the text in which the formulaic sequence under investigation occurs, helps to create a picture of where the phrase is used. This can help in determining boundaries. It can illustrate whether particular variant combinations occur in different contexts.

3.2.3.2. The Spoken Component

The remaining 10million words of the BNC are spoken, and divided roughly equally into a *demographic* section, comprising informal conversations, and a *context-governed* section, containing more formal spoken interactions such as interviews, lectures, and debates. Spoken corpora tend to be smaller than written corpora, being more time-consuming, difficult, and expensive to compile. The spoken component of the British National Corpus is one of the largest spoken corpora of British English available.

As Aston and Burnard chart (1998: 32), the demographic section recorded the conversations and interactions of 124 volunteers, with approximately equal numbers of males and females, of five age groups, and four social classes. A sixth age group (16 and

under) is also included, and this forms The Bergen Corpus of London Teenage Language (COLT), currently the largest corpus of spoken teenage language.

The 762 texts which comprise the context-governed section are divided into four categories: *educational and informative* (144 texts), for example lectures, classroom interaction; *business* (136 texts), such as union talks, or business meetings; *institutional* (241 texts) for example sermons, political talks etc., and *leisure* (187 texts) including sports commentaries, television interaction and so on. The remaining 54 texts were unclassified.

Using corpora for research allows for an empirical view to be formed on language, and parts of that language – the BNC provides more naturally occurring data than could be collected manually. The corpus allows an in-depth quantitative study to be performed on formulaic sequences, thus giving a representative view of their frequency and behaviour.

3.3. Procedure for Preparing the Data

The aim of the project was to look at formulaic sequences in detail in order to examine the boundaries of variation. This section outlines how the sources described in section 3.2 are used to investigate the research aims.

3.3.1. Techniques for Grouping Sequences: The Chaining Process

Section 2.4 outlined previous research methods used for investigating the limits of variation (for example Barkema, 1996c; Moon, 1998a; Howarth, 1996), and illustrated places where these methodologies could be developed, such as collocation technique, size and representative-ness of corpora, and finding the actual full results of flexibility.

The dataset of sequences to be studied was taken from the LID and checked against the BNC.[8] The phrases were entered into the corpus to find frequency of occurrence, variant forms and occurrence details (i.e. domain in which they appeared, year of appearance etc.). To find the maximum variations possible for a sequence before it stops being fixed and

[8] The BNC is also far larger than any of the corpora used by Moon, Howarth or Barkema, and is more representative than the narrow discourse domains used by Howarth or Williams.

becomes context-dependent, a technique I have labelled as the chaining process was used. The process is typically applied as follows, taking the phrase *chance your ARM/luck* as an example.

> **Chance your arm/luck** *BrE*
> SPOKEN to try to do something that is new or involves a risk, even though you doubt you will succeed: *You won't get anything done in life if you don't chance your arm somehow | I didn't know how to spell it, but I chanced my luck and wrote something down.*
> (1998: 7)

The form *chance your arm* is the dictionary or canonical form of the phrase. The form *chance your luck* is the variant form of the sequence as given by the dictionary. When the canonical form is entered into the BNC, all occurrences, both literal and non-literal are recorded, as are the contextual details, e.g.:

i. Enter the canonical form and note all occurrences, both literal and figurative
ii. Repeat with all verbal inflections:

Chance your arm
Chances your arm
Chanced your arm
Chancing your arm

Steps i and ii are routine, involving the researcher entering the search enquiry, and recording the results. Ungrammatical examples, such as *chances your arm,* are included for comprehensiveness, and to include the possibility of colloquial dialects presenting such tokens. The dictionary variant form is then entered and the same details taken:

iii. Repeat for the dictionary variant form:

Chance your luck
Chances your luck
Chanced your luck
Chancing your luck

iv. Allow for determiner changes for both cases:

Chance _ arm	Chance _ luck
Chances _ arm	Chances _ luck
Chanced _ arm	Chanced _ luck
Chancing _ arm	Chancing _ luck

v. Repeat with spaces between verb and noun to allow for insertions. Start with one space between the verb and noun, (as seen in iv) and increase this number, taking into account the verbal inflections and noun plurals.

Chance _ _ arm Chance _ _ _ luck etc.

Steps iii, iv, and v are also routine, involving the researcher searching for the query terms in the BNC and recording the results. To find different variations in the noun, simply *chance your* is entered as a query term. The number of occurrences at each stage is noted, thus frequency of substitutes can be seen. Similarly, to find new verbal variants, *your arm/luck* is entered. Note that verbal inflectional forms, noun plural forms and determiner variants are accounted for.

vi. To find noun variants and plurals enter the verb with different determiners, without a determiner, and with different inflectional forms of the verb, e.g. simply *chance your*:

Chance my/your/his/her/its/their/our/null
Chances my/you/his/her/its/their/our/null
Chancing my/your/his/her/its/their/our/null
Chanced my/your/his/her/its/their/our/null

Noun substitutes found for such searches for *chance your ARM/luck* were: *arms, hand, hands*, and *needle*

vii. To find verb variants, enter the noun and noun plural, with different determiners, and with spaces to allow for insertions, e.g. simply *your arm*:

My/his/her/its/their/our/your/_ arm
My/his/her/its/their/our/your/_ luck
My _/his _/her _/its _/their _/our _/your _ arm
My _/his _/her _/its _/their _/our _/your _ luck
My _ _/his _ _/her _ _/its _ _/their _ _/our _ _/your _ _ arm
My _ _/his _ _/her _ _/its _ _/their _ _/our _ _/your _ _ luck etc.

Verbal substitutes found for *chance your ARM/luck* were: *press, push*, and *try*

Steps vi and vii involve the researcher making a judgement as to which variant forms are acceptable, in semantic meaning and in syntactic structure. Steps vi and vii rely on the researcher's intuition as to what is an acceptable form.

The noun and verb substitutes found are recorded when the variant sequence has the same conceptual meaning as the canonical; this is the semantic criterion of the chaining process. The syntactic criterion is that the V det N structure is maintained; this is a delimitation of the research project. In full, the chaining process would allow for alternative syntactical arrangements to be included. Previous collocational methodologies use the researcher's intuition to enter search sequences. The rules of the chaining process indicate the nouns and verbs of the sequence to be searched for, however, intuition is involved when determining whether a sequence has the same central meaning as the canonical form or not. The concepts of HAVING A GO and NOT NECESSARILY BEING SUCCESSFUL are the criteria for a form being included as having the same meaning for this sequence. Recording the literal and the non-literal occurrences of a sequence means that it is the lexical form of the sequence that is collected. This reduces the reliance on intuition to some extent.

Each new verb variant found is entered with the determiner, and all possible determiner substitutes, in the same way as *chance your*, in order to find new noun substitutes. Similarly, each new noun found is entered in the same way as *your arm/luck* to find new verbal substitutes. This process continues, like a loop, with all new noun substitutes, and

verb substitutes joined into the cycle until no new substitutes are found. All verbal inflected forms, noun plurals, determiner substitutes, and personal pronouns are taken into account. Literal and non-literal forms are recorded: it is the lexemes within or making up the formulaic sequences that are being focussed on, thus proper names have also been included. The limits of variation are found when the chaining process "runs out". The maximum number of substitutions represents the limits of substitution for a sequence. No new substitutes are found, so the number of substitutes found is the limit for that sequence. The frequency of substitutes also determines the limits of the sequence. Frequently occurring substitutes are *central* to the sequence, whereas substitutes which only occur once represent the outer limits of the substitution for the sequence. This means that rare occurrences which may be stylistic, or which may otherwise have been excluded using statistical significance techniques, are recorded for completeness. Excluding marginal variant forms at this stage in the analysis prevents full analysis of the putative boundaries at a later stage. The second criterion for reaching the limits of the chaining process is when the central meaning changes to the extent that variant forms cannot be said to share the same conceptual meaning as the canonical form, even if they share the same lexical items. In the case of *chance your ARM/luck,* the limits of the chaining process are reached when there are no new substitutes found with the same conceptual meaning. Occurrences such as *try your best,* or *earn your luck* did not have the same meaning. The occurrence *press + det + luck* occurred only once, and thus *press* represented the outer limits of the verbal substitution.

One problem with this technique occurred with reference to highly frequent lexemes such as *it,* or *have,* where entering these into the BNC yielded more results than the software allows to be viewed. In these cases, the maximum number of results (2000) that the BNC user software programme (SGML Aware Retrieval Application - SARA) shows was used.

The chaining process consists of a balance of routine application and researcher's judgement. Steps i to v are routine, utilizing the LID surface form of the sequence and its variants. Steps vi and vii require the researcher to make decisions as to what forms are relevant or not. In itself and in giving frequencies of substitutes, the chaining process allows maximum noun, verb and determiner substitutes to be seen, as well as allowing for insertions to be detected. This is a methodical approach to investigating the boundaries of

lexical substitutions, and as will be seen in Chapter 5, allows for central and peripheral members of the lexical set pertinent to the formulaic sequence to be seen. This is in contrast with previous works on variation (see section 2.3), which treat variant forms of statistical significance as being equal.

The chaining process builds on previous techniques (see section 2.4) by being more loop-like, and using new noun and verb substitutes to find new verb and noun substitutes respectively, all with the same underlying meaning until the process runs out. In order to find full actual variation, even those items with a very low frequency are counted. In comparison to previous techniques, the chaining process also permits the recording of contextual details for the investigation of genre and form, for use when distinguishing separate sequences. Use of the technique shows which phrases can be linked together through chaining, for example *chance your arm* chains to *push your luck,* which has its own dictionary entry, yet has the same underlying meaning. The chaining process can thus be used for investigating the boundaries of variation, and between sequences, via its recursive application and by recording frequencies.

The chaining process illustrates the construction of collocational clusters, chains or networks, which demonstrate how formulaic sequences can be grouped together (see section 2.4.5). Networks help to answer Barkema's questions regarding whether two expressions are variants of the same sequence, or whether they are two separate, but synonymous formulaic sequences. The networks presented by Williams and Howarth are located within specific discourse domains. Using the BNC and the LID for selection of sequences, this research is not particular to any genre or discourse domain. Aitchison suggests some rules regarding networks: networks must not be fixed, although some links must be durable. New links must be able to be added, and old ones changed as needed. The modules (networks) should not have rigid boundaries but should overlap adjoining modules (1994: 229). The networks facilitated by the chaining process have these characteristics.

Recording lexemes through the chaining process allows networks to emerge. Each node is a lexical noun or verb which link together to create the formulaic sequence. The basic phrase pattern is recognized when looking at the variant forms. It is important to note that for the purposes of this research project, the determiners are not illustrated; it is thought that the

verb and noun carry the lexical content of the formulaic sequence (see section 2.3.1). Words in the network theory (Aitchison, 1994) link words within a semantic class. In phrasal networks, such as Howarth's and the ones found in this research, the nodes are verbs and nouns linking to form a verbal idiom. In Barkema's research, adjectives and nouns link to form NP formulaic sequences. For Barkema's linear "domino" of *near miss* to *close shave,* the following network can be drawn:

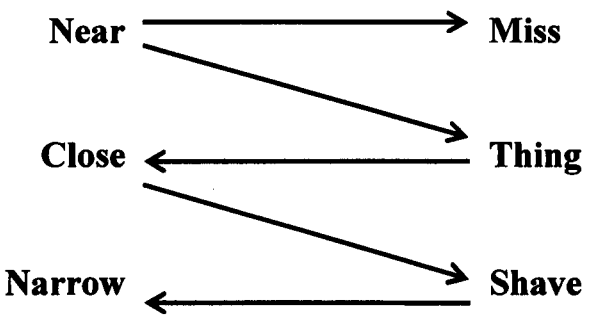

Figure 3-1 – Network for *near miss* to *close shave*

The arrows in Figure 3-1 illustrate the direction of the process. If the chaining process was applied in this case, *near miss* would be the first entry query. Then *near* would be entered to find any collocates in the Adj N structure with the same meaning. This results in the combination *near thing. Miss*, when entered in Barkema's corpus, does not result in any new adjectives in place of *near.* When entering the noun *thing* in the chaining process, the adjective *close* is found. The combination *close thing* has the same meaning as the canonical *near miss.* Entering *close* leads to *shave* which in turn leads to *narrow.* The direction of the diagram illustrates the direction of the chaining process. The canonical sequence occurs at the top and the chaining process results are illustrated progressing downwards through the diagram as new items are found. The direction of the arrows in the diagram illustrates which entry term led to the finding of which node. *Near* leads by collocation to *miss* and *thing. Thing* then leads by collocation and common conceptual meaning to *close* which in turn leads to *shave.* Finally *shave* leads to *narrow.*

In this research project, I have called these networks *basic networks.* These are the same as Moon's idiom schema (see section 2.4.4). They have the same underlying meaning, and a

shared lexical set and syntactic structure (in this case). So Moon's example of *shake in one's shoes* can be interpreted as:[9]

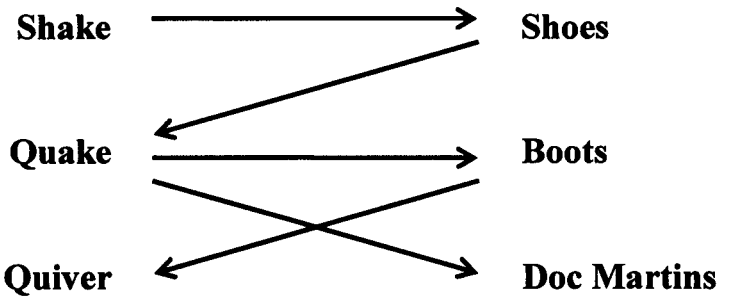

Figure 3-2 – Basic network for *shake in one's shoes*

A further illustration comes from Howarth's overlapping clusters (1996: 44):

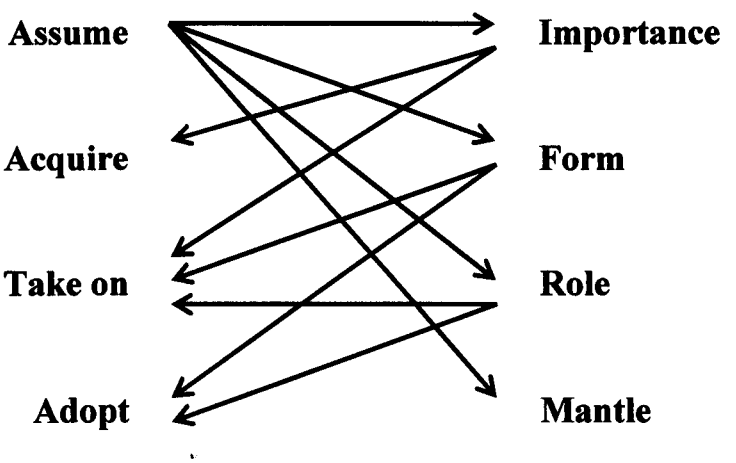

Figure 3-3 – Basic network for *assume importance*

In this example, throughout the chain, there is an expansion in meaning from the original *assume importance*, meaning to have an air of consequence or significance, to the forms *take on a role,* or *adopt a role*, being to accept a position. The same underlying meaning of "to take up" and "position" are present but the particular meaning changes. This will be

[9] The chaining process in this research is restricted to a V det N structure. Moon's example of *fan the flames* in section 2.4.3 would be an example of a full network, i.e. where the syntactic changes are recognized.

known as a *chaining network*, where there is the same underlying meaning and lexical set and structure, but the overlying specific meaning changes.

A further example (as will be seen in section 5.4) is for the sequence *throw a wobbly:*

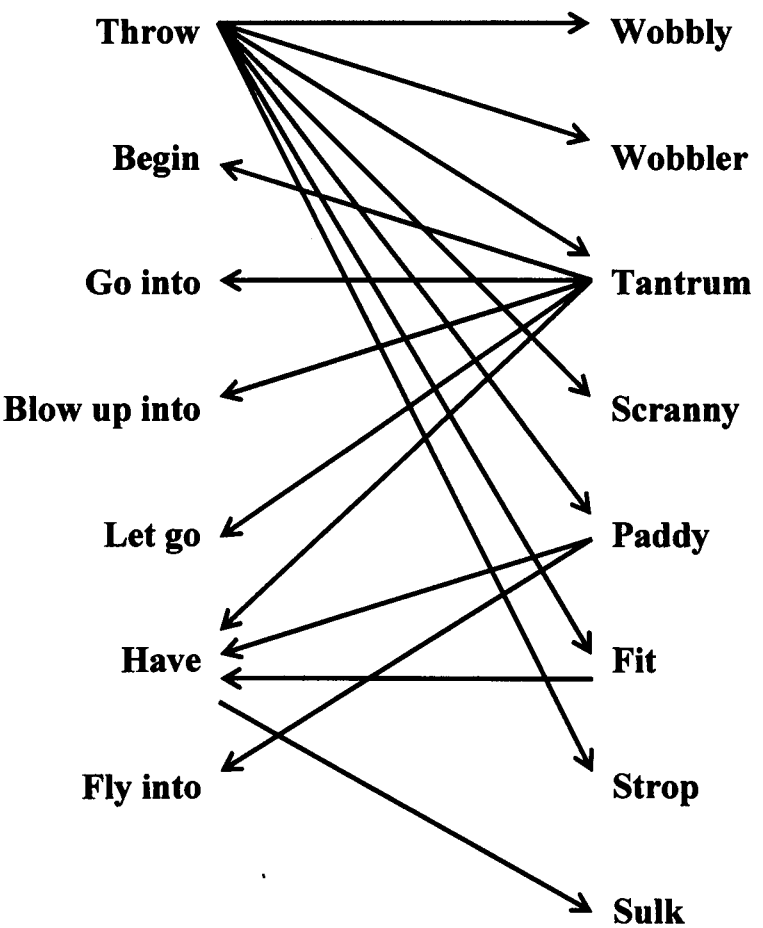

Figure 3-4 – Chaining network for *Throw a WOBBLER/wobbly*

The meaning remains the same of "getting very angry" but the overlying meaning changes from being frustrated very quickly in *throw a wobbly* to longer-term bad mood in *have a sulk.* The basic network is still seen underlying Figure 3-4. The top of the diagram represents the first entries in the chaining process: the canonical *throw* + *det* + *wobbly* and its variant *throw* + *det* + *wobbler.* The application of the chaining process progresses through the diagram, with *tantrum, scranny, paddy, fit*, and *strop* being found as variants for *wobbly.* Towards the end of the "list" of nouns on the right-hand side of the diagram,

the development of meaning is demonstrated. The frustration develops into anger in *fit,* which in turn develops into a long-term emotion in *strop* and *sulk.* In Figure 3-4, the nouns demonstrate the development of meaning.

Howarth says that "gaps" in the networks, for example **have a scranny,* provides an opportunity for looking at the historical development of the network. Overlaps can be seen as the accidental merging of independent collocations. There is no communicative need for the "gaps" to be filled (1996: 44). Each lexeme found using the chaining process, regardless of frequency, is a node in the network. The frequencies yielded in the methodology will be reported in Chapters 5 and 6, where it will also be seen that prototypes, as with schema, are compatible with networks. As Moon said (see section 2.4.4), all sequences form a network; the more fixed the sequence, the smaller the network. Such networks illustrate the limits of the number of substitutes permitted for a formulaic sequence. Such networks are more useful than discrete entries in idiom dictionaries for showing how language is organized, and how sequences can be grouped together.

3.3.2. Normalization

The different domains and different media of the BNC are not of equal size. This means that comparison of results is likely to be skewed. For example, a high number of occurrences in an imaginative domain give the impression that the particular phrase has a preference for occurrence in fictional texts. However, the imaginative domain is the largest in the BNC, so it would be expected that there would be more occurrences in that domain. Conversely, a phrase may only show a few occurrences in the belief and thought domain. However, this is the smallest domain, so a handful of occurrences in the smallest domain could result in a comparison with many occurrences in the largest domain. As it is important to take into account the sizes of the domains and media, a normalization technique has been used. For this research project, only the domains were used for contrasting or comparing different forms of a sequence. The results of the media of the occurrences studied were not used.

The technique used is to compare the domains against the smallest one in each case. Table 3-4 illustrates:

Domain	Texts	Words	Idiom Token	Normalized	Ratios of Domain to Smallest Domain[10]
Imaginative	625	19664309	6554769.7	55.32	11.30:1.00
Natural and Pure Science	144	3752659	1250886.3	66.79	2.16:1.00
Applied Science	364	7369290	2456430.0	85.97	4.23:1.00
Social Science	510	13290441	4430147.0	66.79	7.64:1.00
World Affairs	453	16507399	5502466.3	47.76	9.48:1.00
Commerce and Finance	284	7118321	2372773.7	69.44	4.09:1.00
Arts	259	7253846	2417948.7	62.15	4.17:1.00
Belief and Thought	146	3053672	1017890.7	83.22	1.75:1.00
Leisure	374	9990080	3330026.7	65.16	5.74:1.00
Unclassified	50	1740527	580175.7	50.00	1.00:1.00
Spoken	915	10365464	3455154.7	153.64	5.96:1.00

Table 3-4 - Domains of the BNC, the number of texts per domain, the number of words per domain, and normalizing information

The number of texts and words per domain are taken from the BNC. In terms of number of texts and word count, the imaginative and spoken domains are the largest, then social science, world affairs, and leisure. In the empirical chapters, the occurrences of formulaic sequences are normalized according to the domain word counts per division. As seen in section 2.2.1, the V det N structure of formulaic sequences is a common one, and the average word length of sequences is found to be 3.56 (Moon, 1998a: 78) and 2.7 (Erman & Warren, 2000: 51). Comparing the occurrences of a multi-word unit to the number of single words per domain etc. would give an erroneous result. Due to this, a notional *idiom token* count has been recorded, the notional idiom unit having a word length of approximately three. The idiom token count is the number of each domain word count divided by three to give a rough estimate and impression of the number of three-word units per domain. This allows for a more accurate normalization between the number of occurrences found in the data chapters and the domain sizes.

To normalize the data, the domains are compared to the smallest domain, which is the unclassified domain. The equation used is:

[10] The ratios are of the sizes, in idiom tokens, of the domains compared to the size of the smallest domain; unclassified. Thus the imaginative domain is 11.30 times larger than the unclassified domain.

Number of text occurrences x number of idiom tokens in smallest domain

divided by

number of idiom tokens in domain

So looking at the imaginative domain, there are 625 texts and 19664309 words. There are 6554769.7 idiom tokens. The smallest domain is the unclassified on, which has a total of 1740527 words and 580175.7 idiom tokens. The sum is thus:

(625 x 580175.7)/ 6554769.7 = 55.32

This means that in 580175.7 idiom tokens of the imaginative domain (i.e. comparable size to the smallest domain; the unclassified domain) there are 55.32 texts represented. So looking at the normalized column in Table 3-4, we can see that the imaginative domain is no longer one of the largest in the BNC. It seems that per 580175.7 idiom tokens, the spoken domain is the most represented, followed by applied science, with 85.97 texts per 580175.7 idiom tokens, and then the belief and thought domain. Belief and thought, if only the number of texts were looked at, is not one of the largest domains. In comparison to the smallest domain however, it is one of the most represented.

The ratios column in Table 3-4 also shows the ratios of the idiom token counts per domain to the smallest domain, unclassified. So the ratio of the imaginative domain to the unclassified domain is 6554769.7:580175.7, or 11.30 to one. The ratio can also be found by comparing the original number of texts per domain to the number of texts in the normalized domain, so for example for the imaginative domain, there are 625 texts, but 55.32 texts when the domain is normalized according to the smallest domain. The original number of texts is 11.30 times larger than the normalized number. These ratios will be important when comparing the case studies. For example in the case study *pin your HOPES/faith on* there are 23 occurrences of the phrase in an imaginative domain. However, to take into account the size of domains in the BNC, the ratio of the imaginative domain to the unclassified domain is used. So to work out the number of occurrences of *pin your HOPES/faith on* taking into account relative domain size, the sum is:

23/ 11.30 = 2.04

So there are 2.04 occurrences of *pin your HOPES/faith on* in an imaginative domain once the domain sizes are normalized according to the smallest domain.

This chapter has discussed the tools and techniques used to carry out the research aims. The next chapter illustrates how the data sources and methodology were used to select and prepare the data for investigation in the analysis chapters.

4. Selection and Preparation of the Research Data

4.1. Introduction

This chapter details the selection and preparation of the data used for the research exploration. The formulaic sequences from the LID and were organized in a database, from which a subset was chosen as a dataset. From the examination of this dataset, a selection of five formulaic sequences were chosen for further study with regard to lexical substitutions (Chapter 5) and five were chosen for insertions (Chapter 6).

Section 4.2 of this chapter outlines the creation of the database, sections 4.3.1 and 4.3.2 show the selection of the dataset from the database and its examination, and section 4.4 details the selection and preparation of the formulaic sequences from the dataset chosen for further investigation in the empirical chapters.

4.2. The Database

A database was created to contain all of the sequences from the LID which showed any kind of variation. The aim of the database was to represent all of the information about each sequence; in particular the variation details. There were 2210 entries in total. The use of a database means that a variety of smaller subsets for study can be selected quickly, easily and objectively.[1] This section explains the creation of the database; the layout and the fields it contains, and a detailed look at the variation fields; those which are the most important regarding the research questions.

4.2.1. The Fields

The database of the LID sequences showing any variation has 13 fields. These are shown in Table 4-1. The fields enable a logical layout and also facilitate the selection from the database of smaller subsets, such as written sequences, dialect, headword category etc. The examples of use of each sequence that the LID gives were not shown, as this information

[1] The database was designed for use by the author only for the present research project.

was not felt to be relevant to the purpose of the database: the selection of sequences for investigation of the research questions.

	Field	Description
1	ID	Numerical code for each entry
2	Headword	Headword of the sequence as shown in capitals in the LID
3	Headword Category	Lexical Category of the headword
4	Formulaic Sequence	LID sequence
5	Dialect	LID notation as to whether the sequence was mainly found in British English or American English
6	Variation 1	The first variation occurring in the sequence
7	Variation 2	The second variation occurring in the sequence, if any
8	Variation 3	The third variation occurring in the sequence, if any
9	Number 1	Number of variant forms of Variation 1
10	Number 2	Number of variant forms of Variation 2
11	Number 3	Number of variant forms of Variation 3
12	Medium	LID notation as to whether the sequence was mainly found in written or spoken texts
13	Notes	Any other notes the LID gives for the sequence

Table 4-1 – Fields of the database, and their description

Most of the headwords of the sequences entered into the database were nouns; 1921 of the 2210 entries were nouns. The Notes field contains notes from the dictionary regarding the context of use, or the register. It also contains notes such as whether the sequence had two noun phrases and variation occurs in one of them, the relevant noun phrase which had the substitution is recognized here, along with whether the sequence had more than one definition sense.

An example from the database is the sequence *keep/hold sth at BAY.* This was the first sequence entered into the database, so has an ID of 1. The layout of the fields in the database reads from left to right, fields 1 – 13. The headword is *bay,* which is a noun. There were no notes for dialect, so this field is empty. The first variation in the sequence, reading from left to right, is substitution of the verb, *keep* vs. *hold.* This is noted in the Variation 1 column. This sequence has a second variation, context-dependent substitution of the object *sth.* This is noted in the Variation 2 field. There are no more variations for this sequence as given in the LID, so the Variation 3 field is empty for this sequence. The variation of the verb has two forms, *keep* and *hold.* This number (2) is recorded in the Number 1 field; the

number of variants of variation 1. The Number 2 field refers to the number of forms variation 2 takes; in this case it is context-dependent, and thus potentially unlimited. This is marked by an asterisk. Variation 3 is empty for this sequence, and thus the Number 3 field is too. The dictionary does not suggest whether this sequence is of a particular medium, nor does it give any extra notes, so these fields are both empty. Section 4.2.2 gives more details regarding the variation and number fields. The database layout for *keep/hold sth at BAY* is given in Table 4-3.

4.2.2. Variation

Each sequence was "tagged" for variation category. The category of the variation stems from the type of variation as shown in the dictionary, for example in the LID entry *vanish/disappear into thin AIR,* there is lexical substitution of the verb. In the phrase *drop a BRICK/clanger,* there is lexical substitution of the noun. The phrase *FINDERS keepers losers weepers* shows phrasal deletion by being reduced to *FINDERS keepers,* and *get/come to grips with sth* shows two types of variation: lexical context-dependent substitution and lexical substitution of the verb. The LID does not label the type of variation, just lists what variation occurs. The database labels the category of variations shown in the dictionary. Each variation category in the LID was given a code, for ease when manipulating the database. These facilitated the selection of smaller subsets of the database. They were recorded in the Variation 1, 2, and 3 columns as necessary. Table 4-2 shows the different categories of variation possible, the corresponding code, and an example of each:

Variation Category	Code	Example
AdjP Adjective	AdjP2	Don't get CUTE/smart with me
AdjP Degree Modifier	AdjP1	Not too/very tightly WRAPPED
AdvP Adverb	AdvP2	Almost/nearly burst a BLOOD vessel
AdvP Degree Modifier	AdvP1	
Any Adjective or Adjective Phrase	Any Adj	Be in a ____FRAME of mind
Any Adverb or Adverb Phrase	Any Adv	
Any Noun or Noun Phrase	Any N	It's GOODBYE ____ hello ____
Any Preposition	Any Prep	
Any Verb or Verb Phrase	Any V	Sb couldn't ____ his/her way out of a wet PAPER bag
Conjunction (Coordinator)	Co1	Hit and/or miss
Conjunction (Subordinator)	Co2	When/if PUSH comes to shove
Context-dependent	C.D.	Put a DAMPER on something
Deletion (Lexical)	Del2	Home (away) from HOME
Deletion (Phrasal)	Del1	FOOLS rush in (where angels fear to tread)
Negation (Negative Particle vs. lexical item)	Neg	Don't/never look a gift HORSE in the mouth
NP Determiner	NP1	Die a/the DEATH
NP Modifier Adjective	NP2	Do a vanishing/disappearing ACT
NP Modifier Noun	NP3	Murphy's/Sod's LAW
NP Noun	NP4	Make like a BANANA/atom and split
Opposite Order	Opp2	POACHER turned gamekeeper/gamekeeper turned POACHER
Opposite Rearrangement	Opp1	Be like CHALK and cheese/as different as CHALK and cheese
PP Noun Phrase Determiner	PP2	In the/a BLINK (of an eye)
PP Noun Phrase Modifier Adjective	PP3	Be in the same/right BALLPARK
PP Noun Phrase Modifier Noun	PP4	
PP Noun Phrase Noun	PP5	In the BUFF/raw/nude
PP Preposition	PP1	In/on the front LINE
Syntactical variation	Syn	RING hollow/Have a hollow RING
VP Auxiliary	VP1	HEADS will/shall/must roll
VP Main verb	VP2	Jump/climb on the BANDWAGON
VP Phrasal Verb Particle	VP3	Come up/out smelling of ROSES

Table 4-2 - Variation categories, corresponding codes, and examples

Referring to the example of *keep/hold sth at BAY* as given in section 4.2.1, the verbal substitution listed in the Variation 1 field is recorded as VP2, substitution of a verb phrase main verb. The context-dependent substitution *sth* recorded in the Variation 2 field is noted as C.D.

Not every category listed in Table 4-2 was exemplified in the database. Such categories, e.g. a noun acting as a modifier to another noun within a prepositional phrase (PP4 in Table 4-2) have been included for completeness (also AdvP1, Any Adv, and Any Prep). These categories are also narrow – for example, a substitution of the head noun in a noun phrase (NP4) is categorized separately from the noun in a noun phrase of an overall prepositional phrase (PP5), even though they both represent a head noun variation.

The *any* categories, such as Any N, Any Adj etc. are context-dependent categories. However, they are regarded here as being more specific than the context-dependent (C.D.) category. Examples such as *a ___ too far* have a slot which can be filled by any noun or noun phrase. Examples such as *catch sb in the ACT* are more specific, referring to a noun that is an animate being. The noun *somebody* itself could also be a choice. Examples such as *sth knows no BOUNDS* could also have an embedded clause filling the context-dependent slot.

Of the 2210 formulaic sequences in the database, 593 showed two or more types of variation. For example, the phrase *keep/hold sth at BAY* shows lexical substitution of the verb (VP2), and context-dependent substitution of the object (C.D.). The formulaic sequences were analysed from left to right along the phrase, and categorized in up to three variation columns according to the number of types of variation. The maximum number of variations found was three, with additional dialect details being recorded in a separate field. Another example is the sequence *sb is/gets too big for their BOOTS/britches.* This has context-dependent substitution of the subject noun phrase *sb* so C.D. would be entered in Variation field 1. This has a potentially unlimited number of substitutes, so Number 1 field has an entry of *. The verbal substitutes are *is* and *get* so the Variation field 2 would be VP2 and the Number 2 field has a total of two possible substitutes. The third variation is variation of a main noun in a PP. The third variation is noted in Variation field 3 as PP5 and the Number 3 field also has a total of two possible substitutes. The sequence *sb is/gets too big for their BOOTS/britches* has a dialectal substitute; *boots* is the British English (BrE) version and *britches* the American. The British and American English variants are noted in the Dialect field, and in the Notes field, it is recorded that there are two NPs for this sequence, and the dialectal substitute takes place in the second NP. Having separate fields for each of the three variations possible enables all of the types of variation for a

phrase to be recorded; it does not emphasize one type of variation as being more important or prevalent than the other types. The Number fields allow the formulaic sequences with the most possible variant forms to be seen, thus the more fixed expressions can be distinguished from the more variable phrases. Table 4-3 demonstrates how the two sequences discussed; *keep/hold sth at BAY* and *sb is/gets too big for their BOOTS/britches,* are laid out in the database:

	Field	Example 1	Example 2
1	ID	1	101
2	Headword	Bay	Boots
3	Headword Category	Noun	Noun
4	Formulaic Sequence	Keep/ hold sth at bay	Sb is/ gets too big for their boots/ britches
5	Dialect		Boots = BrE, britches = AmE
6	Variation 1	VP2	C.D.
7	Variation 2	C.D.	VP2
8	Variation 3		PP5
9	Number 1	2	*
10	Number 2	*	2
11	Number 3		2
12	Medium		
13	Notes		C.D. = Noun Phrase1, PP5 = Noun Phrase2

Table 4-3 – Fields of the database with examples

4.3. The Dataset

A dataset was selected from the selection of all of the formulaic sequences from the LID which showed any type of variation for the investigation of lexical substitution and insertions.

4.3.1. Selection of the Candidate Dataset

The database shows all of the sequences in the LID which demonstrate variation. The frequencies of the variation categories in each of the variation columns of the database are shown in Table 4-4:

Variation Category	Variation 1	Variation 2	Variation 3	Total
AdjP Adjective	15	10	2	27
AdjP Degree Modifier	1	0	0	1
AdvP Adverb	8	8	1	17
AdvP Degree Modifier	0	0	0	0
Any Adjective or Adjective Phrase	15	0	0	15
Any Adverb or Adverb Phrase	0	0	0	0
Any Noun or Noun Phrase	91	8	0	99
Any Preposition	0	0	0	0
Any Verb or Verb Phrase	0	2	0	2
Conjunction (Coordinator)	1	0	0	1
Conjunction (Subordinator)	6	0	0	6
Context-dependent	941	208	32	1181
Deletion (Lexical)	28	25	8	61
Deletion (Phrasal)	43	87	19	149
Negation (Negative Particle vs. lexical item)	14	2	1	17
NP Determiner	18	42	5	65
NP Modifier Adjective	48	22	4	74
NP Modifier Noun	6	0	0	6
NP Noun	**126**	**58**	**9**	**193**
Opposite Arrangement	10	4	1	15
Opposite Meaning Lexical	1	32	24	57
PP Noun Phrase Determiner	10	33	4	47
PP Noun Phrase Modifier Adjective	10	7	1	18
PP Noun Phrase Modifier Noun	0	0	0	0
PP Noun Phrase Noun	**72**	**56**	**13**	**141**
PP Preposition	71	37	9	117
Syntactical variation	89	78	27	194
VP Auxiliary	27	5	1	33
VP Main verb	553	33	4	590
VP Phrasal Verb Particle	7	2	2	11
TOTAL =				**3137**

Table 4-4 - Distribution of the different variation categories in the database

	Type of Variation	Components	Total
Lexical Substitution	Adjective	AdjP2 + NP2 + PP3	119
	Adverb	AdvP2	17
	Noun acting as a modifier	NP3 + PP4	6
	Preposition	PP1	117
	Noun	**NP4 + PP5**	**334**
	Verb	VP2	590
	Context-dependent	Any Adj + Any Adv + Any N + Any Prep + Any V + C.D.	1297
Grammatical Variation	Conjunction	Co1+ Co2	7
	Degree modifier	AdjP1 + AdvP1	1
	Negation	Neg	17
	Determiner	NP1 + PP2	112
	Syntactical variation	Str	194
	Auxiliary Verb	VP1	33
	Verb Particle	VP3	11
	Opposites	Opp1 + Opp2	72
Deletion	Deletion	Del1 + Del2	210
TOTAL =			3137

Table 4-5 – Summarising the variation category frequencies into the variation types

The total of all of the variations in Table 4-4 and Table 4-5 is 3137; more than the 2210 sequences. This is because of the 2210 sequences, 1450 had only one variation (1450 variations in total). 593 sequences had two variations (593*2 = 1186 variations), and 167 had three variations (167*3 = 501 variations). Thus the 2210 sequences showing variation showed a total of 3137 variations.

Table 4-4 and Table 4-5 show that lexical substitution is the most common variation type, with context-dependent substitution being the most frequent category, followed by substitution of the main verb in a verb phrase, and of the head noun in a noun phrase. Section 4.2 showed that the headword of formulaic sequences in the database was most frequently the noun. Noun substitutes often carry more lexical content and are more disruptive than verb substitutes (Moon, 1998a: 127). Nouns have importance as headwords, so the phrases demonstrating only one type of variation; lexical substitution of the noun, were taken as a subgroup of the database. The subset is thus NP4 + PP5, i.e. 334 sequences (these are emboldened in Table 4-4 and Table 4-5).

Table 4-4 shows the frequencies in the variation columns. However, the frequencies shown are not those of the variation category alone. For example, the frequency in variation column 1 for NP4 is 126. This means that 126 of the 2210 sequences had NP4 in the variation column 1 field. They could have more than one variation. Similarly, there are 58 sequences which have NP4 in the variation column 2. These could have a different variation category in variation column 1 and could also have a third variant. To look at the figures for the lexical substitution of the noun only, the database must be manipulated so that the variation columns 2 and 3 are empty, and the variation column 1 shows either NP4 or PP5. These frequencies are shown in Table 1, Appendix.

There were 134 phrases that had noun substitution.[2] Of the 134 phrases, there were 77 that had the noun in a noun phrase (NP4). The other 57 were taken from the noun phrase of a prepositional phrase (PP5). The research is focused on the 77 phraseological units showing variation of the noun in a noun phrase (NP4) as the larger of the two groups.

This group of sequences was parsed for syntactic structure, as shown in Table 4-6:

[2] Note that some substitutions such as *not have two CENTS/pennies to rub together* has a lexical substitution which is also dialectal.

Phrase Structure	Frequency
VP NP	27
NP	11
VP NP *	9
VP NP PP	7
NP PP	6
NP VP NP	3
NP VP NP PP	2
AdvP NP VP NP	1
Neg AdjP NP	1
NP AdvP	1
NP PP NP PP	1
NP VP Conj NP VP	1
NP VP NP Conj NP VP NP	1
NP VP NP VP	1
NP VP NP VP NP	1
NP VP PP	1
VP AdjP PP	1
VP NP VP AdvP	1
VP NP VP NP	1
TOTAL =	77

Table 4-6 – Phrase structure frequencies

The syntactic parsing in Table 4-6 was carried out using the Penn Treebank (Brill, 1995) and The Linguist's Search Engine (Resnik and Elkiss, 2003) with the help of Dr. Paul Clough from the Department of Information Studies at the University of Sheffield.[3]

As can be seen, the most common structure is VP NP, with a frequency of 27, and is specifically V det N.[4] The structure VP NP* (nine occurrences) refers to those idiomatic expressions which have the VP NP (V det N) structure, but which have some modification to it, for example negation in the verb phrase, a modifier in the noun phrase, or the possibility of a PP following it. The sequence *stem the TIDE/flow/swell of* has a VP NP structure, but the noun has a possible context-dependent post-modifying prepositional phrase. The sequences *pin your HOPES/faith on* and *set the STAGE/scene for* also have a possible context-dependent post-modifying prepositional phrase. Not all occurrences of these phrases require the post-modifying prepositional phrase; some occurrences are simply

[3] The standard tags used in the Penn Treebank can be found at http://www.cis.upenn.edu/~treebank/home.html

[4] The phrases *do BIRD/time, raise HELL/Cain* and *tempt FATE/providence* have a V det N structure, however the determiner is null.

V det N in structure (as opposed to V det N Prep), for example *stem the TIDE/flow/swell.* The sequences *stem the TIDE/flow/swell of* and *pin your HOPES/faith on,* for example, have an open-slot for the noun phrase following the preposition. In a case such as *play both ends/sides against the middle,* the syntactic structure is VP NP PP and the noun phrase within the prepositional phrase is specified, unlike in *pin your HOPES/faith on.* Section 2.2.2 showed that a common syntactic structure of formulaic sequences is the verb + predicate one, which is supported here in Table 4-6. Section 2.3.3 discussed insertions and outlined that for this research project, anything beyond the V det N structure is an addition. For this reason, the nine occurrences of VP NP* are grouped together with the 27 VP NP cases to form the dataset.

An insertion is material added within the V det N structure. Material within the phrase which interrupts the syntactic structure, such as adjuncts or interjections, is termed *interruption.* Interruptions and additions are not looked at within this project.

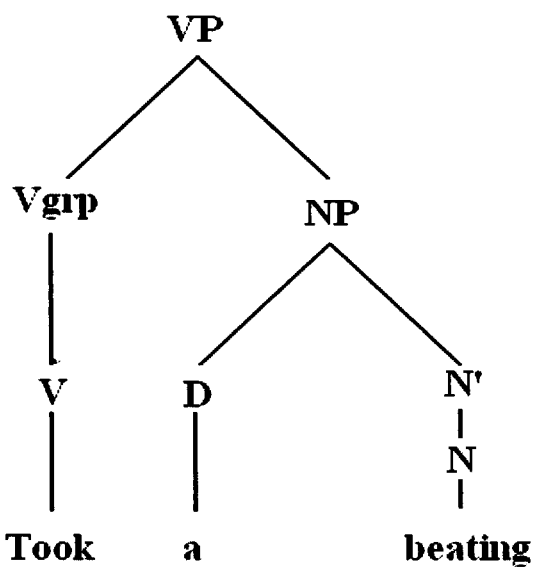

Figure 4-1 - Syntactic tree for the research project

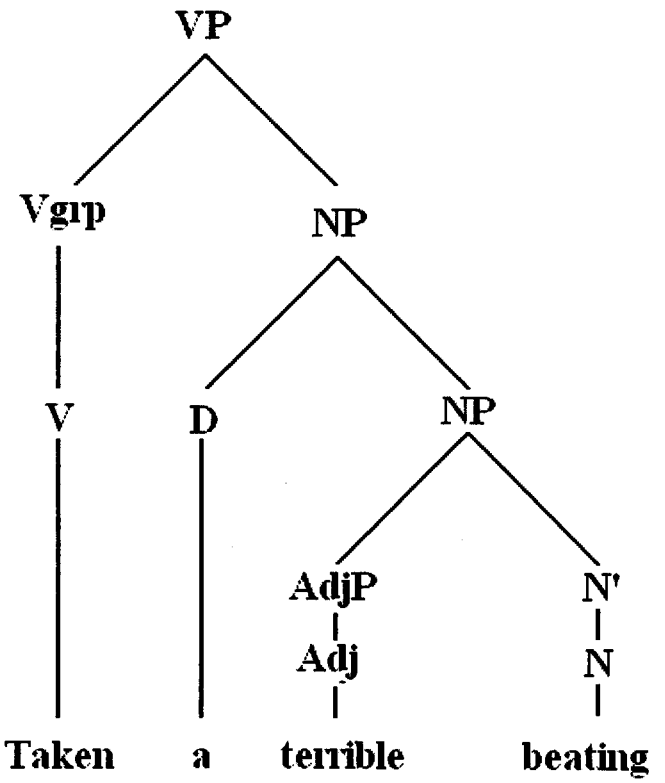

Figure 4-2 - Syntactic tree for the research project demonstrating insertion

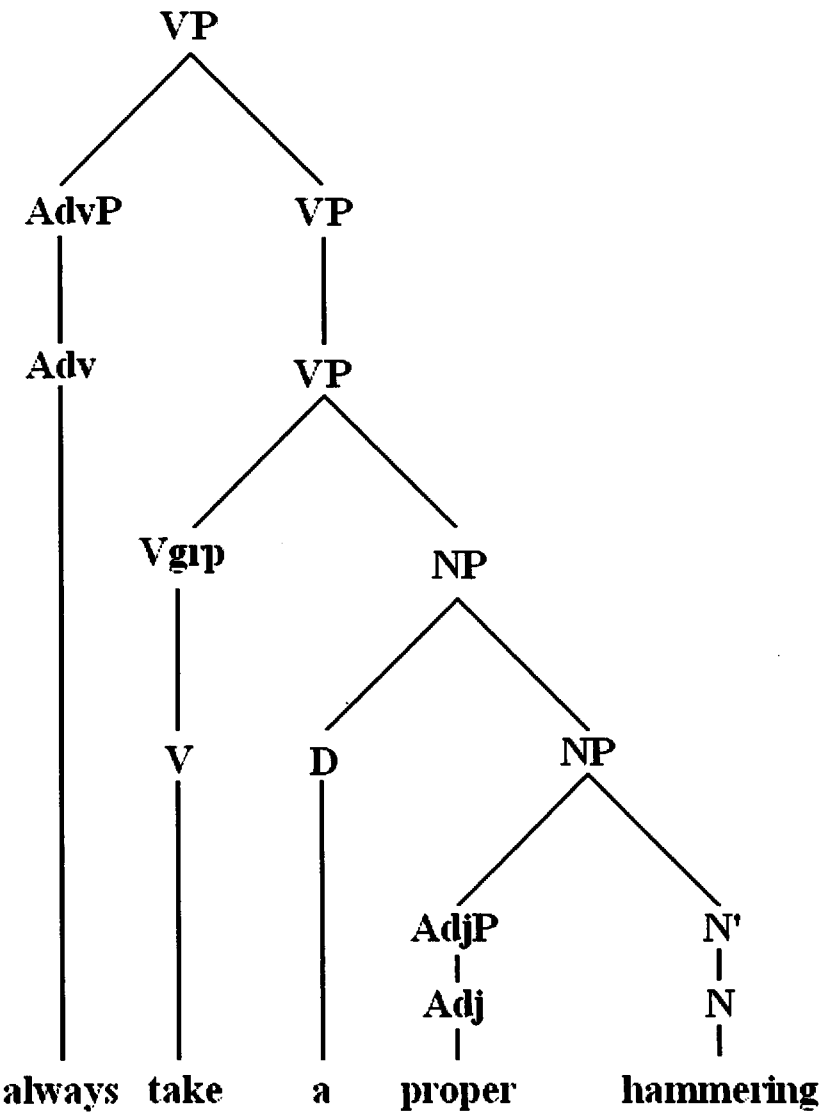

Figure 4-3 - Syntactic tree for the research project demonstrating addition and insertion

Figure 4-1 shows the basic syntactic structure for the phrases in the dataset. In Figure 4-2, the adjectival insertion *terrible* fits into the NP alongside the determiner and the noun and into the overall V det N structure of the phrase. In Figure 4-3, the adverbial phrase *always* is an adjunct, and is a syntactically separate constituent from the formulaic sequence. It is a sister of the sequence, syntactically, and is outside the V det N structure, thus it is an addition.

