

**IDIOMATIC CREATIVITY:
A PRAGMATIC MODEL
FOR CREATIVE IDIOMATIC USES
IN AUTHENTIC ENGLISH DISCOURSE**

THUC ANH VO, MA, BA

Thesis submitted to the University of Nottingham for the degree
of Doctor of Philosophy

MARCH 2011

ABSTRACT

This thesis is a corpus-based investigation from a pragmatic perspective of the phenomenon of idiomatic creativity. The ultimate aim of the thesis is to provide comprehensive empirical evidence of the intertwining relationships between different factors of idiomatic creativity and the effects of such relationships on the reassessment of the determinants of idiomatic creativity in context-specific communicative events.

It has often been suggested in the psycholinguistics tradition that the semantic transparency of some idioms facilitates the creative manipulations of such idioms in discourse, while the lack thereof hinders creativity in some others. In other words, the internal characteristics of idioms, including compositionality, analysability and motivation, which contribute towards transparency, are postulated to determine idiomatic creativity. While theoretical arguments and illustrative examples appear to support such a hypothesis, empirical validation of the hypothesis using large bodies of authentic data is still required for it to gain acceptance in the linguistic community. In the first half of the thesis, therefore, quantitative analyses of corpus data and statistical tests are carried out to empirically validate the psycholinguistic hypothesis concerning the semantic determinants of idiomatic creativity. It is found in the analyses and tests that there are significant trends in the data that point towards the prevalence of transparent idioms over opaque ones in terms of both the range and the frequency of their creative manifestations in authentic English discourse.

In the second half of the thesis, with a view to flesh out the otherwise decontextualised and superficial results from such quantitative analyses, qualitative analyses of individual instances of idiom variants in the data are carried out, thus elaborating on the varying degrees of importance of the underlying factors of idiomatic creativity and, notably, their interaction with one another in specific contexts. The results suggest that, together with idioms' semantic characteristics, external factors, including context, cognitive constraints and phrasological constraints, also contribute to idiomatic creativity. It is further revealed that the roles of the facilitating and constraining factors of idiomatic creativity vary according to the particular context-specific types of variant, in which context has the overarching power to overrule certain constraints as well as to disambiguate unusual creative uses. As such, the true functional profile of idiomatic creativity is argued to be best described and predicted in relation to context.

A context-based model of idiomatic creativity in authentic English discourse is therefore proposed in this thesis in the hope that it will contribute to the existing literature on idiomatic creativity and offer a better understanding of the possibilities and constraints on the phenomenon in real life discourse.

ACKNOWLEDGEMENTS

This thesis could not have been possible without the support, encouragement and inspiration from my supervisor, Professor Ronald Carter. From the beginning to the very end, he has been there, not only to guide me through the process of carrying out and completing this thesis, but also to lend me strength through difficult times. I owe it all to him.

My gratitude to Professor Svenja Adolphs for her helpful feedback at the upgrade stage, which significantly improved the quality of the thesis towards the later stage.

A special thank you to Professor Zoltan Dornyei for giving me valuable comments and suggestions on statistical tests despite his busy schedule.

During the course of carrying out this research, I am also indebted to my dear friends who have offered kind and generous help at various stages: Chris Hull, Waqas Rashid, Ian Forster, Feroud Seeparsand for their help with the transcription of spoken data, Dawn Knight for her assistance during my data collection stage with CANGODE and BNC, Phoebe Lin, Irina Dahlmann, and Phillip Durrant for helpful discussions and materials.

Many thanks to the School of English Studies for letting me use the transcriptions of the Nottingham Multi-Modal Corpus (NMMC), which adds to the spoken data of the thesis.

I would like to extend my heartfelt thanks to my dear friends and family for their support through personal crises and exhaustion, which helped me reach the destination of this rewarding but also long and difficult journey.

TABLE OF CONTENTS

<i>Abstract</i>	Page
<i>Acknowledgments</i>	<i>i</i>
<i>List of tables and figures</i>	<i>ii</i>
<i>Glossary of terms</i>	<i>vii</i>
	<i>ix</i>
Chapter 1: Introduction	1
1.1. A preliminary definition of idioms	1
1.2. Idioms and the struggle between fixedness and creativity	4
1.3. Towards a context-based model of idiomatic creativity	6
1.4. Research questions	7
1.5. Structure of the thesis	8
Chapter 2: The concepts of creativity, idiomaticity and idiomatic creativity	
2.1. Overview of multiword units	9
2.1.1. Terminology	9
2.1.2. Multiword units and the lexicon	10
2.1.3. Taxonomies of multiword units	11
2.1.4. Clines of idiomaticity	15
2.2. Defining features of idioms	17
2.2.1. Formal and semantic features	17
2.2.1.1. Compositeness	17
2.2.1.2. Institutionalisation	18
2.2.1.3. Non-compositionality	19
2.2.1.4. Fixedness	20
2.2.1.5. Condition of use	21
2.2.2. Psychological and phonological features	21
2.2.3. A working definition of idioms	22
2.3. Concept of creativity	23
2.3.1. Linguistic creativity	24
2.3.2. Idiomatic creativity	27
2.3.3. Creative idiom variants	28
2.3.4. Idiom variation strategies	29
2.3.4.1. Lexico-grammatical strategies	29
2.3.4.2. Conceptual strategies	30
2.3.5. Communicative purposes	32
2.4. Recap of a working framework for idiomatic creativity	34
2.5. Conclusions	35
Chapter 3: Theories of the determinants of idiomatic creativity	37
3.1. Orthodox view of idioms and idiomatic creativity	37
3.2. Compositional view of idioms	39
3.3. Idiomatic transparency	40
3.4. Literal well-formedness and its role in idiomatic creativity	42
3.5. Idioms' motivation and its role in idiomatic creativity	43
3.5.1. The concept of motivation	43
3.5.2. Cognitive linguistics – An overview	44
3.5.3. Image-schema theory	45
3.5.4. Conceptual metaphor theory	46
3.5.5. Conceptual metaphors in idioms	47
3.5.6. Conceptual metonymy and other sources of motivation	48
3.5.7. The role of motivation in idiomatic creativity	50
3.6. Analysability and its role in idiomatic creativity	53
3.6.1. The concept of analysability	53
3.6.2. The role of analysability in idiomatic creativity	54
3.6.3. The combined effects of the features	59

3.7.	Cognitive constraints on idiomatic creativity	61
3.7.1.	Functionality	61
3.7.2.	Recognisability	63
3.7.3.	Semantic compatibility and grammaticality	64
3.7.4.	Non-ambiguity, non-conspicuity and evocational autonomy	66
3.8.	Context and its role in idiomatic creativity	67
3.8.1.	Theories of context in the literature	67
3.8.1.1.	Context as a predetermined aspect of communication	68
3.8.1.2.	Context as hearer's selection	69
3.8.1.3.	A combined approach	70
3.8.2.	The role of context in idiomatic creativity	71
3.9.	Conclusions	72
Chapter 4:	Data description and research methodology	74
4.1.	Mixed method research	74
4.2.	The mixed methods used in the thesis	76
4.3.	Rationale of the use of corpora in the thesis	78
4.4.	The role of corpora and the human analyst in the thesis	79
4.5.	Step-by-step technique in data collection	81
4.5.1.	Data collection methods and associated difficulties	82
4.5.2.	Compilation of a small corpus of spoken and written texts	83
4.5.2.1.	Spoken texts	83
4.5.2.2.	Written texts	85
4.5.3.	Manual identification of idioms in the self-compiled corpus	85
4.5.4.	Selection of the idiom sample set via stratified sampling	87
4.5.5.	Eliciting idiom variants from corpora	88
4.5.5.1.	Search by one lemma	90
4.5.5.2.	Search by two or more lemmas	91
4.5.5.3.	Search by construction pattern	92
4.5.5.4.	Search by related lemmas	92
4.5.5.5.	Limitations of the current tools and techniques	94
4.6.	Data analysis procedures	95
4.6.1.	Quantitative analyses of the correlation between idiom transparency and idiomatic creativity	95
4.6.1.1.	Variables	95
4.6.1.2.	Statistical tests: Non-parametric tests	96
4.6.2.	Qualitative characterisation of idiomatic creativity in relation to its facilitating and constraining factors in context	98
4.6.2.1.	Extra data	98
4.6.2.2.	Generic analytical process	99
4.6.2.3.	Inter-raters	99
4.7.	Conclusions	99
Chapter 5:	Empirical evaluation of the correlation between the internal characteristics of idioms and idiomatic creativity	101
5.1.	General description of idiomatic creativity in the self-compiled corpus	101
5.2.	General frequency distribution of idiomatic creativity in the BNC, CANCODE and MICASE	101
5.3.	Literal well-formedness and idiomatic creativity	104
5.3.1.	An analytical framework of literal well-formedness	104
5.3.2.	Charts and discussions	107
5.3.2.1.	Is there a correlation between well-formedness and idiomatic creativity?	107
5.3.2.2.	Does type of anomaly among ill-formed idioms correlate with idiomatic creativity?	109
5.3.2.3.	Does structural complexity correlate with idiomatic creativity?	110
5.3.3.	Conclusions	113
5.4.	Motivation and idiomatic creativity	114
5.4.1.	An analytical framework of motivation	114

5.4.2.	Charts and discussions	116
5.4.2.1.	Does motivation correlate with idiomatic creativity?	116
5.4.2.2.	Does the degree of motivation correlate with idiomatic creativity?	117
5.4.2.3.	Do conceptual metaphors affect idiomatic creativity among motivated idioms?	119
5.4.2.4.	Does the complexity of motivation affect idiomatic creativity?	120
5.4.3.	Conclusions	122
5.5.	Analysability and idiomatic creativity	123
5.5.1.	An analytical framework of analysability	123
5.5.2.	Charts and discussions	124
5.5.2.1.	Is there a correlation between analysability and idiomatic creativity?	124
5.5.2.2.	Is there a correlation between the degree of analysability and idiomatic creativity?	125
5.5.3.	Conclusions	127
5.6.	Transparency and idiomatic creativity	128
5.7.	Conclusions	133
Chapter 6:	A context-based model of idiomatic creativity – Possibilities and constraints	135
6.1.	A context-based model of idiomatic creativity	135
6.2.	An analytical framework of idiom variants	137
6.2.1.	Context-based classification of idiom variants	137
6.2.2.	Lexico-grammatical classification of idiom variants	139
6.3.	Context-dependent idiom variants	140
6.3.1.	Context-dependent substitutions	141
6.3.1.1.	Incompatible context-dependent substitutions	142
6.3.1.2.	Ill-formed context-dependent substitutions	144
6.3.1.3.	Context-dependent substitutions in unmotivated/unanalysable idioms	144
6.3.1.4.	Recognisability constraint on context-dependent substitutions	146
6.3.1.5.	Functionality constraint on context-dependent substitutions	148
6.3.2.	Context-dependent premodifications	150
6.3.2.1.	Incompatible context-dependent premodifications	151
6.3.2.2.	Ill-formed context-dependent premodifications	152
6.3.2.3.	Context-dependent premodifications in unmotivated/unanalysable idioms	153
6.3.2.4.	Recognisability constraint on context-dependent premodifications	155
6.3.2.5.	Functionality constraint on context-dependent premodifications	155
6.3.3.	Minor type of context-dependent variants: Coordinations	156
6.3.4.	Conclusions: Context-dependent variants	158
6.4.	Context-independent variants	159
6.4.1.	Understanding the meanings of context-independent variants	160
6.4.2.	Context-independent substitutions	163
6.4.2.1.	Context-independent substitutions in transparent and opaque idioms	164
6.4.2.2.	Grammaticality and compatibility constraints on context-independent substitutions	167
6.4.2.3.	Recognisability and functionality constraints on context-independent substitutions	168
6.4.3.	Context-independent modifications	169
6.4.3.1.	Context-independent modifications in opaque idioms	169
6.4.3.2.	Grammaticality and compatibility constraints on context-independent modifications	170