In the terminology used in this research, the label *insertions* is used generically for the overall kind of formulaic sequence modification (i.e. additional material to the phrase), and

it is also used in a specific sense for material inserted into the phrase.[5] As will be seen in Chapter 6, most insertions were lexical (adjectives/ adverbs) or of a grammatical (pre-determiner) nature. There were only two occurrences of an interruption.[6]

This study focuses on the largest group of sequences with substitution of the noun (NP4) only – those with the VP NP (27) structure, and with the same structure plus some modification (VP NP*; 9). The structure V det N forms the outer limits of the formulaic sequence. Only those with a noun as a headword were chosen, omitting only *BLOW it/that/him.* This left a group of 35 formulaic sequences which forms the dataset for the investigation of lexical substitution and insertions.

4.3.2. Preparation of the Dataset

The LID was used to select the 2210 sequences showing some form of variation, and these were entered into a database. From this, a dataset of 35 sequences showing substitution of the noun only, and with a structure of VP NP were selected, each having a noun as the headword. These were then searched for in the BNC using the chaining process (see section 3.3.1). This technique allowed the frequency of occurrences to be determined, the different lexical substitutes, both noun and verbal, to be seen alongside frequency, the insertions found, and also the context of occurrence (domain).

Each occurrence of a phrase is known as a *token.* Although the term token traditionally represents a singular word, in this case, looking at phrases, each phrasal unit represents a token. Each different form of the phrase is known as a *type.* Differences in inflections,

---

[5] Some authors use insertions as a defining property of idioms; a pure idiom would not allow for an insertion. However, formulaic sequences with a compositional reading do allow for them.

[6] Only two occurrences of interruptions were found in the whole dataset, and these occurred in the phrases *pin your HOPES/faith on* and *set the STAGE/scene for* between the noun and preposition: *pinned her hopes, **rather desperately, she owned,** on* and *set the scene, **as it were,** for.* Note that Barkema (1996a) found only 1% of 1664 formulaic sequences he studied showed interruptions. There were no examples found of stylistic interruptions inside the V det N structure in this study. Section 6.5 shows that there are 72 occurrences of *pin + det + hopes + on,* and 74 occurrences of *put + det + faith + in.* Of the former, there are 55 tokens where the preposition directly follows the noun, and there are 68 tokens of the latter with a directly following preposition. This shows that there is a high tendency for the preposition to directly follow the noun, and thus insertions between the noun and preposition are rare. The interruption here is adverbial, and there is a reported speech construction "she owned". By the phrase *as it were,* the speaker is putting an emphasis on his use of the phrase *set the SCENE for* both for highlighting the phrase itself and also the use of metaphor.

tense, substitutes, determiners, and insertions yield different types. For this research project, the different lexical noun and verb combinations are known as *lexical combinations.* This means the noun (including all possible inflections) is combined with the verb (with all tense and person markings etc.) So, for example:

Take a beating, take a beating, took a beating, took a beating, took a beating

= five tokens, two types, and one lexical combination with a frequency of five

Take a beating, takes a beating

= two tokens, two types, and one lexical combination, with a frequency of two

Flip your wig, flip your wig, flip their lid, flip their tiny lids, flipped his lid

= five tokens, four types, and two lexical combinations; *flip* + *det* + *wig,* and *flip* + *det* + *lid*

Table 4-7 lists the formulaic sequences in the dataset, showing the headword and the substitutions acknowledged by the dictionary under the column heading *formulaic sequence*. The total tokens and types are shown, as is the number of lexical combinations found for each sequence. The frequency of the LID canonical form is also given. The sequences are listed in alphabetical order according to the headword of the sequence.

Formulaic Sequence	Token	Type	LID Canonical Form	Total Lexical Combinations
Chance your ARM/luck	255	70	16	8
Cover your ASS/butt	83	39	8	10
Kiss my ASS/butt	11	7	9	3
Be left holding the BABY/bag	66	12	9	2
Take a BEATING/hammering	112	90	17	31
Do BIRD/time	103	38	52	4
Miss the BOAT/bus	42	8	27	2
Have a BRAINWAVE/brainstorm	33	21	13	8
Be a BREEZE/doddle	119	43	13	9
Shoot the BREEZE/bull	5	3	4	2
Drop a BRICK/clanger	21	18	3	4
Burn your BRIDGES/boats	23	15	3	3
Cut the CRAP/shit	20	7	18	3
Don't pull that CRAP/shit	0	0	0	0
Hit the DECK/dirt	454	34	18	13
Shut your FACE/trap/cakehole/gob	128	39	7	16
Tempt FATE/providence	40	13	26	6
Drag your FEET/heels	78	28	60	3
Raise HELL/Cain	427	128	16	16
Pin your HOPES/faith on	215	140	72	8
Blow your own HORN/trumpet	29	18	0	3
Flip your LID/wig	400	91	3	13
Overstep the MARK/limits/bounds	289	145	27	17
Be a sore POINT/spot	67	46	13	15
Pack a PUNCH/wallop	35	29	30	4
Hit the SACK/hay	8	5	7	2
Keep your SHIRT/pants/hair on	18	1	0	1
Call the SHOTS/tune	98	13	49	3
Set the STAGE/scene for	321	35	37	4
Kick up a STINK/fuss	379	171	6	14
Be just the TICKET/thing/job	160	55	14	15
Stem the TIDE/flow/swell of	176	46	40	19
Watch your TONGUE/mouth	516	79	7	15
Throw a WOBBLY/wobbler	150	50	4	16
Not mince your WORDS/matters	37	26	36	2

Table 4-7 - Token, type, and lexical combinations for sequences in the dataset

In Table 4-7, the lexical combinations column represents the number of types disregarding any determiner substitutes, tense markers or inflections. A high number of lexical combinations indicates a high number of substitutions; *take a BEATING/hammering* has only 112 tokens, yet 31 combinations, as opposed to *hit the DECK/dirt,* which has 454 tokens, yet only 13 combinations. A high type count may indicate a high combination count, but may also indicate a high degree of tense and inflections involved, as well as determiner substitutes. The phrase *watch your TONGUE/mouth,* however, has a low type count in comparison to its token frequency. This phrase has a high degree of fixedness; of the 516 tokens, 334 take the lexical combination *bite + det + lip,* and of these 204 take the form *bit her lip.* There is less inflectional variation for this phrase, resulting in a lower type frequency. A high token frequency, and a low type and combination frequency is an indicator of fixedness, for example *be left holding the BABY/bag.* Table 4-7 also shows the frequency of the LID canonical lexical form, e.g. the sequence with the headword as the noun. In most cases, this number is low compared to the token frequency, indicating that variant forms are more common than the form as listed by the LID, in particular, for example *blow your own HORN/trumpet;* only examples with the noun *trumpet* were found in the BNC. Exceptions are, for example, *cut the CRAP/shit* and *not mince your WORDS/matters,* which had a high occurrence in the canonical lexical form compared to the number of tokens.

Table 4-8 shows the number of noun substitutes and verbal substitutes found for the dataset sequences, and also sums these to give the total number of lexical substitutes:

Formulaic Sequence	Lexical Substitutes (N)	Lexical Substitutes (V)	Total Substitutes Involved
Chance your ARM/luck	4	4	8
Cover your ASS/butt	4	4	8
Kiss my ASS/butt	3	1	4
Be left holding the BABY/bag	2	1	3
Take a BEATING/hammering	8	16	24
Do BIRD/time	2	3	5
Miss the BOAT/bus	2	1	3
Have a BRAINWAVE/brainstorm	2	8	10
Be a BREEZE/doddle	5	5	10
Shoot the BREEZE/bull	2	1	3
Drop a BRICK/clanger	4	1	5
Burn your BRIDGES/boats	3	1	4
Cut the CRAP/shit	3	2	3
Don't pull that CRAP/shit	0	0	0
Hit the DECK/dirt	6	3	9
Shut your FACE/trap/cakehole/gob	12	4	16
Tempt FATE/providence	3	4	7
Drag your FEET/heels	3	1	4
Raise HELL/Cain	5	9	14
Pin your HOPES/faith on	3	4	7
Blow your own HORN/trumpet	3	1	4
Flip your LID/wig	11	3	14
Overstep the MARK/limits/bounds	6	6	12
Be a sore POINT/spot	4	9	13
Pack a PUNCH/wallop	2	3	5
Hit the SACK/hay	2	2	4
Keep your SHIRT/pants/hair on	1	1	2
Call the SHOTS/tune	3	1	4
Set the STAGE/scene for	4	2	6
Kick up a STINK/fuss	6	8	14
Be just the TICKET/thing/job	5	8	13
Stem the TIDE/flow/swell of	6	10	16
Watch your TONGUE/mouth	8	8	16
Throw a WOBBLY/wobbler	8	7	15
Not mince your WORDS/matters	1	3	4

Table 4-8 - Frequency of noun and verbal substitutes for sequences in the dataset

Table 4-8 shows that the phrase *take a BEATING/hammering* had the most lexical substitutes overall, and the phrase *shut your FACE/gob/trap/cakehole/mouth* had the most noun substitutes. The phrase *keep your SHIRT/hair/pants on* was much more fixed, having only one possible noun (*hair*) and one possible verb (*keep*). The phrase *don't pull that CRAP/shit* showed no occurrences in the BNC.[7]

The lexical substitutions column tallies individual lexical items. Thus all inflected forms of a noun or verb are taken as being one lexical form. These columns include the canonical forms, so for example for the phrase *keep your SHIRT/hair/pants on* there were only occurrences of *keep your hair on* found, with inflected forms of the verb. This then has one noun involved and one verb involved, and two lexical items involved altogether. The syntactic structure of Verb det Noun was maintained throughout.

Section 2.2.2 discussed the notion of weak verbs, and suggested that these delexicalized verbs occur more frequently than lexical verbs which are rarer, and less open to change. The dataset frequencies support this, in phrases such as *take a BEATING/hammering* and *have a BRAINWAVE/brainstorm.* Sequences such as *drag your FEET/heels* and *pack a PUNCH/wallop* support the notion that lexical verbs are open to less substitution. Even in the phrase *stem the TIDE/flow/swell of* which demonstrates a high degree of verbal substitution, there are many more occurrences of the lexeme *stem* than any other verb substitute, indicating a fair amount of preference and stability for this verb.

Table 4-7 and Table 4-8 can thus be read as follows: for the phrase *chance your ARM/luck,* there were 255 tokens, i.e. total occurrences of the phrase including the canonical form, and all possible variant forms, and inflected forms. Only 16 of these were in the lexical form *chance + det + arm.* There were 70 different types (combinations of the lexemes with inflections and tense etc.), and there were 8 different lexical combinations. There were four noun lexemes involved: *arm, hand, needle*, and *luck,* and four verb lexemes: *chance, try, press*, and *push,* leading to a total lexeme involvement of eight.

[7] This could be because it is a sequence in the LID which occurs in the American portion of the Longman Corpus Network.

Table 4-9 shows the dataset with the frequencies of the other type of variation studied; insertions. Insertions within the V det N structure can be either grammatical, such as quantifiers etc. or lexical, such as adjectives. Table 4-9 shows the number of lexical and grammatical insertions found for each sequence, and presents the total number of insertions.

Formulaic Sequence	Insertions Lexical	Insertions Grammatical	Total Insertions
Chance your ARM/luck	1	0	1
Cover your ASS/butt	2	0	2
Kiss my ASS/butt	1	0	1
Be left holding the BABY/bag	2	0	2
Take a BEATING/hammering	37	14	51
Do BIRD/time	0	0	0
Miss the BOAT/bus	0	0	0
Have a BRAINWAVE/brainstorm	5	1	6
Be a BREEZE/doddle	10	5	12
Shoot the BREEZE/bull	0	0	0
Drop a BRICK/clanger	6	0	6
Burn your BRIDGES/boats	0	0	0
Cut the CRAP/shit	2	0	2
Don't pull that CRAP/shit	0	0	0
Hit the DECK/dirt	1	0	1
Shut your FACE/trap/cakehole/gob	13	0	13
Tempt FATE/providence	1	0	1
Drag your FEET/heels	1	0	1
Raise HELL/Cain	72	22	94
Pin your HOPES/faith on	43	24	67
Blow your own HORN/trumpet	0	0	0
Flip your LID/wig	4	1	5
Overstep the MARK/limits/bounds	99	0	99
Be a sore POINT/spot	17	0	17
Pack a PUNCH/wallop	15	5	20
Hit the SACK/hay	0	0	0
Keep your SHIRT/pants/hair on	0	0	0
Call the SHOTS/tune	2	2	4
Set the STAGE/scene for	18	0	18
Kick up a STINK/fuss	76	90	166
Be just the TICKET/thing/job	6	0	6
Stem the TIDE/flow/swell of	20	0	20
Watch your TONGUE/mouth	0	0	0
Throw a WOBBLY/wobbler	19	4	23
Not mince your WORDS/matters	0	0	0

Table 4-9 – Frequency of lexical and grammatical insertions for sequences in the dataset

Not all of the phrases allowed for insertions within the phrase. None of the phrases has a common inserted word mentioned as a note in the LID. Each insertion occurrence is counted separately, even though all occurrences involving an insertion may be the same, for example there were three occurrences of *wreak **such** havoc* in this exact form. They count as three tokens, one type, one lexical combination, and three occurrences of insertion. A null determiner is counted as being a grammatical variant.

It can be seen from Table 4-8 and Table 4-9 that the different formulaic sequences of the dataset accommodate different amounts of lexical substitutions and insertions. Phrases such as *keep your SHIRT/pants/hair on* and *drag your FEET/heels* are much more fixed than such phrases as *take a BEATING/hammering* and *throw a WOBBLY/wobbler*. Note also that there is no correlation between the number of tokens and the number of substitutes or insertions. The phrase *hit the DECK/dirt* has 454 tokens, yet only 34 types and nine total lexical substitutes. It also has only one token showing an insertion. The phrase *take a BEATING/hammering,* by comparison, has 112 tokens, and 90 types, with 24 different lexical substitutes, and 51 total insertions.

The fact that different phrases allow differing amounts of manipulation suggests that the formulaic sequences in the dataset are not of the same classificatory status. The LID classifies all phrases as being *idioms.* The accepted definition of a stereotypical idiom suggests that they are lexically fixed. Table 4-8 and Table 4-9 show that for many phrases in the dataset, and thus by suggestion in the whole database, there is more variation and less fixedness than the definition of a pure idiom allows.

It seems that the phrases which allow a great deal of lexical substitutions and insertions are heading towards the *free phrase* end of the spectrum, whereas the more fixed ones are more like pure idioms. For example, *take a BEATING/hammering* allows for a great deal of variation, all having the same meaning as the canonical form – for example *take a hammering, got a hammering, taken a bit of a bashing, faced another battering, received their biggest thrashing, take a licking* etc. However, *set the STAGE/scene for* is one of the formulaic sequences of the dataset with the most tokens, yet does not allow for such

variation. It is much more fixed, allowing only for the structures *set/set up the scene/stage/mood/tone.* Other such fixed phrases include *miss the BOAT/bus, hit the SACK/hay,* and *tempt FATE/providence.* Whilst the LID gives all of its entries the same status, it is clearly the case that they are not all idioms, and not all of the same classificatory status.

Table 4-8 shows the number of lexemes found to be involved per phrase using the chaining process in the BNC. The formulaic sequences listed are as found in the LID; they show the lexical substitutions the dictionary suggests. This is also represented in Table 2, Appendix. Many of the phrases in the dataset (29) were found to allow for more lexical involvement than shown in the dictionary. The phrases *keep your SHIRT/pants/hair on* and *cut the CRAP/shit* (no BNC occurrences) were found to have less lexical activity than that shown in the dictionary. There was only one example of *wallop* for *pack a PUNCH/wallop.* There were no occurrences of *matters* for *not mince your WORDS/matters.* An aspect of the chaining process is to record the frequency of occurrence of the variant forms, and this means that frequent and rare lexical substitutes are emphasized. So for example, whilst the phrase *stem the TIDE/flow/swell of* had ten different verbs found in the BNC, the lexeme *stem* was much preferred, with 91 of the 176 tokens taking this verb. In Table 4-8 a lexeme was included as being a possible noun or verb substitute even if it was only found in one token. Such rare lexemes were accounted for since they did have an occurrence as a possible substitution, even if they were somewhat tenuous members of the phrase's lexical set.

4.4. Cases for Further Investigation

So far, the database containing all of the sequences showing variation in the LID has been described, as has the selection from it of a dataset of 35 formulaic sequences showing lexical substitution of the noun only. In previous studies (Moon, 1998a) the most common approach has been to investigate a wide selection of formulaic sequences in a general survey. Other studies (Gibbs, 1995; Philip, 2000) investigated one sequence (*spill the beans* in the former, *like a red rag to a bull* in the latter). This research project combines the two approaches. I investigate in detail five formulaic sequences that demonstrate the highest

degree of lexical substitution of the noun and five that show the most insertions.[8] These are presented in chapters 5 and 6.

4.4.1. The Dataset and Lexical Substitutions

A large proportion of phrases in the dataset show substitution of the noun: 32 of the 35 sequences showed substitution of the noun. There were 25 phrases showing insertions. Eight of the sequences showing insertion had only one or two tokens with an insertion. Whereas the insertions had a tendency to occur either in large or small numbers, for example *kick up a STINK/fuss* having 166 tokens with insertions compared to *chance your ARM/luck* having only one insertion token, the substitutions of the noun had a more gradual scale, with those as exemplified in the case studies having eight to 12 substitutes, and grading down to the least substitutions; *keep your SHIRT/hair/pants on,* and *not mince your WORDS/matters.*

4.4.2. The Dataset and Insertions

Ten of the dataset phrases showed no insertions. Of the phrases showing insertions, there were two types of insertions possible. Firstly, there were insertions of an adjectival or adverbial modifier kind, i.e. lexical, for example *dropped a* **massive** *clanger* for the phrase *drop a BRICK/clanger,* or *stem the* **growing** *tide* for the phrase *stem the TIDE/flow/swell of.* The second type of insertion consisted of more grammatical material, specifically pre-determiners, for example *taken* **quite** *a beating* for the phrase *take a BEATING/hammering,* or *having* **a bit of** *a wobbly* for the phrase *throw a WOBBLY/wobbler.* These examples fall under a *specificity and amplification* label suggested by Moon (1998a: 130). This includes examples such as:

> *Have (first) dibs on something*
> *Lead somebody a (merry) dance*
> *On your (own) head be it*
> (cf. Moon, 1998a: 130)

[8] See Table 3, Appendix.

This category tends to consist of modifier material. Moon (1998a: 130) says that this material is often adjectival or adverbial, and examples she gives are *cut the (umbilical) cord* for the former, and *stop someone (dead) in their tracks* for the latter. She suggests that insertions such as these can be seen as exploitations, although they are not necessarily jocular (1998a: 174).

Gläser (1998: 130) suggests that in terms of lexical substitutions there are some substitutes that are systematic. These are frequently occurring variants such as synonyms and hyponyms etc. The substitutes given by the dictionaries are seen as being acceptable alternatives for that phrase. Other substitutes are creative modifications; changes made by an individual for a particular stylistic effect. In terms of insertions, it would seem that grammatical insertions are more systematic and lexical insertions creative. It is also the case that grammatical insertions consist of a smaller set, whereas there is a larger vocabulary for lexical insertions.

The grammatical material found tended to consist of quantifiers. Note that there were some examples where the phrase in canonical form has a null determiner, but occurrences were found with a determiner inserted, for example, the phrase *do BIRD/time* had 103 occurrences, 76 of which contained insertions. However, every example of insertion for this phrase was a determiner; an actual determiner inserted where the canonical phrase had a null determiner. The syntactic structure for this phrase remains V det N, where det in this case is optional. Such examples (for example *done the time, did his bird*) were not counted as insertions, as the presence of a determiner in such cases acts as a determiner substitute rather than an insertion. Similarly, examples of quantifying determiners, for example *some, more* etc. were also classed as determiner substitutes and not insertions.

Most of the lexical insertions found in the case studies were adjectives. It is also possible for a phrase to have both grammatical and lexical insertions at the same time, for example *created* **more of a public** *stink* for the phrase *kick up a STINK/fuss, more of a* being the grammatical insertion, and *public* being the adjective modifier for this occurrence. The number of grammatical insertions was counted separately from the number of lexical insertions, regardless of whether they appeared in the same token or not. So out of the 35

phrases of the dataset, 25 showed lexical insertions, and ten of those also showed grammatical insertions.

4.4.3. Selected Sequences from the Dataset

The sequences selected for further analysis in Chapter 5 (Lexical Substitutions) are: *shut your FACE/trap/cakehole/gob, flip your LID/wig, throw a WOBBLER/wobbly, take a BEATING/hammering*, and *watch your TONGUE/mouth.* For Chapter 6 (Insertions), the phrases *kick up a STINK/fuss, overstep the MARK/limits/bounds, pin your HOPES/faith on, raise HELL/Cain*, and *take a BEATING/hammering* show the most total insertions and will be further examined, noting that the latter will be analysed in both chapters, thus a total of nine case studies were selected for this research project.

5. Lexical Substitution

5.1. Introduction

The research project focuses on variation within formulaic sequences. This chapter concentrates on lexical substitution of items within the phrase; specifically, substitution of the noun within a noun phrase.[1] This chapter is based upon the five expressions from the dataset which presented the highest number of lexical substitutions of the noun. Each case study follows the same organisation:

- Definition and Senses - the dictionary definition and variants, substitutes found in the chaining process
- Networks and Compositionality - discussing the findings from the chaining process with respect to theories of networks and compositionality
- Summary - relating the findings to the research questions

This chapter investigates the boundaries of lexical substitution: how many lexical noun substitutions can occur per formulaic sequence, and at which point the sequence changes from being a formulaic sequence following the idiom principle, to being a free construction following the open choice principle. The boundaries between one formulaic sequence and another, and the boundaries between a literal reading of the sequence and a non-literal one are also explored. The cases studied in this chapter are the formulaic sequences *take a BEATING/hammering, throw a WOBBLY/wobbler, shut your FACE/trap/gob/cakehole, watch your TONGUE/mouth,* and *flip your LID/wig.*

5.2. Overview of Cases

Table 5-1 gives the frequencies for the cases investigated: the total number of occurrences found in the BNC, the number of lexical noun and verb substitutes found, and the sum of the lexical noun and verbal substitutes found. The frequencies help to demonstrate the boundaries of the formulaic sequences; how much substitution they can undergo. For

[1] As opposed to the noun in a noun phrase within a prepositional phrase (see section 4.3.1).

example, the sequence *shut your FACE/trap/gob/cakehole* has only one verb used in the LID, and four noun substitutes. Table 5-1 shows that using the chaining process in the BNC showed 12 noun substitutes, and four verb substitutes before the chaining process was exhausted. The numbers given in the lexical noun and verb represent the limits of substitution for the sequences. No others were found so the limits for the sequence *shut your FACE/trap/gob/cakehole* are 12 nouns and four verbs.

Table 5-1 is taken from Table 4-7 and Table 4-8 in section 4.3.2 which show these figures for all of the phrases in the dataset, and which emphasize that these five sequences have the most lexical substitutions of the noun.

Phrase	Tokens	Types	Lexical Substitutes (N)	Lexical Substitutes (V)	Total Lexemes Involved
Shut your FACE/trap/gob/cakehole	128	39	12	4	16
Flip your LID/wig	400	91	11	3	14
Throw a WOBBLY/wobbler	150	50	8	7	15
Take a BEATING/hammering	112	90	8	16	24
Watch your TONGUE/mouth	516	79	8	8	16

Table 5-1 – Five sequences from the dataset showing the highest number of noun substitutes

Table 5-1 also shows the number of different types and tokens found. Note that the type is the different specific forms of the sequence (see section 4.3.2). With the exception of *take a BEATING/hammering,* all of the case studies have more noun substitutes than verb substitutes. Note also that with the exception of *flip your LID/wig* and *watch your TONGUE/mouth,* the cases have relatively low token frequencies in comparison to others in the dataset (see Table 4-7). It is not necessarily the case (see Čermák, cf. Moon, 1998a: 121) that the sequences showing the highest number of lexical substitutions of the noun have the largest token frequencies.

The cases studied here have similar register distributions. In most of the tokens, author or audience details were unknown, so only general tendencies are given here, and will not be discussed further in the case studies. Domain frequencies are given when relevant in the case studies, where this information helps to determine different formulaic sequences.

In general, written forms were preferred over spoken forms in frequency terms (note that the spoken section of the BNC is much smaller than the written one), however the sequence *shut your FACE/trap/gob/cakehole* also had a high preference for spoken tokens; 53 out of 128 total tokens. *Shut your FACE/trap/gob/cakehole* was the formulaic sequence in the dataset showing the most spoken occurrences. The sequence *throw a WOBBLY/wobbler* also had a relatively high number of spoken occurrences; 35 out of 150 total tokens. A stereotypical notion of formulaic sequences is that they are predominantly spoken, although the cases studies in this chapter (and Chapter 6) do not support this.

Most of the tokens for all of the cases were from the year range 1985 – 1994, with *take a BEATING/hammering, flip your LID/wig,* and *watch your TONGUE/mouth* also having tokens in the 1975 – 1984 range, and *flip your LID/wig, shut your FACE/trap/gob/cakehole,* and *watch your TONGUE/mouth* having tokens in the 1960 – 1974 range. The typical audience of the texts was mixed gender for all of the case studies, although *watch your TONGUE/mouth* had a high proportion of female recipients, and *flip your LID/wig* also had a tendency for a female audience. The level (a subjective assessment of the text's technicality or difficulty) was medium with low level in secondary frequency position apart from *watch your TONGUE/mouth,* which had these two levels in reverse preference order; low then medium. The author age tended to be unknown, with author gender preference being female for *flip your LID/wig,* and *watch your TONGUE/mouth,* male for *take a BEATING/hammering,* and *shut your FACE/trap/gob/cakehole,* and equal for *throw a WOBBLY/wobbler.* Authors tended to be sole in the main for all cases except *take a BEATING/hammering.*

Note that in Table 5-1, lexical substitutes with only one token occurrence were still recorded. Thus forms with few occurrences are not to be excluded as rare examples, but can be seen as peripheral members of the lexical set for that particular phrase, and the outer limits of substitution. Analysing the distribution of the lexical items involved shows that just because the phrase shows lexical substitution, not all variant forms are of the same standing within the phrase.

5.3. Case Study 1: Take a BEATING/Hammering

5.3.1. Definition and Senses

The LID definition for this formulaic sequence is:

> 1. To be severely criticised, badly damaged, or completely defeated
> (Often used in newspapers, on television news etc.)
> 2. To lose money, or become less in value
> (1998: 21)

The layout in the LID suggests that *beating* and *hammering* are equally frequent and interchangeable. This definition is reflected in the occurrences found in the BNC:

> *Example 5-1*
>
> Screenwriter Hart moves the story to into the 20th century by suggesting that Peter was an orphan taken to Great Ormond Street Hospital for sick children in the mid – 1960s by Wendy Darling, played by Maggie Smith, and then adopted by an American family. The film **took something of a battering** from U.S. critics.
> *Daily Telegraph: Electronic Edition: Arts Pages*
> *AJ8 74*

There were 112 tokens of this phrase and its variant forms in the BNC, and of these there were 90 types, indicating that this phrase is not fixed in the same way as, for example, *hit the DECK/dirt,* which has 454 tokens, and only 34 types (see Table 4-6). There were 16 different verb substitutes found: *take, get, give, come in for, deserve, have, hand out, withstand, face, administer, suffer, be in for, spare, be, produce,* and *receive*, and eight noun substitutes: *hammering, bashing, beating, battering, thrashing, drubbing, licking*, and *trouncing*. Grammatical substitutes were *a, the, another, their,* and *an.* Note that the verbs *come in for, hand out*, and *be in for* are phrasal; they require the prepositions in order to act as a complete semantic and syntactic unit. Thus they have been included here, whereas examples such as *with a merciless hammering (CH3 7663)* have not been included, as there was a distinct syntactic change from the required V det N structure to PP det N. Note also that the verb *give* appears. It is a ditransitive verb, with the structure V N det N. Altenberg (1992: 240) says that where there are two complements to a verb in a formulaic sequence, one is fused with the verb and the other functions as a variable lexical slot, and as such is

not recurrent. This is the case with for example *give them a good thrashing (HR5 546)* where *thrashing* is fused with the verb, and the direct object is variable.

Lexical Verb Substitute	Frequency
Take	68
Give	14
Get	8
Be	4
Hand out	3
Have	3
Be in for	2
Suffer	2
Administer	1
Come in for	1
Deserve	1
Face	1
Produce	1
Receive	1
Spare	1
Withstand	1
TOTAL =	**112**

Table 5-2 - Lexical verb substitutes for *Take a BEATING/hammering*

Lexical Noun Substitute	Frequency
Hammering	29
Battering	29
Beating	21
Drubbing	12
Thrashing	10
Bashing	9
Licking	1
Trouncing	1
TOTAL =	**112**

Table 5-3 – Lexical noun substitutes for *Take a BEATING/hammering*

Table 5-2 and Table 5-3 show the different lexical possibilities found for this phrase and their frequency of occurrence using the chaining process in the BNC. For the verbal substitutes, *take* is clearly the favoured verb for this phrase. The frequency distribution of

the nouns is less diverse than for the verbs. Many of the verbs could be described as peripheral members, having only one token each, as can *licking* and *trouncing* in the nouns. In the nouns, the canonical *beating* is slightly less frequent than the dictionary variant *hammering.* The substitute *battering* occurs as many times as *hammering,* yet is not mentioned in the dictionary as a possible variant. The formulaic sequence *take a BEATING/hammering* is not mentioned in the CCDI, and can be found in the CIDI with the definition "to be defeated or to lose a lot of money" (1998: 26) but without any variants.

Take a BEATING/hammering in total has 112 tokens, and 90 types; 90 different permutations of the verb, determiner and noun, including pluralization and verbal inflections etc. (see section 4.3.2). Table 5-4 shows the different lexical combinations of lexis of the phrase; specific determiners are not shown but the lexical forms of the verbs and nouns are shown. In total, this phrase has 31 different combinations (see Table 4-6 section 4.3.2); Table 5-4 shows only those with more than one token occurrence. Of the 31 different combinations, 15 have a token frequency of more than one:

Lexical Combination	Frequency
Take + det + battering	26
Take + det + beating	17
Take + det + hammering	15
Take + det + bashing	6
Get + det + hammering	6
Give + det + drubbing	5
Give + det + hammering	5
Be + det + drubbing	2
Get + det + drubbing	2
Give + det + thrashing	2
Hand out + det + beating	2
Hand out + det + thrashing	2
Have + det + hammering	2
Receive + det + thrashing	2
Take + det + thrashing	2
TOTAL =	**96**

Table 5-4 – Frequent lexical combinations of *Take a BEATING/hammering*

The forms *take + det + beating, take + det + hammering*, and *take + det + battering* are prevalent here. Note that Table 5-2 and Table 5-3 showed this verb and nouns to be frequent. Of the 29 occurrences of the noun *battering,* 26 occurred with the verb *take.* Of the 21 occurrences of the noun *beating,* 17 occurred with this verb. However, the noun *hammering* did not have as great a preference for this verb; only 15 of its 29 occurrences occurred with *take.* So *hammering* occurs as a noun more frequently than *beating* for this phrase, but the combinations *take +det + beating* and *take + det + battering* are more frequently occurring than the combination *take + det + hammering.* The forms *take +det + beating* and *take + det + battering* are more fixed, occurring predominantly with the verb *take*, whilst the combination *take + det + hammering* is more flexible. The nouns *thrashing* and *drubbing* occurred with the greatest number of possible verb substitutes, eight and six respectively. *Hammering* occurred with five different verbs, and *bashing, beating*, and *battering* occurred with four different verbal substitutes.

Section 2.2.2 showed that *take* can be seen as a multi-functional, lexically non-specific verb. Adam Kilgarriff's BNC word frequency lists show that the top ten most frequently occurring verbs in the BNC are: *be, have, do, say, go, get, make, see, know,* and *take.*[2] These have a more functional nature than verbs which carry more lexical content, such as *suffer, spare, withstand* etc. Since such "lexically rich" verbs occur less frequently (as seen in the BNC word frequency lists), they are thought to be less open to substitution. Interestingly, in this case study, the verb *take* is a more functional verb, and thus many verbal substitutions could be expected. There are indeed verbal substitutions, but as seen in Table 5-2, the central verb to this phrase is the verb *take.* Other verbal substitutes for this sequence, such as *administer, face*, and *spare* are peripheral, with verbs such as *give* and *get* falling between the centre and the periphery. This is shown in Figure 5-3.

The definition sense one, as given by the LID, clearly divides into three uses in the BNC occurrences. One use is as shown to be criticism; a non-physical non-literal beating, as seen in Example 5-1. The second frequently occurring use for this definition is complete defeat. Many of the occurrences in the BNC showed this use, 34 of the 112 tokens – a defeat in sport, for example:

[2] http://www.kilgarriff.co.uk/bnc-readme.html

Example 5-2

Ryan is trying to keep a long–term perspective since his appointment as Welsh coach runs to the end of the next World Cup, in the autumn of 1991. But all the plans in the world will mean nothing if Wales **takes another battering.**
Independent: Electronic Edition: Sports pages
A4B 244

Example 5-3

Tranmere on a Tuesday night. Sheer torture. This wasn't defeat, it **was a drubbing**. A demolition job which former United star John Aldridge started with the first goal after just 12 minutes.
Central T.V. News Scripts
K21 1930

Most of the cases demonstrating defeat were concerned with sports, although there were some examples of a defeat in war, and in a political election. Sense two of the LID definition involves economic loss. This was the least common of the uses, with only 13 of the 112 tokens:

Example 5-4

On the UK economy Jim admits small business is **taking a hammering** but blames the banks, and problems in Germany and the U.S., ahead of the Tories.
Northern Echo: Leisure Pages
K52 8388

The uses of criticism, defeat, and economic loss are non-literal; the phrase meaning is not the total sum of the meaning of the parts. The phrase has different targets of use, such as in sports, criticism, and finance, but there are similar meanings, divided by the dictionary into the two senses with the underlying concept of ATTACK. Another use for which this phrase and its variant forms can be applied is in the sense of being damaged or hurt.

Example 5-5

Cleaning your tack properly once a week gives you the vital opportunity of checking that the stitching is in good order and that the parts where metal rests on leather (such as reins and stirrup leathers) are not too worn. Most New Zealand rugs **get quite a hammering.**
Today's Horse
BPB 617

Example 5-6

Morita Killer Wind roars back in Florida. Smashed houses, flying roofs, streets of rubble...once again Florida **takes a beating** from nature.
The Daily Mirror
CH1 9686

The use of the phrase as to damage or hurt can relate to animate beings, in the sense of a physical attack, and to inanimate objects, in the sense of being damaged or well used and worn. This use applied to 44 of the total 112 tokens, and divides into literal (34) and non-literal (10) occurrences. The sense of something suffering a physical attack, as seen in Example 5-5 and Example 5-6 is metaphorical in the use of a literal image of physical damage and hurt. The object in question is not literally being hammered, nor is the sense of *take* as in "to carry away" present. In terms of figurative imagery, the uses of criticism and economic loss are "more figurative" as it is harder to see the analogy between a physical blow and a verbal one, or a financial loss. The use regarding a physical attack is least figurative, as it is easy to see the relationship or motivation (see section 2.2.5) between the literal meaning and the image, and the use of defeat is in-between the two ends of the scale. The target regarding a physical attack is a concrete one; a person or object. The target of use regarding business, for example, in the sense of economic loss, is a more abstract one.

An example of the use of damaged or worn or attacked which takes a literal reading is seen in Example 5-7:

Example 5-7

"Greenheart oak" he said, tapping the hardwood underskid that was pinned beneath the kelson. "That'll **take a bashing** on a hard shore and never a crack in fifty years."
Armada
EWH 413

Note that section 3.3.1 showed that the chaining process allowed for literal and non-literal forms of the same formulaic phrase to be collected; a formulaic sequence is not necessarily non-literal by definition. The rules of the chaining process are automatic rather than interpretative; once a substitute is regarded as a possible replacement for an item in the formulaic sequence according to semantic meaning, for example *thrashing* replacing

beating/hammering, all other occurrences of *thrashing* in the V det N structure are collected without discriminating literal and non-literal occurrences. Collecting literal and non-literal tokens provides examples with which to study the boundaries between a literal and non-literal reading of a sequence.

In Example 5-4 the noun *hammering* occurred in a context of economic loss, whereas Example 5-5 had the same noun but with a meaning of well used or damaged. Likewise Example 5-1 and Example 5-2 had the same noun *battering* but the former with a sense of criticism and the latter with a use of defeat. There was a slight tendency for the nouns *thrashing, drubbing, trouncing,* and *licking* to occur with a use of defeat; seven of the ten *thrashing* occurrences, nine of the twelve *drubbing* occurrences, and the individual occurrences of *trouncing* and *licking* occurred with this sense. There was also a slight preference for *bashing* to convey criticism. However, the numbers for these are small, and the remaining occurrences of these nouns, and the noun *beating,* are distributed throughout the other senses. In section 2.3.1, it was seen that noun substitutes are often systematic, being synonyms, or hyponyms etc. This seems to be the case for *take a BEATING/hammering,* where the substitutes for this phrase are on the whole synonymous, systematic, and interchangeable. Words which have similar beginnings or endings (e.g. *-ing* in this case), and similar rhythmic pattern may be mentally stored together in clumps (Aitchison, 1994: 143).

Note the sequence *take a BEATING/hammering* has a different meaning from *take some/a lot of BEATING* which has the meaning "used in order to say that something is better, more enjoyable etc. than almost anything else of the same type" (Longman, 1998: 21). The difference between these is due to the determiner. Where the indefinite determiner is present (see Example 5-4, Example 5-7 etc.), the phrase describes an attack of some sort. Where the quantifier is *some* or *a lot of,* the sense of the phrase changes from being "attack", to a different sense – that of "being better than"/"winning", almost the opposite polarity from the occurrences in the case study. The formulaic sequence *take some/a lot of BEATING* is a lot less flexible than *take a BEATING/hammering.*

5.3.2. Networks and Compositionality

Networks link lexical nouns and verbs together in combinations which have a common central concept, and shared lexical set. Basic networks were seen to organize noun and verbal substitutes for a particular formulaic sequence into a schema with the same underlying meaning and syntactic structure. Chaining networks create a continuum of meaning typically with one formulaic sequence at one end, through lexical substitutions to a separate formulaic sequence at the other end, with degrees of meaning change throughout.

The formulaic sequence *take a BEATING/hammering* forms a chaining network. There are different uses of the phrase, and slight preferences for particular nouns for different uses for example *bashing* for criticism, and *thrashing* and *drubbing* for defeat. This is demonstrated in Figure 5-1. The sense of criticism starts at the top of the network, through *beating, battering*, and *hammering* which have no preference of use, to the nouns used in a context of defeat. The underlying meaning ATTACK remains the same throughout, and despite the slight use preferences, the nouns were each able to replace each other. Figure 5-1 is thus a combination of a basic network, where all of the combinations have the same meaning, and a chaining network, representing the different uses.

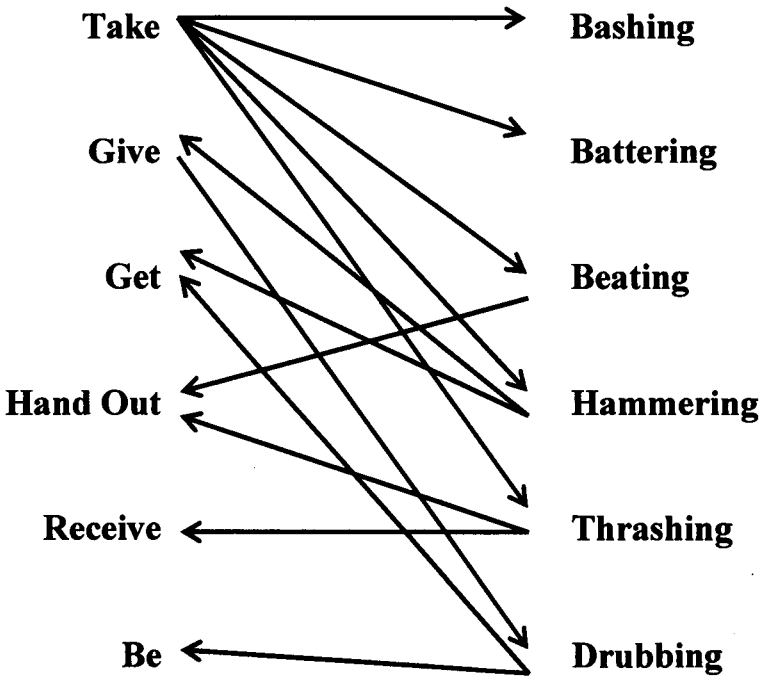

Figure 5-1– Network for *Take a BEATING/hammering*

The arrows in Figure 5-1 show directionality. Starting the chaining process by entering the verb *take* results in combinations with the nouns *bashing, battering, beating hammering,* and *thrashing.* In turn, entering the noun *hammering* leads to the verbs *give* and *get. Beating* gives the verb *hand out. Thrashing* also results in combination with *hand out* and also with *receive.* Searches with the verb *give* leads to the noun *drubbing,* which forms combinations with *be.* The limits of the network in Figure 5-1 were reached when no new verb or noun substitutes were found with the same central meaning and syntactic structure.

Table 5-2 and Table 5-3 indicate that there are some lexical items which appear to be more central to this phrase. The verbal lexeme *take* is prevalent, as are the nouns *battering* and *hammering.*

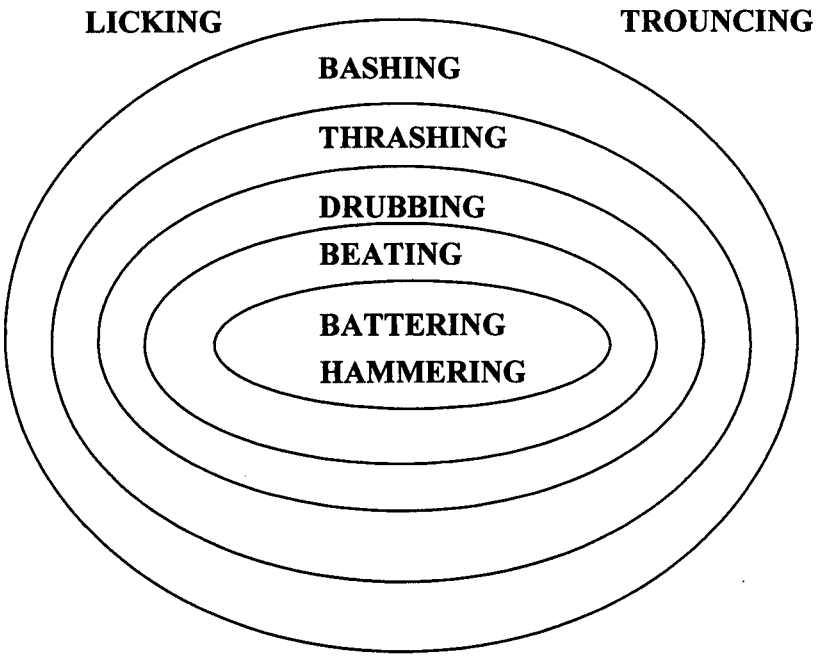

Figure 5-2 - Central and peripheral noun members of the lexical set for *Take a BEATING/hammering*

Figure 5-2 shows the central and peripheral nouns for this phrase. The nouns *licking* and *trouncing* are peripheral, having only one token occurrence each. The further out in the ellipse the noun is, the smaller the frequency of its occurrence. The rarer occurrences demonstrate the outer limits of the substitution. *Battering* and *hammering* are equally central to this phrase. Figure 5-3 shows a similar phenomenon with the verbs for this phrase.