6.4.3.3. Recognisability and functionality constraints on context-independent modifications	171
6.4.4. Minor types of context-independent variants: Constructional adaptations	174
6.4.4.1. Passivisation	174
6.4.4.2. Pluralisation	176
6.4.4.3. Grammatical highlighting	177
6.4.4.4. Rankshift	178
6.4.4.5. Nominalisation	179
6.4.5. Minor types of context-independent variants: Conceptual adaptations	181
6.4.5.1. Coordination	181
6.4.5.2. Allusion	182
6.4.5.3. Idiom blending	184
6.4.6. Conclusions: Context-independent variants	184
6.5. Proposal of a context-based model of idiomatic creativity	187
6.6. Conclusions	188
Chapter 7: Conclusions and outlook	190
7.1. Summary	190
7.2. Contributions of the thesis	191
7.2.1. Content insights	192
7.2.2. Methodological insights	193
7.2.2.1. Methodological strength of the thesis	193
7.2.2.2. Suggestions about corpus design	195
7.3. Limitations of the thesis and suggestions for improvements	196
7.3.1. Data assumption	197
7.3.2. Data size	197
7.3.3. Data representativeness	198
7.3.4. Data balance	198
7.3.5. Individual bias and mistakes	198
7.4. Related research possibilities	201
7.4.1. Teaching idioms and idiomatic creativity	201
7.4.2. Translating idioms	202
7.4.3. Understanding socio-cultural movements with idioms	202
7.4.4. Idiomatic functions and their lexico-grammatical, phonological and nonverbal features	203
<i>Appendices</i>	205
<i>References</i>	228

LIST OF TABLES AND FIGURES

	Page
 <u>FIGURES</u>	
Figure 2.1: Cowie's taxonomy of word combinations	13
Figure 2.1: Howarth's taxonomy of word combinations	14
Figure 4.1: WordSketch list of collocates of 'lid' in the BNC	93
Figure 4.2: Histograms of the normality of distribution of variables	97
Figure 5.1: Well-formedness and variant tokens and types	107
Figure 5.2: Well-formedness and average types and tokens	107
Figure 5.3: Motivation and average token and type	116
Figure 5.4: Average variant token and variant type by analysability	124
 <u>TABLES</u>	
Table 2.1: Types of idiom variation strategies	30
Table 4.1: Tests of Normality	97
Table 5.1: Frequency distribution of the creative types and tokens of variants in the three corpora	102
Table 5.2: Frequencies of creative variants in the three corpora	102
Table 5.3: Invariable versus creative usage of 100 idioms in three corpora	103
Table 5.4: List of idioms with creative tendencies	103
Table 5.5: Descriptive analysis of well-formedness and variant token and variant type	107
Table 5.7: Spearman correlation between well-formedness and idiomatic creativity	108
Table 5.8: Descriptive analysis of types of anomaly and idiomatic creativity	109
Table 5.9: Kruskal-Wallis test for ill-formed idioms and variant token and variant type	110
Table 5.10: Descriptive analysis of number of content words and variant token and variant type	111
Table 5.11: Kruskal-Wallis test for number of content words and variant token and variant type	111
Table 5.12: Spearman correlation between number of content words and variant token and variant type	112
Table 5.13: Descriptive analysis of number of phrases and variant token and variant type	112
Table 5.14: Kruskal-Wallis test for number of phrases and variant token and variant type	113
Table 5.15: Spearman correlation between number of phrases and variant token and variant type	113
Table 5.16: Descriptive analysis of motivation and variant token and variant type	116
Table 5.17: Mann-Whitney test for motivation and variant token and variant type	117
Table 5.18: Spearman correlation between motivation and variant token and variant type	117
Table 5.19: Descriptive analysis of degrees of motivation and variant token and variant type	118
Table 5.20: Spearman correlation between the degree of motivation and variant token and variant type	119
Table 5.21: Descriptive analysis of metaphorical motivation and variant token and variant type	120
Table 5.22: Mann-Whitney test for metaphorical motivation and variant token and variant token	120
Table 5.23: Descriptive analysis of complexity of motivation and variant token and variant type	121

Table 5.24:	Mann-Whitney test for the complexity of motivation and variant type and variant token	121
Table 5.25:	Spearman correlation between the complexity of motivation and variant token and variant type	121
Table 5.26:	Descriptive analysis of analysability and variant token and variant type	124
Table 5.27:	Mann-Whitney test for analysability and variant token and variant type	125
Table 5.28:	Spearman correlation between analysability and variant token and variant type	125
Table 5.29:	Descriptive analysis of degrees of analysability and variant token and variant type	126
Table 5.30:	Kruskal-Wallis test for the degrees of analysability and variant token and variant type	126
Table 5.31:	Spearman correlation between the degrees of analysability and variant token and variant type	127
Table 5.32:	Descriptive analysis of degrees of transparency and variant token and variant type	129
Table 5.33:	Kruskal-Wallis test for degrees of transparency and variant token and variant type	129
Table 5.34:	Spearman correlation of degrees of transparency and variant token and variant type	130
Table 5.35:	Descriptive analysis of transparency groups and variant token and variant type	131
Table 5.36:	Kruskal-Wallis test for groups of transparency and variant token and variant type	131

GLOSSARY OF TERMS

Analysability/Decomposability

A property where an idiom has separate, meaningful units contributing to its overall figurative interpretation, so that a correspondence is recognised between components of the literal meaning of the idiom and those of its figurative meaning

Collocation

A sequence of words that co-occur more often than would be expected by chance

Compatibility

A constraint on idiomatic creativity that requires any formal modification of the literal meaning of an idiom to be in accordance with the idiomatic meaning

Compositeness

The virtue of being made up of groups of words

Conceptual metaphor

A complex idealised cognitive model that is cognitively constructed by mapping a concrete source-domain onto a more abstract target-domain. Conceptual metaphors are a pre-existing, fixed part of our conceptual system of knowledge on the basis of which we can make sense of meanings, particularly of novel, imaginative expressions

Conceptual metonymy

A conventional association by the recognition of contiguity relationship between a source domain and a target domain, including *adjuncts* (things found together) or *functional relationships* (things performing similar functions) etc. Like conceptual metaphors, conceptual metonymies are also part of the cognitive competence of normal speakers and hearers and are readily accessible in particular linguistic and extra-linguistic contexts

Constituent motivation

The type of motivation that links constituents of the literal meaning to components of the idiomatic meaning

Context

A set of elements, ranging from co-textual features, physical surroundings of the speech situation, social situation, participants common background knowledge and the channel or medium, which contribute to the hearer's interpretation of what the speaker means by a given utterance

Context-dependent idiom variant

The type of idiom variants whose meanings cannot be worked out via interpretative strategies due to the lack of underlying motivations between the variants and the original idioms. Instead, the meaning of a context-dependent variant is a function of the meaning of the canonical idiom and the meaning of the

variant in context. Context is therefore essential in the interpretation of this type of idiom variant

Context-independent idiom variant

The type of idiom variants that requires no extra context for the processing of their meaning. Upon the activation of the idiomatic meaning of the original canonical idioms, the variants can be readily understood using inferential skills without the need for extra context. This independence is thanks to the motivated nature of the variation

Creativity

A psychological construct that involves novelty, originality, or newness, achievable by breaking with expected norms

Determinants of idiomatic creativity

The factors that determine how open an idiom potentially is for creative variations in discourse. Facilitating factors include analysability, motivation, well-formedness, while constraining factors include recognisability, functionality, grammaticality and compatibility

Emblem

Stereo-typical cultural information or symbols which can serve as direct bases for idiom motivation, including animals, objects, places, or events which are used for people, virtue, etc. thanks to their symbolic significance or eminent characteristics

Fixedness

The status of being stable in form of an idiom across much of its range of occurrence, showing resistance to internal manipulation. Fixedness in idioms is a matter of degree

Flexibility

The potential to allow for a wide range of variations by an idiom, measured by the number of variant types per idiom in the data. Flexibility is usually assumed to entail productivity, but not necessarily so due to the interference of a number of facilitating factors and constraints

Formulaic language

Ready-made word combinations that are stored as wholes in our mental lexicon and can be drawn from memory for use without having to be built from scratch

Full motivation

The type of motivation that renders an idiom most transparent because there are links between the literal meaning and the idiomatic meaning of the idiom both as a whole and on the individual constituent level

Functionality

A qualitative constraint on idiomatic creativity, requiring that an idiom variant is only valid if it is an intentional adaptation of a canonical idiom with the purpose of fulfilling a specific communicative purpose in a given communicative situation

Global motivation

The type of motivation that links the literal meaning to the idiomatic meaning as a whole, but no correspondence can be found between components of the literal meaning and those of the idiomatic meaning

Grammaticality

A constraint on any variant regarding its formal structures, such that the variant must be constructed in compliance with the grammar of the target language

Idiom

An institutionalised word combination which is semantically non-compositional and whose syntactic, lexical and phonological form is to a greater and lesser degree fixed. An idiom expresses a concept.

Idiom's base-form

The neutral, standard and canonical form and usage of an idiom

Idiom variant

The idiomatic construction resulting from a given formal or semantic alteration of an idiom's base-form

Idiom variation strategies

A variation strategy captures a distinct pattern of how an idiom is adapted in response to the meanings and communicative functions it is intended to encode. Variation strategies are most often described in lexico-grammatical terms (substitution, insertion, permutation, truncation, etc.)

Idiomatic creativity

The linguistic phenomenon in which an idiom undergoes transformations/alterations to produce creative variants in specific communicative events to serve specific communicative purposes

Institutionalisation

The degree of being conventionalised, recognised and familiar such that a phrase is reproduced as a memorised sequence because of its specific meaning or pragmatic function

Linguistic creativity

Language uses that depart from expected norms via metaphoric invention, punning, etc. (pattern-reforming) or echo to create parallels and symmetries (pattern-forming). Linguistic creativity is chronological and cultural specific and fulfills fundamental communicative purposes

Literal well-formedness/ill-formedness

The status of being grammatically, semantically or pragmatically correct, or otherwise, of the literal meaning of an idiom with reference to the grammatical rules of the target language

Mixed method research

A combination of qualitative and quantitative methods within a single research project

Motivation

A speaker's ability to make sense of an idiomatic expression by reactivating or remotivating their figurativity, i.e. to understand why the idiom has the idiomatic meaning it has with a view to its literal meaning. The sources of motivation include, but not exclusively, conceptual metaphor, conceptual metonymy, emblem, knowledge of the world, etc.

Multi-word unit

A recurrent pattern of lexical material sanctioned by usage, which is potentially stored and retrieved as a whole in the mental lexicon, and which is fixed to greater or lesser degree

Non-compositionality

The characteristic of a linguistic string where there is an absence of any discernable relationship between its form and its meaning, i.e. the meaning of the string cannot be derived by adding the meanings of the constituent words according to the principle of compositionality

Principle of compositionality

The meaning of a sentence is determined by the meaning of its component parts and the manner in which they are arranged in syntactic structure

Productivity

The frequency with which a particular idiom is used in creative variant forms in discourse, measured by the frequency of appearance of the creative variants of an idiom in the data

Recognisability

A general constraint on all types of idiom variation that defines the maximum degree to which an idiom variant may deviate from the base-form while still performing its idiomatic function

Stratified sampling technique

A research technique generally used when the population is heterogeneous, or dissimilar, where certain sub-populations can be isolated (strata). A sample is selected from the target population in such a way that it is representative of the population. This is achieved by taking samples from each stratum or sub-group of the population so that the proportion of each stratum in the sample should be the same as in the population.

Transparency

The degree to which the internal semantic structure of an idiom promotes understanding of the literal and figurative meaning of the idiom and their correspondences, thus allowing variations. The degree of transparency of an idiom is dependent on the degree of analysability, motivation and literal well-formedness of that idiom.

CHAPTER 1

INTRODUCTION

Against the common view of vocabulary as an inventory of separate words, modern linguists and language users have noticed a systematic, inter-dependent nature of words in use that involves co-occurrence and conceptual associative organisation (Fernando, 1996). It is noticed that such phrases as *a kettle of fish*, *by and large*, *think outside the box*, etc. are neither freely combined nor easily reconfigured; instead, their components enter into various pragmatic and semantic relationships that render the phrases relatively fixed. The term *fixed expression* has indeed been popular in the literature to refer to a variety of linguistic expressions, among which are idioms, collocations, phrasal verbs, sayings, proverbs and conversational gambits, which share the property of lexical fixedness between the words inside the expressions, thus allowing for potential idiosyncrasies in terms of syntax, semantics and pragmatics.

Over time, however, the so-called lexical fixedness of such phrases has been brought into question due to the overwhelming evidence of the wide range of non-canonical uses in real life discourse (e.g. *a kettle of piranhas*, *by and not so large*, *think outside but press right against the box*). Research into the mechanisms of fixed expression variations inspired by such evidence has indeed seen significant growth in both popularity and depth, particularly since the availability of large electronic corpora.

As a subgroup of fixed expressions, *idioms* were traditionally considered one of the most stable across their occurrences. Nevertheless, recent empirical studies have uncovered a flexibility (to certain extents) of idioms that has not been captured before. Zooming in on idioms and their variations, the current thesis is a corpus-based investigation into the underlying factors of the variable uses of idioms in authentic English discourse. In this introductory chapter, a preliminary definition of idioms is first introduced to familiarise the readers with the concept, followed by a detailed explanation of the rationale and the specific research questions that this thesis addresses.