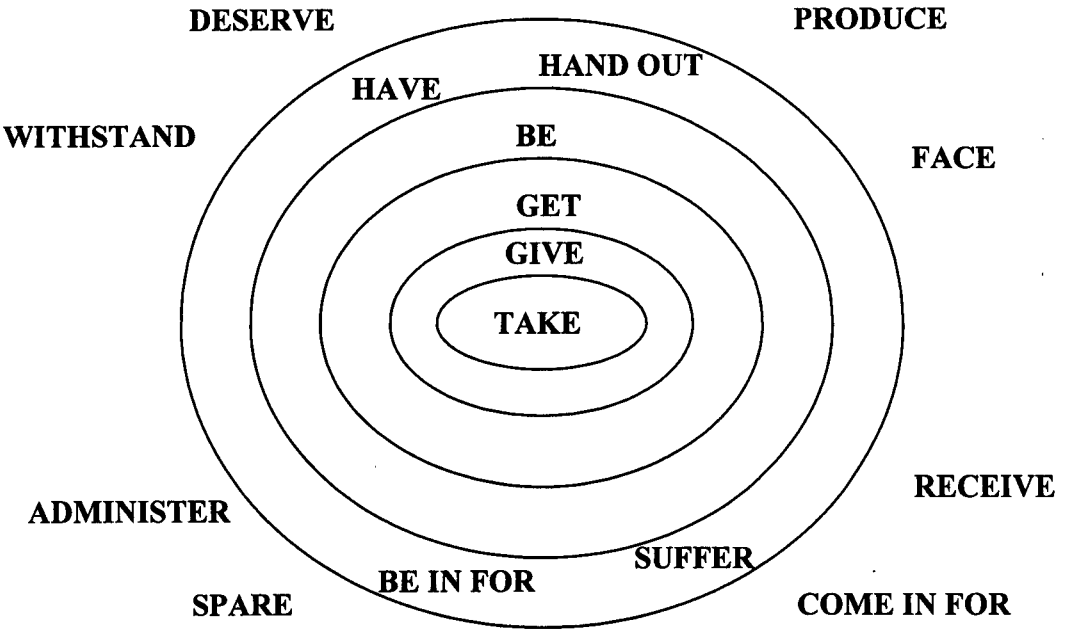

Figure 5-3 - Central and peripheral verb members of the lexical set for *Take a BEATING/hammering*

That there seem to be central and peripheral members of the noun and verb lexical sets for this formulaic sequence is in line with prototype theories. The nouns and verbs for this formulaic sequence (see Figure 5-2 and Figure 5-3), in terms of frequency, can be represented as central and peripheral members of the lexical set pertaining to the sequence *take a BEATING/hammering*. The figure for the verbs (Figure 5-3) demonstrates the key members of the lexical verb set in terms of frequency only; *take* is the most frequently occurring verb, thus it is the *prototype* verb for this phrase. As the verbs decrease in frequency, they become less prototypical of the sequence, for example *deserve,* or *face,* with only one token occurrence are on the periphery. The nouns for this phrase are also shown in terms of frequency (Figure 5-2), *battering* and *hammering* being the most frequently occurring, thus the most important, or prototypical of this formulaic sequence. The nouns *licking,* or *trouncing* are least frequent, these are peripheral members of the lexical set. Note that whereas the verbs differ radically in meaning, such as *give* (ditransitive) vs. *take* (transitive), and thus are prototypical in terms of frequency only, the nouns are more systematic and from the same semantic area, and convey the same emotive force.

The frequencies found for substitutes influence the networks. A high frequency supports an establishment of meaning in a particular combination, for example *take + det + battering, take + det + beating*, and *take + det + hammering* are frequent combinations for *take a BEATING/hammering* (see Table 5-4). Occurrences such as *receive + det + thrashing,* with only one occurrence, represent the outer limits of the network. The frequent combinations support the central concept, and the peripheral combinations represent the outer limits of the concept. The idea of a central concept with outer limits supports the idea that formulaic sequences are not dead metaphors, or stored in an atomistic list (see section 2.2.5). Instead of sequences being organized in a list with a specific meaning, it seems that there is instead a central concept and associated lexical set. Language users wanting to use the concept can select verbs and nouns from the lexical set and apply them to form a combination. Thus combinations with a singular occurrence are not to be disregarded as one-off occurrences, but seen as a language user's knowledge of the concept and its use. Howarth says that overlapping clusters demonstrate gaps, or combinations which are not possible (see section 2.4.5), and for which there is no communicative need. The idea of a central concept and lexical set suggests that gaps occur not where combinations are impossible, but where a language user has not created that particular combination, and so there has been no occurrence of it in the corpora investigated. Peripheral combinations demonstrate a language user's knowledge and creativity of language.

Figure 5-1 shows that verbs such as *give* and *hand out* were found as possible substitutes for the sequence *take a BEATING/hammering.* The central concept behind *take a BEATING/hammering* as given by the LID definition is UNDER ATTACK; someone receiving criticism, being damaged, defeated or having a financial loss. The semantic role of the subject of *take a BEATING/hammering* is as an experiencer. However, when the verb is *give,* or *hand out,* the semantic role of the subject is as an agent; they are giving the attack, not receiving it. Thus the central concept of *take a BEATING/hammering* is ATTACK. Section 3.2.1.2 showed how the LID displays sequences with a main entry and a closely related one; the example given was *sb has been sold a BILL of goods* and *sell sb a BILL of goods.* In a similar way, the sequence *take a BEATING/hammering* could be displayed, taking into account the direction of the attack, for example:

take a beating/hammering
1. To be severely criticised, badly damaged, or completely defeated
(Often used in newspapers, on television news etc.)
2. To lose money, or become less in value
give a hammering

The change in semantic role and directionality of the attack suggests that, in agreement with Stein (see section 2.2.2), verbs such as *take* and *give* do carry some meaning and are not just empty verbs in the sequence *take a BEATING/hammering.* The change of semantic role also supports the notion of a central concept for the sequence; the language user can adjust the sequence using the central concept of ATTACK from experiencer to agent. The combination *be +det + drubbing* is another example of a change in semantic role. In this case, the subject of the sequence is the attack itself, rather than the subject undergoing the attack (*take + det + beating*) or the subject giving the attack (*give + det + hammering*). *Be + det + drubbing* represents an adjustment of the central concept. There is a bigger "shift" in semantic roles in this case, and so *be + det + drubbing* is on the outer boundary of the sequence *take a BEATING/hammering.*

Theories of compositionality suggest that formulaic sequences either have each word being figurative, contributing to the overall meaning, or have one literal part of the phrase, and one part that carries the idiomaticity. This is usually the case with phrases that have the verb *have* or *be* (*be a BREEZE/doddle, be a sore POINT/spot*) - commonly occurring verbs, or primary verbs. The noun is used in a figurative sense (Howarth, 1996: 95 – 97). The sequence *take a BEATING/hammering* acts in this way. *Take* is a common, everyday verb, as are many of its possible variants; *get, give, have.* The verb of *take a BEATING/hammering* seems to be literal, and possibly reduced, in reading. The fact that the noun seems to take idiomaticity, and the verb is literal is also shown in cases where the structure may change, for example from V det N to PP det N:

Example 5-8

Wigan swept into the final of the Greenalls Lancashire Cup **with a merciless hammering** of Oldham.
The Daily Mirror
CH3 7635

The sense is the same, although the structure is different from that specified in this dissertation (V det N). This phrase does not necessarily need a verb to be present to retain the metaphoric meaning.

5.3.3. Summary

In summary, for the sequence *take a BEATING/hammering*, it seems that the noun is open to systematic substitutions of a synonymic nature, leading to central and peripheral prototypical members of the lexical set, whereby frequency of occurrence is a key factor. The lexical substitutes form a network, whereby there is a common lexical set and a central common concept. Frequently occurring combinations support the central concept, and rare combinations are on the periphery, the outer limits of the network. Although the different noun substitutes have slight preferences for different uses, they can be used interchangeably. The noun *battering* is as central to the phrase as the dictionary-given *beating* or *hammering*. A factor which may provide some explanation as to why this phrase is open to many lexical substitutions is the functional nature of the verb. The phrase has an underlying concept of ATTACK with surface uses of criticism, damage, defeat, and economic loss. This phrase, in accordance with the notes suggested by the LID, has a preference for leisure texts, such as newspapers and news reports, as seen in Example 5-6 and Example 5-8.

5.4. Case Study 2: Throw a WOBBLY/Wobbler

5.4.1. Definitions and Senses

Example 5-9

With children's tantrums, you'll find the old child-rearing catchphrase "They'll grow out of it" usually holds true. It may help to keep this in mind next time your two-year old **throws a wobbly** at the supermarket checkout.
Parents
G2T 1496

Example 5-10

The slightest thing would set her off, usually when she was feeling insecure and wanted attention. She loves the telephone, and she **threw a wobbly** when I wouldn't let her play with it.
Good Housekeeping
ED4 1278

Examples such as Example 5-9 and Example 5-10 demonstrate the definition given in the LID:

> To suddenly become very angry, often about something that is not very important. (Spoken)
> (1998: 388)

The phrase *throwing a WOBBLY/wobbler* describes an act of expressing anger and frustration, and the strings it extends to are similar expressions. The targets of such expressions are stereotypically toddlers and children, but, as demonstrated by findings in the BNC (94 tokens referring to adults and 52 to children), can also be of other age ranges.

The Concise Oxford Dictionary (10th ed. 1999: 1643) defines *wobbly* as:

> An uncontrolled outburst of anger and frustration, typically in a young child.

The Oxford English Dictionary (OED) defines *wobbly* in terms of the phrase in which it typically occurs:

> To throw a wobbly: to lose one's self-control in a fit of nerves, panic, temperament, annoyance, or the like; also, to act in an unexpected way, causing surprise or consternation.
> (http://dictionary.oed.com)

As Mountain notes (1999: 3), the human experience of emotions has been traditionally studied by psychologists and sociologists, and only recently has emotionology – the study of emotion – been recognized and studied by linguists. It is thought (Fernando, 1996: 120), that emotions and cognition are largely metaphorical in language. The English vocabulary of emotions, like vocabulary in general, is made up of single words, compounds, phrasal

verbs, phrases, idioms, and so on. Idioms arise from the experience of everyday life. There is a variety of emotions, which have differing degrees of intensity. As Mountain says (1999: 10), the image of a child having a "tantrum" is now commonly used to show anger and frustration in any age group, conjuring up images of childish agitation. Synonymy applies itself well to expressions of emotions. There are many synonyms which apply to for example the emotion of anger: rage, madness, wrath, ire, fury, pique, irritation, annoyance, outrage, etc., although it is clear that the degree of intensity of the emotion varies between these synonyms. This will be seen for the noun substitutes for this case study, in section 5.4.2. Table 5-5 and Table 5-6 show the verb and noun substitutes for *throw a WOBBLER/wobbly:*

Lexical Verb Substitute	Frequency
Have	93
Throw	52
Begin	1
Blow up	1
Fly into	1
Go into	1
Let go	1
TOTAL =	**150**

Table 5-5 – Lexical verb substitutes for *Throw a WOBBLY/wobbler*

Lexical Noun Substitute	Frequency
Fit	73
Tantrum	52
Wobbly	14
Wobbler	4
Paddy	3
Strop	2
Scranny	1
Sulk	1
TOTAL =	**150**

Table 5-6 – Lexical noun substitutes for *Throw a WOBBLY/wobbler*

Note that, as with *take a BEATING/hammering* (see section 5.3.2), there appear to be central and peripheral members of the lexical set; *have* and *throw* are clearly more frequent

than the other possible substitutes. Verbs requiring a prepositional complement have again been included (see section 5.3.1). Similarly, *fit* and *tantrum* are more central than the other nouns. The noun substitutes *tantrum, paddy, wobbly*, and *wobbler* have colloquial tendencies in the OED, and are also synonymic, *tantrum* having the definition of "an outburst of petulance or ill-temper". The OED Online defines *wobbler* as meaning the same as *wobbly,* and suggests it is found frequently in the phrase *throw a wobbler. Paddy* is defined as being a "fit of temper, a rage; a hot temper". The definition for *scranny* given by the OED is different from the use it is found in here; it is described as meaning "crazy, wild or silly", and is described as being slang, whilst in the data, the example found *mister will throw a* ***scranny*** *at you (FLX 526)* is of the same sense of anger as other noun substitutes for *throw a WOBBLY/wobbler.* The nouns *fit,* with a sense of "violent access or outburst of laughter, tears, rage", and *strop*, "a fit of temper, a sulk" also have anger as an underlying component, however, note that the meaning of *sulk* differs in that it means "a state of ill humour or resentment marked by obstinate silence or aloofness from society".[3] Since this also forms part of the definition of *strop*, the meaning of *wobbly* develops through the different noun substitutes in a systematic manner, the degree of intensity of the emotions being a variable factor.

Of the 64 occurrences of the phrase *have* + *det* + *fit,* 23 tokens had a literal, medical sense, for example:

One of the children **had a fit** on the bus this morning	GUR 361
...breathing rapidly and foaming at the mouth, and finally **had a fit**	K41 476
...leaving off my medication and of course they **have a fit**	F8C 107

The remaining 41 had a non-medical sense; a sense of being very angry. In section 3.3.1, it was noted that literal and non-literal examples of the same lexical combinations were collected, for this phrasal type, both literal senses were collected; that of anger and that of a medical fit.

Table 5-7 shows the most common lexical combinations found for this phrase:

[3] http://dictionary.oed.com

Lexical Combination	Frequency
Have + det + fit	64
Have + det + tantrum	25
Throw + det + tantrum	23
Throw + det + wobbly	13
Throw + det + fit	9
Throw + det + wobbler	4
TOTAL =	**138**

Table 5-7 – Frequent lexical combinations of *Throw a WOBBLY/wobbler*

The most frequent combinations of this phrase are the lexical combinations *have + det + fit* and *have/throw + det + tantrum.* The LID has a separate entry for *have/throw a fit* with a suggested British English variant as being *have/throw a blue fit.* It is suggested by the dictionary as being a spoken phrase, as with *throw a WOBBLY/wobbler,* and is given a definition of "to get very angry or upset" (1998: 123). The CIDI has an entry for *throw a WOBBLER/wobbly,* suggested as being informal British or Australian English, and having a meaning of "to suddenly become very angry" (1998: 429).[4] The CIDI has a separate entry for *have/throw a fit* as being informal and meaning "to be very angry" (1998: 140). The CCDI only has an entry for *throw a WOBBLY/wobbler* with the definition:

> If someone throws a wobbly or throws a wobbler, they lose their temper in a noisy, uncontrolled, and childish way, often about something unimportant
> (British English)
> (2002: 403)

So the lexical substitute *tantrum* is not mentioned in any of the dictionaries, despite its high occurrence in the BNC. Reasons for this are explored in section 5.4.2. Note also that whereas the nouns *fit* and *tantrum* can occur with either of the verbs *have* and *throw,* the nouns *wobbly* and *wobbler* only occur with the verb *throw*, bar one occurrence of *wobbly* with *have.* As with the combination *take a beating* in section 5.3.1, it seems also the case here that the canonical form is less flexible in terms of how many nodes it links to than the combinations it chains onto.

[4] Note the reverse head word to the LID.

5.4.2. Networks and Compositionality

The sequence *throw a WOBBLER/wobbly* seems to be compositional. As mentioned in section 5.3.2, the verb *have* is potentially a more functional verb than verbs which seem to carry more lexical content, for example *blow up, fly into, let go* etc. Section 2.2.2 detailed delexicalized verbs such as *have* and *take,* and suggested that where a verb is semantically empty, it may be open to more substitutions. However, looking at Table 5-5 the verb *have* is a central member to this phrase, alongside *throw*. Other less frequent verb substitutes are more lexically rich.

The phrase reading seems to be compositional. Where the verb is *throw,* the figurativeness is in the verb – one cannot literally "throw" a tantrum. Where the verb is *have,* the content of the phrase is carried in the noun. In the cases here where the verb is *have,* the phrase is literal and compositional, with the noun taking its literal sense. Referring to the definition of *tantrum* as was mentioned in section 5.4.1, where this occurs with the functional verb *have,* the meaning is carried by the noun, and the noun has its literal reading. This is also the case where the noun is *paddy.* That the phrase *have a tantrum* does not occur in the idiom dictionaries studied could be due to *tantrum* retaining its literal reading, and *have* carrying little lexical content.

The verb carrying little lexical content can be seen in examples where the verb is not present, and the structure involving the noun is different, but the meaning remains, as seen in Example 5-11:

> *Example 5-11*
>
> If this scenario occurred often enough, you would have taught her that, if whining does not work, **a noisy tantrum** will do the trick. You have only to give in a few times (out of many when you stick to your planned ignoring) to lose the battle.
> *Discipline: A Positive Guide for Parents*
> *B10 736*

As well as having synonymic central and peripheral verb and noun entries, the sequence *throw a WOBBLER/wobbly* seems to be one that forms a network in its variant forms. The network starts with the dictionary entry; the canonical form. This sequence can form a

schema, or a basic network, in which the individual noun and verb lexemes chain to create different combinations which each have the same meaning and use.

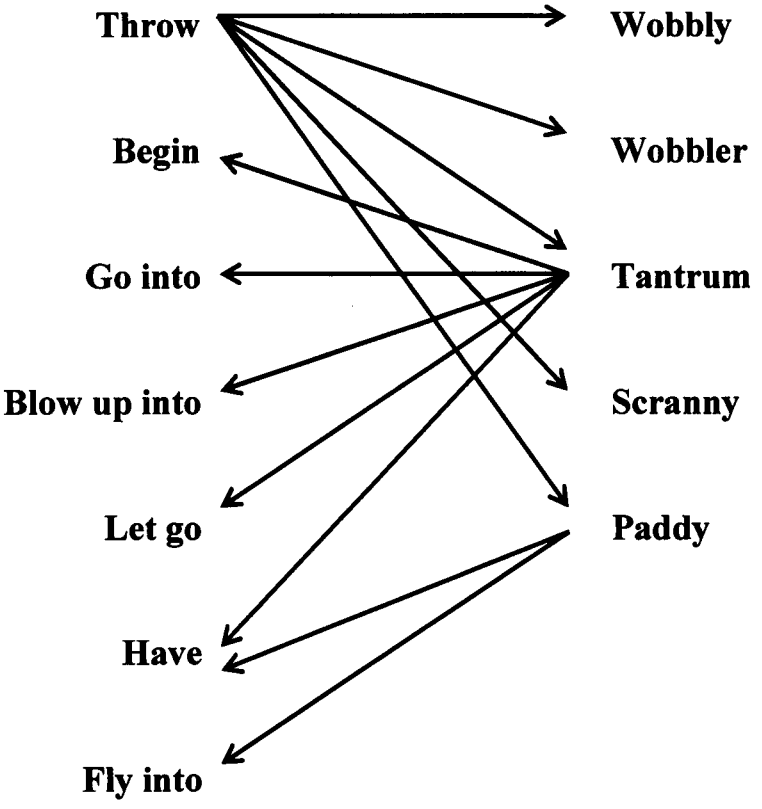

Figure 5-4 – Basic network for *Throw a WOBBLER/wobbly*

As mentioned in section 5.4.1, the meanings of *strop, sulk,* and *fit* are different from those of *paddy, wobbly* etc.

Example 5-12

...the Soviet ambassador to the U N was **having a sulk** and was refusing to attend the Security Council and he therefore persuaded the Security Council to pass a resolution
Durham University: Politics Lecture
JSM 11O

The imagery of childish immaturity in expression of anger and frustration is demonstrated in Example 5-12, and although both the verb and noun are substituted, the semantics from

the canonical sequence are retained. A further modification is that *sulk* seems to be a more internal emotion than *tantrum* (see section 5.4.1). The verb *throwing* emits the emotion *from* the person experiencing it, whereas *having* keeps the feeling *within* the experiencer. Since *sulk* has internal properties, it cannot, or is unlikely, to occur with *throw*.

As mentioned, the chaining process can lead to the noun *fit:*

> *Example 5-13*
>
> "I'm not coming you know, Floyd. We'll have fun here." "We are leaving now." "We are not." When a girl smiles this beautifully, she is about to **throw a fit.**
> *Payback – R. James*
> *J13 947*

The phrasal type used in Example 5-13, with the noun *fit*, has the emphasis on anger and losing one's temper, rather than sense of frustration seen in the canonical *throw a WOBBLY/wobbler.* This phrase, as mentioned in section 5.4.1 has its own entry in the LID. It is also linked in the dictionary to the American version of the phrase *have a cow* (to become very angry and upset about something someone has done, especially someone you know well). There were no occurrences of *have a cow* in the BNC.

The semantics of the example *have* + *det* + *fit* are slightly varied from the original framework. Section 5.4.1 showed that *fit* has two literal senses, one in the medical sense, and one in the sense of anger. In the latter sense, *fit* contains a much greater emphasis on anger, than on frustration, and seems not to have the same association with childlike imagery. This is shown in the definition. In the basic network, the lexical combinations are the same in meaning (to suddenly become very angry), and the lexemes chain to each other to form the network. However, this sequence and its network can be developed in a similar way to Howarth's overlapping collocation clusters (1996: 44), into a chaining network (see section 3.3.1). This is demonstrated in the chain from *throw a wobbly* to *throw a fit;* the different stages within the scale are best understood in relation to each other, thus the scale lexically and semantically develops. The intensity of the emotion develops along the scale.

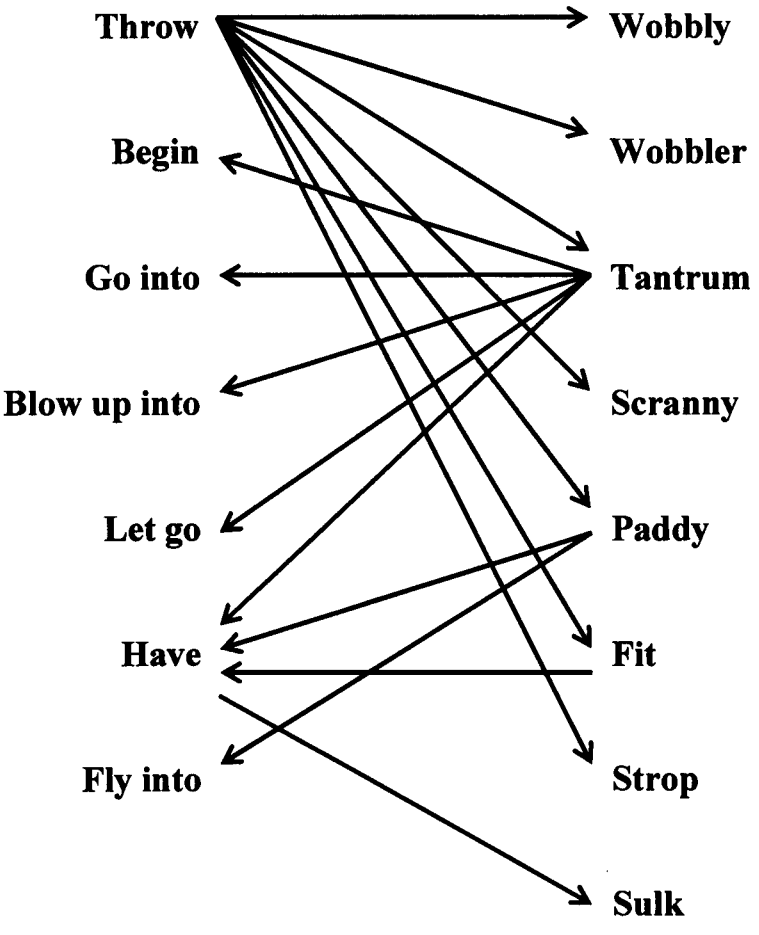

Figure 5-5 – Chaining network for *Throw a WOBBLY/wobbler*

In Figure 5-5, the "path" of the chaining process can be seen, with the diagram beginning with the first entries into the BNC; *throw + det + wobbly* and *throw + det + wobbler,* and progressing as new substitutes are found. The meaning develops through the new substitutes, in particular through the meanings of the nouns.

The difference between *throw + det + wobbly, have + det + tantrum*, and *have + det + fit* can also be shown in terms of register. Table 5-8 gives the domains of the total occurrences for *throw a WOBBLER/wobbly*, with the actual frequency, and the frequency once the domain size within the BNC is accounted for.

Domain	Frequency	Normalized
Applied Science	23	5.44
Arts	3	0.72
Imaginative	59	5.22
Leisure	12	2.09
Social Science	11	1.44
World Affairs	7	0.74
Spoken	35	5.87
TOTAL =	**150**	

Table 5-8 - Frequency of the domain of all of the tokens of *Throw a WOBBLY/wobbler*

From Table 5-8, the top three domains in terms of frequency were imaginative, spoken, and applied science. Once the comparative domain sizes are accounted for, however, it seems that spoken texts were the most favoured for this phrase. Most of the applied science texts were texts referring to child care and parenting skills, and the form of the phrase: *have a tantrum.* There were 25 occurrences of the lexical type *have* + *det* + *tantrum*. Of these, 16 occurrences were applied science texts. Of the remaining seven applied science occurrences, six involved the noun *tantrum.* Nineteen of the 23 applied science texts referred to child care and parents, and came from three different texts. Ten occurrences came from *Discipline: A Positive Guide for Parents*, seven from *Professionals and Parents: Managing Children's Behaviour*, and two from *Behaviour Problems in Young Children*:

Example 5-14

Time out was used whenever Chris **had a tantrum**. At home this meant putting him out into the hall of the bungalow (where there were no stairs).
Professionals and Parents: Managing Children's Behaviour.
CGS 1474

Domain	Frequency	Normalized
Applied Science	16	3.78
Arts	1	0.24
Imaginative	3	0.27
Leisure	3	0.52
World Affairs	1	0.11
Spoken	1	0.17
TOTAL =	**25**	

Table 5-9 - Frequency of the domain of all of the tokens of the combination *have* + *det* + *tantrum*

Table 5-9 shows the domain frequency and frequency once normalized for the type *have + det + tantrum.* Clearly the applied science domain is preferred for this type. So although in terms of overall domain size it looks as though this phrase as a whole has a tendency to occur in applied science texts, there is a bias which has been caused by the type *have + det + tantrum* and the frequency of occurrences from three texts as opposed to many occurrences from many different applied science texts.

Most of the occurrences of the form *have a fit* were from imaginative texts – there were 64 occurrences of this combination and of these, 37 were in the imaginative domain. The domain frequency and frequency once normalized are shown in Table 5-10:

Domain	Frequency	Normalized
Imaginative	37	3.27
Leisure	3	0.52
Social Science	3	0.39
World Affairs	5	0.53
Spoken	16	2.68
TOTAL =	**64**	

Table 5-10 - Frequency of the domain of all of the tokens of the combination *have + det + fit*

Table 5-10 shows the preference for the imaginative domain. The canonical *throw a wobbly/wobbler* did not have such a strong preference for a domain. This provides evidence for a chain from *throw a wobbly* (more figurative) to *have a tantrum* to *have a fit,* lexically the three stages are different, semantically they link, but they also have different uses in context. There is also a change in semantic role from *throw* (agent) to *have* (experiencer), reinforcing that *throw a wobbly* and *have a fit/tantrum* are individual sequences which can be linked by a common central concept and shared lexical set.

If the criterion structure for this phrase was relaxed from V det N to allow for PP det N it seems that the scale can continue resulting in two different phrasemes: *in a mood* and *be in a fit of.* A delimitation in the use of the chaining process was to limit the data to the V det N structure, thus for the sequence *throw a WOBBLY/wobbler, have + det + fit/sulk/strop* marks the end of the chain. There is a change from the quick outburst of anger and frustration in the canonical form to anger in *have + det + fit* to a more long term mood in *have +det + sulk. Have + det + sulk* represents the outer limits of the central concept. The

meaning has developed as far as it can without completely changing and becoming a completely unrelated formulaic sequence, and lexically too, the limits have been reached within the boundaries of the syntactic structure. If the structure was relaxed, *in a mood* and *in a fit of* would be included in the network, and the boundaries would expand. There would still be a central concept of anger and frustration; however the meaning would change from a quick outburst over something trivial, to a more long term emotion.

Using the idea of prototypes as shown in Figure 5-2 and Figure 5-3 in section 5.3, the chaining networks can be further developed. For this sequence, it has been seen that the nouns *paddy, wobbly, wobbler*, and *tantrum* have similar dictionary definitions, whereas *fit, scranny, sulk*, and *strop* differ in emotion and degree of emotion involved. Similarly, the nouns and verbs demonstrate a degree of central and peripheral membership. Cantos-Gómez believes that collocates within the same environment do not build flat lexical frames (or networks, as has been shown here), but develop "complex interrelated hierarchies similar to "constellations"". He suggests that collocates form "lexical conceptual multi-dimensional frames: lexical constellations" (2001: 103). He goes on to suggest that within these constellations, a word need not directly collocate with the node word, but with other words in the same context, which may themselves collocate with the node. This latter point, and the notion of constellations, corresponds to the chaining networks (see section 3.3.1). Combining the idea of central and peripheral lexical members, and the notion that frames need not be flat means that a more developed network than chaining networks can be created; *constellation networks*. Constellation networks allow the frequency of the lexical substitutes to be represented.

Howarth's overlapping collocation clusters (1996: 44) and Moon's schema (1998a: 161) do not account for frequency, and central and peripheral members. Howarth suggests that "there are clusters in the language consisting of some strongly-formed collocations along with many weakly-established combinations", and that such a phenomenon resists a flat diagrammatic approach (1996: 189). This means that previous notions of networks do not take into account prototypical or rare lexical items. The feature of frequency is recorded in the use of the chaining process, and so previous networks need to be developed in order to accommodate it. Constellation networks illustrate networks in a less linear representation:

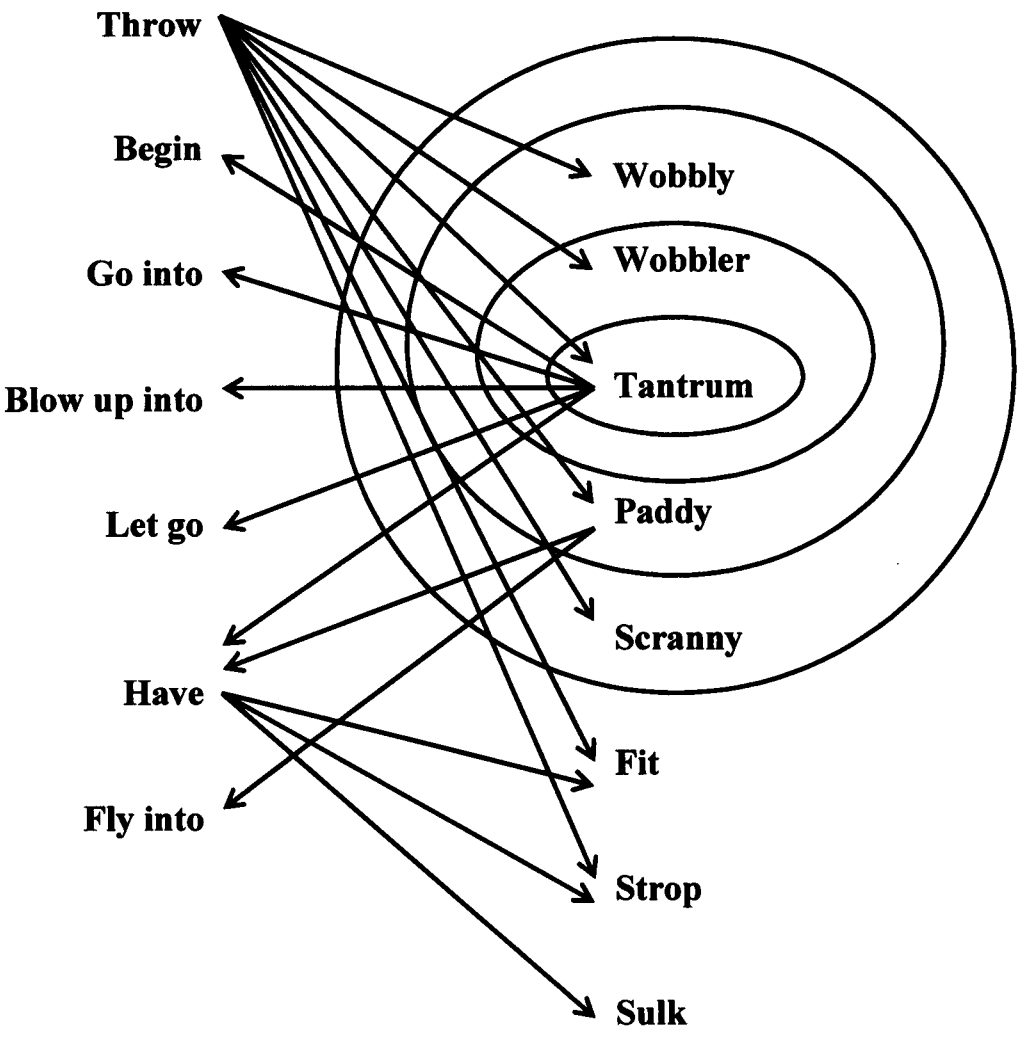

Figure 5-6 – Constellation network for *Throw a WOBBLY/wobbler*

In a network such as Figure 5-6, synonymous members can be grouped together, demonstrating central and peripheral members. Figure 5-3 showed all of the verbs for *take a BEATING/hammering* in a prototype demonstrating the central and peripheral verbs for that sequence. In constellation networks, prototypes group together words of the same meaning, showing central and peripheral forms, as was seen in Figure 5-2. The prototypes overlie the basic network, chaining lexemes and linking combinations which have the same meaning. Again, the first entries in the BNC using the chaining process are seen at the top of the diagram, and progression of new substitutes found follows down the diagram, the arrows showing which node led to which substitute. A constellation network also overlies the chaining network, which is itself a development of the basic network. At the same time,

the position within the constellation network indicates frequency, for example, *tantrum* is more frequent than the peripheral *scranny* or *paddy*. *Fit* has a different meaning than *tantrum* etc., so is not joined within the prototype. The combinations with *fit, strop*, and *sulk* have different meanings than the original *throw a WOBBLY/wobbler,* although these sequences can be grouped together as the meaning develops, using the chaining networks. Note that in section 5.3.2, a network for *take a BEATING/hammering* was given. The prototypes shown in Figure 5-2 could be amalgamated into the network to give a constellation network.

The work here suggests networks of phrases, where each node is a noun or verb of the formulaic sequence, and these link to others by form and meaning, creating lexical combinations. As will be seen in section 5.6.2 and 5.7.2, this also contributes to lexicography, showing how phrases can be objectively grouped by lexical set and underlying meaning, regardless of literal or non-literal reading, and how phrases relate to each other. As Cantos-Gómez says, networks show how lexical knowledge is "schematic and associative" (2001: 197). The networks, or scales, shown here demonstrate the V det N syntactic frame only. The networks would expand if different syntactic frames were included, for example PP det N.

5.4.3. Summary

In conclusion, *throw a WOBBLER/wobbly* demonstrates both central and peripheral members of the lexical set and demonstrates the development of networks; basic networks, chaining networks, and constellation networks, emphasizing the boundaries of the substitution in the frequencies, and the boundaries between one formulaic sequence and another. The underlying concept of the chain is ANGER, with varying degrees of intensity of emotion. The phrase seems to have a canonical non-literal form *throw a WOBBLY/wobbler,* and chains to literal forms *have + det + tantrum* and *have + det + fit.* The literal forms have preferences for applied science childcare texts (*tantrum*), and imaginative and spoken texts (*fit*). As a whole, this phrase has a higher frequency of spoken occurrences than other case studies, despite the relatively small spoken component of the BNC.

5.5. Case Study 3: Shut your FACE/trap/gob/cakehole

5.5.1. Definitions and Senses

The sequence *shut your FACE/trap/gob/cakehole* and the next (section 5.6 *watch your TONGUE/mouth*) together form one large case study. The two phrases individually appear in the dataset, however, it is possible to chain in a network from *shut your FACE/trap/gob/cakehole* to *watch your TONGUE/mouth*, as will be seen in section 5.6.2. Table 5-1 showed that *shut your FACE/trap/gob/cakehole* has 12 noun substitutes and four verb substitutes, and *watch your TONGUE/mouth* has eight noun and eight verb substitutes. However, since the two sequences can join by meaning and lexical set, it should be seen that *shut your FACE/trap/gob/cakehole* has 20 noun substitutes and 12 verb substitutes, and that this is the same lexical set as *watch your TONGUE/mouth.* For the purposes of investigating each of the sequences individually, I have imposed a division at the combination *hush* + *det* + *noise.* Section 5.5 will primarily discuss the phrase *shut your FACE/trap/gob/cakehole* and section 5.6 will discuss the sequence *watch your TONGUE/mouth,* and the links between the two case studies.

The definition for the phrase *shut your FACE/trap/gob/cakehole* is given in the LID as:

> A rude expression used in order to tell someone to stop talking
> (spoken, used especially by children)
> (1998: 112)

Example 5-15

"Have a bit of sense! If you know what's good for you, you'll make yourself scarce before you wake the whole house....Better cut and run before the coppers arrive." "**Shut your face!**" The man took a step towards him, and Connor saw his hands were shaking
Flood Water – P. Ling
FPM 128

As seen in section 5.2, this phrase has the highest proportion of spoken tokens of all the case studies, with 53 out of 128 tokens being spoken. A further 46 tokens occurred in quotation marks within written texts, representing direct speech, leaving only 29 occurrences in complete written form. Table 5-11 shows the distribution of the ages of the speakers of the 53 spoken tokens, with frequency of occurrence and frequency once

normalized. Table 5-11 goes some way towards supporting the dictionary's suggestion that this phrase is used primarily by children, and shows that the age ranges 15 – 24 (teenager/young adult) and younger favour the use of this phrase:

Speaker Age Range	Frequency	Normalized
0 - 14	7	7.00
15 - 24	17	6.83
25 - 34	5	1.56
35 - 44	7	2.22
45 - 59	10	2.77
60+	2	0.84
Unknown	5	N/A
TOTAL =	**53**	

Table 5-11 - Frequency of the age range of speakers of *Shut your FACE/trap/gob/cakehole*

If the conversational occurrences are looked at in more detail, it seems that there are in fact several occurrences of the phrase per text, especially in the texts spoken by younger speakers.

Example 5-16

PS53C Oh that's alright I just had it on too loud. Oh damn that hurt. *PS53J* Sorry I haven't Why d'ya want to why d'ya *PS53C* what you going on about? **Shut your mouth**. Oh that's original now just **shut your mouth**. Shut up. Alex Alex I liked I liked it when there's shut up *PS53J* fuck me fuck me fuck me fuck me PS53K Fuck you?
PS53C – "Danny" 13 student C1 Male
PS53J – "Daniel" 13 student C1 Male
PS53K – "Takeo" 14 student Male
Conversational
KPA 372

Example 5-17

PS556 It's on that, it's brilliant! No. Shut your *PSOOO* She's so *PS556* mouth! Shut *PSOOO* she is brilliant! *PS556* your mouth! *PSOOO* I agree with you. *PS556* Shut, thank you Nadima. *PS555* **Shut your mouth**! *PS556* And you're just a *PS555* a one-way conversation *PSOOO* I like Madonna. *PS556* shut, **shut your mouth**! **Shut your mouth**! *PS555* I like Madonna.
PSOOO – Unknown
PS555 – "Josie" 14 student London C2 Female
PS556 – "Shelley" 15 student London Female
Conversational
KPG 192

Due to natural overlap in conversation flow, there are obviously several more occurrences of the phrase in this text than seen at first glance, for example *shut..., shut your...* This repetitiveness is perhaps due to bickering talk as children grow and develop independent personalities, and fight for dominance within social groups.[5]

Lexical Verb Substitute	Frequency
Shut	123
Hold	2
Shut up	2
Stop	1
TOTAL =	**128**

Table 5-12 – Lexical verb substitutes for *Shut your FACE/trap/gob/cakehole*

[5] The spoken examples of teenage language form The Bergen Corpus of London Teenage Language (COLT). Few of them are very "natural", the speakers being self-aware of themselves and the tape recorder, e.g.:
PS05Y Yeah, they all look the same, they all look the same PS062 er, I didn't mean to *that's on tape* Clare PS05Y I know PS062 I would love to hear that PS05Y Kim **shut your mouth**
PS05Y Clare female 15, student, Welsh, C2
PS062 Kim female, 9, student, Welsh

Lexical Noun Substitute	Frequency
Mouth	55
Gob	13
Trap	9
Face	7
Noise	3
Neck	2
Row	2
Book	1
Cakehole	1
Lick-split	1
Racket	1
Rap	1
TOTAL =	**96**

Table 5-13 – Lexical noun substitutes for *Shut your FACE/trap/gob/cakehole*

Table 5-12 shows the different verbal substitutes found. Clearly this phrase has a strong attraction for the verb *shut,* with the remaining substitutes being peripheral. Table 5-13 also suggests a central noun member; *mouth*, with *gob* and *trap* being more frequent than the other more peripheral nouns. There were 32 occurrences of the combination *shut it* which have been included in the total tokens for this sequence. However, as *it* is not a lexical substitute, it has not been included in the noun count in Table 5-13. The tendency for central nouns is seen in Table 5-14, which shows the main lexical combinations for this phrase. The type *shut* + *det* + *mouth* is the most ubiquitous, with the form *shut it* being next most frequent. Reasons as to why these forms are more frequently occurring will be given in section 5.6.2.

Lexical Combination	Frequency
Shut + det + mouth	54
Shut + it	32
Shut + det + gob	12
Shut + det + trap	9
Shut + det + face	7
TOTAL =	**115**

Table 5-14 – Frequent lexical combinations of *Shut your FACE/trap/gob/cakehole*

Table 5-12, Table 5-13, and Table 5-14 showed that the relative frequencies of the noun variants are different from the suggested frequencies as given by the LID. The dictionary suggests that *face* and *trap* are substitutes of equal frequency, as are *gob* and *cakehole*, the latter pair occurring less frequently than the former.[6] However, the data here suggests that *gob* and *trap* are more frequent, followed by *face*, and that *cakehole* is merely a peripheral member. The phrase does not occur in the CCDI, but occurs in the CIDI with variants suggested as being *face/gob/mouth/trap,* and a meaning being "an impolite way of telling someone to stop talking" (1998: 350). It is suggested that this phrase is very informal. Note that the CIDI suggests *mouth* to be a substitute, but not *cakehole.* In general lexicography, frequency data from corpora are contributory to entries and information about entries. This does not seem to be the case for the LID for this phrase, although the data here is from the BNC only. The LID takes its data from the Longman Corpus Network which as a whole may demonstrate different central and peripheral members than the BNC on its own. There is no example of the form *shut it* in any of the idiom dictionaries.

The form *shut it* has been included as there were examples where the pronoun replaced the noun phrase, and the meaning of telling someone to be quiet was retained for example:

Example 5-18

"but I was hoping to travel on my own. I..." Ryker cut her short. "Shut up," he whispered. "Just **shut it.**" Donna turned to say something to him.
Heathen – S. Hutson
G0P 1810

The pronoun has not been included in Table 5-13 as it is not a "lexical" substitution. *Shut it* refers to *shut* + *det* + *mouth* but there is an intensification of meaning. There is a greater emphasis on the verb *shut* in the combination *shut it.* Formally, *shut it* should not be included in the data as it does not follow the V det N structure. However, there were 32 occurrences of *shut it* in the BNC, and they have been recognized here as they are conceptually relevant to the canonical formulaic sequence, and the networks (see section 5.6.2).

---

[6] The LID layout is **Shut your FACE/trap** (also **shut your gob/cakehole**) see section 3.2.1.2.

There are systematic relationships to be seen between the noun substitutes for this phrase. Most obviously the use of the word *face* here is an example of metonymic expansion from *mouth* – the whole is used to refer to the part. This is also the case for the noun *gob* and by extension *neck.* They are from the same semantic area; body parts. The nouns *noise, rap, row*, and *racket* extend from the meaning of the phrase being to stop talking; they refer to the noise being made. Again, they are from the same semantic area, this time referring to noise. The nouns *cakehole* and *lick-split* are demonstrating rare lexical use on the part of the speaker, being inventive terms for the mouth; *cakehole* being literally "the hole where one eats cake", and *lick-split* similarly referring to the lips and mouth. They are possibly creative colloquial forms, and as such, are rare. They do not occur in the OED, although their meanings are self-evident. The substitute *trap* is more metaphoric. Although the nouns differ in terms of systematicity and semantic area, they are interchangeable, as was seen to be the case for *take a BEATING/hammering* in section 5.3. Changing the noun does not create a difference great enough in meaning to suggest a chain or scale, more that there are prototypes with central and peripheral members. Another creative substitute found was the noun *book:*

Example 5-19

PS0U1 No because no but I Colin, let me say it. PS0TW Well anyway PS000 Bernard, you've got such a country accent at times. PS0U1 It's Belfast. PS000 No it's not. Not bleeding Belfast! PS0TW **Shut your fucking book**! PS0U2 Bernard, aren't you going to ask me? PS0TW We are.