1.1. A preliminary definition of idioms

By definition, true idioms, such as *kick the bucket*, *shoot the breeze*, *bite the dust*, etc., are conventional multi-word units which are semantically opaque and structurally fixed. In other words, the defining characteristics of an idiom have been argued to include the followings:

- (1) Compositeness: the virtue of being 'units of words', or 'composites', consisting of two or more lexical constituents
- (2) Institutionalisation: the conventionalisation or familiarity of the phrase
- (3) Fixedness: the status of having restricted syntactic, morpho-syntactic and lexical variability, which cannot be explained by the general grammatical rules of the given language
- (4) Non-compositionality: the semantic figurativity of the phrase, which means the meaning of the phrase cannot be derived from the meanings of constituent words. For example, *open a can of worms* (to create a situation that will cause trouble or be unpleasant) cannot be interpreted by adding the literal meanings of constituent words (*open, can, worms*).
- (5) Transparency/Opaqueness: the discrepancy between the literal and the idiomatic meaning, such that no connection can be established between the two meanings, causing an idiom to be opaque. The idiomatic meaning of a fully opaque idiom such as *trip the light fantastic* (to dance) can only be acquired via an arbitrary connection. The meaning of a more transparent idiom such as *open a can of worms*, on the other hand, can be seen as 'making sense' thanks to the existence of certain links between the two meanings. The contribution of the verb *open* (create) and the knowledge associated with the concept WORM (unpleasant squirming creeping invertebrates) help the language user see the connections between the literal and the figurative meaning of the phrase. The idiom can then be analysed into meaningful components ('open' = create, cause; 'a can of worms' = troubles, problems), which contribute to the understanding of the figurative meaning of the idiom as a whole.

The opacity or transparency of an idiom can therefore be broken down into three components, including:

- (6) Literal well-formedness/ill-formedness: the presence of potential irregular idiosyncratic lexical items and/or grammatical patterns in the literal meaning of an idiom renders it ill-formed (*trip the light fantastic*), as opposed to grammatically correct phrases (*open a can of worms*).
- (7) Analysability: the possibility of devolving the figurative meaning of an idiom on its constituents through top-down analysis, such that the phrase can be analysed into meaningful components that corresponds to components of the figurative meaning. For instance, the idiom *spill the beans* is analysable because it can be analysed into meaning components that correspond to components of the idiomatic meaning (*spill* = divulge, *the beans* = a secret). Analysability therefore does

not imply the possibility of ‘guessing’ the meanings of unfamiliar idioms through bottom-up composition but describes the process of top-down reanalysis once the idiomatic meaning of a phrase is known.

- (8) Motivation: the discernible links between the figurative and the literal meanings of certain idioms that makes it possible to understand why these idioms mean what they mean. *Spill* meaning ‘divulge’/‘let out’ is motivated by knowledge of the autonomous semantic contents of *spill* (e.g. *spill blood*, *spill one’s guts*), in the same way that *swallow* means ‘receive with reluctance’ in *swallow bitter pills*, *swallow a hard truth*. However, it is unknown why *the beans* means ‘a secret’, hence the lack of motivation for this particular correspondence.

To sum up this preliminary definition and description of idioms, it is necessary to acknowledge that the definitory parameters and/or quintessential characteristics of idioms as introduced above have been treated as ‘descriptive spotlights’, the sum of all of which provides a holistic picture of the nature of the phenomenon, but none of which is a prerequisite due to the clinal nature of these dimensions as well as the diversity of the types of idiomatic constructions. As a result, with reference to how many of these characteristics are emphasized as well as their relative degrees of strikingness, different definitions and classifications of idiomatic expressions can be distinguished in the literature. In this thesis, idioms are understood as:

A construction of two or more words that features figurativity, i.e. its semantic structure is non-compositional. It is institutionalised and structurally and lexically fixed to certain degrees.

In other words, compositeness, non-compositionality, institutionalisation and fixedness are the four characteristics perceived as determining the idiom status of a phrase. Other characteristics (5-8) as discussed earlier are non-defining and may or may not feature in an idiom without changing its status. The presence or otherwise of these characteristics, however, changes the class of an idiom and signals its behaviour in discourse, as we shall see later in the thesis.

With reference to one of the most important characteristics of idioms, i.e. fixedness, despite being characterised as a type of fixed expression, it is an empirical fact that idioms do allow variability. The question, however, remains when and how much of this variability can we expect given the inherent fixed and institutionalised nature of idioms? More detailed discussions of the fixedness of idioms relative to the concepts of creativity and context will be provided in the next sections to introduce the core elements that the thesis revolves around in its quest to find answers to that question.

1.2. Idioms and the struggle between fixedness and creativity

Fixedness is an inherent attribute of all idioms as a result of the development of idiomatic expressions, from ad hoc expressions through pragmatic inferencing and then through salience reduction (Hudson, 1998), at which stage 'the contribution of the parts of the expression to the whole is beyond conceptualisation, and the expression becomes fixed in its realisation' (ibid: 2). This fixedness means that each idiom shows such a strong fossilisation in a particular grammatical structure and lexical arrangement that variations are unrecognizable or simply rejected on the basis of nonstandard usage. From a cognitive perspective, due to the idiosyncrasies in syntax and semantics of an idiom, fixedness in realisation in all uses ensures the prompt, successful processing of the idiom, particularly in the cases of highly opaque ones. In other words, it is traditionally believed that idioms are fixed across most of their occurrences and are preferably used invariably in most contexts due to phrasological preferences and economy of processing.

Contrary to this traditional view of idioms, which underscores their quality of being 'fixed' as such, corpus linguists in recent years have consistently produced evidence of the variability of idioms in discourse, hence the concept of *idiomatic creativity*. Creativity in linguistics is often understood as involving a break with the norms, with an emphasis on the flouting of expectations and the reshaping of patterns to create newness. As a result, idiomatic uses that are not in their canonical forms but refashioned to fit specific communicative contexts and/or to create special effect (humour, sarcasm, etc.) are believed to embrace the essence of creativity, for example:

- [1] *Some pockets were fully or partially lined*
- [2] *Lips lined - so are his pockets*
- [3] *Not all earmarks are meant to line congressional pockets*
- [4] *Tiford and Barwick could apply the soft-talk while lining their elegant pockets with silver*
- [5] *Many a dark cloud can provide a silver lined pocket*

Variations and exploitations such as these are treated as idiomatic creativity because they are more than just alternative wordings of the canonical idiom *line one's (own) pocket*; they serve specific semantic and pragmatic roles in discourse, i.e. 'that of personalizing and making one's language use relevant to the situation in which it is uttered' (Philip, 2006: 95) with added stylistic undertones.

In her corpus analysis of fixed expressions and idioms (FEIs), Moon (1998) reports a large number of idioms to be 'unstable'. To be specific, around 40 per cent of the FEIs in English have at least one lexical or grammatical variant and as many as 14 per cent have two or more variants. More recently, Philip (2006) even claims that creative idiom variants in real life usage have been found in his study to outnumber canonical idioms. Such empirical

evidence of the pervasiveness of idiomatic creativity in discourse has started to highlight the systematic breach of the fixedness in idioms in real life discourse. On the other hand, the presence of a large body of invariable idioms in the data (e.g. around 60% of FELs – Moon, 1998) serves to uphold the fixedness hypothesis to certain extents. There is evidently a dynamic tension between fixedness and creativity, such that either idiomatic fixedness or idiomatic creativity is a matter of degree, depending on various factors.

It is this struggle between fixedness and creativity under the effects of various constraining and facilitating factors that sparks my interest in the subject matter of idiomatic creativity. Ultimately, I strive to build a model of idiomatic creativity that is comprehensive enough to account for the flexibility of idioms, while critical enough to delineate the possible constraints where applicable.

1.3. Towards a context-based model of idiomatic creativity

The factors contributing to the flexibility of an idiom have been postulated in the literature to include aspect of transparency (motivation, analysability, and literal well-formedness). That is, if an idiom is more or less well-formed, analysable and motivated (i.e. fully transparent) (for example *skating on thin ice*), it is *more likely* to have more creative variants in discourse than an ill-formed, unanalysable and unmotivated (i.e. fully opaque) idiom (for instance *shoot the breeze*). While this hypothesis has its merits, it is argued that previous models of idiomatic creativity, which are built on this hypothesis, are too heavily geared towards internal semantic characteristics of idioms, thus missing an important dimension of discourse, i.e. context, and the cognitive constraints on idiomatic creativity, which both need to be incorporated into the model.

To ensure the interpretability and communicability of any novel expression in general and of a novel variant of an idiom in particular, the target expression has to satisfy the cognitive requirements of *recognisability*, *grammaticality*, *compatibility* and *functionality*. That is, a creative variant of an idiom is required, for processing purposes, to be recognisable as deriving from a canonical idiom, to be grammatically correct according to the rules of the target language (English in this case), to be semantically compatible with the figurative meaning of the canonical idiom, and to serve a discernible purpose in the speech event. For example, *the pail has singing by him* _[sic] is not accepted as a valid variant of *he kicked the bucket* because it is ungrammatical according English grammar ('the pail has singing'), unrecognisable as deriving from the original idiom, incompatible with the idiomatic meaning 'to die', and apparently does not serve a function in discourse due to the lack of meaningfulness. In other words, while the semantic structures of transparent idioms might be open for creative manipulations, the possibilities of deriving valid creative variants from them are not open-ended but restricted due to the above constraints.

On the other hand, although theoretically the possibilities of deriving creative variants from opaque idioms are not great, they can be improved by context. In fact, context is argued to be a facilitating factor that increases the chance for creativity not only amongst opaque idioms but also amongst transparent ones by lifting some or all of the aforementioned semantic and cognitive constraints. Context has increasingly been included in any language description/theory that emphasizes usage-based meanings across many disciplines, including pragmatics (Levison, 1983), discourse analysis (Schiffrin, 2003), or corpus linguistics (Adolphs, 2008). Due to the highly deictic nature of language, it is not an exaggeration to claim that meaning depends on context. Defined as a set of elements, ranging from co-textual features, physical surroundings of the speech situation, social situation, participants common background knowledge and the channel or medium (see Auer, 1995; Hymes, 1972), context is essential in meaning comprehension. The following examples of idiomatic creativity are highly context-specific, which begs the question of how much of idiomatic creativity and which parts of it are context-dependent:

- [6] *Holly had fun and painted the town, er, orange*
- [7] *You're barking up the wrong fish*
- [8] *By the end of the night, one by one, they threw in the napkin*

Without contexts, these variants of the canonical idioms *paint the town red*, *bark up the wrong tree*, and *throw in the towel* are ambiguous due to the non-salience of the substituted constituents and the unanalysability of the idioms. Nevertheless, in the context of Holly wearing a striking orange dress on a night out (in [6]), a reply from a woman to man who was trying to flirt with her by talking about fishing (in [7]), and a group of women dining in a restaurant and waiting in vain for a hot waiter (in [8]), the creative variants do not only make sense but also create witty context-specific wordplays. Considering that the three canonical idioms are opaque and unanalysable, these examples suggest that context has the power to increase the creativity potential of idioms beyond what has been accounted for by semantic or cognitive models of the phenomenon. Indeed, Glucksberg (1993) rightly comments that the determinants of idiom flexibility/productivity are complex and feature a combination of both the internal semantics of the idiom and the pragmatics of the discourse context. This relationship between idiomatic creativity and context, however, is far from clear and decidedly under-researched. A complete framework of the role of context in facilitating idiomatic creativity has not been available in the literature. This thesis aims to change this.

In this thesis, therefore, I aim to develop a model that takes context as one of the facilitating factors of idiomatic creativity. The thesis seeks to explore in depth, using corpus data, the potential for creative uses of idioms in relation to the internal semantic characteristics of idioms under the effect of cognitive constraints and context, all of which exist in a multidirectional intertwined relationship with each other as a dynamic network. The interaction between context and other facilitating and constraining factors arguably

represents the truthful state of affairs concerning the phenomenon of idiomatic creativity as it happens in real life discourse. Such a context-informed model, as a result, is believed to be desirable in our quest towards a model that can both *describe* the possibilities and *predict* the potential for creativity among idioms.