PS0TW	None	Irish
PS0U1	None	Irish
PS0U2	None	Irish

Conversational
KE1 3063

Here the noun seems to be used in the same way as *mouth,* or *face* etc. As with *trap,* it seems to be metaphorical. Again its sole appearance in the data suggests that the use of this noun is peripheral and perhaps colloquial.

Notice that the imperative is important to this phrase. This will be seen to be a distinguishing feature between the formulaic sequence *shut your FACE/trap/gob/cakehole,* and the sequence at the other end of the scale; *watch your TONGUE/mouth* (see section

5.6.1). Of the 128 tokens for this phrase, 102 were in the direct imperative form. The remaining 26 examples were in the indirect imperative form. This tendency towards being an imperative is seen in the LID notes for this phrase, and is perhaps related to the high proportion of spoken tokens found.

5.5.2. Networks and Compositionality

As was stated in section 5.5.1, this formulaic sequence chains onto the next case study in this chapter; *watch your TONGUE/mouth,* section 5.6. The chaining process connecting the two sequences will be discussed in section 5.6.2. This section looks at the sequence *shut your FACE/trap/gob/cakehole* only.

Table 5-12 and Table 5-13 showed that there are clear central and peripheral members; the verb *shut* having many occurrences, and the noun *mouth* also being frequent. *Shut your FACE/trap/gob/cakehole* has a high proportion of literal occurrences. The dominant combination *shut* + *det* + *mouth* is literal and compositional. By extension, the occurrences with the nouns *gob, lick-split,* and *cakehole* can be read as being literal. The nouns *face* and *neck* can also be read literally if we recognize them as being metonymic relatives of *mouth.* Note that *shut* + *det* + *mouth* is literal in reading, as in "to close one's mouth", however it also has an extension of meaning to "making no sound". Extra meanings applied to compositional sequences were discussed in section 2.2.2. The nouns *noise, rap, row,* and *racket,* coming from the same semantic set "sound" are non-literal; one cannot "close" a noise. The nouns *trap* and *book* create metaphoric images. The form *shut it* was also literal; of the 32 examples of this type, there were four tokens where the referent was an object such as a door, the rest had a referent of mouth or noise, in the sense of an order to be quiet.

Shut your FACE/trap/gob/cakehole is different from the other case studies in distinct ways. Firstly, it has a clear tendency towards being spoken as an order; an imperative, a mood which is not dominant in the other phrases. Also, *shut your FACE/trap/gob/cakehole* has a much higher number of spoken occurrences than the other case studies, even when the component sizes of the spoken and written parts of the BNC are accounted for. The repetitiveness seen in section 5.5.1 is a feature which is not seen in other case studies. Finally, this phrase is much more open to literal interpretation than other tokens and

sequences. This sequence seems to be a fixed functional expression, as in a formula, rather than an idiomatic one. The term *formulaic sequence* has been used in this research to allow for cases such as *shut your FACE/trap/gob/cakehole*; phrases which are fixed but also not necessarily non-literal. This phrase is more of a spoken convention; a fixed expression, than the other case studies which are more open to non-literal interpretation. This is also the case for the frequently occurring type *shut it.* As such, *shut your FACE/trap/gob/cakehole* has been included in the LID due to overall frequency and familiarity, and its fixedness.

5.5.3. Summary

In conclusion, *shut your FACE/trap/gob/cakehole* has a high tendency to occur in a spoken context, either in conversation, or as reported speech within written texts. As such, it has a high proportion of imperative examples. The phrase is also more prevalent amongst the younger age ranges of the BNC. This phrase has central and peripheral members, and many of the tokens have literal readings, in particular those of the lexical combination *shut* + *det* + *mouth.* This phrase can chain onto the next case study, *watch your TONGUE/mouth,* and the boundaries between the two sequences are discussed in section 5.6.2.

5.6. Case Study 4: Watch your TONGUE/mouth

5.6.1. Definition and Senses

Watch your TONGUE/mouth and *shut your FACE/trap/gob/cakehole* together form one case study sharing the same lexical set. Section 5.6.2 discusses the links between the two sequences. I have divided the two since the LID lists them as separate entries. I treat them as two individual formulaic sequences which link together. I have divided them at the combination *hush* + *det* + *mouth.*

The definition for *watch your TONGUE/mouth* as shown in the LID is:

> Used in order to tell someone to stop talking about something, or to stop being so rude (spoken)
> (1998: 356)

This is seen in an example such as:

Example 5-20

"It usually means a court-martial or a decoration the governor hasn't time for anything else. And never forget that he also means trouble, so **watch your tongue** when you're in his presence. I can tell you, he's got a very short fuse."
As the Crow Flies – J. Archer
K8T 995

There were 516 tokens of *watch your TONGUE/mouth* and its variant forms in the BNC, with 79 types, indicating that this demonstrates more fixedness than other sequences such as *take a BEATING/hammering* (see section 5.3). There were eight different noun substitutes found, and eight different verb substitutes. Table 5-15 and Table 5-16 show the different substitutes found:

Lexical Verb Substitute	Frequency
Bite	396
Hold	63
Watch	35
Mind	8
Button	6
Keep	4
Hush	3
Button up	1
TOTAL =	**516**

Table 5-15 – Lexical verb substitutes for *Watch your TONGUE/mouth*

Lexical Noun Substitute	Frequency
Lip	340
Tongue	117
Peace	24
Language	15
Mouth	14
Word	3
Noise	1
TOTAL =	**514**

Table 5-16 – Lexical noun substitutes for *Watch your TONGUE/mouth*

Table 5-15 and Table 5-16 show that there are some lexemes which are far more frequently occurring than others; *bite* as a verb and *lip* as a noun. The peripheral members for this

sequence are more frequent than peripheral members for, for example *take a BEATING/hammering,* or *shut your FACE/trap/gob/cakehole*, although this could be due to the high number of tokens for this case study. Note that Table 5-16 shows only 514 total nouns. The remaining two occurrences were the definite pronoun *it.* These occurrences were in the imperative form *button it* and have been included as the pronoun clearly replaces a noun phrase such as *your lip* in these cases. They have not been included in Table 5-16 as they are pronouns and do not fit the V det N structure, however, as with *it* in the form *shut it* (see section 5.5.1) they are included for conceptual relevance. Table 5-17 shows the most frequently occurring lexical combinations for this sequence:

Lexical Combination	Frequency
Bite + det + lip	334
Bite + det + tongue	63
Hold + det + tongue	43
Hold + det + peace	20
Watch + det + mouth	13
Watch + det + language	10
Watch + det + tongue	7
TOTAL =	**490**

Table 5-17 – Frequent lexical combinations of *Watch your TONGUE/mouth*[7]

Table 5-17 confirms with Table 5-15 and Table 5-16 that *bite* and *lip* are common to this phrase. The canonical *watch + det + tongue* or *watch + det + mouth* are rare in comparison. Discussion of the different lexical combinations and meaning will follow in section 5.6.2 with reference to the chaining process.

Of the noun substitutes, there is again systematic substitution for parts of the head where sound is produced; *mouth, tongue, lip.* Other nouns are extensions of the sounds that are produced when talking; *noise, language, word.* The noun *peace* is almost antonymic to the rest of the nouns; Table 5-17 shows that *peace* occurred predominantly in the combination *hold + det + peace,* with the meaning to remain quiet, rather than to stop talking.

[7] Of the examples of the combination *hold + det + peace*, the determiner was always a pronoun. This combination is on the boundary for the sequence *watch your TONGUE/mouth,* as a change in determiner from a pronoun to the definite determiner changes the meaning of *hold + det + peace* from "keeping quiet and not saying anything" to "minimizing conflict".

The meaning of *watch your TONGUE/mouth* is similar to that of *shut your FACE/trap/gob/cakehole*, (to stop talking), although it also has an extra sense of not being rude. There is also a tendency for the sequences to occur in different relationships. The sequence *watch your TONGUE/mouth* could be used by a superior to an inferior, whereas *shut your FACE/trap/gob/cakehole* can be used between equals. *Watch your TONGUE/mouth* is also thought to be a spoken sequence, however of the 516 token occurrences found, only 11 were spoken, and only two of these were in the canonical form and LID variant. A further 45 tokens were of reported speech within a written text, although again, there was more of a tendency for the forms *bite + det + tongue* or *hold + det + tongue* to be spoken than the original *watch + det + tongue/mouth.* Whereas *shut your FACE/trap/gob/cakehole* was primarily spoken in domain, this formulaic sequence had a preference for the imaginative domain. This was due to the type *bite + det + lip.* There were 334 occurrences of this type in the dataset, and 201 came from a Mills and Boon text, i.e. romantic imaginative (fictional prose) books of a low audience level written for a predominantly female audience, for example:

Example 5-21

"I shouldn't - well, it was at the party..." She stopped and **bit her lip**, then, seeing Sophie's colour rising, she added, "Edward told me that he had seen you and Robert kissing out on the balcony."
Vets in Opposition – M. Bowring
JYE 1941

Table 5-18 shows the domain for the type *bite + det + lip.* The high occurrences of imaginative fictional prose texts and frequency of tokens biases the results for the case study as a whole. Note that only two occurrences of this type are spoken.

Domain	Frequency	Normalized
Applied Science	1	0.24
Arts	2	0.48
Imaginative	319	28.23
Leisure	2	0.35
Social Science	3	0.39
World Affairs	5	0.53
Spoken	2	0.34
TOTAL =	**334**	

Table 5-18 – Frequency of the domain of all of the tokens of the combination *bite + det + lip*

Unlike the sequence *shut your FACE/trap/gob/cakehole,* the imperative is not so ubiquitous for this phrase; only 53 of the total tokens for *watch your TONGUE/mouth* were in this form. The majority were in the declarative. Contrasting with the verb *shut* in section 5.5, the verbs here showed tense and inflection:

> *Example 5-22*
>
> I would have liked to say that, so far, medical science had taught us nothing about polio, but I **held my tongue.** He looked at me more kindly. "You're thinking about the so-called "miracle cures" of Sister Kenny?"
> *The Other Side of Paradise – N. Barber*
> *CHG 1360*

Even in inflected form, the meaning of keeping quiet is retained, and the sense of it being a self- directed command, although in Example 5-22, the strength of the order is not the same as in the direct imperative cases.

5.6.2. Networks and Compositionality

The formulaic sequence *watch your TONGUE/mouth* forms a network via the chaining process with the sequence *shut your FACE/trap/gob/cakehole.* This section discusses the links between the two, the meanings, and the lexical combinations involved.

Starting with *shut your FACE/trap/gob/cakehole,* and as with all of the case studies, a Basic network can be constructed:

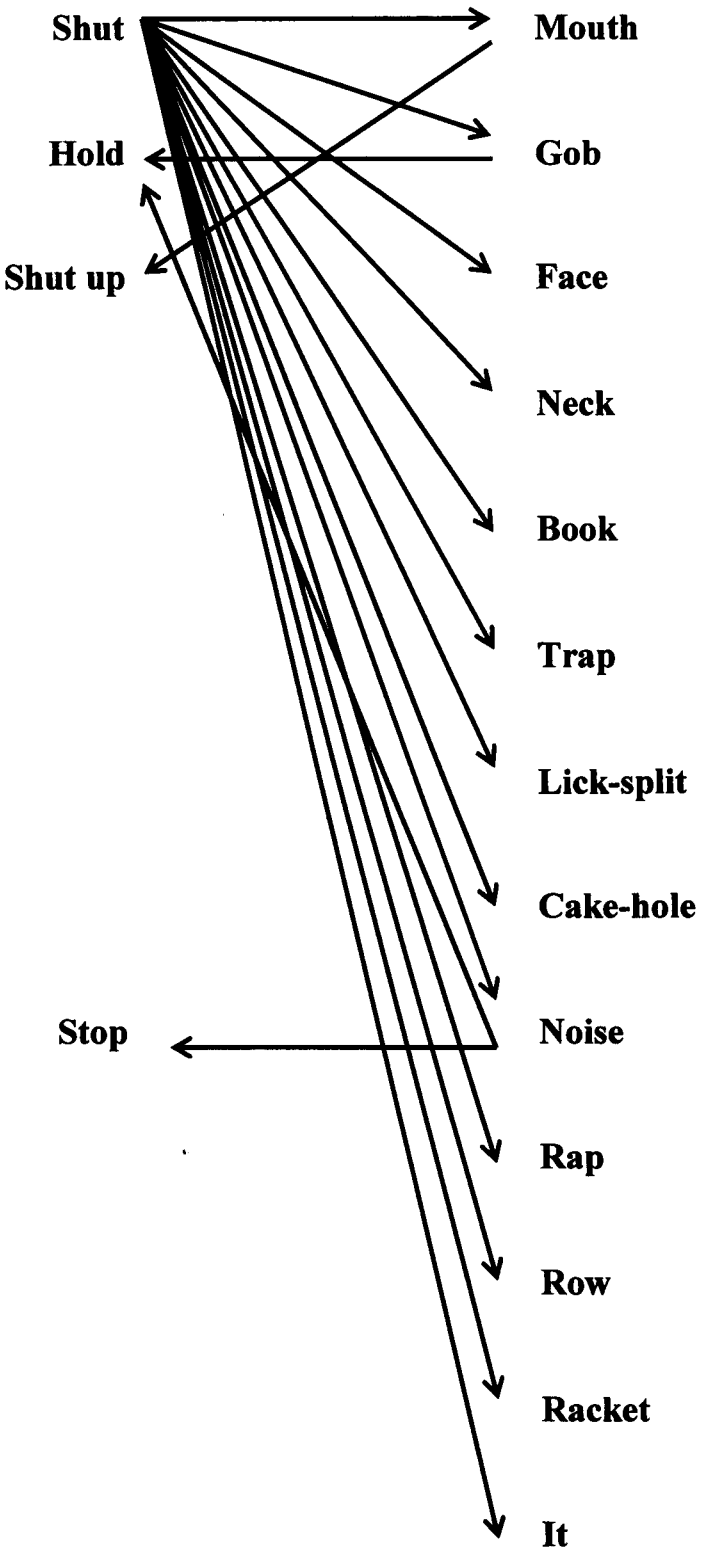

Figure 5-7 – Basic network for *Shut your FACE/trap/gob/cakehole*

With the imperative meaning of "be quiet", these all have the same meaning, and so the schema is formed. *Shut* forms combinations with all of the nouns in Figure 5-7. The noun *mouth* combines with *shut up. Gob* leads to *hold,* which in turn combines with *noise. Noise* also combines with *stop.* The sequence *shut up* is a sequence in its own right. It has 1279 occurrences in the BNC. It has not been included in the data for *shut your FACE/trap/gob/cakehole* as it does not have the V det N structure. The form *shut up* + *det* + *mouth* is an extension of *shut up.* Using the notion of prototypes, central and peripheral members can be imposed on the basic network:

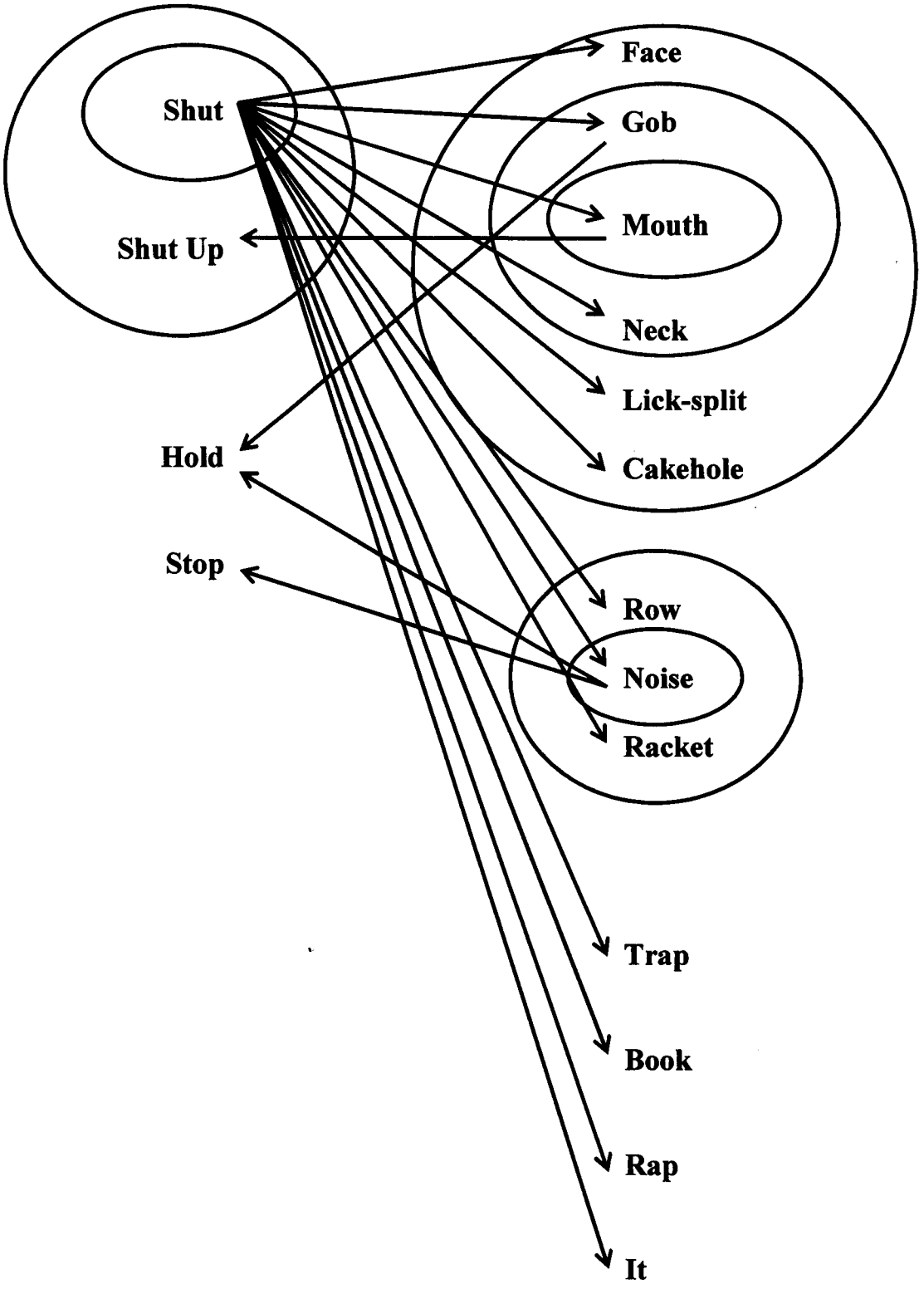

Figure 5-8 – Constellation network for *Shut your FACE/trap/gob/cakehole*

Figure 5-8 shows that the nouns *face, gob, mouth, lick-split, cakehole*, and *neck* are metonymic substitutes from the same semantic field, likewise *row, noise*, and *racket.* The verbs *shut* and *shut up* are also related in a prototype.

A separate process can also be carried out for the sequence *watch your TONGUE/mouth,* although not detailed here. The different nouns, verbs, and major lexical combinations encompassed in these processes were laid out in sections 5.5.1 and 5.6.1 respectively. As mentioned previously, *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* together share the same lexical set, and one large network. There are two nouns I have chosen as a dividing point between the two sequences *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* are the nouns *noise* and *mouth,* and the verb *hold.* These occur as tokens for *shut your FACE/trap/gob/cakehole* as:

Hold yer gob	A6C 1535
Hold yer **noise**	AEB 1245
Stop your **noise**	APW 169
Shut your **noise**	HTN 1547

There were also 54 tokens of *shut* + *det* + *mouth* and one of *shut up* + *det* + *mouth.* Whilst the two LID sequences *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* are united by a common sense of "stop talking", they are lexically linked by the noun *noise.* For the sequence *shut your FACE/trap/gob/cakehole,* there is a lexical link in a basic network:

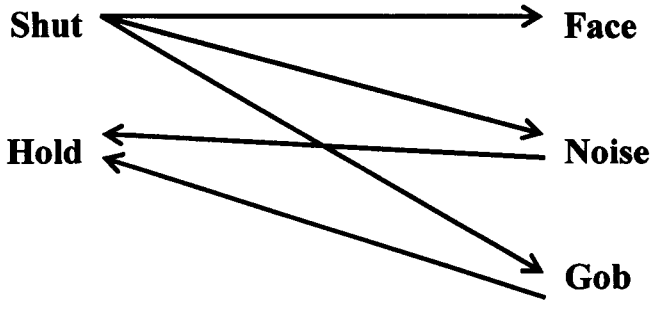

Figure 5-9 – Partial basic network for *Shut your FACE/trap/gob/cakehole*

For the sequence *watch your TONGUE/mouth,* there were examples such as:

Hush your noise	HGG 1733
Hush your mouth	KNV 905
Hush your tongue	HH1 5108

They chain lexically in a basic network from the canonical form:

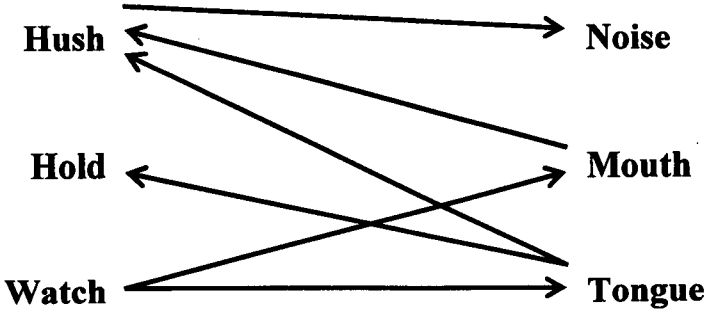

Figure 5-10 – Partial basic network for *Watch your TONGUE/mouth*

In Figure 5-10 the diagram has been drawn in reverse order, such that the first entries in the BNC are at the bottom and thus the diagram chains upwards in terms of entries in the BNC. This is so that the joining of the two sequences will be demonstrated more clearly.

The lexical phrase type *hush + det + noise* can chain onto *hold + det + noise* via the meaning of "keep quiet/stop talking", as seen in Example 5-23 and Example 5-24. I divided the sequences *shut your FACE/trap/gob/cakehole* at *hold + det + noise,* and for *watch your TONGUE/mouth* at *hush + det + noise.* This meant that the two datasets were kept separate for individual analysis, and there was no overlap of data in sections 5.5 and 5.6. *Hush + det + noise* is the cut-off point for dividing the two sections of the chapter only, not for dividing the two sequences.

Example 5-23

"You gonna let her get away with that?" howled Rose. "It stinks ! pooh!" "**Hold yer noise**!" said Thacker. "Ain't you gonna punish her? Hang on the weights! Hang on the weights!"
A Twist of Fate – P. Scobie
AEB 1245

Example 5-24

"About our business and the Lord Owen's," said Julian, with a vicious snap that belonged rather to a wife than a daughter, "and worse pressed than we are, very likely. **Hush your noise** and see to your journeymen, and I'll do the rest."
A Bloody Field by Shrewsbury – E. Pargeter
HGG 1733

The basic network for *shut your FACE/trap/gob/cakehole* develops as shown in Figure 5-11:

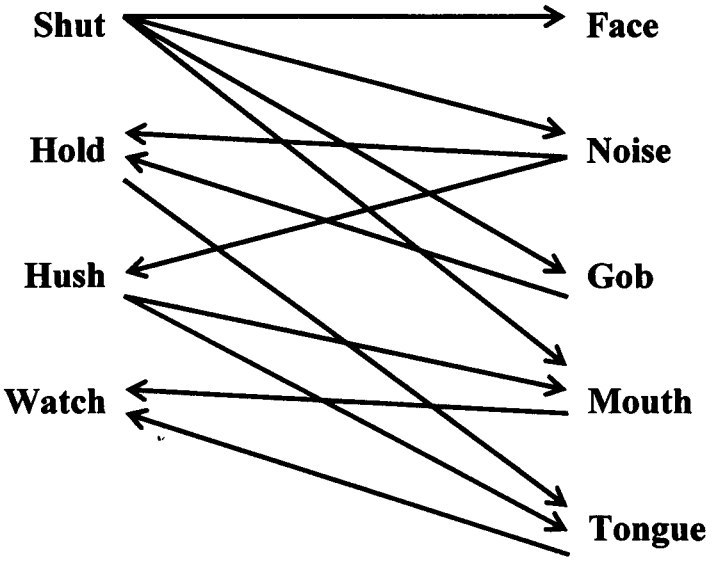

Figure 5-11 – Partial chaining network from *Shut your FACE/trap/gob/cakehole* to *Watch your TONGUE/mouth*

The central concept for this schema is to stop talking. However, as mentioned in section 5.6.1, there is an additional sense to *watch your TONGUE/mouth;* that of stopping being rude, as seen in Example 5-25. This creates the chain; two separate formulaic sequences are linked lexically, and with differing degrees of meaning throughout variant combinations in a network.

Example 5-25

At the rate things were going, Meredith could swallow up the budget for the entire year before the season was a quarter way through. Stella was forced to **hold her tongue** when Dotty or Babs Osborne spoke slightingly of Meredith.
An Awfully Big Adventure – B. Bainbridge
FNU 1353

The sequence *hold* + *det* + *tongue* has its own entry in the LID:

To stop yourself from speaking even though you want to say something, sometimes used to tell someone not to speak (Old fashioned)
(1998: 356)

This form occurs more in the data than the canonical *watch your TONGUE/mouth.* It is also found in the CIDI with the sense of "to stop talking" (1998: 397). It too notes the phrase's use to be old fashioned. It also suggests this phrase is often an order, a proposition not supported by the data here. From the LID and CIDI, the phrase *hold your tongue* is related semantically to *shut your FACE/trap/gob/cakehole,* having the meaning to stop talking.

The CCDI builds on the sense of restraint as shown in the Longman dictionary for this phrase, having an entry for *bite your tongue* and suggesting *hold* as a lexical verb substitute:

If you bite your tongue or hold your tongue, you do not say a particular thing, even though you want to or are expected to, because it would be the wrong thing to say in the circumstances, or because you are waiting for a more appropriate time to speak.
(2002: 373)

This sense of "not saying something which may be inappropriate" links the phrase *hold your tongue* to the phrase *watch your TONGUE/mouth.* The form *bite your tongue* has its own entry in the CIDI, suggested as a commonly occurring idiom with the meaning of stopping yourself from saying something inappropriate, even if you really wanted to say it (1998: 397). Again, this has the meaning of self-restraint. The form *bite your tongue,* also more common in the data than *watch your TONGUE/mouth* (see Table 5-17) lexically allows basic networking to the form *bite your lip:*

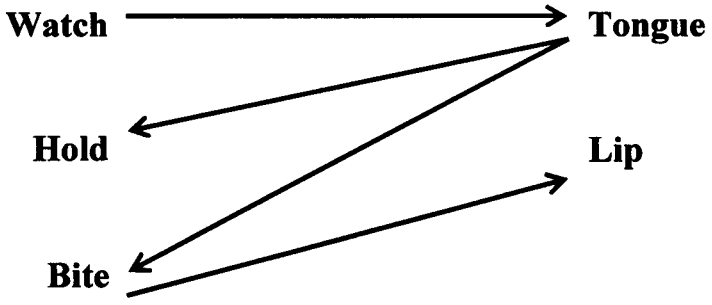

Figure 5-12 – Partial basic network for *Watch your TONGUE/mouth* including *bite* and *lip*

There was no entry for *bite your lip* in any of the three idiom dictionaries, despite Table 5-17 showing this was the most commonly occurring lexical combination. Example 5-26 shows that this phrase has the same meanings of keeping quiet and not saying anything inappropriate as for the examples for *hold* + *det* + *tongue* or *watch* + *det* + *tongue:*

Example 5-26

"Does it matter to you that I should be convinced?" "Yes, of course it does!" she all but yelled back, then **bit her lip**. Dammit, she was doing it again - letting her mouth speak before her brain was properly engaged.
Lover's Charade – R. Elliot
JY5 2064

It also has a sense of regret at one's words, and wishing it was possible to take back what was said. The reason why *bite your lip* may not have been included in the dictionaries is because several of the examples demonstrate a literal reading; literally biting one's lips to stop from talking or making a noise, or as a physical action whilst thinking for example:

Example 5-27

Standing close to him in the hallway, she looked up at him and **bit her lip**. "I need to think about this some more," she said.
Love Over Gold – S. James
GV8 2052

Even in the examples where the meaning is "to not say anything", as seen in Example 5-21 and Example 5-26, there is still a literal reading and image of biting the lips to not say anything, even if the person involved is not actually performing the action. The

combination *bite + det + lip* has been included in the dataset due to examples such as Example 5-21 and Example 5-26 having a meaning of "refraining from saying something". The combination is on the boundary of the sequence *watch your TONGUE/mouth* due to the number of occurrences with a literal reading. Rather than being a sequence with literal and non-literal readings, perhaps *bite + det + lip* is a literal sequence with an extended meaning. Note that combinations on the boundaries of networks usually have very few occurrences. High frequency supports the central concept, with rare examples being on the periphery. In the sequence *watch your TONGUE/mouth,* frequency does not determine the boundary. The combination *bite + det + lip* is on the periphery of the network for *watch your TONGUE/mouth* in terms of meaning and lexical substitutes, however it has a high frequency in contrast to usual boundary combinations.

The phrase *bite your lip* chains onto the form *button your lip:*

> *Example 5-28*
>
> Which is why I suggest you **button your lip**, except to mention that whatever she decides, you are on her side, supporting her through what's going to be a painful experience.
> *The Daily Mirror*
> *CH1 5127*

This has its own entry in the LID, with the lexical verb substitute *zip:*

> Used in order to tell someone not to say something unpleasant, or not to tell a secret (often used by adults speaking to children)
> (1998: 215)

The verb *zip* was not found in the data here. The meaning sense is the same as not saying anything inappropriate, although it also has a sense of not revealing anything. The chaining network is lexically and semantically so far thus:

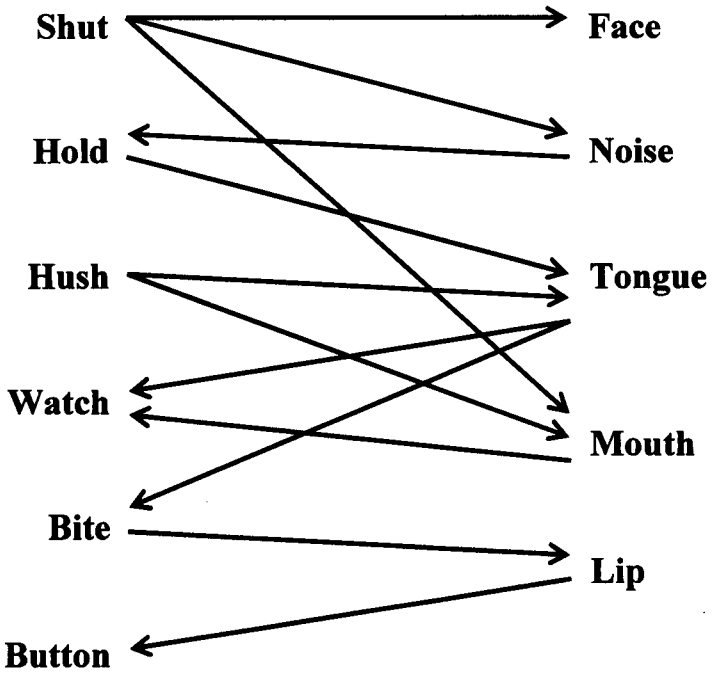

Figure 5-13 - Partial chaining network from *Shut your FACE/trap/gob/cakehole* to *Watch your TONGUE/mouth* including *bite* and *lip*

The meaning changes and develops from "stop talking" in the imperative in *shut your FACE/trap/gob/cakehole* to "be aware not to say anything inappropriate" (non-imperative) in *watch your TONGUE/mouth* to "not saying anything inappropriate and do not let yourself talk" in *bite your lip* to "not saying anything inappropriate or revealing anything" in *button your lip.* Uniting all of the sequences is the lexical set, and also an underlying sense of "keeping quiet", be it to stop talking, not say anything rude, or not to reveal anything. There is also a common concept of restraint, either imposed in an imperative order (*shut your FACE/trap/gob/cakehole*), or self-imposed (*bite your lip*).

The formulaic sequence *button your lip* also has an entry in the CCDI with the meaning "if you button your lip you keep silent about something although you would really like to talk about it" (2002: 233). It also suggests a variant form to be *button it,* which was found in the dataset here. The CIDI also has an entry for *button it,* with a meaning of an informal impolite way to tell someone to stop talking (1998: 56). The CCDI has a secondary sense for this phrase, meaning "to rudely tell someone to be quiet", as seen in Example 5-29, and which links the phrase to other phrases in the chain, such as *shut your face,* or *watch your*

tongue. The form *button it* also provides this linkage point via a connection to *shut it,* both having the meaning of "telling someone to be quiet", and having the imperative form.

Example 5-29

"So you lot better remember that, if you're awake and listening!" Then to Evelyn: "And you **button it.** I want my kip." She rolled on to her back to show that the conversation was finished.
A Twist of Fate – P. Scobie
AEB 861

The dataset also showed examples of *hold your peace,* and *keep your peace* which again had the same meanings of keeping quiet and not saying anything inappropriate, as seen in Example 5-30:

Example 5-30

So I decide to **hold my peace** for a little while longer. When her shift is over, I propel us straight to a Vintage Horror All-Nighter in an old cinema near King's Cross for some spiritual refreshment.
The Dyke and the Dybbuk – E. Galford
HGN 1395

The LID defines *hold your peace* as "to keep quiet and say nothing" (1998: 259), although it does not suggest *keep* as a verbal substitute. This formulaic sequence is not found in the CCDI or CIDI. Note that the dataset had an example of *watch your lip,* with the meaning of "not saying anything rude or inappropriate". The forms *watch/read my lips* have a different meaning; that of telling someone to listen to what you have to say. The combinations *hold + det + peace*, *keep + det + peace*, and *watch + det + lip* are on the boundaries of the network for *watch your TONGUE/mouth.* They represent the language user's knowledge of the central concept of restraint when talking and the associated lexical set. The language user selects the lexical items to fit the syntactic framework with the central concept. The examples of *watch your lip* and *keep your peace* also demonstrate the idea of a concept and lexical set. The speaker understands the sequences *watch/read my lips* and *keep the peace* and can manipulate these sequences and to fit the concept of restraint when talking. This manipulation by the speaker provides evidence against the dead metaphor, or atomistic list theories.

The chaining process thus lexically and semantically links the following sequences and their central and peripheral lexical members.

Shut your face/trap/gob/cakehole
Shut it
Watch your tongue/mouth
Bite/hold your tongue
Bite your lip
Button your lip/it
Hold/keep your peace

The links between the sequences may be tenuous, for example one-off occurrences which may have been counted in previous research as random occurrences. For example, the combination *hold* + *det* + *noise* with its singular occurrence provides a link from *shut* + *det* + *face* to *shut* + *det* + *noise* to *hold* + *det* + *noise* to *hold* + *det* + *tongue* to *watch* + *det* + *tongue.* The LID canonical forms *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* link together via the singular occurrences. This supports what Moon says (1998a: 164 – see section 2.4.3) whereby a metaphor stabilizes, destabilizes, and then restabilizes. The sequence *shut your FACE/trap/gob/cakehole* with its variants has a concept of stop talking in the imperative and with a young speaker age. Towards the edges of this concept, the metaphor destabilizes to occurrences such as *hold* + *det* + *noise, hush* + *det* + *noise* etc., and then restabilizes around the concept of restraint when talking and not saying anything inappropriate with the sequence *watch your TONGUE/mouth.* Combinations such as *hold your peace* and *bite your lip* are on the boundaries of the concept of *watch your TONGUE/mouth.*

As was seen in section 3.2.1.2 the LID, has an idiom activator, a section containing 10 basic concept words, each divided into groups of idioms which have similar meanings, and which is aimed to help people recognize which idioms are similar, and which are different. No information is given as to how these idioms are collected and grouped other than that they have similar meanings. One such concept word is *problem.* The LID did not have any concept words containing the formulaic sequences explained in sections 5.5 or 5.6 unlike

both the CIDI and the CCDI. The CIDI has a theme panel entitled "Speaking and Conversation". This has subdivisions:

> Talking too much
> Talking in a friendly way
> Good at talking
> Not talking
> Making someone tell you something
> Telling someone to be quiet
> (cf. Cambridge, 1998: 452)

Within the "not talking" subdivision, were included the phrases *bite your tongue,* and *hold your tongue,* and within the "telling someone to be quiet" subsection were the phrases *shut your face/mouth* and *button it.* The CCDI has a theme entitled "Revealing and Hiding" which contains the formulaic sequence *button your lip.*

Using the chaining process as shown in this section groups together the phrases *shut your face/trap/gob/cakehole, watch your tongue/mouth, bite/hold your tongue, bite your lip, button your lip/it,* and *hold/keep your peace,* as well as their central and peripheral substitutes, in a systematic manner via the lexical sets involved. The networks demonstrated in this section group sequences with an overarching topic of "Speaking and Conversation". The nodes of the networks are verbs and nouns which link to form combinations which have a similar or developing meaning and lexical set. Each node has the possibility of having prototypical central and peripheral lexical components, thus the network grows as members overlap.

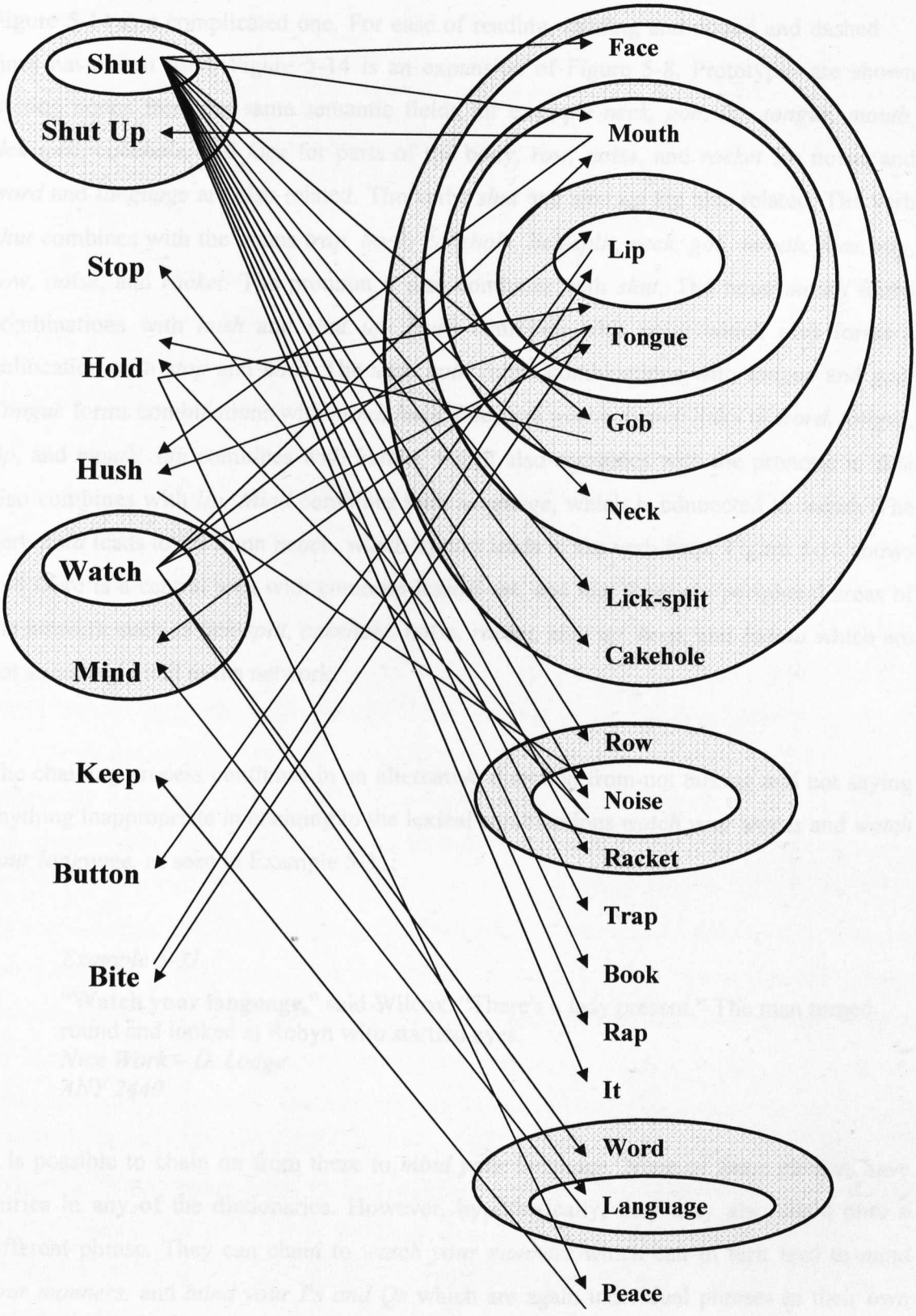

Figure 5-14 – Constellation network for *Shut your FACE/trap/gob/cakehole* and *Watch your TONGUE/mouth*

Figure 5-14 is a complicated one. For ease of reading, shading and dotted and dashed lines have been used. Figure 5-14 is an expansion of Figure 5-8. Prototypes are shown uniting nouns from the same semantic fields for example *neck, gob, lip, tongue, mouth, lick-split, cakehole,* and *face* for parts of the body, *row, noise,* and *racket* for noise, and *word* and *language* are also related. The verbs *shut* and *shut up* are also related. The verb *shut* combines with the nouns *trap, book, cakehole, lick-split, neck, gob, mouth, face, rap, row, noise,* and *racket.* The pronoun *it* also combines with *shut.* The noun *mouth* forms combinations with *hush* and *shut up. Hush* combines with *noise* which also forms a collocation with *stop* and *hold.* The verb *hold* forms combinations with *tongue* and *gob. Tongue* forms combinations with *mind, hush, bite,* and *watch. Watch* links to *word, tongue, lip,* and *mouth. Lip* combines with *button,* which also combines with the pronoun *it. Bite* also combines with *lip. Mind* combines with *language,* which is connected to *watch.* The verb *hold* leads to the noun *peace,* which in turn leads to the verb *keep.* Figure 5-14 shows that there is a central area with common lexical set, and that there are peripheral areas of the network such as *lick-split, cakehole, noise, racket, shut up, keep,* and *button* which are not so strongly tied to the network.

The chaining process continues in an alternative direction from not talking and not saying anything inappropriate in chaining to the lexical combinations *watch your words* and *watch your language,* as seen in Example 5-31:

> *Example 5-31*
>
> "**Watch your language,**" said Wilcox. "There's a lady present." The man turned round and looked at Robyn with startled eyes.
> *Nice Work – D. Lodge*
> *ANY 2440*

It is possible to chain on from these to *mind your language.* None of these phrases have entries in any of the dictionaries. However, hypothetically, they may also chain onto a different phrase. They can chain to *watch your manners* which can in turn lead to *mind your manners,* and *mind your Ps and Qs* which are again individual phrases in their own right, with a different meaning than the starting point.

Example 5-32

Apart from these, his back and legs were completely bare. Willie could not take his eyes off him. "Can I help you, son?" said the postmaster. Willie blushed and slid his card across the counter. The man glanced down at it. "Stayin' with Mr Oakley, eh? You'll have to **watch yer Ps and Qs** there."
Goodnight Mr. Tom – M. Magorian
CAB 726

The dictionaries have no entries for *watch/mind your manners,* but for *mind/watch your Ps and Qs,* they suggest a meaning of being careful about your actions or what you say in order that you do not offend anyone, or act/speak in an inappropriate manner. Where this refers to "speaking inappropriately" it chains from the sequences already seen. However, the meaning here develops into behaviour. The phrase *mind/watch your Ps and Qs* is grouped in the CCDI theme for "Communication and Relationships", and thus moves away from "Speaking and Conversation". The phrases *watch/mind your manners* and *watch/mind your Ps and Qs* have not been included in the dataset as the implication of action and behaviour changes the underlying meanings; the examples within the BNC referred to manners and Ps and Qs with a sense of conduct and etiquette, as seen in Example 5-32, rather than an emphasis on the spoken words, as seen in the rest of sections 5.5 and 5.6. Note that the examples *stop your noise (APW 169)* and *shut yer racket (ACK 2727)* in the case study *shut your FACE/trap/gob/cakehole* have a sense of "making a fuss, ado or commotion" as well as "to be quiet", and thus also have a behavioural tendency.

The combination *watch your language* marks the end of the chaining process. The meaning has developed semantically from the canonical form, and the lexical set has grown. However, continuing the chain would result in linking unrelated formulaic sequences, such as *watch your Ps and Qs* with *watch your tongue* and *shut your face.* So although sequences which have their own separate entries in dictionaries, such as *watch your tongue* and *shut your face,* they are connected in a network by a similar semantic meaning and lexical set. Note that Howarth's overlapping collocation clusters (see section 2.4.4) retain similar meaning but do not join unrelated sequences in a network (1996: 44). Once the lexical combinations lead to a change in the central concept, the end of the chaining process has been reached, and the networks are complete.

With respect to compositionality theory, as has been mentioned in sections 5.5.2 and 5.6.1, many of the lexical phrase types for this case study and *shut your FACE/trap/gob/cakehole* have literal, compositional meanings. The phrases *hold your mouth/tongue* and *watch your tongue/mouth* are literal if the verb senses are "to halt an intended action" in the former, and "to observe carefully" in the latter. Likewise for the verb *mind* if the sense is to be careful about. These refer to the phrase meaning of "not say anything rude or inappropriate". The phrase type *bite your lip,* as mentioned, is open to a literal reading, as are *keep your peace* and *bite your tongue.* The phrase becomes non-literal in the forms *hold your peace,* and *button your lip,* where the literality depends on the verb.

5.6.3. Summary

In summary for this case study, the phrase *watch your TONGUE/mouth* and its variant forms form a chain with *shut your FACE/trap/gob/cakehole,* and together with the variant types such as *bite/hold your tongue, bite your lip, button your lip/it,* and *hold/keep your peace* forms a thematic set with the heading "speaking and conversation" and general concept of restraint and "keep quiet and/or not say anything rude or inappropriate". The chain also has scope to follow on to behavioural phrases such as *mind your manners* and *watch your Ps and Qs.* A large proportion of tokens for this case study took the form *bite + det + lip,* which had a literal reading and was found in imaginative texts. This had a high frequency despite being on the boundary of the concept. The phrase *watch your TONGUE/mouth* and its variant types were found in fictional prose written texts, with very few spoken texts. Alongside *shut your FACE/trap/gob/cakehole,* there was a tendency towards the imperative, or self-command, although the direct imperative was not as common.

The individual LID sequences *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* demonstrate basic, chaining, and constellation networks. They also show how two sequences which are different in surface meaning and use, and thus listed separately in idiom dictionaries, can be connected with common concept and lexical set. They are connected at the boundaries by a rare one-off combination, which in some research projects may have been ignored as a random occurrence. The language user's knowledge of central concept and associated lexical set is exemplified here in peripheral

examples such as *keep your peace* and *watch your lips* which use and manipulate other formulaic sequences (*keep the peace* and *watch/read your lips*). This provides evidence against the dead metaphor theory.

5.7. Case Study 5: Flip your LID/wig

5.7.1. Definitions and Senses

The definition for this formulaic sequence within the LID is:

> To suddenly become very angry (Spoken)
> (1998: 208)

This is reflected in examples such as:

> *Example 5-33*
>
> Ten-year-old Opal (Gaby Hoffmann) and 15-year-old Erica (Samantha Mathis) simmer with resentment at their mother's abandonment and completely **flip their tiny lids** when she takes up with talent spotter Arnold Moss (Dan Aykroyd).
> *Today*
> *CBC 7017*

As will be seen in section 5.7.2, this formulaic sequence forms a chain with its lexical combinations to form an objective thematic group of phrases. There were 400 tokens for this formulaic sequence and 91 types, indicating a fair amount of fixedness. There were 11 noun substitutes and three different verbal substitutes as shown by Table 5-19 and Table 5-20:

Lexical Verb Substitute	Frequency
Lose	348
Blow	46
Flip	6
TOTAL =	**400**

Table 5-19 – Lexical verb substitutes for *Flip your LID/wig*

Lexical Noun Substitute	Frequency
Temper	239
Head	52
Cool	30
Top	27
Rag	16
Composure	11
Fuse	10
Gasket	6
Wig	4
Lid	3
Stack	2
TOTAL =	**400**

Table 5-20 – Lexical noun substitutes for *Flip your LID/wig*

Table 5-19 and Table 5-20 give the different frequencies of the nouns and verbs found. This sequence has a high number of noun substitutes, yet comparatively few verb substitutes; of the 400 tokens, the vast majority have the verb *lost.*

Lexical Combination	Frequency
Lose + det + Temper	238
Lose + det + Head	52
Lose + det + Cool	30
Blow + det + Top	27
Lose + det + Rag	16
Lose + det + Composure	11
Blow + det + Fuse	10
Blow + det + Gasket	6
Flip + det + Lid	3
Flip + det + Wig	3
Blow + det + Stack	2
Blow + det + Temper	1
Lose + det + Wig	1
TOTAL =	**400**

Table 5-21 – Frequent Lexical combinations of *Flip your LID/wig*

Table 5-21 shows the frequencies of all of the different lexical combinations for this formulaic sequence.[8] As will be seen in section 5.7.2, this case study forms a chain linking the three main verbs with their relative noun substitutes. At each stage there are central and peripheral members to the noun substitutes. Note that the combination *lose* + *det* + *temper* dominates this phrase's tokens. The different lexical combinations and their meanings will be discussed further in section 5.7.2.

The noun substitutes and types can be explained by metaphor (see section 2.2.5). A conceptual metaphor, discussed by Gibbs (1993) and relevant here is ANGER IS HEATED FLUID IN A CONTAINER. Phrases which Gibbs suggests relate to this metaphor include:

Blow your stack
Flip your lid
Hit the ceiling
Get hot under the collar
Lose your cool
Get steamed up

Gibbs tested people's descriptions of their mental images for each of these phrases. He expected to find that people would give similar descriptions for their mental images of each of the phrases because of the constraints of the underlying metaphor, for example MIND IS A CONTAINER, IDEAS ARE PHYSICAL ENTITIES, and ANGER IS HEAT. The results showed that 75% of participants had consistent general images, for example for *flip your lid* "...participants specifically imagined for [this phrase] some force causing a container to release pressure in a violent manner" (1993: 68) and 88% showed consistency in causes and consequences of the image:

[8] All 400 tokens are accounted for in Table 5-21 unlike in other case studies, as there were only two combinations with only one occurrence here. Other case studies had large numbers of combinations with only one occurrence, so only the frequent combinations were listed for those.

> When imagining anger idioms, people reported that pressure (i.e. stress or frustration) causes the action, that one has little control over the pressure once it builds, its violent release is done unintentionally (e.g. the blowing of the stack) and that once the release has taken place (i.e. once the ceiling has been hit, the lid flipped, the stack blown), it is difficult to reverse the action
> (1993: 68)

So for this case study, it seems that *flip your lid* and its variant *wig* and the type *blow your top/stack* are metaphoric in the sense of anger being heated fluid in a container. The type *blow* + *det* + *fuse/gasket* also has an image involving the concepts of ANGER and HEAT; as seen in section 5.7.2, if a fuse gets very hot, it blows, or breaks, and causes the equipment it was working in to cease functioning. If the phrase has a conceptual metaphor and a mental image, then the lexical substitutes are also limited to words which fit in with the image and conceptual metaphor. The type *lose* + *det* + *cool* also implies that HEAT leads to volatility. Note that, as was mentioned in section 2.2.5, metaphor is a specific type of figurative language; not all formulaic sequences are metaphoric.

The LID suggested that the canonical phrase tended to be a spoken one. However, the results here suggest otherwise. There were only 17 spoken tokens of the 400, and only one of these was from the canonical form. There were also 98 tokens of reported speech within a written text. This means that most of the tokens were written, and not written as reported speech, which contradicts the dictionary's findings.

5.7.2. Networks and Compositionality

The formulaic sequence *flip your LID/wig* forms basic and chaining networks, and unites separate formulaic sequences, grouping together phrases which have their own entries within idiom dictionaries in a systematic manner. The lexical nodes providing the step to the next lexeme in the network may only be single occurrences. These may be peripheral compared to other more frequent members of the lexical set however they are relevant with respect to the network.

The phrase *flip your LID/wig* was also found in the CCDI and the CIDI. The two definitions are:

> If someone flips their lid or flips their wig, they become extremely angry or upset about something, and lose control of themselves
> (Collins, 2002: 142)
>
> a To become crazy (humorous)
> b To suddenly become very angry (informal)
> (Cambridge, 1998: 225)

The combination *flip + det + wig* networks to the combinations involving the verb "lose" by the combination *lose + det + wig:*

> *Example 5-34*
>
> "Now, now Sherlock, **don't lose your wig**. Tim Wilson wants to see you." "What about?" "Oh, he's heard you're chasing the five grand and wants to make sure you don't know something he doesn't."
> *Murder Forestalled – P. Chester*
> *FAP 1782*

This retains the sense of "be angry"; however it also has a sense of losing control. Note that networking *flip + det + wig* and *lose + det + wig* would be a basic network as far as the LID and CIDI are concerned, as the meaning of being angry and losing control is retained. For the CCDI, this is a chaining network, as the meaning is developing from that of simply being angry to one containing a connotation of losing control.

Table 5-21 showed that the combination *lose + det + wig* had only one occurrence in the BNC. However, this occurrence provides the connection point to the sequence *lose + det + temper.* The connection between *flip + det + wig* and *lose + det + temper* is thus tenuous. If *lose + det + wig* had not occurred, there would be no link. It is a case where the language user recognizes the sequences *lose your temper* (literal) and *flip your lid/wig* and can manipulate the central concept and lexical set.

The non-literal type *lose + det + head* is found in all three of the idiom dictionaries. In the LID and CCDI, it has the sense of losing control; "to behave in an unreasonable way when you are in a difficult or worrying situation" (Longman, 1998: 161). Note that the CCDI suggests this to be a "key" or important idiom. The CIDI definition gives the type a

meaning of suddenly becoming very angry. This "angry" sense is found here in the data in examples such as:

Example 5-35

My language became worse. I then walked across to the photographers and lost my temper, **lost my head.** I thought that I had been set up. I was as angry as hell.
Linford Christie: An Autobiography
BMM 2027

Example 5-35 uses two phrases to express the anger; *lost my temper* and *lost my head.* This repetition emphasizes the emotional feeling. There is also a development of the anger in the use of *lost my head;* not only is it used to show the anger, but it also could mean that the person lost their self-control in the situation.[9] Example 5-36 shows the development of meaning; anger in Example 5-35 to loss of control in Example 5-36. It is possible to replace *lost my head* in Example 5-35 with the canonical *flip you LID/wig,* however *lose your head* in Example 5-36 could not be substituted and the example retain the same meaning. The sense of loss of control was more common throughout the examples of *lose + det + head,* however examples such as Example 5-35 show that it is still used with an angry meaning, which demonstrates the strength of the network relationships.

Example 5-36

Again he ran his left hand through his hair, but this time he felt a tingle of anticipation. There was always something around the corner if you didn't **lose your head.** He had a definite feeling of success about the Carefree deal
Crimson – S. Conran
FPB 113

A similar example where the development of meaning can be seen is in the combination *lose + det + cool.* Example 5-37 and Example 5-38 show this, the former demonstrating anger, the latter self-control:

[9] The rules of the Chaining process also allowed for literal examples such as "...that Louis XVI was held in Paris (and soon to **lose his head**)" *(B7l 1355).*

Example 5-37

In this situation, it hardly ever helps if you start shouting or **losing your cool.** The best approach is to be patient and philosophical. If you arrive late, raise merry hell, and insult the stage crew you will certainly be remembered.
The Rock File – N. York
A6A 114

Example 5-38

He's encountered the weight of my right boot before. "Don't **lose your cool,** Gowie," he says, quickly. "I was only joking. I was only joking. "
An Alternative Assembly Book – M. and L. Hoy
ALH 2621

Lose + *det* + *cool* had the sense of being very angry in all three of the idiom dictionaries. The CCDI and CIDI also added a sense of behaving in an uncontrolled or bad-tempered manner in the former, and of shouting in the latter. Thus the phrase means not only being very angry, but acting angry too. Example 5-37 shows emphasis via repetition; *start shouting* and *losing your cool.* It also uses the phrase *raise hell* with the insertion *merry,* using two formulaic sequences (*losing your cool* and *raise hell*) and an insertion to create an emphatic effect.

The "angry" sense leading to loss of self-control follows in the examples of *lose* + *det* + *composure.* This phrase type has a literal reading, meaning to keep one's calm and poise. There were also examples for this phrase where, as with Example 5-35, there was a dual meaning of being angry and losing self-poise, such as "...Perhaps that would explain it, why he'd **lost his composure** with Antonini" *(JY7 292).*

The combination *lose* + *det* + *wig* also leads to the combination *lose* + *det* + *rag.* All three dictionaries suggest this means "to be very angry", and of typically British English origin. The CIDI also suggests behaviour of shouting too. Again, this is a non-literal type. Most of the *lose* + *det* + *N* examples were of the form *lose* + *det* + *temper.* This has a non-literal reading, and there were 238 examples of the 400 with this form (see Table 5-21).[10] The meaning throughout was of being very angry:

[10] Note that this combination is so commonly occurring that it may be overlooked that it is non-literal, and it is assumed to have a literal reading.

Example 5-39

Losing his temper, Howard roared, "I am that man. I stand here for the President, and there is no spirit good or bad who will hinder me.
Chief Joseph: Guardian of the Nez Perce – J. Hook
ALX 501

So far, there is a chaining network which links the canonical *flip your LID/wig* via *lose + det + wig* to *lose + det + temper,* incorporating a semantic development to phrases such as *lose + det + head/composure,* which also have a sense of loss of self-control as well as being angry:[11]

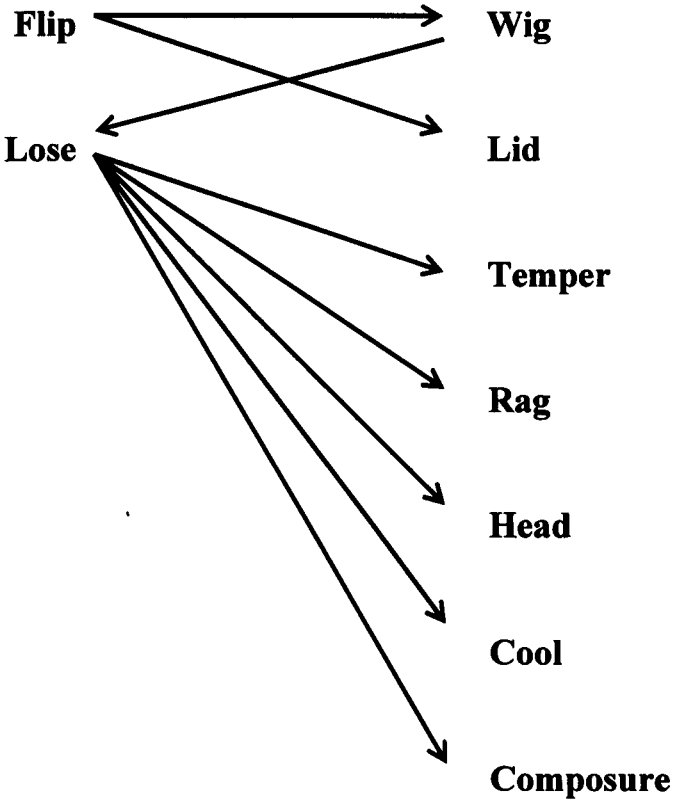

Figure 5-15 – Partial chaining network for *Flip your LID/wig*

[11] Whilst there are central and peripheral lexemes for this sequence, they do not form constellation prototypes in the same way as, for example the nouns *hammering, beating* etc. The nouns for this sequence are separate in meaning from each other so there are no constellations here.

The phrase *lose + det + temper* can chain in a basic network onto the other types as seen in Table 5-21 (*blow your top/fuse/stack/gasket)* via the singularly occurring type *blow + det + temper.*

> *Example 5-40*
>
> Instead of putting them down for the afternoon, I'd like to be able to do something with them. But I do think if I don't have a break, then before long I shall **blow my temper**.
> *Discipline for Parents – M. Herbert*
> *B10 352*

Note that this type has a non-literal reading. Again the singular connection point is weak, yet demonstrates manipulation of the sequences, concept and lexical set. It seems there are three definite formulaic sequences in this case study; *flip your lid/wig, lose your temper* and *blow your fuse/gasket,* each with central and peripheral combinations. It seems this is a case where the metaphor in the original *flip your lid/wig* destabilizes and restabilizes via *lose your temper* to *blow your fuse/gasket.* The connection points *lose + det + wig* and *blow + det + temper* are on the boundaries between the sequences.

The formulaic sequence *blow your top/stack* was found in all three of the idiom dictionaries, and all three maintain the sense of "to be very angry", with the LID again incorporating the feeling of losing poise:

> To get so angry about something that you lose control of what you are saying or doing
> (1998: 357)

The CCDI separates the substitutes into two entries, *blow your top* and *blow your stack,* cross-referencing each other, and also suggests the action of becoming angry may be accompanied by shouting. The CIDI also puts the substitutes in reverse order; *stack* before *top,* suggesting the latter to be a substitute of the former, which is not supported by the data here, as Table 5-19 and Table 5-20 show. The phrase *blow your fuse* was also present in all of the dictionaries, although its substitute *gasket* was only found in the LID and CIDI. The LID definition is of being angry and upset, whilst the CCDI and CIDI entries also suggest

behaving in a violent, uncontrolled manner. The CCDI dictionary also explains the origin of the idiom:

> If you blow a fuse, you suddenly lose your temper, and cannot control your anger
> Note: A fuse is a safety device found in electrical equipment. If this equipment becomes too hot, the fuse blows, or burns. This breaks the electrical circuit, so that the equipment will stop working.
> (2002: 149)[12]

The use of the phrase with variants *fuse, gasket, stack* or *top* is used in the same way as the canonical *flip your LID/wig* or *lose* + *det* + *temper/rag,* as Example 5-41 shows. Note that in this example, the phrase is quoted, emphasizing it for the reader and thus showing the author recognizes it has a metaphorical use:

> *Example 5-41*
>
> Not for nothing was he known as Windy - when upset he really did "**blow his top**" but his bark was worse than his bite - he was an excellent teacher and took both Standards Six and Seven.
> *I Remember, I Remember – A. Maidment*
> *B22 43*

The full chaining network for the sequence *flip your LID/wig* is thus:

[12] Note that this definition includes the non-literal sequence *lose your temper.*

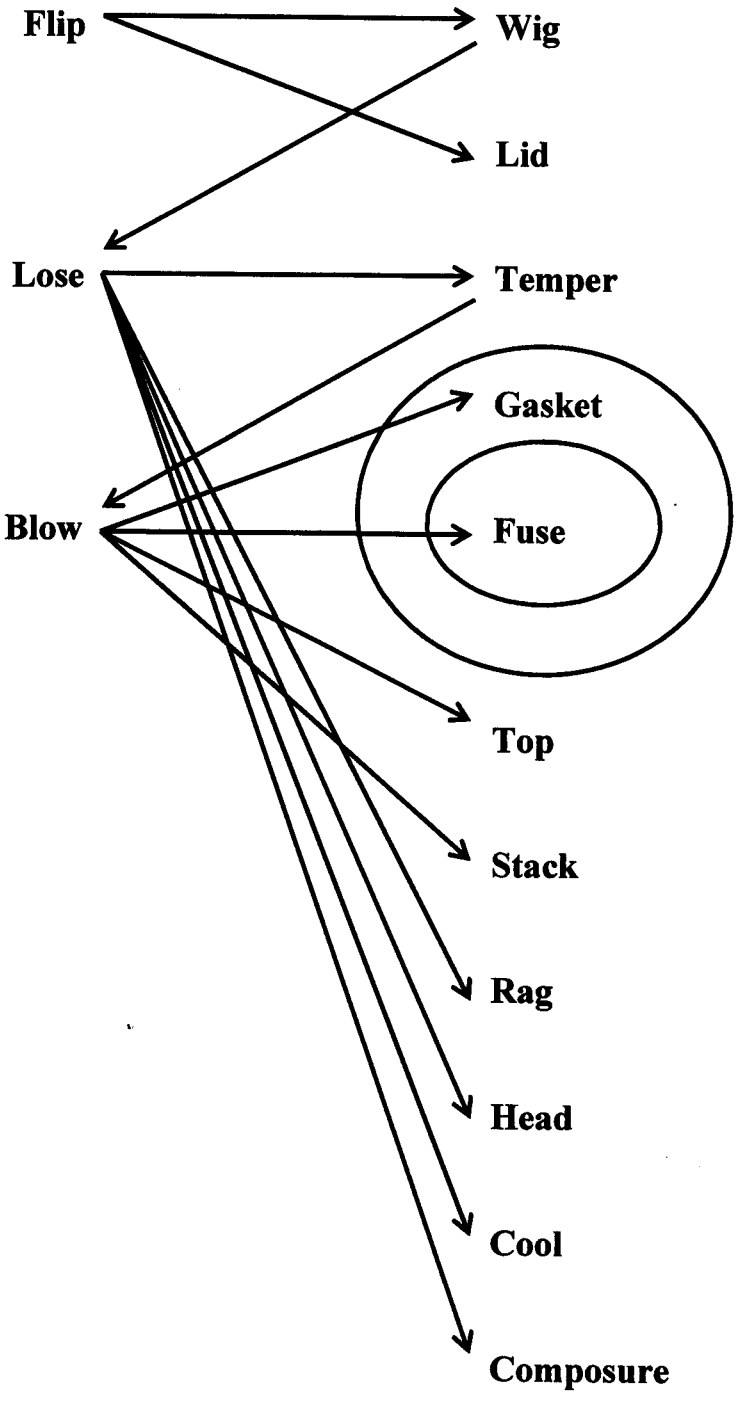

Figure 5-16 – Constellation network for *Flip your LID/wig*

Throughout this network, the same meanings of "being very angry" and "losing self-control" are present. As with *watch your TONGUE/mouth* in section 5.6.2, the chaining process has given a systematic objective method of grouping, or networking formulaic sequences with similar meanings. In all three of the dictionaries, there were thematic panels for the concept of ANGER. The Longman idioms activator has the concept word ANGRY with subdivision "become suddenly very angry", which contains the formulaic sequences *blow a fuse/gasket, blow your top/stack, flip your lid/wig,* and *lose/blow your cool.* Note that it also contains the phrase *have/throw a fit,* linking this case study to section 5.4. Another subdivision "angrily complain about something" contains the case study seen in section 6.7 *Raise HELL/Cain.* The CIDI has a theme panel entitled ANGER, with subdivision "being angry", containing only *blow a fuse* from this case study, but also *have/throw a fit* again. The Collins thematic index has a theme ANGER AND IRRITATION containing the phrases *flip your lid/wig, blow a fuse, lose your rag, blow your stack, blow your top,* and *throw a wobbly/wobbler.* Note that expanding the syntactic requirements of this study from V det N to PP det N would lexically relate *throw a wobbly* to *flip your lid,* via *have/throw a fit* to *in a fit* to *in a temper* to *lose your temper.* The chaining process has a clear end point for this sequence, it "runs out" after *blow your top/stack/gasket/fuse.* Without altering the syntactic structure, there are no further lexemes to chain to.

5.7.3. Summary

In summary for this case study, the canonical formulaic sequence *flip your LID/wig* forms a chain with three distinct points; the form *flip your lid/wig, lose + det + N,* and *blow + det + N.* This chain is open to metaphoric imagery. The chain lexically and semantically links phrases of the same meaning in a more systematic and objective manner than in idiom dictionary thematic panels. The central concept for this case study is ANGER. The chain also matches other chains in having a non-literal start and end point, and in having a highly token-frequent mid-point, as the sequence destabilizes and restabilizes. This phrase was mainly imaginative in domain, influenced by the literal type *lose + det + temper.*

5.8. Conclusion

It is clear from these case studies that far more lexical substitutions occur in use than are accounted for in The LID. The chaining process is a set of rules allowing for a systematic method of investigating the maximum different lexical substitutions in the BNC.

Basic networks link phrases with similar meaning and lexical set, for example *take a BEATING/hammering.* This sequence also demonstrates a product of the chaining process; frequent and rare, or central and peripheral lexical members of the basic network, and these can be demonstrated as prototypical members, or rare occurrences. The rare substitutes demonstrate the outer limits of the sequence's substitution, and also in the cases of the spoken occurrences of *scranny* (*throw a WOBBLER/wobbly)* and *lick-split* (*shut your FACE/trap/gob/cakehole*) demonstrate that the peripheral members may be more colloquial or "slangy" than the central lexical members.

Chaining networks are a development on basic networks, grouping phrases of similar and developing meaning and lexemes, such that sequences which have separate entries within dictionaries can be grouped in a systematic manner, for example *flip your lid* with *blow your stack* etc. Constellation networks unite the chaining networks and the concept of central and peripheral lexical members, and develop the theories of networks from those given in section 2.4.4 by illustrating the frequent and less frequent lexical substitutes. The networks also provide evidence as to the boundaries between one formulaic sequence and the next, showing that there are no clear-cut boundaries for sequences.

The use of networks contradicts the traditional view of formulaic sequences forming a list, each having to be learned and stored separately. There is more variation than provided for in such atomistic theories, and the networks demonstrate that sequences are much more interactive with each other, and not separate comprehensive units. It is possible that idiomatic expressions are stored in a similar way to words in semantic fields. This is being developed in idiom dictionaries through thematic panels (see section 3.2.1.2). There were groupings found here for the cases *watch your TONGUE/mouth* and *flip your LID/wig.* Conceptual metaphors were seen to be beneficial in grouping formulaic sequences.

The networks suggest productivity of language users. They recognize the central concepts, associated syntactic structure and lexical set and can manipulate them. This creates innovative combinations, such as *blow + det + temper* or *hold + det + noise.* These one-off occurrences provide the opportunity to create networks connecting sequences. In terms of frequency, the one-off occurrences are at the boundary between sequences. The boundaries are reached when there are no new lexical items found, and also when there is a big enough change in the common central concept, such as from restraint and not saying anything inappropriate to behaving appropriately for the sequence *watch your TONGUE/mouth.* Combinations not found in networks do not demonstrate gaps or impossible combinations. Instead they represent where a user has not created that particular combination in the BNC data.

The findings also provide evidence for compositional theories of idiomatic phrases. None of these cases have a meaning derived in the same way as, for example *kick the bucket* – a completely different meaning from the words that make up the phrase. The case studies here show that the meaning can be built up along the phrase, but that individual parts of the phrase may be idiomatic, whilst others retain their literal meaning, for example *take a beating* has a literal verb, but idiomatic reading of the noun.

This chapter has built on themes discussed in Chapter Two of compositionality and networks, using five case studies of phrases from the dataset showing the most lexical substitutes. The next chapter develops the notions of compositionality, and the boundaries of variation, by investigating inserted items into formulaic sequences.

6. Insertions

6.1. Introduction

This second data chapter focuses on another type of variation found in formulaic sequences; that of insertion of elements within the sequence. This chapter investigates lexical and grammatical insertions, and focuses on the five expressions from the dataset which showed the highest number of total insertions. Previous research in the field of phraseology has focussed predominantly on substitutions; there has been comparatively less work done on insertions. Each case study follows the same organisation:

- Definition and Senses - the dictionary definition and variants, substitutes found using the chaining process
- Networks and Compositionality - discussing the findings from use of the chaining process with respect to theories of networks and compositionality
- Insertions - discussing the lexical and grammatical insertions found
- Summary - relating the findings to the research questions

The boundaries of variation are investigated in this chapter on the following sequences: *take a BEATING/hammering, overstep the MARK/limits/bounds, pin your HOPES/faith on, kick up a STINK/fuss*, and *raise HELL/Cain.*

6.2. Overview of Formulaic Sequences

Table 6-1 shows the frequencies for the formulaic sequences studied in this chapter:

Phrase	Tokens	Types	Lexical Insertions	Grammatical Insertions	Total
Kick up a STINK/fuss	379	171	74	92	166
Overstep the MARK/limits/bounds	289	145	99	0	99
Raise HELL/Cain	427	128	72	22	94
Take a BEATING/hammering	112	90	37	14	51
Pin your HOPES/faith on	215	140	15	24	39

Table 6-1 - Five sequences from the dataset showing the highest number of insertions

Table 6-1 is taken from Table 4-7 and Table 4-8 in section 4.3.2 which show these figures for all of the phrases in the dataset, and which emphasize that these five sequences have the most total insertions. Table 6-1 also shows the individual frequency of lexical insertions and grammatical insertions for each phrase. The phrase *overstep the MARK/limits/bounds* has a high number of lexical insertions, yet no grammatical insertions. As will be seen in section 6.4, most of the tokens of this phrase with insertions have a literal reading. There are fewer tokens of this phrase which have a figurative reading and also a lexical insertion. In the non-literal reading, there is a definite determiner and few examples of insertions. This supports the findings of Nicholas (1995) (see section 2.3.3). With a literal reading, there are fewer constraints on modifications.

In a general overview, the five sequences are primarily written. The sequence *pin your HOPES/faith on* has only 8 spoken occurrences out of 215 (3.27% of the total tokens for the sequence are spoken), *overstep the MARK/limits/bounds* has 15 spoken tokens in 289 (5.19%), and *take a BEATING/hammering* has 7 tokens out of a total 112 in the spoken domain (6.25%). The sequence *kick up a STINK/fuss* had 44 spoken tokens in its overall 379 tokens (11.61%), whilst *raise HELL/Cain* had 27 spoken tokens in 427 total occurrences (6.32%). The case studies had the majority of occurrences in the BNC year range 1985 – 1994, with *raise HELL/Cain* and *kick up a STINK/fuss* having occurrences in the 1975 – 1984 range. *Overstep the MARK/limits/bounds, raise HELL/Cain*, and *kick up a STINK/fuss* also had tokens in the range 1960 – 1974.

In terms of general distribution, such as audience age and sex (adult, mixed gender), education level (low), and author details (male, aged 35 – 59), these phrases are very similar. Such details do not provide useful information here, so are not focussed on within the case studies. More in-depth differences with regard to context and use will be looked at in the following sections, as relevant in determining one formulaic sequence or variant form from another.

In most cases here, the actual insertion, either lexical or grammatical, may only occur once per phrase. This is not enough to be representative as a commonly occurring insertion; however the fact that it is an insertion means that it should be taken account of. As will be seen, as with substitutions, there are frequently occurring insertions, specifically grammatical insertions; lexical insertions are more individual occurrences. Thus there can be central and peripheral insertions. As with substitutions, the limits of variation – insertion in this case – are reached when no new occurrences are found. Frequently occurring insertions are more central, whereas singular occurrences represent the outer limits of insertions.

6.3. Case Study 6: Take a BEATING/Hammering

Take a BEATING/hammering was one of the top five phrases in the dataset with regard to number of lexical substitutions of the noun, and it is also one of the top five phrases with regard to number of insertions. This phrase was investigated in terms of definition, lexical scales, frequent and peripheral members, composition, and register in section 5.3. In this section, it will be investigated in terms of insertions, and the issues regarding compositionality will be extended. This section thus only has two subdivisions compared to others in this chapter; *insertions* and *summary*.

6.3.1. Insertions

There were 112 token occurrences of the sequence *take a BEATING/hammering* in the dataset, with 37 lexical insertions and 14 grammatical insertions. The grammatical insertions form a smaller "set" in the vocabulary in general than the adjectives available for insertions. Of the grammatical insertions found for *take a BEATING/hammering,* there

were only five different forms found (see Table 6-2). These insertions are a mix of pre- and post- determiners (Greenbaum and Quirk, 1990: 77). The phrasal types *a bit of a, a hell of a,* and *something of a* are quantifiers. As Greenbaum and Quirk say (1990: 77), when preceded by *a,* as in *a hell of a,* there is a positive meaning, i.e. "some/several/a lot", whereas when the initial indefinite article is missing (as in *something of a*), there is a negative meaning, such as "a little/hardly any".[1]

Grammatical Insertion	Frequency
A bit of a	5
A hell of a	2
Quite a	3
Something of a	2
Such a	2
TOTAL =	**14**

Table 6-2 – Frequency of grammatical insertions for *Take a BEATING/hammering*

In contrast, of 37 lexical insertions, there were 26 different types. The adjectives *biggest, fair old, terrible, fair, good,* and *severe* all occurred more than once. The grammatical insertions demonstrate more central properties, whereas the lexical insertions, each occurring relatively few times, demonstrate more peripheral properties. Quirk and Greenbaum (1973: 125) suggests a sub-categorizing scheme for adjectives, and the order they occur in within a phrase:

A Intensifying adjectives
B Post-determiners and limiter adjectives
C General adjectives susceptible to subjective measure
D General adjectives susceptible to objective measure
E Adjectives denoting age
F Adjectives denoting colour
G Denominal adjectives denoting material or resemblance to material
H Denominal adjectives denoting provenance or style

[1] The phrasal quantifier *a bit of a* is an exception. This acts as a hedge to create a negative, or weakening effect.

Such categories are revised by Dixon (1991: 78 – 79), adding classes such as "Speed" (e.g. *fast, slow*), "Human Propensity" (with subgroups such as "angry" *angry, jealous, mad,* "clever" *clever, stupid, kind,* and "happy" *keen, thankful, happy*), and "Similarity" (*like, similar (to), different (from)*). The most useful scheme is Quirk and Greenbaum's, due to the category of intensifying (emphasizing or downtoning) adjectives.

Of the insertions found for this phrase (note the lexical insertions were all adjectival), most of the occurrences (28 tokens) were of a general nature, susceptible to subjective measure, for example:

Take a **proper** hammering	ABS 71
Get a **right** hammering	ALL 747
Had a **good** bashing	FPR 1074
Taken a **severe** thrashing	CBE 1759

However, at the same time as being subjective, these insertions were also intensifying; having a heightening (increasing) or lowering (decreasing) effect on the noun they are modifying. These would be called *emphasizers* by Quirk and Greenbaum (1973: 121). They have a general intensifying effect. Quirk and Greenbaum call *amplifiers* those intensifiers which denote the extreme heightening or lowering effect.

Three tokens were general, open to an objective use using Quirk and Greenbaum's scheme, for example *administered a **56-13** thrashing (CBE 1759).* This is objective as it reflects the actual size of the loss. Two of the tokens reflected the style of the loss, e.g. *took a **physical** battering (CU1 1694),* and two tokens reflected the age of the beating: *took a* **new** *battering (CH2 2206).*

Of the 37 lexical insertions, 29 were single adjectives. The remaining eight tokens were two word insertions, e.g.:[2]

[2] The insertion *fair old* occurred three times.

Took **yet another** hammering	KCL 5397
Suffering a **truly horrific** hammering	K3A 782
Took a **fair old** hammering	K1B 3082
Had a **right old** hammering	KDM 9581

The examples *yet another* and *truly horrific* consist of an adverb modifying the adjective insertion. As Quirk and Greenbaum say (1973: 127), often, the adverb intensifies the adjective, which seems to be the case here. The adverbs seem to be susceptible to subjective measure and emphasize the subjective adjective. Such subjective insertions seem to be attitudinal as compared to objective or concrete insertions. Objective insertions are specifiers rather than emphasizers.

The examples *fair old* and *right old*, whilst appearing to have a subjective adjective modifier preceding an age adjective, act as being two subjective modifiers. The *old* here does not refer to the age of the beating as the *new* does in *took a* **new** *battering* (CH2 2206). Instead, it seems to be colloquial collocation with an intensifying adjective, as demonstrated by the severe criticism seen in example:

> *Example 6-1*
>
> We opt for The Baker's Wife at the Phoenix and all vote it a definite winner, and I wonder again at the jaundiced palate of critics who **gave it a fair old drubbing** when it first trotted out
> *The Guardian: Electronic Edition: Leisure Pages*
> *AAV 691*

As seen previously with the order in which modifying adjectives occur, *fair old* and *right old* follow the correct order having an intensifying adjective preceding the general subjective adjective. Evaluative modifiers occur before objective ones. In reference to section 2.3.3, insertions such as *old, proper, right* or *new* would be classed as being internal. They modify the noun. All but one of the insertions for *take a BEATING/hammering* were internal. Only the token *gave the British a* ***public*** *drubbing (ACS 1760)* could class as being external; modifying the whole sequence.

The sense of being completely defeated was most popular in lexical insertions, and occurred in reference to sports games, generally in newspapers, as Table 6-3 shows:

Sense	Lexical Insertions	Grammatical Insertions	Total
Severely Criticized	7	7	14
Complete Defeat	13	3	16
Economic loss	6	0	6
Damaged/Well-Used	11	4	15
TOTALS =	**37**	**14**	**51**

Table 6-3 – Senses of tokens of *Take a BEATING/hammering* which had insertions

The results here match what was seen in section 5.3.1; the senses of damage and defeat were most common, whilst economic loss was the least frequent sense of all the tokens. Of the nouns involved, all but *licking* and *trouncing* were found to take lexical insertions, although these nouns each only occurred in one token. Only the nouns *hammering, bashing, beating*, and *battering* were found to take grammatical insertions. *Thrashing* and *drubbing* were seen in section 5.3.1 to have a slight tendency towards a sense of defeat.

Both literal and non-literal examples can take insertions, either lexical or grammatical. Example 6-1 and Example 6-3 demonstrate a non-literal reading (criticism in the former, defeat in the latter) with lexical and grammatical insertions respectively. Example 6-2 gives an example of a literal reading with lexical insertion:[3]

Example 6-2

...some of the playing and "rat-at-tat" rhythms sometimes suggest Prokofiev (the pianos **take a fair bashing)**
CD Review
BMC 2489

Table 6-2 shows the grammatical insertions found. One such example can be seen in:

[3] Example 5-5 Chapter 5 demonstrated the singular example of a literal reading with grammatical insertion.

Example 6-3

Had been appalled to discover, through the industry grapevine, that I'd been allowed to continue with the assignment after **taking such a beating** from Ben Issachar.
The Dyke and the Dybbuk – E. Galford
HGN 145

Example 6-3 is one example of a grammatical insertion with a sense of being severely criticized. The effect of the insertion is the same as that in Example 6-1; the effect is of modification, and magnification of the phrase; it is not just criticism, it is a severe criticism that is given. Stylistically, however, use of functional words is not so "colourful" here as in lexical insertion examples. Although Ernst (1981 – see section 2.3.3) focuses on adjectives only, grammatical insertions could be seen to act internally; again on the noun rather than on the phrase as a whole.

Quirk and Greenbaum (1973: 428) describe the insertion *such a* as being a non-correlative determiner group; it occurs in colloquial language, and adds exclamatory force, or emphasis to the statement. The remaining grammatical insertions would be classed by Leech and Svartvik (1994) as being expressions of degree. *A bit of a* indicates a small quantity, as does *something of a,* whilst *a hell of a* represents the other end; a large quantity. *A hell of a* is interesting as it uses the lexical word *hell* as a colloquial substitute for *lot*, and this substitution emphasizes the quantity. The quantity grammatical insertions act in the same way as the intensifying (positive and negative) lexical insertions. Quantity also relates to determiners such as *more* or *some. Quite a* seems able to have two uses. Quirk and Greenbaum (1973: 218) suggest that *quite* may be a downtoner, having a lowering effect, especially in British English, and the example given is:

I **quite** enjoyed the party, but I've been to better ones

The examples found for this phrase in the BNC were:

New Zealand rugs *get **quite a** hammering* BPB 617

My pride *took **quite a** battering* GUE 3375

Both [bottle of alcohol and packet of cigarettes] had *taken **quite a** beating* FAP 3118

Leech and Svartvik, in a more recent grammar than Quirk and Greenbaum's 1973 text, suggest that *quite* can mean "considerably" and the example they give is:

It's **quite** warm today

This has a positive aspect of "raising" the adjective *warm* (1994: 115). In all three of the examples given above, it seems that the downtoning effect is not seen; rather these emphasize the effect of the action.

The grammatical insertions seem to have an epistemic effect; "to signal stronger and weaker commitment to the factuality of statements" (Saeed, 1997: 125). This is especially seen in the case of *a bit of a*, which often acts as a hedge in language. It acts to strengthen or weaken the effect of the phrase. This is an attitudinal effect; it is subjective according to the author, in agreement with the subjective lexical insertions found. Understatements and hedges act as face-saving strategies, either to save the hearer's face, or to defend the speaker's (Thomas, 1995: 173).

Example 6-4

Lake, who is getting out, has **taken a bit of a battering** and feels that her spontaneous outpourings the other day were probably misconstrued
The Guardian: Electronic Edition: Sports Section
A9H 596

In Example 6-4, the use of *a bit of a* is a face-saving hedge, saving the author from making a direct face threatening act (*taken a battering*) and weakens their propositional attitude. In examples such as Example 6-4, the hedge also implies sympathy on the part of the speaker. Compare this with Example 6-5, showing a lexical insertion:

Example 6-5

No force had **taken a worse drubbing** in the first mad onrush of Plan XVII than the Second Army
The Price of Glory: Verdun 1916 – A. Home
K91 43

In this example of complete defeat, the army has taken a severe beating, and the insertion *worse* only emphasizes what a great defeat it was. This epistemic effect could perhaps explain why the use of complete defeat was not found so frequently in grammatical insertions as it was for lexical insertions. "Complete" defeat is a more definite statement of a great loss than *something of a,* for example. Formulaic sequences are often used in texts to attract the reader's attention. They in themselves are thus "colourful" language, and so adding a lexical insertion only adds to the emphatic effect. This perhaps is why there are more occurrences of lexical insertions than grammatical insertions in the dataset.

Hübler recognizes a condition in between statement and hedging, and calls this *indetermination.* He says that:

> ...the difference between the indeterminate sentence (i.e. what is actually said) and the determinate sentence (i.e. what is actually meant) must show a significant *qualitative contrast.* This contrast ensures that the hearer cannot feel harmed by the indeterminate formulation in the same way, or in nearly the same way, as the determinate formulation.
> (1983: 21)

Indetermination has been seen in examples such as Example 6-4, where *a bit of a* acts as a face-saving device.

6.3.2. Summary

It seems that all types of insertions are used for amplification or specification, but it seems that lexical insertions have more of a definite emphatic effect than grammatical insertions. The lexical insertions in this case intensify or emphasize the action and the grammatical insertions quantify it. Most of the insertions for this phrase seem to be attitudinal in force; subjective according to the author/speaker. They also seem to be internal, acting on the noun of the sequence. For *take a BEATING/hammering,* it seems there are more insertions allowed for than idiom dictionaries would suggest, and in particular, several of the

grammatical insertions occur more than once (see Table 6-2). Insertions do not seem to prefer one variant or another for this sequence, nor do they occur more in literal or non-literal readings. They do support evidence given in section 5.3 that this sequence shows a tendency for leisure and periodicals; there was the highest number of insertions found with a sense of defeat. This sense tended to occur in newspapers.

6.4. Case Study 7: Overstep the MARK/Limits/Bounds

6.4.1. Definition and Senses

The LID gives the figurative definition of this phrase as:

> To offend people by doing or saying things that you should not do or say (1998: 224)

This sense is represented by occurrences found in the BNC, for Example 6-6:

> *Example 6-6*
>
> His arrogance goaded her beyond bearing "And who do you think you are, to tell me what I can and can't do? Don't **overstep the mark** too far, Mr Calder!"
> *Destined to Love – J. Taylor*
> *JXV 2390*

This sense is also found in other idiom dictionaries such as the CCDI. The LID suggests that the noun *mark* is found in British English, and *limits/bounds* are the American English variants.

There were 289 occurrences of the phrase and its variants found, with seven different verbal substitutes: *overstep, leap over, spill over, cross, pass, leapfrog over*, and *step over,* and six different noun variants; *measures, bounds, mark, limits, boundary*, and *line.* The determiners found in the occurrences were: *null, his, the, it, these, your, all, a, an, that,* and *their*, and *Driffield's.*[4] Note that some of the verb substitutes consist of a verb and a preposition. Whilst this changes the syntax from V det N to V Prep det N, it seems that these verbs require the preposition to fulfil them, and thus act as prepositional verbs rather

[4] *His* is a possessive pronoun acting as a determiner.

than transitive verbs. These verbs semantically collocate with the preposition *over*, and the following noun phrase is a complement rather than an adjunct, and thus these verbs have been included as verb substitutes, as opposed to cases such as *in a tantrum* from the canonical *throw a WOBBLY/wobbler*, where the prepositional phrase acts as an adjunct, and the preposition does not fulfil a verb's needs.

Lexical Verb Substitute	Frequency
Cross	211
Overstep	61
Step over	8
Pass	5
Spill over	2
Leap over	1
Leapfrog over	1
TOTAL =	**289**

Table 6-4 – Lexical verb substitutes for *Overstep the MARK/limits/bounds*

Lexical Noun Substitute	Frequency
Line	146
Boundary	93
Mark	27
Limits	12
Bounds	10
Measure	1
TOTAL =	**289**

Table 6-5 – Lexical noun substitutes for *Overstep the MARK/limits/bounds*

Table 6-4 and Table 6-5 give the frequencies of the different lexical substitutes found for this phrase, for the verb and the noun. In accordance with the LID, there are more occurrences of *mark* than for *limits* or *bounds*, although the frequencies for these are relatively small. The LID suggests that *limits* and *bounds* are found more in American English and *mark* is the British variant. This is demonstrated in the frequency figures shown in Table 6-5.

Table 6-6 gives a guide to the frequency of the verb and noun lexical combinations – grammatical information such as determiners and also insertions are not seen on this table.

It shows that *cross + det + line* is a popular combination of this phrase. The canonical *overstep + det + mark* is not as frequently occurring in the BNC. The chain from *overstep + det + mark* to *cross + det + line* will be discussed in section 6.4.2.