1.4. Research questions

As discussed earlier, the underlying facilitating factors of idiomatic creativity have been suggested to include aspects of the semantic characteristics of idioms, including the grammaticality of their literal meanings (literal well-formedness), the availability of the conceptual links between the literal and figurative meanings (motivation) and the possibility of analysing the idioms into meaningful components (analysability), the lack of one or more of which will result in difficulties in manipulating idioms. However, there is little research into how much and in what manner these elements affect idiomatic creativity in relation to the constraining factors (recognisability, functionality, grammaticality, compatibility) and, particularly, to context. As a result, via careful analyses of corpus data, the research questions read:

- *Is there an empirical correlation between idiomatic transparency (determined by literal well-formedness, analysability, motivation) and idiomatic creativity?*
- *Do cognitive constraints (including recognisability, functionality, compatibility and grammaticality) apply to all types of idiomatic creativity?*
- *How does context facilitate idiomatic creativity and to what degree?*
- *Are these facilitating and constraining factors equally applicable to all types of idiom creativity in all cases? Or is there an order of priority that is case-specific?*

Answers to the above questions serve to build a more comprehensive account of idiomatic creativity that takes into consideration a wide range of factors and, more importantly, their interaction with one another in real life discourse.

‘Real life discourse’, as we have reiterated so far, means that the variants will be elicited from real texts and corpora and treated in relation to the contexts they are embedded in. Many linguists admit that the existing literature on the transformation potentials of fixed expressions and idioms are ‘marred by a lack of authentic data’ (Moon, 1998: 105; Kovecses and Szabo, 1996). Indeed, studies in the area of idiomatic creativity and its determinants have been carried out using psycholinguistic experiments, survey questionnaires and introspection. I am convinced that the way forward in idiomatic creativity studies is to analyse large bodies of authentic written and spoken texts so that the underlying mechanisms can emerge with minimal human manipulation of the data.

1.5. Structure of the thesis

The thesis revolves around six central chapters. Except for the first chapter, being the introduction to the thesis, detailing the aims, the research questions, and the shape of the thesis, chapters 2 and 3 provide general theoretical background for the thesis. Chapter 2 looks at the concepts of *idiomaticity* and *creativity* in separation prior to discussing their contributions as parts of the concept of *idiomatic creativity*. A working definition and framework of idiomatic creativity is proposed at the end of the chapter to serve as the theoretical departure point for the analytical procedures in the second half of the thesis. Chapter 3 reviews the hypotheses in the literature about the determinants of idiomatic creativity, thereby drawing a picture of the current debates in the field concerning the subject matter and identifying the gaps in such debates where the thesis fits. Chapter 4 is dedicated to a careful description of the data and methodology used in the thesis as well as the advantages and disadvantages associated with such data and methodology. Chapters 5 and 6 are the main analytical chapters, addressing various aspects of the research questions. Chapter 5 examines the correlation between literal well-formedness, motivation, analysability and idiomatic creativity via statistical analyses, charts and graphs. The results are expected to reveal the extent to which semantic characteristics are determinants of idiomatic creativity. Via detailed qualitative analyses of the data, chapter 6 continues the investigation into the roles of literal well-formedness, motivation and analysability in idiomatic creativity but under the effect of the cognitive constraints and particularly in relation to contextual elements. Together, the results from chapters 5 and 6 form a context-based framework of idiomatic creativity, one that takes into account the interactions between the internal semantic structures of idioms, cognitive constraints and context. Chapter 7 concludes the thesis and explores the possibilities for upgrading existing corpus tools to better support idiomatic creativity studies in the future.

CHAPTER 2

THE CONCEPTS OF IDIOMATICITY, CREATIVITY, AND IDIOMATIC CREATIVITY

In this chapter, the concepts of idiomaticity and creativity will be introduced to lay the foundation for subsequent discussions and analyses of the phenomenon of idiomatic creativity in the thesis. The first half of this chapter is dedicated to providing a comprehensive description of idioms and their status relative to other multiword units, outlining their key features within the broader context of multiword units and discussing their defining characteristics as portrayed within different frameworks. The second half of the chapter discusses the concepts of creativity in general and linguistic creativity in particular to set the scene against which the essence of the phenomenon of idiomatic creativity is anatomized. In-depth discussions of the prominent features of idiomatic creativity as well as explanations of related concepts in use in this thesis are provided to pave the theoretical path for the rest of our idiomatic journey.

2.1. Overview of multiword units

2.1.1. Terminology

In the history of fixed expression studies, there are major disciplinary variations of terminology that can confuse readers and researchers alike, the most popular of which include *fixed expressions*, *multiword units*, *formulaic sequences*, *prefabricated chunks*, etc. The different terms in effect highlight different aspects of the phenomenon and indicate the authors' perceived importance of a feature of the strings over others. In using terms that place the emphasis on 'formulaicity', such as *formulaic sequences* (Schmitt, 2004), *formulaic language* (Wray, 2002), or *prefabricated speech* (Bolinger, 1975), the authors imply that the holistic storage and retrieval of these expressions are the most important features. The term *lexical phrases* (Nattinger, 1980) is based on the assumption that language-in-use is primarily recurrent word combinations that are stored, accessed and processed as if they were single words (Lewis, 1993). *Fixed expression* (Carter, 1998; Hudson, 1998), *fixed expressions and idioms (FEIs)* (Moon, 1998), on the other hand, are intended to denote fixed and recurrent patterns of lexical material sanctioned by usage, thus emphasizing lexico-grammatical stability. Placing the emphasis on the importance of co-occurrence, Sinclair (1991) uses the term *collocation* to refer to all word combinations that have statistical significance with regard to the occurrence of their constituents within a short space of each other in a text. In essence, these different labels all address one linguistic phenomenon in common. Even though each term carries one or more features that are deemed defining, they do not reject the possible existence of other features. The adoption of a particular term varies to suit each individual's

purposes and reflects his/her own approach to the subject matter. In this thesis, *multiword unit* will be adopted for general usage, referring to:

a recurrent pattern of lexical material sanctioned by usage, consisting of two or more words which co-occur in a more or less fixed arrangement, familiar among users of a target language, and having a specialised meaning

The adoption of the term *multiword unit* in this thesis, therefore, places the emphasis on the compositeness of the strings, while avoiding the controversies around the notion of 'fixedness' in the term *fixed expression*, as well as indicating a disinterest in the storage and retrieval aspects of the phenomenon as expressed in such terms as *formulaic sequences* or *prefabricated chunks*. As it will become clear later on, multiword units are by no means fixed; rather, they *can* be instantaneously refashioned to adapt to specific situations in everyday life conversations and to achieve specific communicative purposes.

2.1.2. Multiword units and the lexicon

The evidence of the ubiquity of multiword units in every aspect of our verbal behaviour has been well documented (Coulmas, 1981; Peters, 1983; Sinclair, 1991; Pawley and Syder, 1983; Biber, Conrad and Reppen, 1998; Moon, 1998). Altenberg (1990: 134), for example, reports that 70% of running words in the London-Lund corpus form part of recurrent word combinations of some kind. This has led many linguists towards theories of language in which language knowledge is seen as consisting of prefabricated phrases and sentences memorized from previous encounters with them (Bloom, 1973; Bolinger, 1976; Fillmore, Kay and O'Connor, 1988). The basis for such an approach to language representation and usage is the recognition that 'just as we are creatures of habit in other aspects of our behaviour, so apparently are we in the ways we come to use language' (Nattinger and DeCarrico, 1992: 1). Bolinger (1976: 2) also argues that the 'human mind is less remarkable for its creativity than for the fact that it remembers everything'. As a result, multiword units in these theories are treated as central to human verbal behaviour.

Taking a different approach, Chomsky (1965) argues for an opposite theory that is least sympathetic to multiword units as principal features of the language structure. According to him, human beings have the inherent capacity to generate and understand novel sentences that they have never encountered before. Our language knowledge, therefore, consists of words and deep structures from which different surface structures and sentences can be constructed. The language of a normal adult native speaker is fully generated at the time of production and fully analysed in comprehension. As such, prefabricated items such as idioms and other multiword units are pushed to the periphery as minor exceptions to the general language rules.

Although both approaches to language representation have their own merits, it is often believed that to argue for only *one* of the two is to ‘underestimate the great capacity of the human mind to remember things while overestimating the extent to which humans process information by complex processes of calculation rather than by simply using prefabricated units from memory’ (Lamb, 1998: 169). As a result, a middle ground is suggested, such that in our language knowledge co-exists the capacity to use prefabricated multiword units *and* the ability to create and understand entirely novel strings (Wray, 2002). With reference to idiomatic creativity, I find this approach particularly fitting. The human mind is capable of manipulating root idioms to create novel, context-specific variants that might not have been encountered before, but this is dependent on memorised information about such idioms concerning their meanings, conventional usage, etc. Both capacities of memory and creativity are therefore recognised as essential to human’s linguistic skills involving idiomatic creativity.

2.1.3. *Taxonomies of multiword units*

Thanks to its power in describing in a systematic and clear-cut manner the categories of multiword units, the taxonomical approach has been popular with linguists who are concerned with the identification of the linguistic criteria for distinguishing one type of multiword unit from another. The scope of the phenomenon of multiword units, however, is as fuzzy as its terminology, depending on the criteria according to which they are distinguished from literal combinations. The following list by Howarth (1998) illustrates the range of expressions that could be treated as multiword units:

[1]	<i>princely sum</i>	restricted collocation
[2]	<i>gin and tonic</i>	binomial collocation
[3]	<i>put pen to paper</i>	figurative idiom
[4]	<i>kick the bucket</i>	pure idiom
[5]	<i>by and large</i>	asyntactic idiom
[6]	<i>how do you do?</i>	routine formula
[7]	<i>by the way</i>	conversational gambit
[8]	<i>hello, good evening, and welcome</i>	catchphrase
[9]	<i>go to work on an egg</i>	slogan
[10]	<i>more haste less speed</i>	proverb
[11]	<i>when in Rome ...</i>	abbreviated proverb
[12]	<i>1992 and all that</i>	allusion
[13]	<i>you've never had it so good</i>	quotation

These categories are not as clear-cut as they look. This is partly due to the fact that a lot of items can be seen as belonging to different categories if viewed from different perspectives (formal, functional or stylistic), making the task of classifying word combinations a hard nut to crack.

Russian lexicologists are often acknowledged as the founders of a view of phraseology that 'restricts the scope of the field to a specific subset of linguistically defined multiword units [...] with the most opaque and fixed ones at one end and the most transparent and variable ones at the other' (Granger and Paquot, 2008: 28). Vinogradov (reviewed by Cowie, 1998 and Weinreich, 1963) proposes a classification of phraseological units into three types, *idioms* being the first type which are unmotivated, unanalysable, semantically opaque and structurally fixed (also termed *phraseological fusions*). The second type is *phraseological unities*, partially motivated phrases, and semantically semi-opaque. Lastly, *phraseological combinations* or *collocations* are phrases of two open-class words, one of which maintains its literal sense, while the other is used figuratively. The criteria for categorisation in Vinogradov's taxonomy, as a result, are quite close to the modern takes on idiomaticity, i.e. motivation, analysability, semantic opacity, and structural fixedness. Amosova (1963), similarly, uses fixedness and opacity as classificatory criteria to separate *pure idioms* (completely fixed and opaque strings) from *phrasemes* (frozen collocations/compounds in which one element has a meaning unique to the combinations), *phraseoloids* (restricted collocations where there is limited paradigmatic variability) and *fixed combinations* (transparent formulae or fixed collocations) (see Moon, 1998: 12-13).

Among the many characteristics that idioms might have, the defining feature setting idioms apart from other types of fixed phrases is argued by Mel'cuk (1998) to be non-compositionality. Multiword units (or *phrasemes* as he calls them) are of two broad types, including *pragmatemes* (conversational formulas, clichés, proverbs, etc.) and *semantic phrasemes*, of which idioms, collocations and quasi-idioms constitute major subgroups. Idiomaticity, in Mel'cuk's framework is to be distinguished from 'collocation' on account of the different restrictions on them. That is, while the former is characterized by 'a strong restriction on the selection of a subsense', the latter is defined by 'a high degree of contextual restriction' (Weinreich, 1969: 44). In other words, an idiom (such as *kick the bucket*) is an expression whose meaning does not include the meanings of its component words – its selection of meaning/subsense is restricted. A collocation (such as *face the truth*), on the other hand, is a phrase in which one constituent is free on the basis of the meaning (*truth*), but the other is dictated by the free constituent (*face*).