Lexical Combination	Frequency
Cross + det + line	136
Cross + det + boundary	74
Overstep + det + mark	27
Overstep + det + boundary	12
Overstep + det + bounds	9
Overstep + det + limits	8
Step over + det + line	6
Overstep + det + line	4
Pass + det + limits	3
Spill over + det + boundary	2
Step over + det + boundary	2
Pass + det + bounds	2
Overstep + det + measures	1
Leap over + det + boundary	1
Leapfrog over + det + boundary	1
Cross + det + limits	1
TOTAL =	**289**

Table 6-6 – Frequent lexical combinations of *Overstep the MARK/limits/bounds*

The examples in the BNC seemed to divide into two senses. First, there was the dictionary definition of "going too far with your actions and upsetting/offending someone". However, the other sense found was a more literal sense of "crossing a line", either an actual demarcation, or a metaphorical one. Table 6-7 shows the frequencies for the senses:[5]

Sense	Frequency
To go too far and upset/offend someone	56
To cross a demarcation	232
Other	1
TOTAL =	**289**

Table 6-7 – Frequency of the senses for *Overstep the MARK/limits/bounds*

For the chaining process, versions of the phrases with both literal and figurative readings were collected; not all formulaic sequences are non-literal. This phrase had a high

[5] Note that "other" represents an occurrence where the phrase is used as a proper name, e.g. "that's the ceremony of **Crossing the Line**" (H7H 872).

proportion of occurrences with a literal reading, such as the actual demarcation being walked across in Example 6-7:

Example 6-7

In order to keep the boys separate from the girls he used to draw a chalk line down the middle of our meeting room, and when we got carried away and **crossed the line** he would burst into a ferocious rage, pick up anything that was to hand - usually a book - and throw it at the offender
Peace and War: Growing up in Fascist Italy – W. Newby
G3B 204

In total, for this phrase, there were 132 occurrences which had a figurative reading, and 156 occurrences with a literal reading. It seemed that of the occurrences with a literal reading, it was the form *cross* + *det* + *line* that contributed significantly to the frequency. Of the 136 occurrences of *cross* + *det* + *line,* 114 had a literal reading and only 21 a figurative reading. The remaining one was a proper name example. Compare the lexical combination *cross* + *det* + *line* with *cross* + *det* + *boundary* which had 45 occurrences with a figurative reading, and 29 with a literal reading, and *overstep* + *det* + *mark,* which had 25 occurrences with a figurative reading, and two with a literal one. Where the sense was of actually crossing a demarcation, the referent was either an actual physical line, or it could be a more figurative one:

Example 6-8

It was a constant battle, fuelled by the basic survival instinct which kept men moving ahead over thousands of years towards what we are pleased to call modern civilization today. In Baldersdale that process gathered no speed at all and indeed, never even **crossed the finishing line.**
Seasons of my Life – H. Hauxwell
BN6 21

There were also examples where the line to be crossed was between two conceptual objects:[6]

Example 6-9

The Irish Prime Minister, Mr Albert Reynolds, said that Unionists had "**crossed the line** to vote for peace" by backing Dr Hendron.
Daily Telegraph: Electronic Edition: Foreign News Pages
AK2 170

Where the type was *cross* + *det* + *line,* and mostly of literal reading, there were two main types of referent. The first was that the demarcation was an actual sports marking. Most of the sports cases referred to a finishing line in a race however other sports lines, such as the line to cross when scoring a goal in football or rugby, for instance, are also found, as in Example 6-10:

Example 6-10

They scored a good try by internationalist Craig Lawson who **crossed the line** from a short pass from his scrum-half, David Paterson, son of the Scottish selection committee
The Scotsman: Leisure Pages
K5J 1957

The second type of line is a different sense of the word *line;* it is that of a railway line, as opposed to a marking line:

Example 6-11

Gorse blossomed gold on magnesium limestone embankments and, with our hearts in our mouths, and the necessary British Rail Permit in our guide's pocket, we **crossed the line.**
The Birdwatcher's Handbook
F9H 1530

[6] In this case, the Unionists have crossed the theoretical line from the decision to settle the dispute violently, to choosing a more peaceful option. In fact, this text goes on the show that the Unionists also had to walk over a physical line in that they had actually crossed over the entry to the election halls to vote.

A third type of line was that of a boundary between two countries, or land areas. This was less frequently occurring than the other types. Example 6-12 demonstrates crossing the line between safe and enemy territory in war:

Example 6-12

It flew very slowly. They **crossed the line** at 3,500 feet. As soon as the anti-aircraft fire began to thin out the F2B dived to begin photographing below cloud level.
Goshawk Squadron – D. Robinson
HRA 2368

Where an occurrence had a meaning sense of crossing an actual demarcation, there was a difference in the sense of the border to be crossed. In several cases, instead of the action taking place from one side of the border to another, there was diffusion over many boundaries. Example 6-13 shows that the role of a community carer now traverses traditional roles; it overlaps several job descriptions. Whilst many of the occurrences for this phrase demonstrate "going over" a line, or represent a change from one thing to another, there were also tokens found where the action was overlapping several items.

Example 6-13

The Griffiths report (1988) encourages the idea of 'community carers' who would **cross the traditional boundaries** between social work, health service, voluntary and private tasks in the domiciliary care of disabled people.
Dementia: Sharpening Local Plans: Priorities for the 90s – J. Kileen
FTY 75

6.4.2. Networks and Compositionality

This phrase seemed both to form a scale and also to have peripheral and central members. From Table 6-4 and Table 6-5 it is clear that some of the lexemes occur more frequently than others; *overstep, cross, line* and *boundaries* seem to be central lexical members. *Leap over* and *measures* seem to be marginal members of the lexical set for this formulaic sequence. It seems the forms *overstep + det + boundary* and *cross + det + line* can be used for the same meanings. The meanings are those of "going too far and offending someone", and also for "crossing a literal or figurative demarcation". The form *cross + det + line* occurs more frequently in the BNC, and also occurs more in a literal context. The forms

overstep + *det* + *mark* and *cross* + *det* + *boundary* seem to occur more with a non-literal reading.

A network for this formulaic sequence seems to connect lexemes rather than meaning, i.e. a basic network:

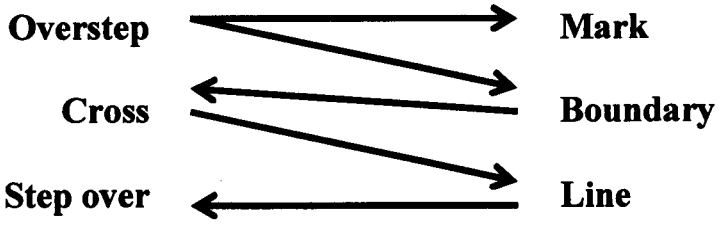

Figure 6-1 - Basic network for *Overstep the MARK/limits/bounds*

The phrasal types for this formulaic sequence form a schema, where the variant lexical items can be recognized as belonging to the group. In Figure 6-1, the direction of the chaining process is illustrated in the direction of the arrows; *overstep* links with *mark* and *boundary. Boundary* occurs in lexical combination with *cross. Cross* also forms a lexical combination with *line,* which in turn links with *step over.*

However, it seems there is a clear semantic difference between the forms *overstep* + *det* + *mark* and *cross* + *det* + *line*. Where the verb is *cross,* there is more of a leaning towards a literal reading. Where the verb is *overstep,* or a verb plus preposition combination, with *over* as the preposition, it seems the tendency is for more figurative reading occurrences. Note, however, that *cross* is by far the most frequently occurring verb of this set in the BNC, with more than 10,000 occurrences of the verb, compared to 83 for the lexeme *overstep.*

The meanings of the two central verbs for this lexical set are also slightly different. The OED cites meanings of *overstep* to be:

> 1. a. To step over or across; to travel beyond or to the other side of; to pace over (land) in order to make measurements (obs.). Also fig.
>
> b. To go beyond (a limit, as of what is considered socially acceptable, feasible, etc.); to violate (a rule or standard of behaviour).
>
> c. in to overstep the mark (also line).
>
> d. To act beyond what is proper or authorized; to transgress. rare.
>
> (http://www.oed.com/)

For *cross* one sense is given as:

> To pass over a line, boundary, river, channel, etc.; to pass from one side to the other of any space
> (http://www.oed.com/)

So the meanings of the two verbs are different; *overstep* meaning "to go beyond", and *cross* meaning "to go from one thing to another". Whilst the underlying concepts of TRAVERSE and DEMARCATION are common to phrases involving either verb, it seems that this difference contributes to *overstep* + *det* + *mark/boundary* being one formulaic sequence and *cross* + *det* + *line* being another, particularly if the literal readings are taken into account – note the literal "lines" mentioned in connection with *cross* + *det* + *line* in section 6.4.1; sports lines, or railway lines are demarcations to be crossed one way once only. There is no need to "go beyond" these demarcations.

That these two combinations *cross* + *det* + *line* and *overstep* + *det* + *mark* are separate is further discussed in section 6.4.3.

6.4.3. Insertions

There were 289 occurrences of the sequence *overstep the MARK/limits/bounds*, and 99 of these had an insertion. Table 6-1 showed that all of the insertions were lexical; there were no grammatical insertions. Most of the examples showing insertions had a literal reading. Of the 99 occurrences with insertions, only four had a sense of "to go too far and offend or upset someone". The rest all had a sense of crossing a delineation. Most of the examples with insertions occurred in the form *cross* + *det* + *line,* which had a high frequency of meanings of passing over a demarcation, and a high number of literal readings. There were 46 instances of *cross* + *det* + *line* with an insertion, and 40 of these had a literal reading. The following examples are with regard to a railway track and sports demarcation respectively:

Crossing the **main** line	CE9 664
Crossed the **finishing** line	AAU 5

Many of the inserted examples served to indicate what line was being crossed; specification as opposed to emphasis (compare with *take a BEATING/hammering*). This was the same for the next most frequent form which allowed for insertions; *cross* + *det* + *boundary.* This had 44 occurrences, although 26 had a figurative reading, and 18 literal. The definite action of crossing from one item to another, or going beyond a point could explain the lack of grammatical insertions. It would not be possible in many literal instances to "cross *a bit of* a line", or to "overstep *something of a* boundary".

There were several examples where the insertion shows that the author recognizes the phrase to have possible non-literal uses. In Example 6-14 the use of the insertion *invisible* shows that it is not an actual line being crossed, but more a metaphorical one:

Example 6-14

The problem can more easily arise for young academics, who are close in age to their students and do not know when they are **crossing an invisible line.**
The Scotsman: Foreign News Pages
K5D 522

With the occurrences of *cross + det + boundary,* it seems that either the phrase takes an insertion, or it has no insertion, but takes a post-modifying prepositional phrase, for example:

> *Example 6-15*
>
> ...despite the gains made by repeal feminism [sic] women still risked censure and social ostracization when they **crossed the boundaries** from the private to the public sphere.
> *Dangerous Sexualities – F. Mort*
> *GOD 796*

Of the 74 total occurrences of this type, 15 had a post-modifier, 44 had an insertion, and six had both. This specification as to the type of delineation does not occur to the same degree in other types. Possibly this is due to meaning. Crossing a sports line or a railway track is one action, and not beyond, as in the difference between *cross* and *overstep.* Also, the referent of the line, as mentioned in section 6.4.2 is usually a sports line, a railway line or an area border, so in examples with no insertions, the meaning is clear from context what is being traversed. With *overstep + det + mark,* the phrase tended to have a different sense; that of "going too far and offending someone", and thus did not require specification. It seems that *boundaries* has a slight tendency to require specification as to the starting and end points of the action.

In terms of compositionality, it seems there is a difference between the two main types found here; *cross + det + line* and *overstep + det + mark.* Both forms function synonymously in literal readings, and can have either sense. However, it seems that where the verb is *overstep,* the meaning is predominantly "to go too far and upset someone". Where the verb is *cross,* the meaning seems to be more compositional: traverse + a + demarcation, and the reading can either be literal or not. This is also seen in insertions. There were no occurrences of *overstep + det + mark* which had an insertion. Compositionality theory would suggest that a phrase which is compositional would be more open to insertions than one which is non-compositional. If *overstep + det + mark* has a meaning of "going too far and upsetting or offending someone", then this meaning cannot be read compositionally from the component lexemes; i.e. *Overstep + det + mark* may correspond compositionally to "go too far", but it is difficult to see what the rest of the

meaning would correspond to in the lexemes. So it seems that *cross + det + line* is more compositional than *overstep + det + mark*. The chaining process links them together with the same underlying meaning of TRAVERSE + A + DEMARCATION, and there are occurrences of both acting with a sense of "go too far and offend someone" and "cross an actual boundary, be it literal or non-literal", however, it seems that in many ways, the two are distinct formulaic sequences. It seems that the network for this case study reveals two formulaic sequences which behave independently. The fact that *cross + det + line* takes insertions but *overstep + det + mark* does not, supports this division.

The LID has its own entry for *cross the line:*

> To start behaving in a way that people consider offensive, immoral, dangerous or extreme, though it may be only slightly different from the way you behaved before. (1998: 213)

The dictionary also suggests that a variant for this phrase would be *step over the line.* This meaning is very similar to the one given for *overstep the mark.* The two sequences can act in the same way in some cases, but also, clearly, retain their own individual status in how they behave in terms of compositionality and insertions.

Looking at the insertions themselves, there is a contrast between the lexical insertions found here, and those found for *take a BEATING/hammering.* Whereas section 6.3.1 showed the lexical insertions to be subjective, those found for *overstep the MARK/limits/bounds* were more objective, for example:

Crosses the **county** boundary	K1V 3096
Crossing the **winning** line	CBG 3961
Cross the **picket** line	ANY 1910
Overstep the **area** boundary	AOM 524

They occurred in compositional forms which had a literal reading. Such insertions act as specifiers, compared to the attitudinal emphasizers seen in *take a BEATING/hammering.* Following Quirk and Greenbaum's semantic sets for adjectives (1973: 125), the insertions for this phrase would be classed as:

Denominal adjectives relating to provenance or style	2
Adjectives denoting age	1
Adjectives denoting colour	4
General adjectives susceptible to objective measure	42
TOTAL =	**49**

The objective adjectives, such as those seen in the previous examples, are internal modifications, acting only on the noun. Of the remaining 50 insertions, there are two tokens where the insertion is an extension of the verb phrase; a coordination of two verbs:

Crossing **but negating** the boundaries	J86 184
Cross **or straddle** a line	FM1 635

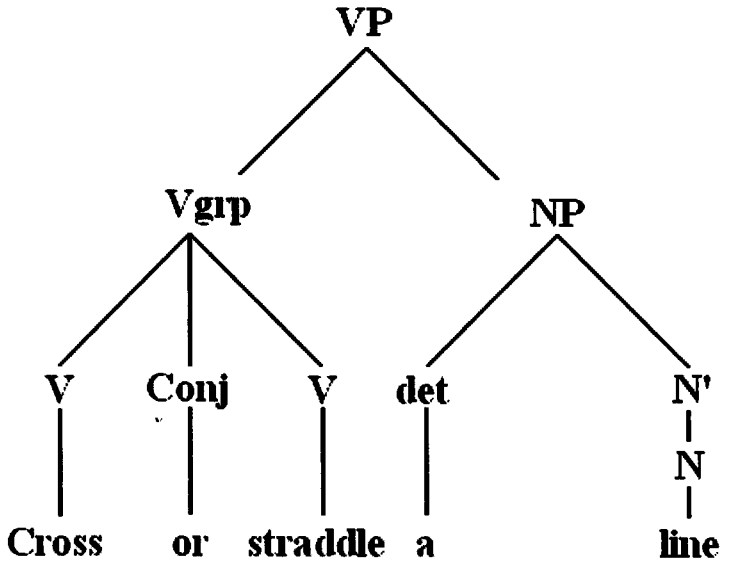

Figure 6-2 – Syntactic structure demonstrating verbal coordination for *Overstep the MARK/limits/bounds*

The coordination of the verb examples are exceptions within the criteria for inserted materials. They are not additions to the V det N structure in the same way as an adverb or a post-modifying PP; there are simply two lexemes with the V status. They are classed as insertions as they are extra material within the V det N structure. The effect of the insertion

in these cases is neither to emphasize, specify nor add attitude, but is to add extra information.

The remaining 48 insertions were also susceptible to objective measure. However, in 41 cases, there was a noun insertion modifying the noun of the formulaic sequence, for example:

Cross the **railway** line	EED 218
Crossed the **tram** line	CBK 495
Cross these **agency** boundaries	BO1 745

There were two occurrences where the noun was a proper noun, for example *crossed the **Danube-Sava** line (FSU 1183).*[7] In seven cases, there was a verb taking an –ing suffix to become a participle, and six of these cases were in the context of sports, e.g. *crossed the **finishing** line (BO3 83).* Given that this phrase has 156 tokens in total with a literal reading, and of the 99 insertions tokens, there are 62 with a literal reading, nouns acting as modifiers are more likely to be found in a literal formulaic sequence than one with more figurative tokens.[8] The nouns acting as modifiers for this phrase's insertions are specifiers. They add more information to the literal interpretation of the sequence. The general adjectives susceptible to objective measures were found in the tokens with figurative reading. As mentioned earlier, with the literal tokens for this phrase, and the meaning of the action of crossing, this phrase is open to specification, and not subjectivity or emphasis.

The general adjectives susceptible to objective measures, such as *international* and *departmental* could be said to be external adjectives – they act as domain delimiters and modify the whole sequence. Other insertions, such as *district* or *county,* seem to be domain delimiters although they are not external. Examples such as *before the disease crosses the **county** boundary" (K1V 3096)* show that the noun modifier is modifying the head noun only. Other internal lexical insertions (both noun and adjective) include e.g. *railway, tram, finishing,* and *picket.*

---

[7] Denominal adjective denoting providence.
[8] Compare this with *take a BEATING/hammering,* which had no noun insertions acting as modifiers, and a relatively high proportion of non-literal tokens.

Of the 99 insertions, there were 77 different types. There were 13 forms which had more than one occurrence. As seen in Table 6-7, there were 232 occurrences of the sense to cross a delineation. Of these, 67 referred to sports, and 16 to crossing a railway line. The remaining 148 referred to crossing a land boundary, or conceptual boundary. These preferences are reflected in Table 6-8, with *finish* and *finishing* referring to sports, *main* and *railway* referring to trains, and the rest referring to land or conceptual boundaries. *Departmental* and *finish* are more central members of the insertions for this sequence.

Insertion	Frequency
Cantonal	2
Central	2
County	2
Departmental	5
Disciplinary	2
District	2
Finish	5
Finishing	4
International	2
Invisible	2
Main	2
Picket	3
Railway	2
TOTAL =	**35**

Table 6-8 - Frequent lexical insertions of *Overstep the MARK/limits/bounds*

There were 21 tokens where there was more than one modifier present, excluding the two examples with conjoining verbs. For example:

Oversteps **proper stylistic** boundaries	ED6 1996
Crossed the **Serbian front** line	HWU 1496
Crossing a **painted red** line	ED1 220

Such examples follow the pattern of the pre-modification sequence as laid out by Quirk and Greenbaum (1973: 404): *proper stylistic* having a general adjective preceding one of provenance, *Serbian front* having an adjective of provenance preceding a noun modifier, and *painted red* having a general adjective preceding one of colour.

There were five tokens with the adjective modifiers having coordinators, for example:

Crossing **functional and divisional** boundaries	CR7 2167
Cross **department or sectional** boundaries	GOU 1393

In two of these examples, the coordinator was not realized, for example *cross **system/trade** boundaries (BP2 804).*

Finally, there was one example with three modifiers:

Example 6-16

Their writings and comments have a confessional tone, and no wonder. They are **stepping over the invisible, moralistic Maginot line** of the old culture of opposition.
Marxism Today
G2J 1635

Example 6-16 has asyndetic coordination (Crystal, 1998: 198) of two adjectives, *invisible* and *moralistic*, and an adjective of provenance. The use of the adjective *invisible* here also shows the author is using metaphorical language when referring to the line of the opposition. That the literal examples have insertions that are more "concrete" noun modifiers, and those with a more figurative reading have more adjectival modifiers provides support for the notion that there is division within this case study of the formulaic sequences *cross* + *det* + *line* and *overstep* + *det* + *mark.*

6.4.4. Summary

In summary, *overstep the MARK/limits/bounds* seems to form a constellation network, with all of the variants able to demonstrate the same underlying meaning, and peripheral members such as *leap over, pass, measure*, and *bounds.* However, on closer investigation, it seems clear that there are two distinct formulaic sequences involved; the non-literal *overstep* + *det* + *mark/boundary* and the literal *cross* + *det* + *line/boundary*. The two are linked lexically and by non-literal occurrences of *cross* + *det* + *line/boundary,* thus demonstrating that there are no clear boundaries of literal and non-literal reading for this case. There is overlap between the two readings. The insertions support the distinction. There were only lexical insertions found, and these tended to be objective adjectives, or

noun modifiers. They were more "concrete" insertions, modifying a noun which had a concrete referent, be it actual (e.g. railway line) or conceptual (e.g. agency boundaries). The insertions were found more frequently in the form *cross* + *det* + *line/boundary.* The form *overstep* + *det* + *mark* is perhaps less compositionally analysable than *cross* + *det* + *line,* and thus less open to insertions.

6.5. Case Study 8: Pin your HOPES/Faith on

6.5.1. Definition and Senses

This case study represents one of the more lexically and semantically fixed formulaic sequences in the dataset. The meaning as given by the LID is:

> To think that one particular person, thing, or event will make you happy, successful etc.
> (1998: 174)

This can be seen in examples such as:

> *Example 6-17*
>
> "I do hope she remembers," Juliet murmured fervently as they sat down. "Don't **pin your hopes on** it," warned David.
> *Hearts in Hiding – A. Grey*
> *JYO 3531*

There were three noun substitutes and four main verb substitutes for this formulaic sequence. The verbs took a prepositional complement and the prepositions tended to be; *on, to, in, into*, or *upon.* The verbs required the prepositional phrase to complete the action of the verb, as was seen with *step over* and *leap over* in section 6.4.1 however there was flexibility as to where the prepositional phrase occurred:

> *Example 6-18*
>
> This is the project **on which astronomers in the US can pin their hopes**, for it will undoubtedly be the single most important telescope of the 1980s
> *New Scientist*
> *B7N 1195*

The verb *pin* requires a prepositional phrase to satisfy it, as do the other verb substitutes for this phrase; *place, put*, and *fasten.* For the purposes of this study, the phrase *pin your HOPES/faith on* is treated as being a block comparable to that of *take a BEATING/hammering.* The structure is V det N Prep. Following the preposition, there is a context-dependent gap for a noun phrase. For the purposes of this research, anything following the noun is being treated as an addition and is not being focussed on (see section 2.3.3). Thus the only insertions looked at are those within the V det N structure for this case study. In Example 6-18, there is grammatical rearrangement of the prepositional complement, and the formulaic sequence is found within a relative clause. This example is classed as a token for this case study, but it is not included as an example of an insertion. Insertions examples are further discussed in section 6.5.3.

Table 6-9 and Table 6-10 show the frequencies of the different noun and verb lexemes for this phrase:

Lexical Verb Substitute	Frequency
Pin	91
Put	83
Place	40
Fasten	1
TOTAL =	**215**

Table 6-9 - Lexical verb substitutes for *Pin your HOPES/faith on*

Lexical Noun Substitute	Frequency
Faith	128
Hope	86
Dream	1
TOTAL =	**215**

Table 6-10 - Lexical verb substitutes for *Pin your HOPES/faith on*

Table 6-9 and Table 6-10 show that this phrase has two clear noun members, *faith* and *hope,* and two clear verb members; *pin,* and *put,* with *place* being a secondary substitute. The nouns *dream* and *fasten* are peripheral members for this set. That *pin* and *put* are central verbs is supported by Table 6-11 which also shows that *pin on* and *put in* are popular combinations for this phrase, followed by *place in.*

Lexical Combination	Frequency
Put + det + faith in	74
Pin + det + hopes on	72
Place + det + faith in	34
Pin + det + faith on	8
Pin + det + faith in	4
Put + det + hopes in	4
Put + det + faith into	4
Place + det + hopes on	3
Pin + det + hopes to	2
Pin + det + faith to	2
Place + det + hopes in	2
Pin + det + hopes upon	1
Pin + det +dreams on	1
Put + hopes + into	1
Fasten + det + hopes on	1
Pin + det + faith in	1
Place + det faith upon	1
TOTAL =	**215**

Table 6-11 - Lexical combinations of *Pin your HOPES/faith on*

Table 6-11 shows all of the different lexical combinations for all of the occurrences for this sequence, and shows that there are two clear phrasal members for this sequence: *pin + det + hopes on* and *put + det + faith in.*

Example 6-19

So, like hundreds of others, he **put his faith in** a franchise. Bank managers tend to be far from impetuous in putting their own money at risk
Daily Telegraph: Electronic Edition: Business Section
AHT 470

Example 6-19 shows, as with Example 6-18, that both of these types of the phrase demonstrate the dictionary's definition – thinking that one particular event or action can bring success or happiness.

So in terms of lexemes, this formulaic sequence demonstrates a degree of fixedness that is higher than those of the other case studies. In terms of types, Table 6-1 showed this phrase to have 215 tokens and 140 types. This type count does not look like a fixed phrase;

especially in comparison to a sequence such as *hit the DECK/dirt,* which has 434 tokens, yet only 34 different types. However, as Table 6-9 and Table 6-10 show, *pin your HOPES/faith on* has relatively few noun and verb substitutes; three and four respectively. In terms of the lexemes involved, it is more fixed than one with more substitutes, such as *take a BEATING/hammering.* The sequence *pin your HOPES/faith on* has a high type frequency in spite of its relatively fixed lexical set. This is due to the number of tokens with insertions, and also the determiner. The determiner for this phrase is a pronoun, and so is open to more determiner substitutions than the definite article. (See section 2.3.3). *Hit the DECK/dirt* only has examples with the definite article.

Table 6-12 shows the top ten sequences of the dataset in terms of determiner substitutes:

Formulaic Sequence	Determiner Frequency
Pin your HOPES/faith on	16
Do BIRD/Time	12
Kick up a STINK/fuss	12
Overstep the MARK/limits/bounds	11
Chance your ARM/luck	10
Raise HELL/Cain	9
Shut your FACE/gob/trap/cakehole	9
Cover your ASS/Butt	8
Drag your FEET/heels	8
Flip your LID/wig	8

Table 6-12 – Ten sequences from the dataset showing the most number of determiner substitutes

Of these phrases, six have a personal pronoun in the determiner position, two have a null determiner, and one has an indefinite article. Only *overstep the MARK/limits/bounds* has a definite article. The determiner substitutes found for this sequence were: *his, their, its, your, her, little, much, any, my, no, null, one's,* and *such,* although *overstep the MARK/limits/bounds* has a high number of compositional literal forms, which may account for the high determiner substitute frequency.

The other factor in giving a raised type frequency is the different inflectional forms of the verb found. If a sequence has a high number of lexical substitutes, determiner substitutes, insertions and verbal inflections, then it will have a high type frequency. Note that the sequence *keep your SHIRT/hair/pants on* had only 18 tokens and only one type; *keep your*

hair on. There were no other substitutions of any kind, insertions or inflections. Thus this was the most fixed sequence in the dataset.

Out of the 215 tokens for *pin your HOPES/faith on*, the vast majority had the sense of thinking one particular thing or event can create happiness or success. A small number of the occurrences had similar underlying concepts, but had an extra meaning added on. Twenty-five of the tokens had a meaning of "to believe in", and nineteen of these had the form *put + det + faith in.* These occurrences tended to occur in religious examples:

Example 6-20

FISHERMEN in general may **put their faith in** St Peter, but the patron saint of fish-hooks - and therefore of anglers in particular
Daily Telegraph: Electronic Edition: Leisure Pages
AJY 1113

This still has the meaning of "to think one particular thing will make you happy", however it also has an added sense of "to believe in" or "trust". This is also seen in examples such as Example 6-21:

Example 6-21

Great mother, now the gods have gone. We **put our faith in** you alone
The Hunt by Night – D. Mahon
HR5 1390

All of the tokens found for this case study had the same underlying meanings of BELIEF/TRUST, PARTICULAR OBJECT/EVENT/PERSON, and BRINGING HAPPINESS OR SUCCESS. This means that this formulaic sequence is more semantically fixed than other case studies, as well as being more lexically fixed.

6.5.2. Networks and Compositionality

As seen in section 6.5.1, the phrase *pin your HOPES/faith on* is one of the more lexically and semantically stable sequences in the dataset. It seems to have two major types: *pin* +

det + *hopes on* and *put* + *det* + *faith on* which have the same underlying meaning. The noun *dream* and the verb *fasten* seem to be peripheral members for this lexical set.[9]

The constellation network for this phrase would be:

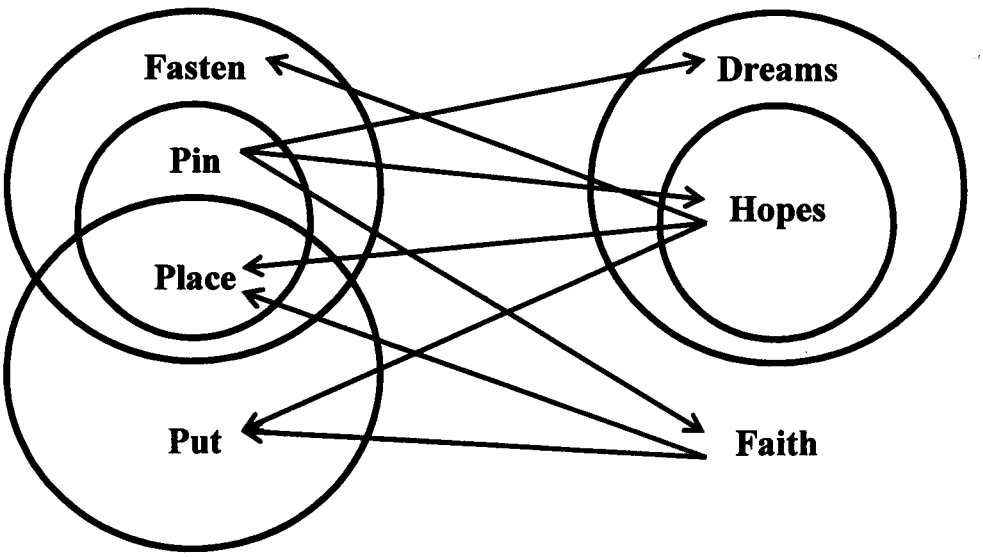

Figure 6-3 - Constellation network for *Pin your HOPES/faith on*

There are two main lexical combinations for this phrase, with the same meaning, and each with a non-literal reading. There is no semantic development from *pin hopes* to *put faith,* although the latter has a specialized use with regard to religion. It seems that the verbs can be used interchangeably and are broadly synonymous, as are the nouns. It seems however from the strong frequencies of occurrence, that *pin* + *det* + *hopes on* is a sequence, as is *put* + *det* + *faith in.* Both are linked lexically, and they share the same underlying meaning. It is interesting to note that whilst there are 72 occurrences of *pin* + *det* + *hopes on* and 74 occurrences of *put* + *det* + *faith in* in the BNC, the LID only recognizes *pin* + *det* + *hopes on,* suggesting only *faith* as a substitute for *hopes.* There is no suggestion that *put* is a highly occurring verb for this phrase. This could be, however, that *put* is classed as being a primary verb (see section 2.2.2) and as such does not carry as much lexical content as *pin.*

---

[9] The prepositions for this phrase were: *on* occurring 84 times and *in* 121. There were four occurrences each of *to* and *into* and two occurrences of *upon.* As Table 6-11 shows, there is a preference for the combination *put* + *det* + *faith in,* and also *pin* + *det* + *hopes on.*

The distinction between *pin + det + hopes on* and *put + det + faith in* can be seen in terms of the register.

Domain	Frequency	Normalized
Imaginative	5	0.44
Natural and Pure Science	1	0.46
Applied Science	6	1.42
Social Science	2	0.26
World Affairs	14	1.48
Commerce and Finance	8	1.96
Arts	9	2.16
Belief and Thought	6	3.43
Leisure	17	2.96
Spoken	6	1.01
TOTAL =	**74**	

Table 6-13 – Frequency of the domain of all of the tokens of *put + det + faith in*

Domain	Frequency	Normalized
Imaginative	6	0.53
Applied Science	8	1.89
Social Science	6	0.79
World Affairs	19	2.00
Commerce and Finance	14	3.42
Arts	3	0.72
Belief and Thought	3	1.71
Leisure	11	1.92
Spoken	2	0.34
TOTAL =	**72**	

Table 6-14 – Frequency of the domain of all of the tokens of *pin + det + hopes on*

Table 6-13 and Table 6-14 show that there is a distinct register difference in use of the two forms. The normalization figures emphasize the frequency differences in the occurrence of the sequence in particular domains. *Pin + det + faith on* occurs more in belief and thought when the domains are normalized, whilst *pin + det + hopes on* occurs more in commerce and finance texts. That *pin + det + faith in* seems to have an "extra" meaning, and a different distribution suggests that these two forms are overlapping, yet separate members.

This phrase does not seem to form a progressive network nor have central and peripheral members in the same way as other case studies. It seems to have two major types which have the same meaning, and the verb member *place* is a synonymous substitute. In terms of figurative and literal readings, this formulaic sequence was again stable. All of the tokens had a non-literal reading; the conceptual nature of the noun meaning that a literal reading was impossible.[10]

The phrase seems to have a non-compositional reading: "To think that one particular person, thing, or event will make you happy or successful" does not seem to correspond to individual components of the phrase, and this meaning does not seem to be the sum of *pin + hopes + on.* However, the fact that this phrase scores highly in terms of insertions would suggest there is an element of compositional analysability to this phrase after all.

6.5.3. Insertions

Of the 215 occurrences for this phrase, there were 39 occurrences of this phrase which had an insertion. There were 15 examples of lexical insertion, and 24 occurrences of grammatical insertion; this was the second highest phrase in the dataset in terms of grammatical insertion frequency. In terms of form and insertions, there was a slight preference for insertions to occur with the noun *faith*, with nine of the lexical insertions and 14 of the grammatical ones occurring with this noun. *Pin your HOPES/faith on* showed either lexical insertions or grammatical insertions; there were no occurrences with both a lexical insertion and a grammatical one.

As mentioned in section 6.5.1, the insertions included for this case study were inside the V det N structure. There were 13 examples (e.g. Example 6-18) where the formulaic sequence occurred in a relative clause due to movement of the prepositional complement, and there were also 15 examples of a post-modifying prepositional phrase to the noun between the noun and the prepositional complement, e.g. *pin his hopes **for the future of man as a spiritual being** on (CL6 122).* However, for the purposes of this research, insertions are limited to material within the V det N structure, such as:

[10] Religious examples such as Example 6-20 and Example 6-21 could be seen as having literal readings.

Example 6-22

The company **pins great hopes on** its new small AC drive Dinverter.
Liverpool Daily Post and Echo: Business Section
K3N 56

Of the 15 lexical insertions, there were two examples of a noun modifying a noun, e.g. *pins its **desktop** hopes on (CST 116)* and *pinned their **redevelopment** hopes on (CD9 1191).* The 13 remaining tokens can then be said to consist of two tokens of general adjectives susceptible to objective measure (Quirk and Greenbaum, 1973: 125), e.g. *placed the **same** faith in (HAO 179), placed his **own** hopes on (CS5 796),* and there was one example of an adjective for subjective measure *putting **undue** faith in (AYJ 2167).* The remaining ten tokens were intensifying adjectives, e.g. *place **greater** faith in (CRW 220), place their **greatest** hopes on (B7L 2455).* Of these nine, six were of the lexeme *great.* The intensifying adjectives would here be said to be emphasizers following Quirk and Greenbaum's classification (1973: 121), whilst the adjectives *greatest* and *absolute* could be said to be amplifiers. They scale upwards, denoting the upper extreme of a scale. The adjectives *excessive* and *enormous* e.g. *places **excessive** faith in (CN5 901)* and *placed **enormous** faith in (A6X 597)* could also be said to be subjective, giving the writer's view as to the degree of faith given.

The grammatical insertions tended to provide emphasis as to the degree of belief involved, as in Example 6-23:

Example 6-23

Christian need [sic] to worry about the placebo effect, he also needs to be aware that he will want to please his therapist, having **placed so much faith in** him
Paganism and the Occult – K. Logan
B2G 1153

All of the 24 grammatical insertions had a quantifying effect, for example *all, so many, a lot of.* As with the adjectival lexical insertions, there was also a small set of grammatical insertions found; there were nine types. The insertion *all* was most prevalent, having ten tokens, e.g. *pin **all** your hopes on (BM1 1592), pinned **all** her hopes on (H8J 222),* and *place **all** my faith in (ED6 100).* The use of *all* as a universal quantifier means that it has a more emphatic meaning than pronunciation would otherwise allow (Hogg, 1977: 121).

As with the intensifying adjectives found for the lexical insertions, the grammatical insertions of quantity demonstrated a high degree of the action. In particular, there were two examples of a quantifying phrase with a mass noun; *an inordinate amount of* and *a great deal of.* These add degree to the action, and especially as, in both cases, the noun is itself pre-modified by an intensifying adjective. The lexical and grammatical insertions both seem to demonstrate internal modification, as the insertions modify the noun, and cannot act to modify the whole phrase.

Pin your HOPES/faith on also had a high number of quantifying determiners. Section 6.5.1 showed that this phrase had the highest number of determiner variants in the whole dataset (see Table 6-12); variants including pronouns and articles as well as quantifying determiners such as *much, more, some, little, no*, and *any.* There were 13 tokens out of the total 215 involving such a determiner (i.e. not a possessive pronoun or article), and none of these had any other form of insertion.

Example 6-24

Our findings are disappointing. The reason the Americans got it wrong was they were **putting more faith into** blood tests to see if the drug worked
Central T.V. News Scripts
K1K 1215

Note that the lexical and grammatical insertions had an intensifying effect, whereas quantifying determiners such as *little, no or any* had a more negative effect. The two examples of *much* occurred in tokens where the verb was negated. Of the 13 occurrences, only five tokens had a heightening degree effect, e.g. *more, such,* and *some*, however this was still not so great an effect as with the grammatical insertions. The remaining nine tokens here had a decreasing or negative effect, for example *puts **no** faith in (G2V 1381)* and *pin **little** hope on (CAH 860).* Perhaps this number of negative examples in the quantifying determiners, which is not present in the rest of the tokens, relies on the phrase meaning; it is a positive phrase concerning abstract notions of hopes and dreams; the canonical occurrences and those with insertions would tend to reinforce this, whereas only the quantifying determiners can allow negative tokens to occur.

6.5.4. Summary

The phrase *pin your HOPES/faith on* has two main types; *pin + det + hopes on,* and *put + det + faith in.* Both of these can be used interchangeably, although there is also evidence for the latter having a specialized use, which is reflected in the domain of occurrence. It could be argued that when this form means "to believe in" that it could have a literal meaning. There was a slight tendency for insertions to occur in *put + det + faith in,* however this could be due to the fact that *put* is a primary verb, and as such may be thought to carry less lexical content. Adding an insertion thus adds extra information to the sequence.

The sequence *pin your HOPES/faith on* demonstrates that the boundaries between sequences are not clear. At a glance, it looks as though this sequence and its two main forms should make a basic network, however further investigation shows they have different properties. The two still have occurrences where they have the same meaning and use, and such occurrences cause an overlap in both form and reading. This shows that *pin + det + hopes on* and *put + det + faith in* could not be treated completely separately within an idiom dictionary, but must be dealt with with reference to each other.

6.6. Case Study 9: Kick up a STINK/Fuss

6.6.1. Definition and Senses

Kick up a STINK/fuss shows both grammatical and lexical insertions. It is the formulaic sequence in the dataset with the most grammatical insertions (see Table 6-1). Example 6-25 demonstrates the definition of this formulaic sequence as given by the LID, the definition being:

> To complain loudly when you are annoyed about something
> (1998: 327)

Example 6-25

A spokesman for the band, at No 11 with Fire says: "We've played dozens of gigs in recent months and we haven't brought the roof down yet. "The only reason they're **kicking up a stink** is because they don't like rave music."
The Daily Mirror
CH1 8140

This sequence is found in the CCDI, suggesting *row* and *fuss* as possible noun substitutes. It also occurs in the CIDI, with the same sense, and with the suggestion that the type *kick up a stink* is informal and British English in dialect, and the types *make/raise a stink* are informal American types. There is no suggestion of a noun substitute in the CIDI.

Table 6-15 and Table 6-16 show the number of substitutes found in the BNC. There were 379 tokens of this sequence found in total, with six noun substitutes and eight verb substitutes. The different nouns found were: *scene, fuss, stink, row, shindy,* and *rumpus,* and the verb substitutes were: *make, kick up, create, cause, put up, start, raise,* and *be.* Some of the verb substitutes have a preposition attached. The substitutes here; *kick up* and *put up,* have the preposition as a verb particle, and thus the phrase structure remains VP NP.

Lexical Verb Substitute	Frequency
Make	291
Cause	41
Kick up	28
Create	13
Be	3
Put up	1
Raise	1
Start	1
TOTAL =	**379**

Table 6-15 – Lexical verb substitutes for *Kick up a STINK/fuss*

Lexical Noun Substitute	Frequency
Fuss	286
Scene	58
Stink	18
Rumpus	14
Shindy	2
Row	1
TOTAL =	**379**

Table 6-16 – Lexical noun substitutes for *Kick up a STINK/fuss*

Table 6-15 and Table 6-16 show that there are more occurrences for *fuss* and *make* than the other possible noun or verb substitutes respectively. This is supported by Table 6-17 which shows the frequent lexical combinations for the phrase. This table focuses on the lexical verb and nouns involved in the types; insertions are embedded:

Lexical Combination	Frequency
Make + det + fuss	247
Make + det + scene	42
Kick up + det + fuss	16
Cause + det + fuss	13
Cause + det + rumpus	9
Create + det + fuss	9
Cause + det + stink	7
Kick up + det + stink	6
TOTAL =	**349**

Table 6-17 – Frequent lexical combinations of *Kick up a STINK/fuss*

Thus from Table 6-15, Table 6-16, and Table 6-17, it is clear that the most common type for this phrase is *make + det + fuss.* With 247 occurrences in 379 tokens, it is a type which cannot be overlooked. It is interesting, then, to note that this type does not appear in any of the idiom dictionaries in its own right, despite appearing in the BNC more frequently than the canonical *kick up a STINK/fuss.* Reasons for this will be discussed with reference to chaining and compositionality in section 6.6.2.

The sense found for this formulaic sequence was clearly that of the dictionary definition; to complain loudly when you are annoyed about something. However, there were noticeable

gradations involved in the tokens. Firstly, there was a sense of "to make a loud noise and ado to attract attention" as demonstrated in Example 6-26:

Example 6-26

"But with Grandma in my bed I'm safe", I wanted to shout. Instead I wedged a chair under the door handle so he couldn't get in. He tried, I heard him; he couldn't **make a fuss**, Mum would have heard.
Unpublished Creative Writing: Prose
HJC 401

There was also a sense of "To cause complaints, uproar, and commotion":

Example 6-27

In the cell he shouted a bit to bring them back, but only got shouted at in his turn by two drunken fellow inmates. Later a woman brought him a plate of bacon and eggs and a cup of tea, but said it was no good **making a fuss** now, it was too late.
Who, Sir? Me, Sir? – K. Peyton
AT4 4123

The sense of "to complain loudly when you are annoyed about something" is still inherent in the tokens with "to make a loud noise and ado to attract attention" and "to cause complaints, uproar, and commotion", and the underlying concepts of COMPLAINT and LOUD unite the meanings. However, there seems to be a scale of the exaggeration of the action:

Complain loudly	→	make a loud noise and ado to attract attention	→	cause uproar and commotion

There is a development of the subject complaining loudly when something dissatisfies them, to complaining loudly in order to attract someone's attention to complaining loudly and thus triggering others to join in and cause a commotion.

The senses of "to complain loudly when you are annoyed about something" and "to cause complaints, uproar, and commotion" are related via their underlying concepts of COMPLAINT and LOUD. The sense of "to make a loud noise and ado to attract attention"

has the underlying concept LOUD in common; it also has a negative behavioural connotation, although not necessarily one of COMPLAINT. It can have the underlying concepts of ADO and ATTRACTING ATTENTION. Thus Table 6-18 shows the senses of this formulaic sequence:

Sense	Frequency	Total
To complain loudly when you are annoyed about something/ To cause complaints, uproar, and commotion	74 134	208
To make a loud noise and ado and attract attention	101	101
To make a smell	2	2
To look after and care for	66	66
Other	2	2
TOTAL =	**379**	**379**

Table 6-18 – Frequency of the senses for *Kick up a STINK/fuss*

There are other senses shown in Table 6-18 alongside the ones already discussed. The sense of "to make a smell" occurred in both tokens in the type *causing a stink,* in one example referring to someone's body odour, and another to the aroma of rotting meat left in a rubbish bin. Example 6-28 plays on this literal sense, whilst actually meaning economic unhappiness, competition, and complaints:

Example 6-28

Perfume price war… it's **causing a stink** in the High Street.
Central T.V. News Scripts [Headlines]
K1L 2064

The sense of "other" also had two tokens *create a scene* and *making a scene.* These were included by the rules of the chaining process which allowed for literal occurrences. They had senses referring to "producing a beautiful panorama". The final sense of "to look after and care for" will be discussed in section 6.6.2.

The senses of "to complain loudly when you are annoyed about something", "to cause complaints, uproar, and commotion" and "to make a loud noise and ado to attract attention" had predominantly non-literal meanings. *Create + det + fuss, kick up + det + fuss, kick up + det + stink,* and *make + det + scene* do not have meanings of V + det + N; a "fuss" or

"scene" is not literally built.[11] Note however, that where the verb was *cause*, the reading was more literal: *cause* + *det* + *fuss* could be read literally and can be analysed compositionally; the noun *fuss* having a meaning of:

a Needlessly nervous or useless activity; commotion

bi A state of excessive and unwarranted concern over an unimportant matter

bii An objection; a protest.

c A quarrel.

d A display of affectionate excitement and attention

(http://www.dictionary.com)

Likewise, the types *create* + *det* + *scene* and *make* + *det* + *scene* also have a literal compositional reading when the definition of *scene* is:

A public display of passion or temper
(http://www.dictionary.com)

The noun *scene* occurred 58 times, and of those, 47 tokens had the sense of "to make a loud noise or ado to attract attention", in accordance with the dictionary definition. Where the sense was of "making a smell", "to look after and care for" or "other", the reading was literal.