Outside of the Russian lexicology circle, many other linguists are also inclined to adopt categorical classifications of multiword units. Criteria for classification might vary from researcher to researcher, depending on their particular goals and disciplinary influences. Becker (1975) and Nattinger (1980) both use a system of six types of multiword expressions on the basis of four criteria, namely 'form', 'function', 'meaning' and 'provenance':

- polywords (idioms being one of them): short phrases with extremely low variability whose meanings exist apart from syntax: *my old man*, *at my place*
- phrasal constraints: *by sheer coincidence*

- deictic locutions (Nattinger) or meta-languages (Becker): patterned routines to direct the flow of the conversation by marking attitudes: *as far as I'm concerned, if I were you*
- sentence builders: *(person A) gave (person B) a (long) song and dance about (a topic)*
- situational utterances: to encode an appropriate social or interpersonal response: *nice to meet you, are you sure*
- verbatim texts: *better late than never*

Using different criteria, Cowie's (1988, 1994) classification of multiword expressions is based on the meanings they convey and the structural levels at which they operate. *Pragmatically specialized expressions* (formulae) are phrases whose meanings 'are largely a reflection of the way they function in discourse' (Cowie, 1988: 132), including greetings, invitations, gambits, etc. *Semantically specialized expressions* (composites), on the other hand, include those whose meanings are 'more or less referential by virtue of their use as invariable units in grammatical constructions' (ibid: 133). Cowie et al. (1983, 1993) further classify semantically specialized expressions into *pure idioms*, *figurative idioms*, and *restricted collocations*, which are to be distinguished from *open collocations* on the basis of their 'opacity'.

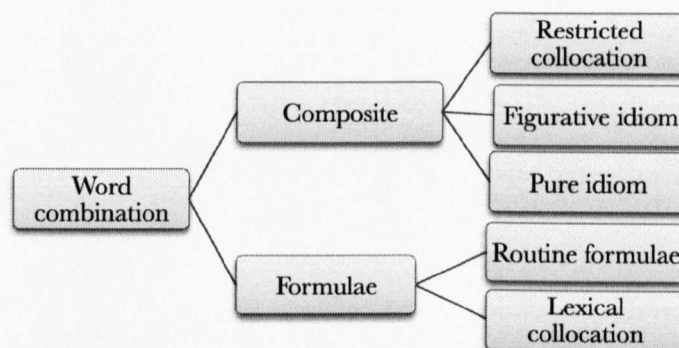


Figure 2.1. Taxonomy of word combinations (adapted from Cowie, 1988; Cowie et al., 1983, 1993)

As such, 'pure idioms' (e.g. *blow the gaff*) are the most fixed and opaque, featuring the strictest sense of idiomaticity: they are perceived as 'the end point of a process by which word combinations first establish themselves through constant re-use, then undergo figurative extension and finally petrify or congeal' (ibid: xii) (cf. Fernando and Flavell, 1981). Quite close in terms of characteristics to pure idioms are 'figurative idioms' (e.g. *a close shave*,) which are slightly less opaque and relatively fixed, in the sense that variation is seldom found and pronoun substitution unlikely. At the other end of the scale, collocations are least fixed and least opaque. 'Restricted collocations' include phrases in which one element is used in a familiar, literal sense (hence more transparent than idioms) with only one element being figurative (e.g. *jog one's memory*). 'Open collocations', on the other hand, are freely recombining (i.e. variable) and transparent because each element is used in its literal sense.

Howarth's (1998) system is very similar to Cowie's, established on the basis of the string's discourse functions and syntactic functions within the clause. *Functional expressions*, which corresponds to Cowie's category of 'formulae', are complete phrases or utterances whose discursal functions include discourse-structuring devices, conversational gambits, routine formulae, proverbs, catchphrases or slogans (e.g. *tell me about it, by the way*). *Composite units*, on the other hand, are only part of a clause or utterance/sentence and are usually in the forms of phrases (noun phrases, prepositional phrases, verb phrases, etc.) whose functions are to convey a concept (e.g. *grasp the nettle, go to work on an egg*). Composite units are further divided on the basis of formal realisation into grammatical and lexical collocations. Lexical collocations are word combinations of two open-class words (N + N, Adj + N, etc.) (e.g. *bread and butter, reach a conclusion*), whilst grammatical collocations are those having at least one closed-class word in their structures (V + particle, Prep + N, etc.) (e.g. *set up, on sale*). Last but not least, each of these categories of expressions comprises of non-idiomatic expressions and idiomatic ones.

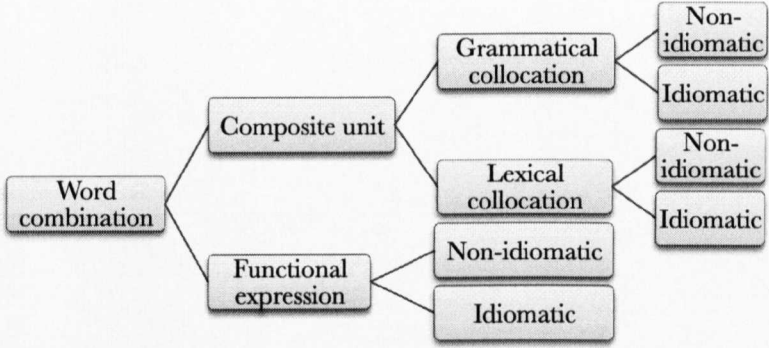


Figure 2.2. Taxonomy of word combinations (adapted from Howarth, 1998)

Multiword units, as a result, are generally acknowledged as a 'special' class of lexical items with idiosyncratic forms and functions in discourse. On the basis of the level at which the idiosyncrasies occurs, Moon (1998) derives three macro-categories, namely *anomalous collocations*, *formulae* and *metaphors*. 'Anomalous collocations' include multiword units which are problematic in lexicogrammatical terms, i.e. they are syntagmatically or paradigmatically aberrant, for example syntactically ill-formed collocations (e.g. *at all, by and large*), cranberry collocations (e.g. *in retrospect, kith and kin*), defective collocations (e.g. *in effect, foot the bill*) and phraseological collocations (e.g. *in/into/out of action, on show/display*). Secondly, 'formulae' are pragmatically anomalous, i.e. the strings have special discourse functions that cannot be normally inferred, for example formulae/sayings (*alive and well, to be or not to be*), proverbs (*you can't have your cake and eat it, enough is enough*), and similes (*as good as gold*). The class of 'metaphors', last but not least, embraces semantic idiosyncrasies that cause the phrases to be non-compositional. Metaphors can be transparent (*behind someone's back, pack one's bags*), semi-transparent (*on an even keel, pecking order*), or opaque (*bite the bullet, kick the bucket*). Idioms in Moon's system are a subtype of metaphors which are either semi-transparent or opaque.

In the aforementioned classifications, idioms are considered a subtype of multiword units. In some other systems of classification, however, they can be elevated to a superordinate position to cover the entire spectrum of multiword units (so that *idiom* means *multiword unit* in general). Makkai (1972), for example, divides the whole field of word combinations into *idioms of decoding* (i.e. expressions which force the listener/reader to decode the meanings because they involve semantic irregularities) and *idioms of encoding* (i.e. expressions with lexico-grammatical peculiarities, which are interpretable but arbitrarily conventionalized, e.g. *answer the door*). Among idioms of decoding, Makkai distinguishes *lexemic idioms* (including phrasal verbs, pure idioms and opaque compounds) and *sememic idioms* (including pragmatic and socio-cultural peculiarities such as proverbs and formulaic greetings). The use of the term 'idiom' in Makkai's system, as a result, is broad and subsumes all types of multiword units discussed so far.

On balance, the most popular systems of classification of multiword units have relied on one or more of the following criteria: functionality, non-compositionality, structural fixedness, motivation, and unanalysability. However, it is clear that there is no consensus among researchers with regard to the question of '*what are idioms?*' It seems that *idiom* is sometimes meant to refer to a subgroup of multiword units which presents the strictest sense of the state of fixedness, i.e. completely frozen with restricted collocation, irregular, and opaque. Other linguists allow idiom a broader coverage. In earlier years, some authors used idiom in an even broader sense, including non-compositional polymorphemic words and collocations (Makkai, 1972), single morphemes (Hockett, 1958), conversational formulae and certain indirect speech acts which are conventionalized (Sadock, 1972, 1974; Morgan, 1978; Gibbs, 1986). The existence of the numerous systems of classification in the field is the result of the difficulty with which multiword units and idioms have been identified and defined.

2.1.4. Clines of idiomaticity

Despite the diverse state of affairs due to the lack of agreement between linguists, taxonomies such as those discussed above always seem attractive to the researcher for their neat categorization of the types of multiword units and a clear indication of where idioms come in. The problem with the taxonomical approach, however, is that the categories are somewhat questionable. The potential overlap of the features/criteria used to construct taxonomies undoubtedly muddies the waters to some degree. Hudson (1998: 13), for instance, criticizes taxonomical categories as being 'neither discrete nor comprehensive', and that 'many fixed expressions cannot be accommodated in any of the categories that have hitherto been defined' (ibid: 34). As a result, many linguists have suggested that the features of multiword units are better described as clines. Carter (1998), for instance, proposes viewing fixed expressions and idioms along the three continua of collocability, syntactic structure, and semantic opacity instead of as categories:

Collocational restriction:	Unrestricted collocation
	Semi-restricted collocation
	Familiar collocation
	Restricted collocation
Syntactic structure:	Flexible
	Regular with certain constraints
	Irregular
Semantic opacity:	Transparent
	Semi-idioms/metaphor/idiomatic similes
	Semi-transparent
	Opaque (overt and covert)

An expression such as *fat chance* is characterised as irregular syntactically, opaque semantically and restricted collocationally, and is thus more intractable than *take a chance*, which is less extreme in all three respects. Carter (1998: 72) also notes that ‘the clines have never been unequivocally clear’ especially with semi-idioms and other transitory types.

Taking the same approach, Fernando (1996) uses lexico-grammatical invariability and semantic non-literality (i.e. opacity) as clines to distinguish between *habitual collocations* and *idioms* and various transitory types according to their relative positions on the clines of variability and non-literality. Fernando emphasizes that sometimes an idiom can have similar collocational behaviour as a habitual collocation, but the salient difference is that idioms ‘enters into further linear relationships not only strengthening the interconnections between the parts of a text but also creating various mediations of meaning at the local and global levels’ (ibid: 59).

The relationship between the clines of features of multiword units, however, is of controversial nature. This is because fixedness/collocational restriction, non-compositionality, motivation, unanalysability are not independent features of multiword units but exist in a causal relationship. The cline of ‘collocational restriction’ is argued to be determined by the other clines. (Detailed discussions of existing theories on this relationship are available in chapter 3). It is part of the mission of this thesis to unveil this relationship using corpus data.

Summary

In this idiomatic jungle of terminologies and frameworks, this thesis progresses on the premise that human beings’ language knowledge comprises both the capacity to produce and understand novel strings in various contexts and the capacity to use prefabricated chunks for economy of processing. The prefabricated part of our lexicon is made up of multiword units, i.e. phrases and expressions which are characterised by the co-occurrence of their constituent words and their institutionalisation due to repeated use. As a

subcategory of multiword units, idioms boast highly idiosyncratic features that set them apart from the rest. Each author has used a different set of criteria that include one or more of the following features: non-compositionality, fixedness, structural anomaly, and functionality. These features of idioms are not straightforward and clear-cut in a black and white fashion but are of a highly complex nature with degrees attached to them. Before a working definition of idioms is proposed for the purpose of this thesis, it is useful to unveil the complexity of each of these features and their power in defining idioms, which serve as the platform for our ultimate destination: to settle on a definition of idiom that will be employed for the rest of the thesis.

2.2. Defining features of idioms

The task of defining idioms has been known in the literature to be problematic. Indeed, multiword units in general, instead of being captured by a single definition, are usually characterised by a bundle of features, any or all of which an idiomatic expression may possess but none of which is individually defining or necessary (Wray, 2002). With idioms in particular, the situation is of a similar nature: since there is no agreement within the linguistic community as to which feature is both necessary and sufficient to define idioms, lists of a combination of features are used. In this section, formal and semantic features are discussed in detail. They are also the features that are used in this thesis. Psychological and phonological features, although valuable for certain research areas, are not suitable for the purposes of this thesis, hence a very brief review at the end.

2.2.1. Formal and semantic features

The most popular definitions of idioms in the literature have been formulated using formal and semantic criteria, including *compositeness*, *institutionalisation*, *non-compositionality*, and *fixedness*. The differences between the definitions, however, lie in the varying degrees of importance attached to each feature in each definition.

2.2.1.1. Compositeness

Formally, all multiword units are set apart from literal strings by the virtue of being ‘units of words’, or ‘composites’. Idioms, as a subset of multiword units, are also subject to this criterion of compositeness. Idioms are almost universally defined as institutionalised *groups of words* with set meanings that cannot be calculated by adding up the separate meanings of the parts.