6.6.2. Networks and Compositionality

Table 6-15 and Table 6-16 show that there are peripheral members of this phrase's lexical set; *row* and *shindy* being marginal noun substitutes, and *put up, start, raise*, and *be* being subsidiary verb substitutes. For *kick up a STINK/fuss,* a network of types seems to form. There are central and peripheral noun members, *fuss, rumpus, shindy*, and *row* being broadly synonymous, as are *make* and *create*:

[11] The combination *make* + *det* + *fuss* will be discussed in section 6.6.2.

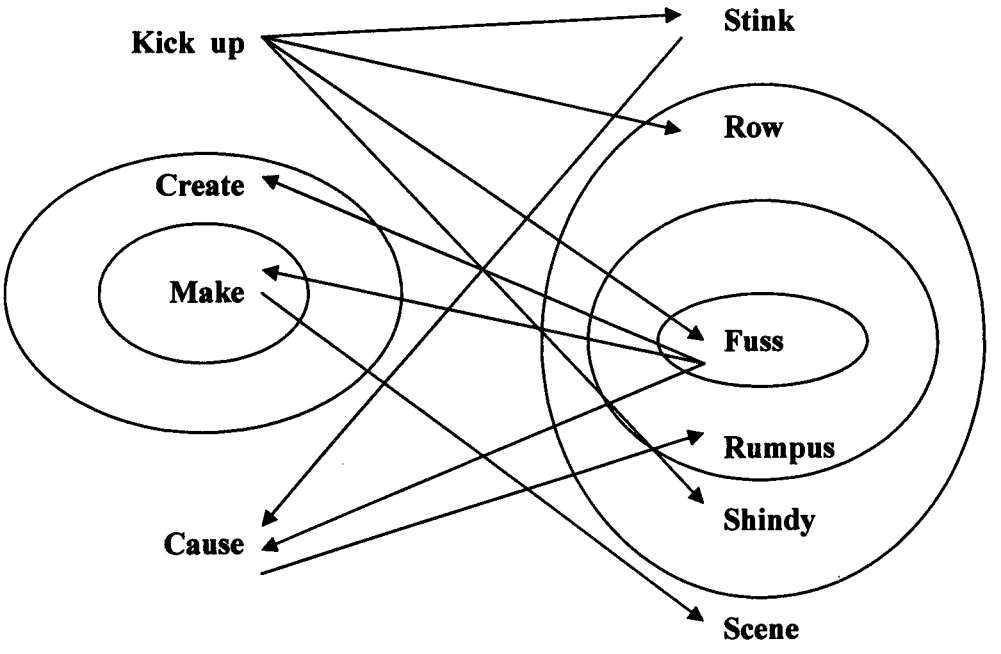

Figure 6-4 - Constellation network for *Kick up a STINK/fuss*

Figure 6-4 is a constellation network. The meaning progresses from "to complain very loudly" to "make a loud noise for attention" to "cause uproar and commotion". The same underlying concepts of COMPLAINT and LOUD are present throughout the progression from the combination *kick up* + *det* + *stink* to *make* + *det* + *fuss.* However there are two possible endings to the chain. The first ending is as shown in Figure 6-4; the chain stops at the meaning of uproar and the form *make* + *det* + *scene*. The second ending is:

Make a fuss → make a fuss of

The meaning of *make a fuss of* is the sense "to look after and care for". It does not have the same underlying concepts of COMPLAINT and LOUD. The phrase *make a fuss* thus has three potential readings:

1. to complain loudly when you are annoyed about something/to cause complaints, uproar and commotion
2. to make a loud noise and ado to attract attention
3. to look after and care for

Example 6-29 demonstrates the sense of "to look after and care for":

Example 6-29

A couple of cats hung around number forty-three, but they avoided Simon-who in any case never fed them. Mick **made a fuss** of them and the girls brought home tins of catfood from the supermarket for them.
A Season for Murder – A. Granger
CEB 2155

There are 247 occurrences of the phrase *make + a + fuss*. Of these, 137 had the sense of "to complain loudly when you are annoyed about something/to cause complaints, uproar, and commotion". 44 occurrences had the sense of "to make a loud noise and ado to attract attention". All 66 of the sense "to look after and care for" had the form *make + a + fuss*. Table 6-19 looks at the post-modifying prepositions which were found to follow the noun *fuss*:

	Post-modifying Preposition					
Sense	**Null**	**About**	**Of**	**Out of**	**Over**	**TOTAL**
1	82	48	1	0	6	137
2	29	11	0	1	3	44
3	10	2	52	0	2	66
					TOTAL =	**247**

Table 6-19 - Post-modifying prepositions for the *Make + det + fuss*

Sense 1 = to complain loudly when you are annoyed about something/to cause complaints, uproar, and commotion

Sense 2 = to make a loud noise and ado to attract attention

Sense 3 = to look after and care for

From Table 6-19 it is clear that the lexical combination *make* + *det* + *fuss* is found predominantly with the dictionary definition, and has a high collocation with "about" as a post-modifying preposition. The same combination is also favoured for the lesser sense of making a loud noise and ado. Of the third sense, this shows a high preference for the form *make* + *a* + *fuss* + *of*. Of the 66 total occurrences of this sense, 52 of those were followed by the preposition *of*.

It is interesting to note that *make* + *det* + *fuss* has negative connotations of COMPLAINT and LOUD or ADO and ATTRACTING ATTENTION, the form *make* + *det* + *fuss* + *of* has an opposite connotation; it is a positive one of looking after someone or something and caring for them. Moon suggests that formulaic sequences exist which may be polysemous (1998a: 187). They have two or more non-compositional meanings in addition to any literal one, and they are "often associated with different collocations or realizations of subject or object" (1998a: 189). As will be seen, the reading of *make* + *det* + *fuss* is literal, regardless of sense, the readings are associated with different prepositions, which vary according to the object of the action; *of* applies here to an animate referent and sense 3, whereas the null or *about* post-modifier refers to sense 1 or 2.

The form *make* + *det* + *fuss* + *of* has a predominantly literal reading. Take senses 1, 2a, and 4 from the definition of *fuss* as shown in section 6.6.1:

a Needlessly nervous or useless activity; commotion

bi A state of excessive and unwarranted concern over an unimportant matter

d A display of affectionate excitement and attention

(http://www.dictionary.com)

It is literally possible for a person to make a display of affectionate excitement and attention, or to make a needlessly nervous or useless commotion over someone.

Taking the meanings of the individual lexemes then and applying the remaining senses of *fuss* from section 6.6.1, it is possible to see the form *make + det + fuss* as being able to be analyzed compositionally. The same cannot be said for *kick up + det + fuss/stink, create + det + scene, make + det + scene* or *create + det + fuss.*[12] It seems that *make a fuss* has a frequently occurring verb which is open to delexicalization (see section 2.2.2). This is supported by the fact that the dictionary definition for the word *fuss* has senses in common with the meaning of the whole phrase. The fact that the verb carries less lexical content in this case is also seen by examples in the BNC such as *don't fuss, without a fuss*, and *no fuss*; it is the noun in these cases which carries the meaning. Note that *make* can be a primary or delexicalized verb of high frequency.[13] It is also possible that the verb could take its literal sense, and the sequence *make + det + fuss* has three possible readings depending on the sense of the noun.

Whilst the senses of the variant types for *kick up a STINK/fuss* and *make + det + fuss of* may be different, there is a common link between them. In all three of the senses shown in Table 6-19, there is an impression of exaggerated behaviour and activity, be it in a negative shade of complaining and making a noise, or in the more positive light of generating excessive concern or affection.

6.6.3. Insertions

Kick up a STINK/fuss has 166 tokens demonstrating insertion, 92 of which are grammatical insertions, and 74 of which are lexical. There were five tokens which had both a lexical and a grammatical insertion present. This phrase has by far the largest number of grammatical insertions of all the phrases in the dataset (see Table 6-1), and also has the highest number of insertions in total. This could be thought to be due to the number of tokens for this phrase, however, the phrases *watch your TONGUE/mouth* and *hit the DECK/dirt* are the top two phrases in terms of tokens (see Table 4-7 and Table 4-8 section 4.3.2), and yet the former has no insertions, and the latter only one token with an insertion. *Kick up a STINK/fuss* has a high number of tokens with a literal reading. There were 73 of the 92

[12] *Scene* with the sense of "to cause a commotion", not the literal sense of "panorama".

[13] Perhaps this explains why the form *make + det + fuss (of)* was not found in any of the idiom dictionaries, despite its high frequency; it is too compositional and literal in reading.

grammatical insertion tokens which had a literal reading, and 57 of the 74 lexical insertion tokens. As was mentioned with regard to *overstep the MARK/limits/bounds,* perhaps the literal readings allow for more insertions than the tokens with a non-literal reading; the literal readings are analysed in a compositional manner, and thus are not affected by the insertion, particularly in cases where the verb is delexicalized. The non-literal readings are often analysed non-compositionally and thus an insertion affects the reading of the phrase. The high number of occurrences of this phrase with a literal reading explains to some extent the high number of occurrences with an insertion.

From the 74 lexical insertions, there were 40 different types. Types which had a frequency of more than one were *big* (12 tokens), *great* (eight tokens), *public* (six tokens), *huge* (four tokens), and *awful, real,* and *embarrassing* with three tokens each. Using Quirk and Greenbaum's semantic sets for adjectives (1973: 125), the lexical insertions can be divided into:

Intensifying adjectives	24
General adjectives susceptible to subjective measure	32
General adjectives susceptible to objective measure	18
TOTAL =	**74**

Most of the insertions were of a subjective internal and intensifying nature, for example:

Kicking up an **awful** stink	KDS 962
Make a **big** fuss	CC7 362
Made the **most dreadful** fuss	A7W 584

Kick up a STINK/fuss, in the form *make* + *det* + *fuss,* has a high number of literal readings, as does *cross* + *det* + *line* in section 6.4.3. However *cross* + *det* + *line* allowed for general adjectives susceptible to an objective nature, whereas the insertions for *make* + *det* + *fuss* are more subjective. Also, in contrast with *cross* + *det* + *line,* which had a high number of noun modifiers, the lexical insertions here are adjectival. This is due to meaning; a *fuss* can be emphasized and exaggerated, whereas the "lines" talked about in section 6.4 are

concrete physical referents of specific kinds, e.g. sports or railway, and so are open to more concrete descriptors, such as noun modifiers.

The 92 grammatical insertions also show an emphatic attitudinal function. There were 21 types of grammatical insertions, and all were insertions of degree. They seem to be epistemic and attitudinal, and remove the author from ownership of the phrase. Again, these act as face-saving acts and hedges as opposed to emphasis, as in *such as.*

Kick up **some kind of a** rumpus	AB3 1743
Kicks up **quite a** fuss	BOP 1546
Make **a bit of a** fuss	FMF 9

Insertions such as *some kind of* are compromisers in Quirk and Greenbaum's scheme (1973: 218); they are downtoners which have only a slight lowering effect on the action of the verb.

There was a high preference in the grammatical insertions for *such.* There were 41 tokens that took this insertion, as Example 6-30 shows:

Example 6-30

"I can't think of any other woman of my acquaintance who would **kick up such a fuss** just because a man wanted to buy her some new clothes. Most of the girls I know-"
Double Fire - M.Lynch and M. Lyons
JXX 1188

Of these, 33 took the form *make a fuss,* and four had the form *make a fuss of.* Of these, 20 had the sense of complaining loudly, seven had a sense of making a loud noise and ado, and six had the sense of to look after and care for. The insertion *such a* is one which adds information about the speaker/author's emotion and attitude. Leech and Svartvik (1994: 1930) call such insertions ways of giving emotional emphasis. *Such* usually occurs before the indefinite article with singular count nouns, such as *fuss.* In total, of the 92 grammatical insertions, there were 64 tokens of *make* + *det* + *fuss,* and 12 of these had a prepositional post-modifier *of.* Of the lexical insertions, there was a similarly high number of occurrences

of *make + det + fuss;* 42 of the 74 occurrences, with 8 of those being *make + det + fuss + of.*

Again, as with *take a BEATING/hammering,* there are insertions such as *a great deal of* and *a hell of a* which emphasize the degree of the fuss being made. Such intensifiers can be referred to as *boosters* (Quirk and Greenbaum, 1973: 216). They are amplifiers which denote a high point on a scale, rather than the extreme position.

There were five tokens where there was both a grammatical and lexical insertion, for example:

Created **more of a public** stink	GXG 1197
Made **such a great** fuss	CAT 963

Of these, three tokens had the grammatical insertion *such a*, and two had *more*, again having an emphasizing attitudinal effect, particularly in the example ***make such a damn*** *fuss (KRL 128),* where the colloquial adjective *damn* also adds to the emphasis and attitude. This is also the case for the five tokens where there was a modifier to the lexical insertion, for example:

Makes a **lot** less fuss	C9L 1745
Made the **most** dreadful fuss	A7W 584

These modifiers were quantifiers, and were used for emphasis of the adjective, which was also subjective and attitudinal.

Kick up a STINK/fuss also has a high number of tokens with a quantifying determiner. There were 21 tokens with a quantifying determiner and no insertions, all of which had the noun *fuss,* and 17 having the form *make + det + fuss.* Whilst *pin your HOPES/faith on* was open to determiner substitutions due to the pronoun determiner, this has an indefinite article, which can also be open to change; more so than the definite article. Of these determiner substitutes, again a sense of degree is seen, with six tokens of *more* and four tokens of *much.* There were three occurrences where a low quantity was seen with the

determiner *no*. All of the quantifying determiners occurred with the noun *fuss*. *Fuss* is a non-count noun; however, the meaning of *fuss* is if not countable, gradable. It is possible to quantify *fuss* in non-numeral terms with terms such as *some, more*, and *much* etc. (Burton-Roberts, 1997: 57).

6.6.4. Summary

Kick up a STINK/fuss forms a chaining constellation network with phrasal members as the canonical form, the prevalent literal form *make + det + fuss* and the form *create + det + scene*. The insertions found for *kick up a STINK/fuss* were mainly intensifying, subjective or attitudinal. They were internal, modifying the noun. Although there were many more examples of *make + det + fuss,* there were examples of insertions with other nouns and verbs, suggesting that insertion is not restricted to *make + det + fuss*. The type *make + det + fuss* had a clear meaning distinction between *make a fuss* (with underlying of concepts LOUD and COMPLAINT) and *make a fuss (of)* (to look after and care for someone). *Make + det + fuss* can thus be seen to be a polysemous sequence, with different meanings depending on the sense of the noun. Both forms are literal in reading. There is also a high tendency for *make + det + fuss* to take the grammatical insertion *such.* The fact that *make + det + fuss* can be polysemous in terms of its specific sense demonstrates that the boundaries of meaning may be "fuzzy", and cannot be neatly pinned down. Section 6.6.2 showed that there was a meaning of exaggerated behaviour and activity that is an underlying sense uniting the "surface" meanings, which may be beneficial in organising these forms rather than listing each atomistically.

6.7. Case Study 10: Raise HELL/Cain

6.7.1. Definition and Senses

Raise HELL/Cain has a high preference for lexical insertions. It also has a high number of tokens with a literal reading. Table 4-6 in section 4.3.2 showed that this sequence, as with other sequences in the case studies, had a low number of canonical tokens in comparison to the total tokens. This will be discussed in section 6.7.2.

The definition for *Raise HELL/Cain* in the LID is given as:

a to complain a lot about something in an angry or noisy way because you are determined to get what you want

b to cause trouble by behaving in an irresponsible way, and being noisy or violent

(1998: 168)

The dictionary also notes that sense one is primarily a spoken one, and that *Cain* is considered to be a more polite version than *Hell.* It is also mentioned that this phrase generates the compound *hell-raiser.* This definition and compound are demonstrated in Example 6-31:

Example 6-31

At some point during the fifties the press coined a phrase, not entirely new but then fashionable, for any actor who emulated the lifestyle of Errol Flynn, if only in part. All they had to do was live life through an alcoholic haze and generally **raise hell** - hence the term "Hellraiser".
Hollywood Rogues – M. Munn
CDG 357

The CCDI elaborates on this definition:

A If you say that someone raises hell, you mean that they cause trouble by behaving badly in public, for example by getting drunk and breaking things or upsetting other people

B If someone raises hell about a situation, they protest strongly and angrily about it in order to persuade other people to correct it or improve it

(2002: 193)

The CCDI also acknowledges a *hell-raiser* to be someone who frequently causes trouble by behaving badly in public. The CIDI suggests the same two senses, and proposes that the sense of behaving badly is a mainly American one.

The formulaic sequence *raise HELL/Cain* and its variant forms has 427 tokens in the BNC and 128 types. There are five different possible noun substitutes: *Hell, Cain, havoc, chaos,*

and *mayhem,* and nine different verb substitutes: *raise, play, create, cause, produce, unleash, do, reap*, and *wreak.* There were also ten different determiner substitutes possible: *null, such, her, a, the, some, more, further, little*, and *enough.* Table 6-20 and Table 6-21 show the frequencies of the different noun and verb substitutes:

Lexical Verb Substitute	Frequency
Cause	163
Wreak	100
Play	75
Create	60
Raise	21
Produce	3
Reap	2
Unleash	2
Do	1
TOTAL =	**427**

Table 6-20 – Lexical verb substitutes for *Raise HELL/Cain*

Lexical Noun Substitute	Frequency
Havoc	262
Chaos	102
Hell	36
Mayhem	22
Cain	5
TOTAL =	**427**

Table 6-21 – Lexical noun substitutes for *Raise HELL/Cain*

Table 6-20 and Table 6-21 show that the canonical forms *raise, Hell* and *Cain* are not the most frequent lexemes for this formulaic sequence. The nouns *havoc* and *chaos* are clearly numerous, as are the verbs *cause, create, wreak*, and *play*. This is also clear from Table 6-22, which shows the frequencies of the different types for this phrase. There was a null determiner in this formulaic sequence.

Lexical Combination	Frequency
Wreak + havoc	97
Cause + chaos	82
Cause + havoc	67
Play + havoc	61
Create + havoc	33
Raise + Hell	16
Create + chaos	16
Cause + mayhem	14
Play + hell	14
Create + hell	6
Create + mayhem	5
Raise + Cain	5
Wreak + mayhem	3
Reap + havoc	2
Unleash + chaos	2
Produce + chaos	2
Do + havoc	1
Produce + havoc	1
TOTAL =	**427**

Table 6-22 – Frequent lexical combinations of *Raise HELL/Cain*

Table 6-22 shows that the canonical forms *raise Hell* and *raise Cain* are infrequent compared to the types *wreak havoc, cause chaos, cause havoc*, and *play havoc.* In the LID, CCDI, and CIDI, *play hell,* or *play (merry) hell (with)* are entries in their own right. The LID quotes *play (merry) hell (with)* as having the definition "to make a lot of trouble or cause a confusing situation". The CCDI suggests *play hell* or *play merry hell* to mean causing trouble by behaving badly or complaining about something, as does the CIDI. Interestingly, the CCDI and CIDI have separate entries for *play (merry) hell* and *play (merry) hell with.* The definitions given for the latter are "to stop something from working as it should" (CIDI) and "if you say that one thing plays hell with another, you mean that the first thing has a bad effect on the second one or causes great confusion" (CCDI). Of the results found in this research, there are fourteen examples of *play + hell.* Ten of these take the form *play + hell + with,* and two of the ten have the structure *play + merry + hell + with.* There are also two examples of the form *raise + hell* which also take the insertion *merry.* Only the LID has an entry for the type *wreak/play havoc,* with the meaning "to cause a lot of trouble or confusion". None of the dictionaries have an entry involving *chaos.*

Of the 61 tokens of *play havoc,* 59 were followed by the preposition "with", suggesting that this type is actually *play havoc with.* In fact, this is noted in the OED Online. Under the sense for *havoc*, *play havoc* is mentioned and it is suggested it occurs frequently with *with.* The verb *play* can be intransitive requiring no object complement, or transitive, needing an object to satisfy it. In these tokens, the verb is transitive, and requires the noun to satisfy it. However, in the phrase *play havoc* the verb seems to be ditransitive, requiring both the noun and qualification of what is being played with. Most of the occurrences have an inanimate qualifying object in the prepositional phrase, e.g. weather or politics are being played havoc with.

There were several meaning senses found from the tokens in the BNC. Firstly, there was the sense of "to complain about something in an angry or noisy way because you are determined to get what you want", as shown in Example 6-32:

Example 6-32

You'd be quite comfortable, I assure you," the man went on, and by his pleading tone Breeze guessed that his employer was used to having his own way - and **raised Cain** when he didn't get it.
The Distance Enchanted – M. Gervaise
BMU 346

It is interesting to note that four of the five examples of *raise + Cain* took this sense; the canonical form taking the primary sense. It could be suggested that the forms *raise Hell* and *raise Cain* are not variants of the same phrase; the verb *raise* has a different sense in each. It has a sense of "create" in the former and "resurrect" in the latter. The motivation between the literal readings and the figurative ones are different. However, the surface meaning of complaining or causing trouble is the same for both forms.

The second sense found was that of "causing trouble by behaving in an irresponsible way, and being noisy or violent". This is shown in Example 6-33 :

Example 6-33

Finch disapproved of Flynn coming to work in such a state, but when drunken Flynn **caused chaos** on the set, Finch knew that the fading swashbuckler was incapable of filming for that day

Hollywood Rogues – M. Munn
CDG 509

The third sense was one of "causing damage to", as demonstrated by Example 6-34:

Example 6-34

Beyond them, everything was chaos. It was mid-morning, and the streets were thronged with people and animals. The flood had **caused tremendous havoc**, breaking down many buildings.
Frankenstein Unbound – B. Aldiss
HGS 2193

Most of the occurrences with this sense took the form *cause/wreak havoc.* This sense is the one given by the CCDI and CIDI as being the definition for the phrase *play (merry) hell with;* a sense of one thing having a bad effect on another. None of the examples of *play (merry) hell (with)* found here took this definition; they all had meanings involving noise, trouble or complaint.

The fourth sense found was similar to the second sense in that it was "to make a lot of trouble or cause a confusing situation". It was associated again with the forms *cause* or *wreak havoc,* and also with the types *cause chaos* and *play havoc:*

Example 6-35

Nobody else was hurt, but surrounding houses were scorched by heat from the blast. The explosion **created chaos** on surrounding roads when traffic-lights were put out of action.
The Belfast Telegraph
HJ4 8720

Table 6-23 summarizes the frequencies of the senses found:

Sense	Frequency
1	21
2	37
3	52
4	317

Table 6-23 – Senses of all of the tokens of *Raise HELL/Cain*

Sense 1 = to complain about something in an angry or noisy way because you are determined to get what you want

Sense 2 = to cause trouble by behaving in an irresponsible way, and by being noisy or violent

Sense 3 = to cause damage to; stop something from working as it should; to have a bad effect on something else

Sense 4 = to make a lot of trouble or cause a confusing situation

The underlying concept for all of these is TO CAUSE TROUBLE, either in complaints, actions, or causing trouble to something else. The four senses differ in terms of specifying the type of trouble being caused.

Lexical Combination	Sense1	Sense 2	Sense 3	Sense 4
Cause chaos	1	3	5	73
Cause havoc	1	3	7	56
Cause mayhem	0	6	2	6
Create chaos	0	1	0	15
Create havoc	1	8	0	24
Create hell	4	0	0	2
Play hell	5	1	0	8
Play havoc	0	0	6	55
Raise Cain	4	1	0	0
Raise Hell	5	10	0	1
Wreak havoc	0	0	32	65
TOTAL =	21	33	52	305

Table 6-24 - Senses of frequent lexical combinations for *Raise HELL/Cain*

Table 6-24 , taking the same senses as shown in Table 6-23, demonstrates the frequencies of the different senses per type found for this case study. Sense one; that of complaining loudly has a preference for the lexical noun *hell*. Sense three; that of having a negative

effect on another has 52 tokens in total, and 32 of these occur in relation to the form *wreak havoc.* The type *raise hell* has a preference for sense two, that of behaving in an irresponsible or noisy way. Sense four; that of causing trouble or a confusing situation occurs frequently, but seems to have a preference for the types *create/cause havoc/chaos, play havoc,* and *wreak havoc*; i.e. not the canonical form *raise hell/Cain.*

6.7.2. Networks and Compositionality

From the discussion in section 6.7.1, it is seems there is a developmental path both lexically and semantically. The starting point is *raise Hell,* with a tendency to mean behave irresponsibly, noisily or violently. *Raise HELL* chains to *raise Cain,* which seems to favour a meaning of complaining loudly. The chain follows on to *create/cause havoc/chaos, play havoc* and *wreak havoc,* with a meaning of causing trouble or confusion, and finally *wreak havoc* has a secondary sense of one thing having a bad effect on another. The underlying meaning throughout is of CAUSING TROUBLE.

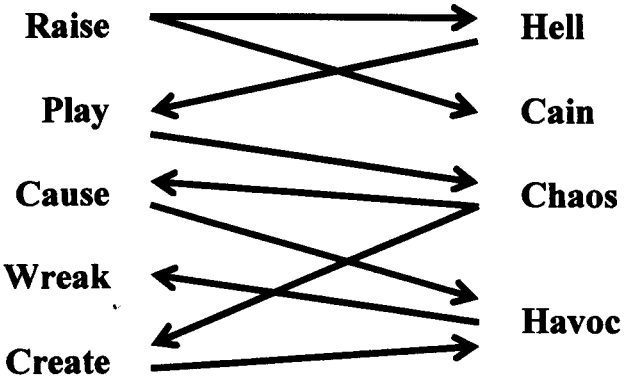

Figure 6-5 – Chaining network for *Raise HELL/Cain*

In terms of compositionality, this phrase acts in the same way as in the case study *kick up a STINK/fuss* and its variant type *make a fuss* (see section 6.6). The canonical form and its variant *raise Hell/Cain,* could be said to be non-compositional, i.e. that it has a meaning other than the meaning of the parts it is made up of. The meaning of *raise Hell* is not literal. The behaviour of Hell has been raised to Earth by someone's behaviour. A similar imagery

meaning base could also be found with *raise Cain;* "Cain" referring to the first biblical murderer. These types have a non-literal reading.[14]

Where the verb is *create, cause,* or *wreak*, the meaning is more compositionally analysable, when the noun is *chaos, mayhem*, or *havoc*. It is possible to cause or create havoc, chaos or mayhem; *havoc* has the OED definition of devastation, destruction or in a weakened sense, confusion, disorder, and disarray. *Chaos* is defined as being a state resembling primitive chaos; utter confusion and disorder, and *mayhem* receives a similar description. As with *make a fuss* in section 6.6, such forms have not been included in the CIDI or CCDI possibly due to their literal, compositional nature. The sequence *wreak/play havoc* has been included as an entry in the LID. This could be due to many factors: frequency, the relative rareness of the verb *wreak* and its obvious collocation with *havoc*, and the non-literal interpretation of *play havoc.* It seems that the use of the chaining process has here led to a progression from the non-literal canonical *raise HELL/Cain* to more literal combinations such as *cause/create chaos/mayhem/havoc.* The boundaries overlapping between non-literal and literal language can be seen in this network, showing integration of the formulaic sequence into natural language; traditionally non-literal language was kept separate from literal language (see section 1.1). The literal and non-literal combinations are not necessarily distinctly separate formulaic sequences as they share the same underlying meaning and lexical set however it seems the different combinations for the canonical formulaic sequence demonstrate the boundaries between literal and non-literal language.

The meanings of the phrase "causing trouble by behaving in an irresponsible way, and being noisy or violent" and "to complain about something in an angry or noisy way because you are determined to get what you want" are very close to the meanings found for the case study *kick up a STINK/fuss,* and that there is possible scope to create a link in the chain from *create chaos/havoc* with these meanings to *create a scene,* and thus to *kick up a fuss.*

[14] The OED Online suggests that the phrase *raise hell,* with a meaning of "to create a disturbance; to cause great trouble", could possibly have originated in the slogan "Kansas should raise less corn and more hell" by Mrs Mary Ellen Lease (1853 – 1933), although there is little evidence to support this.

6.7.3. *Insertions*

Raise HELL/Cain has a high number of lexical insertions, and fewer grammatical insertions; there were 72 tokens with lexical insertion and 22 grammatical insertions. Most of the insertions came from the most frequent forms as seen in Table 6-22. Table 6-25 demonstrates the frequency of these common types and the type of insertion they had:

Lexical Combination	Lexical Insertion	Grammatical Insertion
Wreak havoc	13	7
Cause Chaos	25	3
Cause Havoc	4	2
Create havoc	4	4

Table 6-25 – Frequent lexical combinations and their insertions for *Raise HELL/Cain*

The non-literal type *play havoc* was seen to be frequent in Table 6-22 however there was only one occurrence of this type with any insertion; a grammatical insertion. Most of the insertions were found with a literal reading. All of the grammatical insertion tokens were non-literal, and only ten of the lexical insertion tokens. The tokens with insertions which had a non-literal reading had the noun *hell*, and the verbs *create, raise* or *play*. This concurs with what was seen in section 6.7.2 regarding literal verbs readings and compositional analysability. It is also in concurrence with the cases *kick up a STINK/fuss* and *overstep the MARK/limits/bounds*, in that there is a high number of literal tokens and a high number of insertions.

The lexical insertions for this case study presented a phenomenon not seen in the other case studies; some tokens were open to noun coordination. There were 19 tokens where the noun was coordinated, as Example 6-36 demonstrates:

Example 6-36

The group may become self-perpetuating and exclusive, failing to encourage the involvement of the majority of parents; or be badly led by a bunch of incompetents who **cause chaos and disaster** with every move they make.
Marketing your Primary School – M. Sullivan
AND 1023

The coordination here works to economize the language used. However, the example of verb coordination given in *cross or straddle the line* in section 6.4.1 was a method of adding information regarding the manner of crossing the line. The examples of coordination found for *Raise HELL/Cain* tended to use the coordination for emphasis, as well as adding information, for example:

Create **chaos and anarchy**	HKV 200
Caused **chaos and fear**	CBF 5995
Caused **"mayhem and destruction"**	CBF 1989

In the same way as the verb coordination was an insertion in section 6.4.3, so the noun coordinates here are lexical insertions:

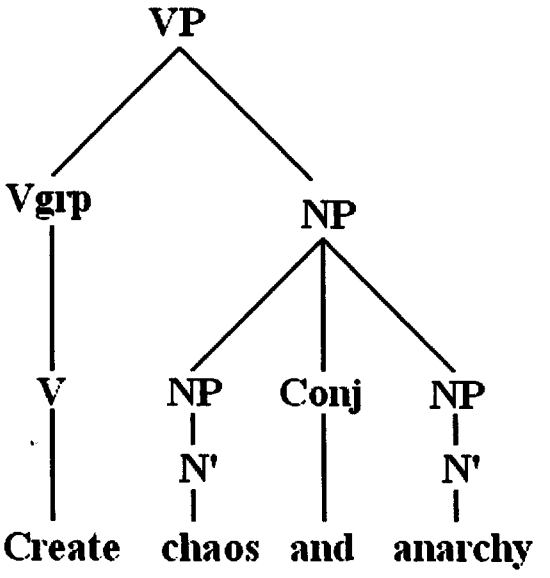

Figure 6-6 - Syntactic structure demonstrating noun coordination for *Raise HELL/Cain*

Of the 19 tokens with noun coordination, there were two occurrences of *mayhem and destruction*, and two occurrences of *anarchy and chaos*; *chaos and anarchy*, and *anarchy and chaos*. There were also cases of lexical and grammatical insertions seen in these noun co-ordinations; seven had an adjectival insertion, and one had a grammatical insertion, for example:

Created **such havoc and destruction** AR8 983

Cause **total chaos and huge problems** HHV 22628

Wreak **revenge, mayhem and generally unbelievable havoc** CEK 6013

In this last example, there is asyndetic coordination of three nouns, two of which are in the lexical set of nouns for this phrase. Also, the adjective *unbelievable* is itself modified by the adverb *generally*, these effects giving emphasis to this phrase.

So of the 72 lexical insertions, there were 19 tokens involving noun coordination. Seven of these also had adjectival insertions; there were 60 occurrences with adjectival insertions. There were 12 tokens of the lexical insertions where the modifier of the noun was another noun. Ten of these twelve were related to traffic, and seven of these used the actual word *traffic*, for example:

Caused **traffic** chaos K2N 857

Caused **motorway** chaos CH1 8133

Whilst these are seemingly domain delimiters, such insertions act on the noun only. Their effect is a reduction in communicative form from "cause chaos on the motorway" to a fronting of the salient point *traffic*. All but one of these occurrences occurred in tabloid or local newspaper texts, with the remaining one being spoken in a television news broadcast. Of all of the tokens of this sequence, the world affairs and leisure domains were the most frequent, having 80 and 90 token occurrences respectively. Such domains contained journals and periodicals such as journalistic texts. A technique used in journalism is to play with words, especially formulaic language, using insertions etc. This is also the case with another noun modifier; a proper name of a month:

Cause **February** mayhem CFT 214

In the adjectival lexical insertions, there was an example of adjectival asyndetic coordination:

Wreaked **social, economic and political** havoc CAJ 1157

There were also three other occurrences with more than one lexical insertion; i.e. having an adjectival modifier, creating a larger emphatic effect:

Caused **widespread traffic** chaos	K5M 4263
Cause **quite unprecedented** chaos and upheaval	K5D 1405
Wreak **revenge, mayhem and generally unbelievable** havoc	CEK 6013

The 60 adjectival insertions could be grouped as follows, according to Quirk and Greenbaum's classification scheme (1973: 125):

Intensifying adjectives	14
General adjectives susceptible to subjective measure	17
General adjectives susceptible to objective measure	29
TOTAL =	**60**

As with *overstep the MARK/limits/bounds,* this phrase is open to more insertions of an objective measure, with adjectives such as *travel, economic*, and *seasonal*, alongside the noun modifiers as mentioned. As with *cross the line,* where the modifier specified the kind of line being crossed, so the objective insertions here aim to specify where or on what the "chaos" was being "created". Unlike *cross the line*, which cannot be emphasized, this phrase can be open to intensifying or attitudinal objective insertions, as people can also show the degree of "chaos", as well as specifying it. In examples such as *seasonal, environmental*, and *economic,* the lexical insertion specifies the domain of the action. Such insertions are external, acting on the sequence as a whole. Seven of the lexical insertions were external adjectives. External adjectives are much rarer than internal insertions, as Ernst found (1981: 52).

There were 57 different types of lexical insertions, and of the grammatical insertion tokens, there were seven different types, indicating that this phrase is not fixed in terms of lexical insertion types. One of the grammatical insertions is the aforementioned *created such havoc and destruction,* with the grammatical insertion of emphasis *such.*

The remaining grammatical insertions also served as degree intensifiers, emphasizing the degree of *chaos* or *havoc* being made, for example *so much, rather more*, and *even more*. Of the 22 tokens of grammatical insertions, there were 13 tokens with the emphatic *such*. There is a null determiner in this sequence, the emotional emphasis of *such* remains as in *such a.*

6.7.4. Summary

Raise HELL/Cain is a non-literal sequence, with few canonical occurrences in the BNC. It develops gradations of meaning such as "to complain loudly", "to cause trouble by being noisy and violent" "to cause damage" and "to cause a confusing situation". These have the same underlying meaning of CAUSE TROUBLE. The Networks seen for this sequence demonstrate that the meaning of the sequence and those of the lexically related forms it chains to, are difficult to pin down precisely. There is overlap in the use of different forms. As has been seen in previous case studies, the most frequent combinations for *raise HELL/Cain* were those with a literal compositional reading. These are distinguished from the canonical form by being open to more insertions. The overlaps between form, meaning and reading, as well as the fact that there are more substitutions and insertions allowed for suggest that sequences such as *raise HELL/Cain* need to be considered alongside their variant literal and non-literal variant forms.

6.8. Conclusion

These case studies show that there are far more insertions occurring in actual use than the LID allows for. It is perhaps the case that as well as specifying sequences with context-dependent slots, idiom dictionaries need to make a note of the fact that many formulaic sequences (not just those with open-slots) can take insertions.

The sequences here were seen to give further examples of constellation and chaining networks. Unlike the sequences in Chapter 5, many of the canonical non-literal cases in this chapter led to a highly frequent form that had a literal reading. It was also seen, e.g. for *kick up a STINK/fuss,* and *raise HELL/Cain,* that it was the types with the literal reading which were open to the most insertions. It seems clear that this evidence supports previous work

regarding compositionality (see section 2.2.2). The literal types, being compositional in reading, were able to take insertions for specification or emphasis, but the non-literal canonical forms were not. In many cases, e.g. *cross + det + line,* the presence of an objective insertion such as *railway,* helped to determine a compositional literal form from a non-literal one. The vast majority of insertions, both lexical and grammatical, were internal, modifying the noun only. Perhaps external modifiers are only accepted by non-decomposable sequences, where the adjective modifies the sequence as a whole, e.g. "With that dumb remark at the party last night, I really *kicked the social bucket*" (Ernst, 1981: 1). Ernst also points out that many adjectives can have properties of internal and external adjectives.

The sequences selected were open to different types of insertions. All were open to both lexical and grammatical insertions; apart from *overstep the MARK/limits/bounds* which only showed lexical insertions. *Pin your HOPES/faith on* and *kick up a STINK/fuss* also showed a high number of quantifying determiners. The grammatical insertions showed a degree of commitment to the action of the formulaic sequence, whilst the lexical insertions seemed to be more stylistically creative, or specificatory. The grammatical insertions were also in the main emphasizers, adding to the effect, however, there were examples such as "a bit of a" which acted as face-saving hedges. Most of the lexical insertions found were adjectival, although there were some noun modifiers found, especially for the literal specificators in *overstep the MARK/limits/bounds,* and *raise HELL/Cain. Pin your HOPES/faith on* had a high number of intensifying adjectives. Most insertions were subjective or intensifying adjectives. Much fewer demonstrated "definite" adjectives such as age, colour or material, unless the sequence had a literal reading.

The high degree of literal occurrences indicated that the boundaries between a non-literal formulaic sequence and its literal counterpart are not clear. There is a great deal of overlap in terms of form and meaning, for example there are some forms, e.g. *cross + det + line* which may take a literal or non-literal reading. The literal reading is also not too dissimilar from the non-literal reading, for example the literal (compositional) reading of *make + det + fuss* and the non-literal reading of *kick up + det + fuss* are more similar than for example the literal and non-literal readings of *kick the bucket.* This provides more support for Network theories. Non-literal sequences cannot be stored alone and individually. They

must be treated in regard to their lexical substitutions and literal forms that this substitution leads to. Formulaic sequences also need not be non-literal; a sequence such as *make + (such) + det + fuss (of)* seems to be a frequent prefabricated unit in its own right that is related to *kick up a STINK/fuss.* Again, this gives more support for ideas such as thematic panels in idiom dictionaries; grouping sequences with a common underlying concept rather than individually.

7. Conclusion

7.1. Introduction

The research problem investigated in this thesis was "what are the limits of variation in formulaic sequences?" This broad research problem was further specified as three research questions:

- What/where are the boundaries of lexical substitution and insertion?
- What/where are the boundaries between one formulaic sequence and another?
- What/where are the boundaries between the literal and non-literal interpretations of a formulaic sequence?

These questions were examined by taking a dataset of formulaic sequences from the LID, and investigating them using the BNC. The dataset consisted of 35 sequences, each of the structure V det N (VP NP), and each demonstrating only one kind of variation; lexical substitution of the noun. I developed a set of rules for use when entering terms in the BNC, and I called these rules the chaining process. The use of the chaining process allows for a systematic method of collecting all possible variants: lexical substitutes and insertions in this research project. The five phrases showing the most lexical substitutions of the noun, and the five phrases demonstrating the most insertions (lexical and grammatical) were identified for case studies. The case studies investigated were *take a BEATING/hammering, throw a WOBBLER/wobbly, shut your FACE/trap/cakehole/gob, watch your TONGUE/mouth,* and *flip your LID/wig* for lexical substitutions in Chapter 5 and *take a BEATING/hammering, overstep the MARK/limits/bounds, pin your HOPES/faith on, kick up a STINK/fuss*, and *raise HELL/Cain* for insertions in Chapter 6.

This chapter summarizes the findings of the research, and addresses each research question. Section 7.3 discusses the implications of the results, showing how they connect to wider fields of study. Section 7.4 details areas where there were delimitations made and where further improvements or developments could be made.

7.2. Research Questions

7.2.1. Research Question 1: What/Where are the Limits of Lexical Substitutions and Insertions?

The limits of a formulaic sequence are reached when the lexical substitution or the insertion to the sequence causes a substantial change in meaning. A change in underlying meaning means that the new combination no longer shares the same central concept as the canonical form. The boundaries of the formulaic sequence have been reached.

As part of the chaining process, the frequency of each variant form for each entry in the BNC was recorded. Thus the total frequencies of substitutes and insertions for each sequence were seen. The most frequent and more peripheral substitutes and insertions for each of the formulaic sequences investigated were revealed. The number of substitutes in actual use in the BNC is greater in most of the cases than those listed by the LID.[1] Lexical substitutions of the noun occur on a gradual scale; the highest number of lexical noun substitutes found was 12, and the lowest was for *keep your SHIRT/hair/pants on* occurring only with the noun *hair,* and *not mince your WORDS/matters* with only the noun *words.* Insertions had a larger scale; *kick up a STINK/fuss* has 166 tokens showing insertion. In contrast *chance your ARM/luck,* and *tempt FATE/providence* have only one token with insertion. There were ten sequences in the dataset showing no insertions.

All of the sequences investigated show the frequencies of each lexical substitute found, and those in Chapter 6 show the frequencies of both the lexical substitutes and the insertions. Each substitute (or insertion) was recorded, even if it only occurred once. This allowed for central and peripheral members of the lexical set to be seen. Section 5.3.1 for example, demonstrated that for the sequence *take a BEATING/hammering* nouns such as *battering* and *hammering* were frequent members of this sequence's lexical set, having 29 occurrences each. Nouns such as *licking* and *trouncing* were peripheral, having only one occurrence each. Similarly for the verbs, *take* was central for this sequence with 68 of the total 112 occurrences taking this verb. The verbs *receive, produce,* and *face* were rare, having only one occurrence. The substitutes with very few occurrences (only one or two)

[1] The sequences in the dataset were selected showing substitution of the noun only, thus any insertions found through use of the chaining process were not mentioned by the LID.

demonstrate the limits of the substitution. Peripheral variations demonstrate the outer limits of the variation for the particular formulaic sequence.

The frequencies of substitutions and insertions also allow fixedness of sequences to be emphasized. For the sequence *pin your HOPES/faith on*, there were only three noun and three verbal substitutes found, indicating lexical fixedness. The substitutes were also semantically closely related, for example, the verbs *pin, place, fasten*, and *put*, and the nouns *hopes, faith*, and *dreams*. Prototypes of the substitutions can be seen, as was seen in section 5.3.1 for the nouns for *take a BEATING/hammering*. These nouns (and verbs for *pin your HOPES/faith on*) have the same meanings, so the central and peripheral members of the lexical sets can be represented in terms of prototypical members for the particular sequence. Prototypes and frequencies can also be a factor in linking sequences together and progressing from one sequence to another, as will be seen in section 7.2.2.

7.2.2. Research Question 2: What/Where are the Boundaries between one Formulaic Sequence and Another?

Central underlying concepts provide the basis for grouping formulaic sequences. Networks identify the boundaries of the sequences. The boundary of the network is reached when substitution of the noun or verb substantially changes the meaning of the formulaic sequence. This is inevitably a subjective decision on the researcher's behalf. Often, at the boundary of a sequence, there is a combination of low frequency.

The sequence *take a BEATING/hammering* gave an example of a basic network. Each noun substitute found was synonymous, and so changing the noun did not affect the reading of the sequence. The underlying meaning of defeat/damage/loss – ATTACK remained the same throughout, and the structure was kept as V det N. Section 2.4.3 outlined Moon's idiom schema (1998: 161). Such schemas were found to have links with prototype and conceptual metaphor theories. Schemas, or basic networks, were found in the research here in all of the cases. Lexical substitution was found throughout, and so schemas could be set up in this way for each formulaic sequence.

The more substitutions a sequence is open to, the larger the network. Thus the size of the network demonstrates where a sequence lies on the collocation continuum; a sequence such as *kick the bucket* would have a linear network:

Kick → bucket

The sequence *pin your HOPES/faith on* is constructed in a restricted network. Figure 6-3 showed that this sequence is closer to the fixed end of the spectrum lexically than the larger network *take a BEATING/hammering.* The end, or the limits, of a basic network is reached when no new substitutes are found using the chaining process. This was the case, for example, with *pin your HOPES/faith on* and *flip your LID/wig.*

The basic networks developed the work on idiom schemas in the fact that for some of the sequences investigated in the case studies, there was a development of meaning from the canonical form. One example was the sequence *throw a WOBBLY/wobbler,* which began with a meaning of "suddenly become very angry, often about something that is not very important" and "chained" to the form *have a fit,* which has more emphasis on anger rather than on the frustration inherent in the canonical form. The lexical combinations are combined in a network with the underlying concept of SUDDENLY BECOME VERY ANGRY, and a common lexical set. Such a network is called a chaining network, building on Howarth's ideas of overlapping clusters.

The end of the network for *throw a WOBBLY/wobbler* is reached through a change in common central concept. The concept of SUDDENLY BECOME VERY ANGRY changes after *have* + *det* + *sulk* as the meaning changes from a short term mood over something trivial to a longer term emotion. Likewise the end of the network for *shut your FACE/trap/gob/cakehole* and *watch your TONGUE/mouth* is reached when the core concept changes from restraint and not talking or saying anything inappropriate to having a meaning of to behave appropriately. In many of the cases investigated, the canonical form had few token occurrences. There were more occurrences of a literal "blanched meaning" form, which then led back to a less frequent non-literal form (for example, *overstep the mark → cross the line → step over the line*). This supports Moon's work on schema.