Occasionally, particularly in earlier publications on the topic, it is argued that compositeness does not necessarily refer to combinations of *words*. Palmer (1981), for instance, includes suffixes (but not prefixes) such as *-ish* in *boyish* as a type of idiom because they also have limited collocations like idioms do (not *dogish**). From transformational-generative grammar perspective, Katz & Postal (1963) extend the boundary of idioms to

include polymorphemic single words. Idioms are understood as ‘any concatenation of two or more morphemes whose compound meaning is not compositionally derived from the meanings of the concatenated morphemes’ (ibid: 275). Their definition, as a result, includes what they call ‘lexical idioms’, i.e. lowest syntactic categories such as nouns, verbs, adjectives, etc., which are polymorphemic or compounded. In an even broader view of idioms, Hockett (1958) brings morphemes as well as extended clauses and exchanges such as poems, novels or the Bible into the realm of idioms. However, as Skandera (2004) observes, the comprehensive view of idiom as mentioned in this paragraph is almost universally rejected in modern linguistic theories. Compositeness meaning word combinations is the contemporary and popular understanding of the concept that the thesis subscribes to.

2.2.1.2. *Institutionalisation*

Arguably, the first thing that sets multiword units apart from literal strings is their status of being conventionalised, recognised and familiar phrases (Grant and Bauer, 2004). In sociolinguistic terms, institutionalisation describes the process ‘by which a standard construction becomes more current [...] and is reproduced as a memorised sequence because of its specific meaning or pragmatic function’ (Barkema, 1996: 135).

Due to its subjectivity, institutionalisation is difficult to determine and measure. Nunberg (1978), based on Lewis’s (1969) description of convention, attempts to formally describe different stages of conventionalisation amongst lexical items, ranging from stage I and II (‘free,’ ‘predictable’ uses) to progressively more conventionalised stages of III, IV and V, where idioms are argued to come in. Given a speaker S, who is aware of the regular use R of a word *w* to refer to *a*, and of general conformity to another regular use R’ of *w* to refer to *b*, the stages of conventionalisation are described as follows:

Stage I: S believes that it is normally believed that, given R’, R is licensed by normal beliefs (*newspaper*)

Stage II: S believes that it is normally believed that, given R’, R is licensed by beliefs to which members of the community would normally conform (*book price, juvenile*)

Stage III: S believes that it is normally believed that, given R’, R is licensed by beliefs that may be accounted normal by some members of the community, but to which members of the community would not normally conform (*nylon to refer to stockings*)

Stage IV: S believes that it is normally believed that, given R’, the beliefs that license R are not conformed to by any members of the community, though these beliefs might still be normally accessible (*make book*)

Stage V: S believes that it is normally believed that, given R’, there exist beliefs such that R would be licensed by them (*kick the bucket, storm the fort*)

Stage VI: S believes that it is normally believed that there is no regularity R' in the use of *w* such that R might be predicted (*bear, inside dope*)

(Nunberg, 1978: 116)

Due to the highly formal analyses involved in Nunberg's model, it is not always adopted as a research method, since native speaker intuition is a much more accessible method of determining institutionalisation. Granted that intuition can be biased or limited depending on the individual's knowledge and general familiarity with figurative language and multiword units, a reliable account of the state of institutionalisation of idioms can still be achieved if enough informants are used. In this thesis, dictionaries of idioms are consulted to assist my intuition about the institutionalisation of idioms, assuming that these dictionaries are reliable authorities on the subject matter since they have been compiled using native speaker's intuition and expert knowledge on the subject matter.

2.2.1.3. *Non-compositionality*

The term 'non-compositionality' is related to the *principle of compositionality* in linguistic processing, which states that:

The meaning of a sentence is determined by the meaning of its component parts and the manner in which they are arranged in syntactic structure (O'Grady et al., 1997: 260)

Idiom processing, however, is an exception to this principle. Idiomatic meanings cannot be derived by adding the literal meanings of individual constituents, hence the 'non-compositionality' of idioms. For example, the conventional, figurative interpretation of *shoot the breeze* (i.e. to chit-chat) cannot be determined through combining the literal meanings of its constituents ('shoot', 'breeze' in V+NP structure). The non-compositionality as such is argued in many frameworks to be the 'essential defining feature' (Cowie, 1988) or 'key criterion' (Grant and Bauer, 2004) of idioms, setting them apart from other types of multiword units (c.f. Sweet, 1899; Sinclair, 1991; Bolinger, 1981; Carter, 1998).

What sets idioms apart from most other fixed expressions is their "non-logical" nature, that is, the absence of any discernable relationship between their linguistic meanings and their idiomatic meanings (Glucksberg, 2001: 68)

The state of non-compositionality among idioms is thought to be the result of the gradual reduction in the meaningfulness of the component words within idioms due to the fixedness in form over time. In the extreme cases, meaning reduction leads to the state of 'petrification' (Cowie, 1988: 135), in which 'the evolved meaning of the whole is no longer traceable to the original meanings of the parts' (ibid) (e.g. *kick the bucket, shoot the breeze*). Some idioms, such as *spick and span* or *to and fro*, contain unique words which do not have their own literal meanings to enter into compositional analysis, in which case the state of petrification

is absolute. In other cases, however, partial meaning recovery might still be possible thanks to the contribution of a literal constituent within the idioms (e.g. *stop someone in their tracks*, meaning ‘stop someone abruptly, right where they are’).

Non-compositionality is sometimes found to be difficult to decide, considering that even ‘highly opaque idioms are often more transparent from a diachronic point of view’ (Skandera, 2004: 25). *Kick the bucket*, for instance, might stem from an old method of slaughtering pigs, which renders the idiom transparent through its metaphorical extension (Gibbs, 1998: 101). This etymology, nevertheless, is not widely accessible to the modern language user. The opacity or otherwise of this idiom, as a result, might depend on an individual’s subjective judgement and familiarity with its origin.

2.2.1.4. *Fixedness*

In earlier studies of idioms and multiword units, particularly before the availability of corpora, it was often argued that idioms in particular and multiword units in general were quite stable in form across much of their range of occurrence (Cowie, 1988). Occasionally, some idioms were found to tolerate variations, but these variations were regular and predictable by native speakers (ibid). Resistance to internal manipulation, as a result, was generally acknowledged in these studies to be characteristic of idioms.

In recent years, however, it is conceded that fixedness, or frozenness, of idioms is a matter of degree. That is, some idioms are almost entirely invariable whilst some others are more flexible. Increasingly, evidence about the flexibility of idioms in discourse has accumulated to refute the fallacious view of idioms as being fixed (e.g. Moon, 1998). It is suggested, as a result, that the stability of idioms is not to be assumed because it changes over time. This stability, consequently, is best not seen as lexical fixedness but *conceptual fixedness* instead (ibid). From a cognitive linguistic point of view, Langacker (1987: 25) also explains idioms as having *established configurations* instead of fixed lexical forms:

To regard an idiom as opaque or as primarily a fixed phrase is therefore simplistic. It is more accurately seen as a complex of semantic and symbolic relationships that have become conventionalized and have coalesced into an established configuration. We can plausibly suppose that this configuration of relationships might be recognized even when anaphora disrupts the normal shape of an idiom, or when its component words are split up and used in grammatical constructions such that they do not form a contiguous linear sequence: an idiom may be recognized as a unit that is to some degree independent of a specific overt morphemic arrangement, even if one such arrangement is far more similar and hence more ‘usual’ than the others.

A conceptual approach, therefore, is proposed to be capable of revealing more about the nature of fixedness than traditional models of semantic and syntactic variability have done (Hudson, 1998). Nevertheless, I argue that the nature of idiomatic fixedness, while apparently not purely lexical, is not purely conceptual, either. For example, while institutionalised variants of *in the nude* include *in the buff*, *in the nuddy*, *in the raw*, such alleged creative forms as **in the bare* or **in the naked* are not accepted (Moon, 1998) even though they are synonymous with the canonical expression. Similarly, *make a scene* ('create a loud/embarrassing situation') is analysable and its figurative meaning can be passivized (a loud/embarrassing situation was created), the idiom itself cannot take a passive variant (**a scene is made by two drunks outside the pub*) due to the 'unnaturalness' of the expression. As a result, idioms are argued in this thesis to be both conceptually, grammatically and lexically controlled.

A detailed analysis of the constraining factors of idioms that entail their state of fixedness as well as the facilitating factors that relieve the effects of such constraints to allow certain degrees of flexibility in idioms is given in chapter 6.

2.2.1.5. *Condition of use*

Fixedness, compositeness, institutionalisation, and non-compositionality as discussed above are often cited as the most important defining features of idioms (cf. Cowie and Mackin, 1975; Cowie et al., 1983; Carter and McCarthy, 1988; Sinclair et al., 1989; Fernando, 1996; Carter, 1998; O'Keeffe, McCarthy and Carter, 2007). However, a further discriminating criterion is proposed in this thesis for the exclusion of phrasal verbs (*to turn up*, *to come across*), conversational routines (*how do you do*), or proverbs (*a rolling stone gathers no moss*) from the class of idioms, although they fall within the boundaries set by the four aforementioned criteria. The difference between these types of phrase and idioms in the sense we use in this thesis lies in their 'condition of use' (Schmitt and Carter, 2004: 9), according to which idioms are those 'typically used to express a concept' (ibid). Proverbs, though very similar in most levels to idioms, typically 'state some commonly believed truth or advice' (ibid). Phrasal verbs and conversational routines are also excluded on the basis of this condition of use.

2.2.2. *Psycholinguistic features and phonological features*

Employing a corpus linguistic methodology, this thesis does not focus on psychological and phonological features of idiomaticity. However, a short review will be provided here to ensure a complete introduction of the approaches to idiomaticity.

In terms of psychological features, when a multiword unit is referred to by the terms '*formulaicity*' (Wray, 2002; Schmitt, 2004) or '*prefabricated chunks*' (Bolinger, 1975), it is implied

that the focus is placed on the *manner of storage* rather than the external observable formal and semantic characteristics. Wray (2002: 9) defines a formulaic sequence as

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.

The definition places the emphasis on two important defining features of multiword units, including 'sequences' and 'representation as wholes.' Based on the fact that they are (potentially) stored as wholes and retrieved as such during speech, formulaic sequences are thought to be identifiable via their 'phonological coherence' (Hickey, 1993; Wray, 2002; Read and Nation, 2004). The underlying assumption for this approach is that the difference between holistic and analytical processing does carry subtle phonological cues that help hearers differentiate formulaic and non-formulaic sequences during listening tasks (Van Lancker and Canter, 1981). Formulaic language, indeed, has been found to be processed faster (during reading tasks) than its literal equivalent, which was taken as the proof for its holistic storage and retrieval in our mental lexicon. If we accept this premise, i.e. they are retrieved whole from memory (or at least with minimal on-line processing), formulaic sequences are expected to be produced more 'fluently', potentially measured by the speed with which they are uttered, indicated by the number of words per unit of time or the number of words between pauses (Pawley and Syder, 2000). It is also found that lexicalised phrases such as idioms do not have pauses or other hesitation phenomena (drawls, repeats, false starts) in the middle of the phrase (Pawley, 1986; Goldman-Eisler, 1968; Raupach, 1984). Literal readings, on the other hand, take longer and contain more pauses.

In short, phonological features such as fluency, intonation patterns, stress, and speed of articulation, are all perceived as potential indicators of formulaicity in spoken language and are increasingly used to characterise spoken formulaicity.

Psychological and phonological features, although valuable in many studies, will not be used in this thesis, because I do not approach idioms from these angles. Instead, a definition of idioms based on the criteria of form and meaning will be formulated in the next section to guide our corpus-based investigation of idiomatic creativity.

2.2.3. A working definition of idiom

To ensure systematicity and consistency for the rest of the thesis, a working definition of idioms is needed to delineate them from other types of multiword units. With all five features (i.e. compositeness, institutionalisation, non-compositionality, fixedness and condition of use) being perceived as necessary but none sufficient on its own, the definition is formulated on the basis of the combination of these features as follows:

An idiom is an institutionalised word combination which is semantically non-compositional and whose syntactic, lexical and phonological form is to a greater and lesser degree fixed. An idiom expresses a concept.

It seems to be a contradiction that idiomaticity (characterised by relative fixedness and institutionalisation) and creativity (defined by novelty and newness) can be fused into one concept (i.e. 'idiomatic creativity'). However, idiomatic creativity, as will be seen in the next section, is an empirical fact. In what follows, 'idiomatic creativity' will be explained within existing theories of creativity and in relation to theories of idiomaticity to highlight its dual characteristics of being both 'creative' and 'idiomatic'.