Networks can thus group together formulaic sequences, for example the sequence *flip your LID/wig* formed a network systematically grouping together this sequence with other LID entries such as *blow your TOP/stack.* Such networks allow the boundaries between one formulaic sequence and another to be seen, thus illustrating the research aim. In section 5.7 there are three sequences: *flip your lid/wig, lose your temper*, and *blow your top/stack.* The "connecting" points in the network are one-off occurrences *lose + det + wig* and *blow + det + temper.* These peripheral combinations should not be disregarded as random occurrences (although, of course, they could be "incorrect" forms produced by an inexperienced or non-native speaker). They demonstrate that the speaker has a concept in mind with an associated lexical set and syntactic structure. They can select the required lexical items as needed. Formulaic sequences are thus not stored in a list with a specific form and meaning. They have a central concept, and common lexical set. The more frequent combinations establish the concept, and are selected as the dictionary entry. Peripheral combinations demonstrate a speaker's innovative use of language, and they also represent the outer limits of the concepts. The more substitutes a sequence has, the more "fuzzy" the boundaries may be. Places where a combination does not occur, for example *blow + det + composure* are not areas where the combination is unacceptable, rather it is that within the corpus there has been no occurrence of a speaker using that particular combination.

A third type of network was used to develop basic and chaining networks. I called this a constellation network. This takes the frequencies of substitutes discussed in section 7.2.1, and incorporates them with the networks. This combination of prototypes and networks creates a less-linear representation of sequences and the links between nouns and verbs, rather than the flat frames or schema. Using this technique reveals that substitutes found do not need to be directly related to a canonical node (noun or verb in this case), but can be related to nodes which are related to the canonical nodes.

The "path" of the network is the same here as in previous research, although it is used here to its limits; until no new substitutes are found, and including those found with only one or few token occurrences. The limits are also when the central concept changes. Within the networks there is a general central concept, but the different lexical combinations have

subtly different surface meanings. This supports the notion of a network surrounding a concept, rather than an atomistic list of idioms.

The networks are based on previous work by Howarth (1996), Williams (2002), Cantos-Gómez (2001), and build on the idiom schema proposed by Moon (1998a) in providing a systematic method of collecting the network nodes. The collocations used in the networks here are sequences of a V det N structure, as opposed to the two-word collocates used by Williams, Cantos-Gómez, and Järvi *et al.* (2004). The networks here are taken from the BNC, which is more broadly representative of British English than the academic writing corpus used by Howarth, the biology field used by Williams, or the marketing communication field of Nokia (Järvi *et al.*, 2004).

7.2.3. Research Question 3: What/Where are the Boundaries between the Literal and Non-Literal Interpretations of a Formulaic Sequence?

The meaning of the noun and or the verb is critical to the distinction of the literal or non-literal interpretation of the sequence. Factors which influence the reading of the lexical components are the determiner and insertions in the sequence.

7.2.3.1. The Status of the Verb

Whether the verb was multi-functional and of high general frequency, such as *take, make,* and *get* or lexical and less frequent, such as *throw, pin,* or *raise,* had an effect on the sequence. If the verb is functional, as is the case of *take* in the sequence *take a BEATING/hammering,* it carries less lexical content and acts grammatically, fulfilling the syntactic structure. As it carries little meaning, the noun is important to the literality or non-literality of the sequence. Such phrases can be decomposable, as the meaning is in the noun component only, and thus the verb can move, or be substituted to a higher degree than a "lexical" verb, and insertions can occur without affecting the meaning of the sequence, for example *make + det + fuss* for the sequence *kick up a STINK/fuss.*

7.2.3.2. The Determiner

The determiner of a phrase is often taken to have little effect on the variation of a sequence, and only the lexical components are seen as important. However, this is not the case. Whether the determiner is indefinite or definite has a bearing on the variability potential of the sequence. Nicolas (1995: 249) showed that phrases with the definite determiner are open to much fewer variations than those with the indefinite or null determiners. This is evident in sequences such as *kick the bucket.* Of the sequences in the analysis chapters, only the sequence *overstep the MARK/limits/bounds* had a definite determiner. In the 35 sequences of the dataset, the majority (12 sequences) had a pronoun determiner, whilst eight had an indefinite determiner, 11 had a definite determiner, three had a null determiner and one had the demonstrative *that.* Fellbaum (1993: 287) says that there are few VP idioms within dictionaries that have an indefinite article in the dictionary form. The listener is likely to interpret an indefinite article literally. The definite article draws the hearer's attention to the non-literal reading of the phrase by default; none of the standard readings being applicable.

Fellbaum (1993: 273) notes that of the total possible range of determiners available in literal language, only a small subset can be found in formulaic sequences. There are few canonical sequences with numerical determiners, negatives, or quantifiers such as *every,* or *many.* This is reflected in the dataset. Previous work by Nicolas (1995) and Fellbaum (1993) focuses in particular on indefinite and definite determiners. In the dataset here, context-dependent pronoun determiners are more prevalent, which causes a high degree of possible determiner substitution. Ten phrases in the dataset were seen to show the greatest number of determiner substitutions. Most of the possibilities found were the null determiner, the articles, demonstratives (*this, that, these*) and possessive personal pronouns. Possible quantifier alternatives included *some, any, little,* and *much,* although none of these were found in the canonical form, and these were rare compared to the other determiners found. Fellbaum suggests that the "flexibility of the determiner is found only in compositional idioms, where it indicates the nouns referentiality. Noncompositional, unanalyzable idioms have no determiner variation" (1993: 272). From the dataset, sequences such as *miss the BOAT/bus, shoot the BREEZE/bull,* and *hit the SACK/hay* have a definite article, and are open to only two noun substitutes, and no insertions. They have

no determiner substitutes. Compare this to a sequence such as *kick up a STINK/fuss,* which has six possible noun substitutes, and had 166 tokens demonstrating insertions. The exception found was the sequence *do BIRD/time,* which had 12 different possible determiner substitutes, yet only had two possible noun substitutes and no occurrences with insertion.

Fellbaum argues that the determiner of a sequence is not arbitrary (1993: 272). The determiner is an important factor in the study of formulaic sequences, and is as significant as the lexical components in terms of variation and analysis.

7.2.3.3. Insertions

Insertions within a formulaic sequence can help to answer the question regarding the boundaries between the literal reading of a phrase and the non-literal reading. The sequence *overstep the Mark/limits/bounds* had 99 tokens showing an insertion, and all of these involved lexical insertion. Of these, only four had the LID meaning of "to go too far and offend someone", the remaining 95 had a sense of crossing a literal (or metaphorical) line. There were 46 tokens with an insertion which took the lexical combinatory form *cross + det + line,* and 44 the form *cross + det + boundary.* Of these, 40 of the examples of *cross + det + line* had a literal reading. 18 of the *cross + det + boundary* tokens referred to crossing an actual demarcation (literal) and the remainders referred to a metaphorical boundary. Section 6.5.2 demonstrated that where the verb was *cross,* the reading of the sequence was literal (even if the actual boundary was metaphorical) whereas when the verb was *overstep* the meaning was that given by the LID. There was also a difference in meaning between *cross* "to go from one side to another" and *overstep* "to go beyond". In terms of insertions, there were only three examples of *overstep + det + mark*; most of the insertions occurred with the verb *cross* and were either objective adjectives, or noun modifiers such as *departmental, finishing, railway,* or *tram.* These are specificatory insertions and modify the noun of the sequence internally; they are "concrete" modifiers giving extra information to the sequence. In the case of *take a BEATING/hammering,* objective insertions aided a more literal reading, whereas subjective, external insertions lend themselves to a non-literal reading.

The differentiation between objective and subjective insertions could also be seen for the sequence *raise HELL/Cain,* where forms of the sequence with a literal reading, for example *wreak/cause havoc/chaos*, were more open to insertions. The literal combinations showed a great tendency for noun modifiers such as *traffic* or *motorway,* and adjectives susceptible to objective measure, e.g. *political.* The sequence *pin your HOPES/faith on* was lexically and semantically the most fixed of the cases investigated. This had a high degree of grammatical insertions, which acted more externally (i.e. the degree of the verb, rather than specifying the noun) than lexical insertions, and the lexical insertions present were of an emphasizing type. It seems that where the insertion is lexical and objective, or a noun, the action of the insertion is internal and operates on the noun of the formulaic sequence. This lends itself to a literal interpretation. With grammatical insertions and lexical emphatic insertions, these act more externally over the whole sequence and lend themselves to a non-literal interpretation.

The insertion within a sequence acts in a way which determines, or aids a sequence's literal or non-literal interpretation, thus identifying these boundaries for a formulaic sequence. However, the boundaries between a literal sequence and its non-literal counterpart are not clear cut. There is some overlap between the form and meaning, for example there were occurrences of *cross + det + line* with the meaning of "to go too far and upset someone" as well as "to cross an actual line or demarcation". The differences between the literal and non-literal readings for most of the case studies here is not so clear cut as that between traditional "pure" sequences such as *kick the bucket* or *spill the beans.* The literal combinations are more frequent than the non-literal counterparts. Where there is a network with a central concept and lexical set, the numerous substitutions and frequencies cause there to be central and peripheral combinations, and different permutations of lexical substitutes create the overlap between literal and non-literal reading.

7.3. Broader Implications

The findings of the dissertation have implications for other areas of linguistic research. This section discusses how research of the boundaries of variation of formulaic sequences may link to areas of lexicography, cognitive linguistics, language use and metaphor theories.

7.3.1. Lexicography

Section 3.2.1.2 detailed the LID and discussed its idiom activator; the ten concept words and related formulaic sequences. The members within each concept are linked by underlying meaning only; they are not necessarily lexically or syntactically related. Section 3.2.2 showed how the CIDI and CCDI also have such grouping systems; theme panels and thematic index respectively. Again the formulaic sequences are grouped in panels by meaning only. The networks found in this research provide a systematic method of grouping sequences which are of the same meaning, and share the same lexical set and syntactic structure. The frequencies recorded for variant forms of a sequence demonstrate the boundaries of the networking. Highly frequent combinations support the central concept of the network. Less frequent combinations are on the boundaries.

Recording the frequency of variant forms also has implications for lexicography. The LID displays variant forms either on the same level form substitutes which are equally frequent, or the more frequent form is listed above as the less frequent one. Recording the actual frequency shows which exact form is more frequent and which is rare or peripheral. Section 5.3.1 showed that the noun *battering* actually occurred more frequently than the LID-given *hammering* or *beating.* The recorded frequencies show that there are more insertions allowed in actual use than noted by idiom dictionaries.

7.3.2. Cognitive Theories of Mental Organization

The network theory here for phrases, and those of two-word collocates is similar to ideas of networks such as Word Grammar (Hudson, 2002) which suggests that words are stored in knowledge along with a vast set of associations, which integrates with other knowledge. This builds on work done by Aitchison (1994) on a network within the lexicon, involving relationships such as synonymy, collocation, super-ordination, and co-ordination. There are also such computational reference systems such as FrameNet and WordNet which organize words and describe the frames or underlying conceptual structures in the former and which organize English lexical items into synonym sets in the latter.[2] The phrasal networks found here provide evidence and support for these theories of grouping and organizing language.

---

[2] http://wordnet.princeton.edu/ http://framenet.icsi.berkely.edu/

Networks also provide evidence against traditional list or atomistic theories concerning the lexicon and phrases. The way that substitutes and insertions can be illustrated and the way in which chaining and constellation networks show sequences develop and link to each other suggests that formulaic sequences between the ends of the collocational continuum are not stored in a separate mental list. Formulaic sequences seem to have a central concept, syntactic structure and associated lexical set from which speakers select or create a combination.

7.3.3. Style and Use

An important factor affecting the possible amount of variation is style; individual use, and use of sequences by the media create innovative variations.

Partington suggests prefabricated language is used to maximize economy of effort, but using variations for stylistic purposes makes communication more effective, especially in headlines, as it engages people's attention (1996: 139). Moon also suggests that exploitation of formulaic sequences is most often associated with journalism (1998a: 170), an example being "Partners *held over a barrel*" (a story about two pub lessees and the quadrupling of their rent) (1998a: 290). As the wealth of media types and styles increases, there will be different ways of using language, and so formulaic sequences will be open to more manipulation. This will thus expand the networks. Phraseological allusion, the process of reference to a formulaic sequence by imagery (Nacisione, 2003: 28) allows for much play on sequences, an example with allusion to the canonical *chance your ARM/luck* being:

> *Example 7-1*
>
> If you fancy *trying your needle* at tapestry but feel daunted by mammoth prices and Bayeaux complexity, Myriad tapestry kits may be the solution.
> *Country Living*
> *A7D 188*

The allusion shown in Example 7-1 supports the notion that speakers recognize a central concept and can adapt the lexical substitutes according to need.

As well as manipulation by the media, there are other stylistical variations. There may be personal exploitations, related to experience and memory. Many of these personal sequences may be shared with others. Wray says:

> Specifically it is proposed that formulaic language is more than a static corpus of words and phrases which we have to learn in order to be fully linguistically competent. Rather, it is a dynamic response to the demands of language use, and as such, will manifest differently as these demands vary from moment to moment and speaker to speaker
> (2002: 5)

This is supported by Naciscione, who says that formulaic sequences are not often found in the canonical form; they change according to the need of the discourse (2003: 29).

The stylistics of formulaic sequence use (who uses it, in which context, and its purpose in use) is important in terms of formulaic sequence variation and analysis. As Naciscione says, "discourse offers innumerable instances and a great variety of stylistic use, which call for enhanced understanding and appreciation of figurative language" (2003: 29). It is impossible to determine the exact boundaries of formulaic sequences, as there will always be possibilities for change due to style and media. As Gläser says (1998: 143) "The stylistic potential of the phrasicon is unchallengable". The use and stylistics of formulaic sequences is an area worthy of further investigation.

7.3.4. Metaphor

A common view surrounding formulaic sequences is that they are dead metaphors (Gibbs, 1992). These are sequences where the original reason behind the meaning has been lost, and so the sequence becomes frozen and fixed with no motivation. However, as Naciscione says "Phraseology is not simply a list of dictionary entries; it comes alive in use" (2003: 25). The phrasal networks developed in Chapters 5 and 6 provide evidence against the dead metaphor view. Networks have a central concept and associated lexical set. In relation to idiom schema, there is an underlying conceptual metaphor which allows for changes in the surface form whilst alluding to the canonical formulaic sequence. This means that variation can occur and change to fit the particular circumstance, and thus supports the notion that formulaic sequences are not "frozen" but alive in use.

7.4. Delimitations and Further Study

Regarding the tools used, there is more written data in the BNC than spoken data. The BNC is a decade out of date. Due to language change over that time, studying such phrases in a more recent corpus such as the monitor corpora Bank of English or CIC may yield different results. As an expansion of the project, it would be interesting to cross-check the findings with the other dictionaries and corpora. Another source of language data which yields a large word frequency, and more genres (although written, not spoken ones) is the World Wide Web. New research such as that done by Kilgarriff (2001) aims to use the web as a corpus, and theoretically could provide a valuable resource for the study of idioms.

The chaining process used builds on previous research methodologies in providing a systematic and robust means of collecting lexical substitutes and insertions, and emphasizing the boundaries of variations, and boundaries between sequences. However, there are difficulties when a network leads to a node which is a pronoun, or a very common verb, for example *it* or *have*. The BNC software limits the amount of results viewable to 2000, so in cases where the word frequency of occurrence in the corpora is in the tens of thousands, it is not possible to view all of the occurrences, and thus there is a possibility of breaks in the network. The BNC, for such cases, gives options such as viewing the first 2000 results, or taking a random sample from all of the results. It is this latter approach which is the action that had to be taken here.

A second issue regarding the chaining process is that formulaic sequences are open to different interpretations depending on the reader's knowledge and cultural background. Intuition is needed to determine whether a combination should be included in the dataset or not. Although through the application of chaining process, and by including literal and non-literal forms of a sequence, I aim to be objective, there is a possibility that sequences described here would not receive the same analysis by other researchers.

I consider the application of the chaining process to be comprehensive for the purposes of this research. However, only I have used the process. The procedure needs to be tested for reliability in future research. The same investigation needs to be carried out by other

researchers using this technique in order to investigate any possible differences in data-collection between the results presented here, and those collected by others.

The variables and delimitations within this research project could also be altered to provide further analysis of formulaic sequences. The dataset was chosen according to the sequences having only one type of variation; lexical substitution of the noun. It was also decided to concentrate on lexical substitution and insertions. Removing these delimitations means that different datasets could be selected based on different criteria. It is also possible to investigate other forms of variation, such as grammatical variation or register. As far as the research questions here are concerned, one such dataset to be explored would contain formulaic sequences which demonstrate context-dependent variation. Such an exploration would add to conclusions regarding the limits of substitution.

Another delimitation imposed was restricting the data to the V det N structure. Removing this limitation means that more data about each sequence can be collected, and the relevant networks can expand. Removing the syntactical factors allows for syntactic variations to be collected in the data preparation. The data investigated was examined for its lexical components. As section 7.2.3.2 showed, the determiner is an essential part of the sequence, and thus for further investigation, this component requires equal examination as the noun and the verb.

As can be seen from Chapters 5 (lexical substitutions) and 6 (insertions), all of the phrases act differently to each other. The possible variables have been reduced as far as possible in the selection of the dataset and case studies, and in the methodology, but it is impossible to analyse each formulaic sequence in the same way as the next. Differing factors in terms of compositionality, verb, number of insertions, networks formed, motivation etc. mean that an idea for developing this work would be to concentrate fully on one formulaic sequence.

Finally, for full analysis, a holistic approach is needed; phraseology has obvious links with a range of fields such as pragmatics, sociolinguistics, semantics etc. which it has not been possible within the limits of this project to explore fully.

7.5. Summary

This work is of interest in that it develops work on the untidy patterns thrown up by word association. The technique and corpus used provide a logical process of selecting cases for investigation, reducing variables and preparing data for analysis within a large representative data source. Investigating the boundaries of variation, literality, and those between separate phrases demonstrates that formulaic sequences are not logical, not peripheral and their continued study is of relevance throughout the networks of different linguistic areas.

Appendix

Variation Category	Code	Variation 1 (Variation 2 and 3 Empty)
AdjP Adjective	AdjP2	11
AdjP Degree Modifier	AdjP1	1
AdvP Adverb	AdvP2	7
AdvP Degree Modifier	AdvP1	0
Any Adjective	*Adj	11
Any Adverb	*Adv	0
Any Noun	*N	71
Any Preposition	*Prep	0
Any Verb	*V	0
Conjunction Coordinator	Co1	1
Conjunction Subordinator	Co2	6
Context Dependent	C.D.	646
Deletion Lexical	Del2	25
Deletion Phrasal	Del1	37
Negation (Negative Particle vs lexical item)	Neg	11
NP Determiner	NP1	14
NP Modifier Adjective	NP2	31
NP Modifier Noun	NP3	5
NP Noun	**NP4**	**77**
Opposite Arrangement	Opp2	4
Opposite Meaning Lexical	Opp1	1
PP Noun Phrase Determiner	PP2	5
PP Noun Phrase Modifier Adjective	PP3	3
PP Noun Phrase Modifier Noun	PP4	0
PP Noun Phrase Noun	**PP5**	**57**
PP Preposition	PP1	40
Syntactical Variation	Str	63
VP Auxiliary	VP1	15
VP Main verb	VP2	305
VP Phrasal Verb Particle	VP3	3

Table 1 - The variation category frequencies of variation column 1 when the variation 2 and 3 fields are empty (i.e. frequencies of sequences with one variation category only)

Phrase	Lexical Variants (V)	Lexical Variants (N)	Total Lexemes Involved	Total Longman Noun Variants	Total Lexemes Involved Longman
Chance your ARM/ luck	4	4	8	2	3
Cover your ASS/ butt	4	4	8	2	3
Kiss my ASS/ butt	1	3	4	2	3
Be left holding the BABY/ bag	1	2	3	2	3
Take a BEATING/ hammering	16	8	24	2	3
Do BIRD/ time	3	2	5	2	3
Miss the BOAT/ bus	1	2	3	2	3
Have a BRAINWAVE/ brainstorm	8	2	10	2	3
Be a BREEZE/ doddle	5	5	10	2	3
Shoot the BREEZE/ bull	1	2	3	2	3
Drop a BRICK/ clanger	1	4	5	2	3
Burn your BRIDGES/ boats	1	3	4	2	3
Cut the CRAP/ shit	2	3	3	2	3
Don't pull that CRAP/ shit	0	0	0	2	3
Hit the DECK/ dirt	3	6	9	2	3
Shut your FACE/ trap/ cakehole/ gob	4	12	16	4	5
Tempt FATE/ providence	4	3	7	2	3
Drag your FEET/ heels	1	3	4	2	3
Raise HELL/ Cain	9	5	14	2	3
Pin your HOPES/ faith on	4	3	7	2	3
Blow your own HORN/ trumpet	1	3	4	2	3
Flip your LID/ wig	3	11	14	2	3
Overstep the MARK/ limits/ bounds	6	6	12	3	4
Be a sore POINT/ spot	9	4	13	2	3
Pack a PUNCH/ wallop	3	2	5	2	3
Hit the SACK/ hay	2	2	4	2	3
Keep your SHIRT/ pants/ hair on	1	1	2	3	4
Call the SHOTS/ tune	1	3	4	2	3
Set the STAGE/scene for	2	4	6	2	3
Kick up a STINK/ fuss	8	6	14	2	3
Be just the TICKET/ thing/ job	8	5	13	3	3
Stem the TIDE/ flow/ swell of	10	6	16	3	4
Watch your TONGUE/ mouth	8	8	16	2	3
Throw a WOBBLY/ wobbler	7	8	15	2	3
Not mince your WORDS/ matters	3	1	4	2	3

Table 2 - Total number of noun, verb and total lexemes involved in the BNC using the chaining process, compared to the number of noun substitutes and total lexical substitutions shown by the LID

Formulaic Sequence	Noun Substitutes	Formulaic Sequence	Insertions
Shut your FACE/ trap/ cakehole/ gob	12	Kick up a STINK/ fuss	166
Flip your LID/ wig	11	Overstep the MARK/limits/ bounds	99
Throw a WOBBLER/ wobbly	8	Raise HELL/ Cain	94
Take a BEATING/ hammering	8	Pin your HOPES/ faith on	67
Watch your TONGUE/ mouth	8	Take a BEATING/ hammering	51
Stem the TIDE/ flow/ swell of	6	Throw a WOBBLY/ wobbler	23
Kick up a STINK/ fuss	6	Stem the TIDE/ flow/ swell of	20
Overstep the MARK/ limits/ bounds	6	Pack a PUNCH/ wallop	20
Hit the DECK/ dirt	6	Set the STAGE/ scene for	18
Be a BREEZE/ doddle	5	Be a sore POINT/ spot	17
Raise HELL/ Cain	5	Shut your FACE/ trap/ cakehole/ gob	13
Be just the TICKET/ thing/ job	5	Be a BREEZE/ doddle	12
Set the STAGE/ scene for	4	Have a BRAINWAVE/ brainstorm	6
Be a sore POINT/ spot	4	Drop a BRICK/ clanger	6
Drop a BRICK/ clanger	4	Be just the TICKET/ thing/ job	6
Chance your ARM/ luck	4	Flip your LID/ wig	5
Cover your ASS/ butt	4	Call the SHOTS/ tune	4
Kiss my ASS/ butt	3	Cover your ASS/ butt	2
Burn your BRIDGES/ boats	3	Be left holding the BABY/ bag	2
Cut the CRAP/ shit	3	Cut the CRAP/ shit	2
Tempt FATE/ Providence	3	Chance your ARM/ luck	1
Drag your FEET/ heels	3	Kiss my ASS/ butt	1
Pin your FAITH/ hopes on	3	Hit the DECK/ dirt	1
Blow your own HORN/ trumpet	3	Tempt FATE/ providence	1
Call the SHOTS/ tune	3	Drag your FEET/ heels	1
Hit the SACK/ hay	2	Shoot the BREEZE/ bull	0
Pack a PUNCH/ wallop	2	Do BIRD/ time	0
Shoot the BREEZE/ bull	2	Miss the BOAT/ bus	0
Have a BRAINWAVE/ brainstorm	2	Burn your BRIDGES/ boats	0
Miss the BOAT/ bus	2	Don't pull that CRAP/ shit	0
Do BIRD/ time	2	Blow your own HORN/ trumpet	0
Be left holding the BABY/ bag	2	Hit the SACK/ hay	0
Keep your SHIRT/ pants/ hair on	1	Keep your SHIRT/ pants/ hair on	0
Not mince your WORDS/ matters	1	Watch your TONGUE/ mouth	0
Don't pull that CRAP/ shit	0	Not mince your WORDS/ matters	0

Table 3 - Frequency order of the phrases in the dataset, and the total number of noun substitutes involved in the lexical set for each phrase. Also showing the frequency of the phrases and the total number of tokens allowing lexical insertions.

References

Primary Sources

Cambridge International Dictionary of Idioms (Cambridge University Press) 1998

Collins Cobuild Dictionary of Idioms (HarperCollins Publishers) 2002

Oxford Dictionary of Current Idiomatic English Volume One: Verbs with Prepositions and Particles ed. by Cowie, A. P. and Mackin, R. (Oxford: Oxford University Press) 1985a

Oxford Dictionary of Current Idiomatic English Volume Two: Phrase, Clause and Sentence Idioms ed. by Cowie, A. P., Mackin, R. and McCraig, I. R (Oxford: Oxford University Press) 1985b

Dictionary of Idioms and their Origins Flavell, L. and Flavell, R. (Kyle Cathie Ltd.) 2000

Longman Idioms Dictionary (Longman) 1998

The Concise Oxford Dictionary (10th ed.) Pearsall, J. (ed.) (Oxford: Oxford University Press) 1999

http://www.titania.bham.ac.uk
The home of the Bank of English

http://www.longman.com/dictionaries/corpus/index.html
Information about the Longman Corpus Network

http://khnt.hit.uib.no/icame/manuals/flob/INDEX.HTM
Manual of Information to Accompany the Freiburg - LOB Corpus of British English (FLOB)

http://www.cis.upenn.edu/~treebank/home.html
The Penn Treebank Project

http://www.kilgarriff.co.uk/bnc-readme.html
BNC Database and Word Frequency Lists

http://dictionary.oed.com
The Oxford English Dictionary

http://www.dictionary.com
Multi-source dictionary search service produced by Lexico Publishing Group (LLC)

Secondary Sources

Abeillé, A. 1995. 'The Flexibility of French Idioms: A Representation with Lexicalized Tree Adjoining Grammar' in: Everaert, M. *et al.* 1995a, 15 - 42

Aisenstadt, E. 1981 'Restricted collocations in English lexicology and lexicography' in: *ITL (Instituut voor Toegepaste Linguistiek) Review of Applied Linguistics*, 53:53-61. Cited in: Howarth, P. A. 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making* (Max Niemeyer Verlag)

Aitchison, J. 1994. *Words in the Mind: An Introduction to the Mental Lexicon* (Blackwell)

Aitchison, J. 1997. *The Language Web* (Cambridge University Press)

Akimoto, M. 1983. *Idiomaticity* (Aoyama Gakuin University: Shinozaki Shorin)

Altenberg, B and Eeg-Olofsson, M. 1990. 'Phraseology in Spoken English: Presentation of a Project' in: In *Theory and Practice in Corpus Linguistics*, ed. by Aarts, J and W Meijs (Rodopi), 1 - 26

Altenberg, B. 1992. 'Recurrent Verb - Complement Constructions in the London - Lund Corpus' in: *English Language Corpora: Design, Analysis and Exploitation. Papers from the 13th International Conference on English Language Research on Computerised Corpora,* ed. by Aarts, J., Haan, P. d. and Oostdijk, N. (Rodopi), 227 - 245

Aston, G. and Burnard, L. 1998. *The BNC Handbook: Exploring the British National Corpus with SARA* (Edinburgh University Press)

Barkema, H. 1993. 'Idiomaticity in English NPs' in: *English Language Corpora: Design, Analysis and Exploration. Papers from the 13th International Conference on English Language Research on Computerised Corpora,* ed. by Aarts, J., Haan, P. d. and Oostdijk, N. (Rodopi) 257 – 279

Barkema, H. 1994a. 'The Idiomatic, Syntactic, and Collocational Characteristics of Received NPs: Some Basic Statistics' *HERMES Journal of Linguistics,* 13: 19 - 41

Barkema, H. 1994b. 'Determining the Syntactic Flexibility of Idioms' in: *Creating and Using English Language Corpora: Papers from the fourteenth International Conference on English Language Research on Computerized Corpora,* ed. by Fries, U., Tottie, G. and Schneider, P. (Rodopi), 39 - 53

Barkema, H. 1996a. 'The Effect of Inherent and Contextual Factors on Grammatical Flexibility', in *Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora (ICAME 16),* ed. by: Percy, C., Meyer, C. F. and Lancashire, I. (Rodopi) 69 - 83

Barkema, H. 1996b. 'Idiomaticity and Terminology: A Multi - Dimensional Descriptive Model' *Studia Linguistica,* 50: 125 - 160

Barkema, H. 1996c. 'The Grammatical Freedom of Lexicalised Noun Phrases' *Corpus Based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME),* (Rodopi) 17: 49 - 68

Benson, M. 1985. 'Collocations and Idioms' in: *Dictionaries, Lexicography and Language Learning,* ed. by Illson, R. (Pergamon Press), 61 - 68

Brill, E. 1995. 'Transformation-based Error-driven Learning and Natural Language Processing: A Case Study in part of Speech Tagging' *Computational Linguistics,* 21: 543 - 565

Burton-Roberts, N. 1997. *Analysing Sentences: An Introduction to English Syntax* (Longman)

Cacciari, C. and Tabossi, P. (eds.) 1993. *Idioms: Processing, Structure and Interpretation* (Lawrence Erlbaum Associates, Publishers)

Cacciari, C. and Glucksberg, S. 1995. 'Imagining Idiomatic Expressions: Literal or Figurative Meanings?' in: Everaert, M. *et al.* 1995a, 43 - 57

Cameron, L. and Low, G. 1999. 'Metaphor' *Journal of Language Teaching,* 32: 77 - 96

Cantos-Gómez, P. 2001. 'An Attempt to Improve Current Collocation Analysis' *Proceedings of Corpus Linguistics, 2001,* UCREL 13: 100 - 108

Carter, R. 1987. *Vocabulary*, London, Allen & Unwin. Cited in: Partington, A. 1996. *Patterns and Meanings: Using Corpora for English Language Research* (John Benjamins Publishing Company)

Čermák, F. 1994. 'Idiomatics' in: *Prague School of Structural and Functional Linguistics,* ed. by Luelsdorff (Amsterdam: John Benjamins), 185 - 95 Cited in: Moon, R. 1998a. *Fixed Expressions and Idioms in English* (Clarendon Press)

Čermák, F. 2001. 'Substance of Idioms: Perennial Problems, Lack of Data or Theory?' *International Journal of Lexicography,* 14: 1 - 20

Church, K. W., Gale, W., Hanks, P., Hindle, D. and Moon, R. 1994. 'Lexical Substitutability' in: *Computational Approaches to the Lexicon,* ed. by Atkins and Zampolli (Clarendon Press), 153 - 177

Cowie, A. P. (ed.) 1998a. *Phraseology: Theory, Analysis and Applications* (Oxford University Press)

Cowie, A. P. 1998b. 'Introduction' in: Cowie, A. P. 1998a, 1 - 23

Cowie, A. P. 1998c. 'Phraseological Dictionaries: Some East - West Comparisons' in: Cowie, A. P. 1998a, 209 - 229

Crystal, D. 1998. *Rediscover Grammar: Revised Edition* (Longman)

Cserép, A. 2002. 'The Analyzability of English Idioms', in *The European Society for the Study of English, 6th Conference - Strasbourg,* ed. by: Hamm, A. (Ranam: Recherches Anglaises et Nord Améicaines) 67 - 78

Cutler, A. 1982. 'Idioms: The Colder the Older' *Linguistic Inquiry,* 13: 317 - 320

d'Arcais, G. F. 1993. 'The Comprehension and Semantic Interpretation of Idioms' in: Cacciari, C. and Tabossi, P. 1993, 79 - 98

Degand, L. and Bestgen, Y. 2003. 'Towards Automatic Retrieval of Idioms in French Newspaper Corpora' *Literary and Linguistic Computing,* 18: 249 - 259

Dixon, R. M. W. 1991. *A New Approach to English Grammar, on Semantic Principles* (Clarendon Press)

Erman, B. and Warren, B. 2000. 'The Idiom Principle and the Open Choice Principle' *Text,* 1: 29 - 62

Ernst, T. 1981. 'Grist for the Linguistic Mill: Idioms and 'Extra' Adjectives' *Journal of Linguistic Research,* 1: 51 - 68

Everaert, M., Van der Linden, E., Schenk, A. and Schreuder, R. (eds.) 1995a. *Idioms: Structural and Psychological Perspectives* (Lawrence Erlbaum Associates, Publishers)

Everaert, M., Van der Linden, E., Schenk, A. and Schreuder, R. 1995b. 'Introduction' in: Everaert, M. *et al.* 1995a, 1 - 13

Fellbaum, C. 1993. 'The Determiner in English Idioms' in: Cacciari, C. and Tabossi, P. 1993, 271 - 297

Fernando, C. 1996. *Idioms and Idiomaticity* (Oxford University Press)

Ferrer i Cancho, R. and Solé, R. V. 2001. 'The Small-World of Human Language' *Proceedings of the Royal Society of London B,* 268: 2261 - 266

Fillmore, C., Kay, P. and O'Connor, C. 1994. 'Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*' *Language,* 64: 501 - 538

Fontenelle, T. 1998. 'Discovering Significant Lexical Functions in Dictionary Entries' in: Cowie, A. P. 1998a, 189 - 209

Fraser, B. 1970. 'Idioms within a Transformational Grammar' *Foundations of Language,* 6: 22 - 42

Gibbs, R. 1992. 'What do Idioms Really Mean?' *Journal of Memory and Language,* 31: 485 - 506

Gibbs, R. 1993. 'Why Idioms are not Dead Metaphors' in: Cacciari, C. and Tabossi, P. 1993, 57 - 79

Gibbs, R. 1994. *The Poetics of Mind: Figurative Thought, Language and Understanding* (Cambridge University Press)

Gibbs, R. 1995. 'Idiomaticity and Human Cognition' in: Everaert, M. *et al.* 1995a, 97 - 117

Gibbs, R. and O'Brien, J. E. 1990 'Idioms and Mental Imagery: The Metaphorical Motivation for Idiomatic Meaning' Cited in: Cacciari, C. and Glucksberg, S. 1995. 'Imagining Idiomatic Expressions: Literal or Figurative Meanings?' in: Everaert, M. *et al.* 1995a, 43 - 57

Gibbs, R. and Nayak, N. 1989a. 'Psycholinguistic Studies on the Syntactic Behaviour of Idioms' *Cognitive Psychology,* 21: 100 - 138

Gibbs, R., Nayak, N., Bolton, J. and Keppel, M. 1989b. 'Speakers' Assumptions about the Lexical Flexibility of Idioms' *Memory and Cognition,* 17: 58 - 68

Gläser, R. 1988. 'The Grading of Idiomaticity as a Presupposition for a Taxonomy of Idioms' in: *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics,* ed. by Hüllen, W. and Schulze, R. (Max Niemeyer Verlag), 264 - 279

Gläser, R. 1998. 'The Stylistic Potential of Phraseological Units in the Light of Genre Analysis' in: Cowie, A. P. 1998a, 125 - 145

Glucksberg, S. 1993. 'Idiom Meanings and Allusional Content' in: Cacciari, C. and Tabossi, P. 1993, 3 - 27

Glucksberg, S. 2001. *Understanding Figurative Language: From Metaphors to Idioms* (Oxford University Press)

Grant, L. and Bauer, L. 2004. 'Criteria for Re-defining Idioms: Are we Barking up the Wrong Tree?' *Applied Linguistics,* 25: 38 - 61

Greenbaum, S. and Quirk, R. 1990. *A Student's Grammar of the English Language* (Longman)

Hanks, P. 2004. 'The Syntagmatics of Metaphor and Idiom' *International Journal of Lexicography,* 17: 245 - 275

Hillert, D. and Swinney, D. 2001. 'The Processing of Fixed Expressions during Sentence Comprehension' in: *Conceptual and Discourse Factors in Linguistic Structure,* ed. by Cienki, A., Luka, B. J. and Smith (M. B. CSLI Publications), 107 - 121

Hogg, R. 1977. *English Quantifier Systems* (North Holland Publishing Company)

Horn, G. 2003. 'Idioms, Metaphors and Syntactic Mobility' *Journal of Linguistics,* 39: 245 - 273

Howarth, P. A. 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making* (Max Niemeyer Verlag)

Hübler, A. 1983. *Understatements and Hedges in English* (John Benjamins Publishing Company)

Hudson, R. 2002. 'Word Grammar' in: *Studies in Word Grammar,* ed. by Sugayama, K. (Kobe: Research Institute of Foreign Studies, Kobe City University of Foreign Studies), 7 - 32

Irujo, S. 1986. 'Don't Put your Leg in your Mouth: Transfer in the Acquisition of Idioms in a Second Language' *TESOL Quarterly,* 20: 287 - 304

Jackendoff, R. 1995. 'The Boundaries of the Lexicon' in: Everaert, M. *et al.* 1995a, 133 – 167

Jackendoff, R. 1997. *The Architecture of the Language Faculty: Linguistic Inquiry Monograph 28* (MIT Press). Cited in: Horn, G. 2003. 'Idioms, Metaphors and Syntactic Mobility' *Journal of Linguistics,* 39: 245 - 273

Järvi, P., Magnusson, C., and Vanharanta, H. 2004. 'The Evaluations of the Collocational Methods in the Marketing Context.' In: *Directions in Business and Administrative Sciences. Academy of Business and Administrative Sciences 2004 International Conferences. Program and Book of Abstracts. July 16-18, 2004 Tallinn, Estonia.* ed. by Korukonda, A. R. 1-7

Keysar, B. and Bly, B. 1995. 'Intuitions of the Transparency of Idioms: Can one keep a secret by Spilling the Beans?' *Journal of Memory and Language,* 34: 89 - 109

Kilgarriff, A. 2001. 'Web as Corpus' *Proceedings of Corpus Linguistics, 2001* ed. by P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja 13: 342 - 345

Kilgarriff, A. and Grefenstette, G. 2003. 'Introduction to the Special Issue on the Web as Corpus' *Computational Linguistics,* 29: 333 - 348

Kjellmer, G. 1990. 'Patterns of Collocability' in: *Theory and Practice in Corpus Linguistics,* ed. by Aarts, J. and Meijs, W. (Rodopi), 163 - 173

Kövecses, Z. and P. Szabó (1996). 'Idioms: A View from Cognitive Semantics' in: *Applied Linguistics* 17(3): 326 - 354.

Kytö, M. and Rissanen, M. 1993. ' 'By and by enters [this] my artificiall foole...who, when Jack beheld, sodainely he flew at him' Searching for Syntactic Constructions in the Helsinki Corpus' *Early English in the Computer Age: Explorations through the Helsinki Corpus* ed. by M. Rissanen, M. Kytö, M. Palander-Collin, 11: 253 - 267

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By* (The University of Chicago Press)

Leech, G. and Svartvik, J. 1994. *A Communicative Grammar of English* (Longman)

Makkai, A. 1972. *Idiom Structure in English.* Janua Linguarum Series Major 48 (Mouton). Cited in: Akimoto, M. 1983. *Idiomaticity* (Aoyama Gakuin University: Shinozaki Shorin)

Mel'čuk, I. 1998. 'Collocations and Lexical Functions' in: Cowie, A. P. 1998a, 23 - 55

Moon, R. 1998a. *Fixed Expressions and Idioms in English* (Clarendon Press)

Moon, R. 1998b. 'Frequencies and Forms of Phrasal Lexemes in English' in: Cowie, A. P. 1998a, 79 - 101

Mountain, D. 1999. 'Expressions of Anger or Frustration' (Unpublished Thesis, University of Sheffield) 1 - 37

Naciscione, A. 2001. *Phraseological Units in Discourse: Towards Applied Stylistics* (Latvian Academy of Culture)

Naciscione, A. 2002. 'Cohesion in Phraseology' *Proceedings of the Tenth EURALEX International Congress,* ed. by A. Braasch C. Povslen (Center for Sprogtennologi) 2: 533 - 539

Naciscione, A. 2003. 'Phraseological Metaphor: Dead or Alive?' *European Society for the Study of English ESSE 6 - Strasbourg 2002, 2 - Linguistics,* Recherches Anglaises et Nord - Américaines No. 36: 23 - 30

Nattinger, J. R. and DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching* (Oxford University Press)

Newmeyer, F. J. 1973. 'The Regularity of Idiom Behavior' *Lingua,* 34: 327 - 342

Nicolas, T. 1995. 'Semantics of Idiom Modification' in: Everaert, M. *et al.* 1995a, 233 - 253

Nunberg, G., Sag, I. and Wasow, T. 1994. 'Idioms' *Language,* 70: 491 - 538

Partington, A. 1996. *Patterns and Meanings: Using Corpora for English Language Research* (John Benjamins Publishing Company)

Pawley, A. 1985. 'Lexicalization' in: *Language and Linguistics: The Interdependence of Theory, Data and Application,* ed. by Tannen, D. and Alatis, J. (University of Georgetown), 98 - 120

Philip, G. 2000. 'An Idiomatic Theme and Variations' *Words in Context* ed. by Heffer and Sauntson, 221 - 233

Popiel, S. and McRae, K. 1988. 'The Figurative and Literal Senses of Idioms, or All Idioms are not used Equally' *Journal of Psycholinguistic Research,* 17: 475 - 487

Pulman, S. 1993. 'The Recognition and Interpretation of Idioms' in: Cacciari, C. and Tabossi, P. 1993, 249 - 270

Quirk, R. and Greenbaum, S. 1973. *A University Grammar of English* (Longman)

Renouf, A. and Sinclair, J. M. 1991. 'Collocational Frameworks in English' in: *English Corpus Linguistics,* ed. by Aijmer, K. and Altenberg, B. (Longman), 128 - 143

Resnik, P. and Elkiss, A. 2003. 'The Linguist's Search Engine: Getting Started Guide Technical Report: LAMP - TR - 108/CS - TR - 4541/UMIACS - TR - 2003 – 109', (Technical Report, University of Maryland, College Park)

Ruhl, C. 1975. (1975) 'Primary verbs', in: *The First LACUS Forum*, ed. by. Makkai, A. and Makkai, V. (Hornbeam Press), 436-445. Cited in: Akimoto, M. 1983. *Idiomaticity* (Aoyama Gakuin University: Shinozaki Shorin)

Rundell, M. 1995. 'The Word on the Street: New Insights on Spoken English from the British National Corpus' *English Today,* 11: 29 - 35

Saeed, J. 1997. *Semantics* (Blackwell Publishers)

Schenk, A. 1995. 'The Syntactic Behaviour of Idioms' in: Everaert, M. *et al.* 1995a, 253 - 273

Schraw, G., Trathen, W., Reynolds, R. E. and Lapan, R. T. 1988. 'Preferences for Idioms: Restrictions due to Lexicalization and Familiarity' *Journal of Psycholinguistic Research,* 17: 413 - 424

Schweigert, W. A. and D. R. Moates (1988). 'Familiar Idiom Comprehension' in: *Journal of Psycholinguistic Research* 17(4): 281 - 296.

Sinclair, J. 1991. *Corpus, Concordance, Collocation* (Oxford University Press)

Stein, G. 1991. 'The Phrasal Verb Type 'To Have a Look' in Modern English' *IRAL,* XXIX: 1 - 29

Strässler, J. 1982. *Idioms in English: A Pragmatic Analysis* (Gunter Narr Verlag)

Tabossi, P. and Zardon, F. 1995. 'The Activation of Idiomatic Meaning' in: Everaert, M. *et al.* 1995a, 273 - 283

Thomas, J. 1995. *Meaning in Interaction: An Introduction to Pragmatics* (Longman)

Ullmann, S. 1962. *Semantics: An Introduction to the Science of Meaning* (Basil Blackwell)

Van Lancker, D. 1981. 'Idiomatic vs. Literal Interpretations of Diatropically Ambiguous Sentences' *Journal of Speech and Hearing Research,* 24: 64 - 69

Vannebo, K. I. 2003. 'Ta Og Ro Deg Ned Noen Hakk: On Pseudocoordination with the Verb ta 'take' in a Grammaticalization Perspective' *Nordic Journal of Linguistics,* 26: 165 - 193

Vega-Moreno, R. E. 2003. 'Relevance Theory and the Construction of Idiom Meaning' *UCL Working Papers in Linguistics,* 15: 303 - 323

Voort, M. V. d. and Vonk, M. 1995. 'You don't Die Immediately when you Kick an Empty Bucket: A Processing View on Semantic and Syntactic Characteristics of Idioms' in: Everaert, M. *et al.* 1995a, 283 - 301

Wierzbicka, A. 1982. 'Why can you *Have a Drink* when you can't **Have an Eat*?' *Language,* 58: 753 - 799

Wierzbicka, A. 1996. *Semantics: Primes and Universals* (Oxford University Press)

Williams, G. C. 1998. 'Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles' *International Journal of Corpus Linguistics,* 3: 151 - 171

Wray, A. 2002. *Formulaic Language and the Lexicon* (Cambridge University Press)

Zgusta, L. 1967. 'Multiword Lexical Units' *Word,* 23: 578 – 587