2.3. Concept of creativity

The history of creativity study cannot be described as long, time-wise. Indeed, although a topic for discussion since the eighteenth century, creativity as a topic of study has only captured the public's attention and been actively pursued in the later half of the twentieth century after it appeared in the 1920s as an official academic subject within the fields of education and psychology (Pope, 2005: 19). However, achievement-wise, the field is not devoid of valuable research findings. During these past fifty years or so, research into the concept of creativity within psychology, socio-cultural studies and linguistics paradigms has been extremely fruitful, offering an ever-growing understanding of this fascinating phenomenon.

Studies into creativity from psychological perspectives are characterised by their converging interests in identifying 'the characteristic cast of mind and the properties of the brain associated with creative acts' (Carter, 2004: 31). Psychological studies into creativity seek answers to such questions as *How does a creative product happen? Which cognitive processes are involved in creative thinking? How to measure creativity? What are the biological and evolutionary bases of creativity?* and so on (Meyer, 1999). Key aspects associated with creativity in psychological research include:

- *Creativity equals novelty and originality*, i.e. breaking with conventions, presenting 'newness.' This perception of creativity is in full accord with the 'genplore' (generate and explore) model by Finke et al. (1992), which emphasises the two most important factors of creativity, i.e. the capability to generate new ideas, and the ability to see and develop the significance of those ideas (cf. Finke, 1995; Ward et al., 1999).
- *Creative abilities are displayed through problem-solving, divergent thinking.* The creative process is regarded as involving the formation of novel analogies or combinations between unassociated conceptual elements (Carter, 2004: 47) through which a solution or identification of problems is uncovered.

- *Creativity is associated with intelligence.* Creativity and intelligence are seen as sharing the key component of 'being able to combine ideas together in novel ways in abstraction from any immediate environmental stipulation' (Carruthers, 2002: 230).

It is important to remember, however, that although popular in today's creativity theories, the idea that creativity should mean originality or that it should break with conventions as discussed is not universal. In the Renaissance, creativity in art was considered to be the ability to imitate nature (Smith, 1961: 18). It was not until the late 19th century that the shift from representational to non-representational creativity was brought about as a response to the increasing mechanization of the society in which everything was mass produced. In many traditional cultures until today, especially the East, creativity means 're-visioning and remembering existing ideas' (Carter, 2004: 30), displaying 'the adherence to established norms and the emphasis on successful achievement through reproduction of those norms' (ibid). Such social-cultural aspects of creativity indeed remind devoted psychologists that socio-cultural perspectives are needed if a well-rounded description of the phenomenon is ever to be achieved, simply because creativity should not be seen as a universally scientific concept but a culturally and historically specific one as well.

Socio-cultural perspectives on creativity, therefore, rapidly gained popularity, tirelessly searching for influences of social groups and movements, historical and cultural characteristics on creativity, leading more researchers to acknowledge that 'culture and traditions strongly influence creative interests, the degree of creativity and the forms it can take' (Yusuf, 2007: 3). Various studies have in fact been carried out to examine the effects of a happy home, a healthy physical body, the deep understanding of the subject matter, social conditions, cultural influences, etc. on an individual's creativity (cf. Murnane and Levy, 1996; Autor et al, 2001; Hanushek, 2002, 2004; Glewwe et al, 1999; Bloom, 2000; Frey and Stutzer, 2002; Jones, 2005; Fleming and Marx, 2006). For example, whilst the creativity tendency in some societies might be geared towards systematic scientific enquiry grounded in formal rules of logic, of proof, and of the empirical validation of hypotheses, in some others it might be expressed more extensively through art, literature, craft, or music. Such differences are argued to be dependent on the specific culture's deep-rooted traditions, mentality and thinking patterns. Indeed, the differences between the East and the West are extensively investigated by such authors as Lubart (1999), Lau et al. (2004) or Landes (2006). Nevertheless, the results are often inclusive and refutable by subsequent studies, which further indicate the complex nature of creativity and its relationship with surrounding entities.

2.3.1. Linguistic creativity

The study of creativity in linguistics can be seen as having started, albeit indirectly, with studies of 'literariness' in poetry and literature in the early 20th century, considering that

creativity is a companion that is 'not easily separated from the nature of literariness in language' (Carter, 2004: 81). Early inherency models of literariness posit a distinction between 'literary/ poetic' language and 'ordinary language', whereby the *differentia specifica* of poetic language in relation to ordinary language is believed to be identifiable in the formal characteristics of the verbal signs themselves (e.g. metre, formulas). This approach is most famously adopted in the theories of 'defamiliarisation' (Shklovsky, 1917) or 'deviation' (Jacobson, 1960). In other words, literary language essentially shows formal departures from 'normal' rules or patterns of language.

Later on, social influences are observed to have consistently brought 'a cultural politics to literary studies' (Rice and Waugh, 1989: 4), forming a body of post-structural works that moved away from the speaker and the text while bringing the reader and social ideology to the fore. For instance, taking the role of the addressee into account, the 'reader response theory' claims 'it is possible to determine [a work's] artistic nature by the nature and degree of its effect on a given audience' (Jauss, 1970). This reader-oriented schema theory is later incorporated into Cook's (1994: 182) model of language play, which posits that literary texts typically carry out the function of challenging/altering existing schemata in the reader, which is possible via deviations at the linguistic-structural level.

Pragmatic approaches to creativity started to gain popularity since Austin's (1962) and Searle's (1969) Speech Act Theory. Ohmann (1971) calls literary speech acts 'quasi-speech acts' because, as he points out, they do not exist outside of the literary contexts. The criterion for literary discourse is, therefore, argued by Ohmann to be the 'lack of illocutionary forces' (ibid: 14) that would normally attach to the utterances. Pratt (1977), however, argues that non-literary categories such as hypothesis, pretence, fantasy, joking, etc. can also display this characteristic, effectively refuting Ohmann's arguments. Pratt herself tries to analyse literary language in terms of its violation and flouting of Austin's and Searle's felicity conditions and Grice's (1975) conversational maxims.

In more recent years, interests in creativity from cognitive semantics tradition also took a rise with various attempts at making a link between creative linguistic forms and the cognitive activities involved. Linguistic creativity is believed to be the result of the process of activating and coordinating of a set of conventional linguistic units, usually through *elaboration* or *manipulation* of the node of a construction in accordance with sanctioned constructional schemas and coding parameters (Langacker, 1987). For example, the computation of the novel construction *crayon sharpener* from the conventional *pencil sharpener* is nothing but an elaborative form of analogy (Langlotz, 2006: 88). This approach reminds us of what Stankiewicz (1960: 75) initially touched upon, that poetic language needs not violate any rules of language but rather materializes via '*manipulations* of available linguistic material and the *skilful utilization* of the possibilities inherent in the spoken language' (emphasis added).

With the availability of corpora, in recent years, it has been noticed that many of the criteria set out exclusively for literary language can be applied to ordinary language as well. Creativity/literariness is argued to be not exclusive to literary texts; it also exists outside of the literature realm. The poetic/ordinary language distinction, therefore, is brought into question and is criticised as unhelpful (Carter, 2004). In the latter half of the 20th century and the beginning of the 21st century, interests in creativity in everyday, non-literary texts grew stronger, particularly since the birth of electronic corpora. Findings from corpus-based creativity studies have further validated earlier observations of the ubiquity of creative language in everyday contexts. Such authors as Tannen (1989), Lecercle (1990), Norrick (1993, 2000), Cook (2000), Hall (2001) and Carter (2004), when looking at creativity in everyday conversations using naturally occurring data, highlight the poetic and creative language in everyday talk in the forms of metaphor, pun, riddle, verbal duelling, conventional joking and even gossip. The boundary between literary language and everyday language is thus 'very blurry and should be conceived as a gradation instead of an absolute division' (Carter 2004: 81). In other words, creativity is argued to be neither trivial, nor literature-exclusive. Creativity is inherent in everyday speech but is still special in certain ways. As Carter and McCarthy (2004: 83) put it, creativity is 'not simply a capacity of special people but a special capacity of all people'.

Pursuing the 'everyday' creativity in discourse, Carter (2004) proposes a theory of creativity with two components, including '*pattern-reforming*,' i.e. creativity by displacement of fixedness, reforming and reshaping patterns of language, and '*pattern-forming*,' i.e. creativity via conformity to language rules rather than breaking them, creating convergence, symmetry and greater mutuality between interlocutors:

There are clearly two levels of 'creative' interactions: first, more overt, presentational uses of language, open displays of metaphoric invention, punning, uses of idioms and departures from expected idiomatic formulation (*pattern-reforming*); second, less overt, maybe even subconscious and subliminal parallelisms, echoes and related matchings which regularly result in expressions of affective convergence, in signals of intimacy and in implicit symmetries of feeling (*pattern-forming*). The categories of pattern forming and pattern re-forming in speech are not purposeless non-pragmatic embellishments. They fulfill, as we have seen, fundamental communicative purposes (Carter, 2004: 109).

Pattern-reforming creativity, with its emphasis on the flouting of expectations and reshaping patterns, more or less corresponds to the orthodox view of creativity that involves newness, breaking of moulds and deviating from the norms of language to create schema refreshing effects in the reader. This type of creativity has been found predominantly (but not exclusively) in literary texts and written genres. Pattern-forming creativity, on the other hand, is typical of spoken genres in which interlocutors interact to create messages. An emphasis is also placed on the communicative functionality of the creative language to

distinguish it from erroneous expressions or slips of the tongue. It can be seen that Carter's model of creativity is comprehensive and encompassing the two most important aspects of creativity as it manifests in both literary texts and casual everyday discourse, which is also the definition of linguistic creativity that this thesis subscribes to.

2.3.2. *Idiomatic creativity*

The term 'idiomatic creativity' has been used in the literature to address the fact that many idioms are creatively reshaped and used in variable forms in specific contexts instead of their canonical forms. Examples of the reshaping/reforming of idioms include:

- Insertions: *getting our Christmas act together*
- Lexical substitutions: *burns the candle at five ends*
- Replacement with more formal/literary or euphemistic synonyms: *the certainty of sylvan ursal defecation*
- Reversion/inversion: *a sheep in wolf's clothing*
- Punning through rhyme and assonance: *trip/skip the light fantastic*
- Co-text: *final judgement must wait until the Government has explain why the £38m payment was simultaneously under the counter but above the board*

(Moon, 1998: 121-177)

As can be seen in these cases, and in other creative idiom manipulations, idiomatic creativity is predominantly pattern-reforming, featuring the displacement of the fixedness of idioms (Carter, 2004).

Sharing this view of idiomatic creativity as *pattern-reforming*, linguists have been treating the phenomenon as a 'unique refashioning of a fixed phrase' (Prodromou, 2007: 17), or 'the ingenious manipulation of idiomatic expressions normally taken as fixed' (He, 1989: 150). An emphasis is placed on the pragmatic purposes and the requirement for cultural or literary awareness to carry out such a procedure of 'refashioning' or 'manipulating' idioms. In a broader context, the strategy for creative play with chunks of language in general is to 'play with these larger patterns by deviating from what is expected, creatively disfiguring them to produce and play with perspectives and ways of seeing' (Carter, 2004: 95). The key concepts emerging from these definitions of idiomatic creativity are:

- (1) refashioning/manipulation/play/deviation/disfiguration
- (2) inventive/unique/ingenious/creative
- (3) pragmatic purposes

The first elements of the definition refer to the reforming/reshaping idioms, i.e. creating idiom variants via certain idiom manipulation strategies. The second elements relate to the 'creative' nature of the manipulation. Apparently, the fixedness of an idiom can be

disrupted by various lexico-grammatical or semantic operations, but only variants that are ‘creative’ can qualify as idiomatic creativity. The last elements, pragmatic purposes, are emphasised in this thesis to distinguish intentional creativity from unintentional variants/slips-of-the-tongue. That is, idiomatic creativity features intentional attempts to play with idioms to create elements of surprise, humour and/or subtle ways of seeing/perspectives. Glucksberg (1993) and Keysar and Bly (1999) even claim that it is the most important factor governing lexical flexibility in idioms. A valid idiom variant is only achieved when it ‘produces a comprehensible difference in interpretation’ compared to the original idiom meaning (Glucksberg, 1993: 21), so that a reasonable communicative intention/purpose can be inferred by the use of an idiom in some form other than the original.

In short, in this thesis, ‘idiomatic creativity’ refers to:

the linguistic phenomenon in which an idiom undergo transformations to form creative variants in specific communicative events to serve specific communicative purposes.

Such a definition entails the need for further elaboration of the concepts of ‘creative idiom variants’ and ‘communicative purposes’, which will be discussed in detail in the next sections.

2.3.3. Creative idiom variants

While the term ‘idiomatic creativity’ denotes the abstract concept, ‘creative idiom variant’ is used to refer to the concrete linguistic expression that embraces ‘idiomatic creativity’. In other words, an idiom variant is the effect of idiomatic creativity in linguistic form. To explain ‘creative idiom variants’, it is necessary to assume that for each idiom there is an *idiom base-form* (Barkema, 1996; Langlotz, 2006). Ideally, this term should suffice to denominate the *neutral*, *standard* and *canonical* form and usage of an idiom. However, Moon (1998) notices that a large number of FEIs, as found in her corpus, do not have a fixed standard form but boast a number of equally institutionalised variants (e.g. *whet/wet someone’s appetite*), which makes it difficult to determine their legitimate base-forms.

Acknowledging this indeterminate nature of the base-form approach in certain cases, for practical reasons, the current study still needs to rely on the concept of ‘canonical’ forms of idioms so that variants can be analysed accordingly. The base-form of an idiom in this study will be assumed to be the dictionary entries, taking for granted that these entries have been recorded according to their higher frequency and familiarity than other variants. In the cases of equally familiar/institutionalised variants, all of the variants will be considered base-forms.

With the term *base-form* denoting the standard usage of an idiom according to its neutral, canonical form, *idiom variant* can be defined as uses of idioms that deviate from this standard via different alteration processes. In other words, *idiom variant* or *variant of an idiom* refers to the idiomatic construction resulting from a given formal or semantic variation/alteration of the base-form (Langlotz, 2006).

Such a loose definition raises questions of how variants are created (manipulation strategies) and how far the deviations could go (constraints on idiom variants). These two aspects of idiom variants are investigated respectively in the next sections.

2.3.4. *Idiom variation strategies*

2.3.4.1. *Lexico-grammatical strategies*

Alteration strategies to create idiom variants have traditionally been described in lexico-grammatical terms. That is, the construction of variants is described in terms of the ways in which units are brought together to form grammatical sequences. The order of such units is also treated in terms of positional norms. For example, Fernando (1996) identifies four ways in which idioms can be manipulated, namely replacement/substitution (which can be either structural or lexical, e.g. *any red herring will do* or *burn one's boats/bridges*), additions (e.g. *with his tongue only partly in his cheeks*), permutations (i.e. rearrangements/conversion of idiom constituents, e.g. *ice-breaker*), and deletions (usually the reduced forms functioning as allusions to their original idioms, e.g. *waves trade carrot* as a short form of *dangle a carrot before the donkey*).

Moon's (1998) lexico-grammatical description of variation strategies discriminates between regular variations and exploitations. In her data, regular variations refer to the variations of an idiom that are equally familiar/institutionalised (e.g. *run rings around someone* vs. *run circles around someone*, *steal someone's thunder* vs. *steal the thunder from someone*, *in the nude* vs. *in the buff*, etc.) These variants have become recognised and accepted as equally canonical as the originals thanks to repeated use. Exploitations, on the other hand, are deliberate one-off instances of idiom manipulation for stylistic purposes, 'to provide some sort of defamiliarisation and typically providing humour' (ibid: 170). Exploitations in Moon's data are manifested in the forms of lexical insertions, lexical substitutions, replacement, reversion/inversion, punning through rhyme and assonance, and co-text (see examples in 2.3.2).

Langlotz (2006) distinguishes between systematic idiom variants and conspicuous wordplay according to the meanings of the variants created via one of the formal or semantic strategies (see table 3.1 below). Unlike Moon, however, Langlotz deliberately excludes creative wordplay from his framework of idiomatic creativity, focusing on systematic idiom variations instead. Such context-specific variants as *covers a multitude of chins* (shaving cream ad) and *Swindon were home and well almost dry* (context of bad weather during a ball game),

although highly creative, are exceptions to Langlotz’s cognitive-linguistic account of idiomatic variation on the basis that they are one-off uses that are not systematic enough.

Table 2.1. Types of idiom variation strategies		
Formal alterations	<i>Morpho-syntactic</i>	Inflectional variants of one (or several) idiom constituents, incl. verb inflection, noun inflection, flexible use of determiners and quantifiers <i>c.g. had to swallow some bitter pills</i>
	<i>Syntactic</i>	Changes in constructional organisation of base-form, incl. modifications, passivisation, clefting <i>c.g. all the Ardakkean beans have been spilled</i>
	<i>Lexical</i>	General alterations of the lexical constituents of an idiom, incl. synonymous or antonymous variants <i>c.g. a tough row to hoe</i>
Semantic alterations	<i>Polysemy</i>	Idioms which have a lexically invariant form but two conventional meanings <i>c.g. come a cropper = fall / fail</i>
	<i>Ambiguation</i>	Occasional meaning adaptation/playing with the semantic potential of idioms <i>c.g. if you like wine, have a good nose</i>
	<i>Meaning adaptations</i>	Modification combined with corresponding lexico-grammatical alterations, incl. intensification, antonymy, specification, perspectivisation <i>c.g. overturn the most ponderous applecart of all</i>

2.3.4.2. Conceptual strategies

In most cases, the changes on the lexical surface of idioms serve specific communicative purposes. With the purpose of capturing the ways an idiom can be adapted in response to the meanings and communicative function it is intended to encode, Langlotz (2006) discusses five variation strategies, including constructional adaptation, literal-scene manipulation, topic indication, topic-related literal-scene manipulation, and ambiguation. The first three types, i.e. constructional adaptation, topic indication, and literal scene manipulation, are argued by Langlotz to be principles of creating systematic variants. Constructional adaption includes operations such as verb inflection or passivisation. This group corresponds to ‘permutation’ as identified elsewhere in the literature (cf. Fernando, 1996; Moon, 1998). Literal-scene manipulation, on the other hand, reconstructs the literal scene in such a way that the consistency of literal scene is maintained, for example *the Chancellor had a narrow tightrope to walk*. The modifier is fully compatible with the literal scene, and, on idiomatic level, evokes a richer image than the original idiom (*a very difficult situation*)

in which the original idiom is elaborated and intensified. Last but not least, topic indication involves a topic indicator, usually via lexical substitution or modification, for example *it still left the Romanians treading a financial tightrope*. It can be seen that topic indication is characterised by inconsistency in literal meaning (*financial tightrope*) but fully consistent and grammatical in terms of idiomatic meaning (*financial problems*). All these three types of manipulations are argued to be systematic because they do not require incorporating contextual information into the processing of the meanings of new variants. They can be interpreted using linguistic knowledge of the meanings of the original idioms and the meanings and usage of the types of construction as dictated by the rules of the English grammar.

The last two types, i.e. topic-related literal-scene manipulation and ambiguation/punning are in the realm of wordplays because they involve multi-referentiality due to ambiguation. Topic related literal scene manipulation is similar to topic indication but instead of using a topic indicator, the speaker would adapt the literal scene to relate to the actual context of conceptualisation by ‘finding a linguistic unit that can be associated with both the usage-context and the conventional scene’ (Langlotz, 2006: 212) (e.g. *spill the pasta* in the context of Italian-mafia-gangster, Moon, 1998:51). The lexical modification/substitution in this case cannot be attributed a corresponding figurative meaning but has to be interpreted literally only, for instance *having had such fun pulling his cross-gartered leg for so long* (Ernst, 1981: 52). The modifiers in this case only provide additional information about the idiom’s context-specific referents. Langlotz, as a result, classes this type of variation as fully context-dependent and can be classified as wordplay due to the ambiguity involved. Last but not least, ambiguation/punning/wordplay is of the most creative nature, characterised by stylistic markedness, ambiguity and context-dependence. Ambiguation involves the creative evocation of multiple referentiality, available through such techniques as rhyme and resonance. Without the knowledge of the immediate contexts of utterance, the meanings of these variants would have been almost impossible to be inferred.

Summary

The implications of these descriptions of idiom variations are multifold. One, there is ‘massive evidence of the instability of the forms of fixed expressions and idioms’ (Moon, 1998: 309). Moon, as a result, urges more corpus studies into the classification and correlation of different kinds of lexical and syntactic variations of idioms. Indeed, the question that motivates this thesis is what facilitates such a state of flexibility amongst idioms, and if there are any constraints on this flexibility. These are the foci of the thesis, which will be explored via empirical analyses of corpus data in chapter 5 and 6.

Two, creative play with idioms can involve either formal, semantic or phonological alterations, or all of them. Amongst them, studies into the creative manipulations of form and meaning have been prevailing in the literature over creative sounds and rhythms,

partly due to the fact that corpora of spoken language have been less available and more difficult to compile. However, with the rapid developments of spoken corpora, it is expected that creativity at the phonetic and phonological levels will catch up with its formal and semantic counterparts. In this thesis, creative idiom variants are described both in terms of forms, semantics and phonological features. Admittedly, phonological features used for analyses in this thesis are those available in written texts and transcripts of spoken texts (phonemes, morphemes, etc.) but not the acoustic features of speech (i.e. intonation, stress, etc.), although it is envisaged that they will soon be popular topics of research in the field of idiomatic creativity when appropriate tools become available.

Three, formal alterations to the idiom base-forms are obvious, involving overt linguistic tools, while semantic and phonological variants are more subtle. Modifications to the meanings and sounds of an idiom can be introduced with or without overt lexico-grammatical alterations. When creative uses of idioms are created without lexico-grammatical alterations, for example the creative play with the ambiguity of sight-sound relations (polysemy, homograph, homonym) or with extended co-texts (e.g. *it's not raining cats and dogs, it's raining birds – but why?*), the idiom base-forms are not disrupted. However, the meanings of these variants are obviously exploited for puns. As a result, the definition of idiomatic creativity in this thesis is not restricted to overt idiomatic variants (deviating from base-forms) but extends to other non-deviating uses of idioms provided that they serve special communicative purposes.

Four, the highly creative form of idiomatic creativity, wordplay, characterised by stylistic markedness, ambiguity and context-dependence, has traditionally been marginalised because they do not fit easily in a theoretical framework:

Methodologically, the ability of people to play with words is outside the scope of a theory of idioms proper; therefore, data involving word games cannot play a role in a theory of idioms. (Schenk, 1995: 258)

This thesis aims to change this tradition by incorporating context-specific puns/wordplay into its data pool so that a comprehensive picture of idiomatic creativity can be produced. More importantly, wordplay represents the essence of creativity at its best, triggering interesting change of perspective and displaying a playful and surprising form of communication. In this thesis, it is the systematic, predictable and uncreative variants that are excluded from analyses.

2.3.5. *Communicative purposes*

The *communicative purpose*, or *functionality*, of a creative expression is emphasized as one of the requirements of linguistic creativity in general (Carter, 2004; Stockwell, 2006), which distinguishes intentional creative language play from slips-of-the-tongue. The fact that

speakers/writers so often recruit creative words, phrases or expressions instead of the neutral equivalents suggests a special quality or certain advantages creativity has over routine language in terms of discursal effects. Amongst them, wit and humour are found to be the most frequent functions of creativity in language (Carter, 2004). The connection between *creativity* and *verbal play* is scrutinized by Chiaro (1992), Crystal (1998), Dunbar (1996), and Cook (2000), whose work is built around the *theories of play* by such psychologists as Bühler (1918) and Huizinga (1939) (see Berlyne, 1969). Verbal play, in these studies, is seen as the denominator of creativity in spoken language. It is worth noticing that many authors, taking this direction, proceed to analyse the connection between *creativity*, *verbal play*, and *humour* (Berlyne, 1969; Nash, 1985; Chiaro, 1992; Berger, 1998; Ross, 1998). From a psychological point of view, there is a strong relationship between creativity and humour to the point that individuals who are found to be creative are 'markedly addicted to playful and imaginative jocularity' (Getzels and Jackson, 1962, quoted in Berlyne, 1969: 812).

Thanks to its property of being novel/innovative and its humour function, creative language, as is used so generously in headlines and other media texts, is an 'attention-grabber' (Redfern, 1984) or 'attention-seeking device' (Lipka, 2000). As Howarth (1998: 29) explains it, when a language use is not as expected, there is a need to refer to the context, which slows down the processing speed. This, he believes, explains why deviation is such an effective tool in discourse when intended.

As well as grabbing attention, creative language is also seen as a device to engage and amuse/entertain readers because it provides defamiliarisation, playfulness and irony (Renouf, 2007; Rua, 2007). Cook (2000) particularly emphasises the social functions of language play amongst both children and adults - it creates bonds between group members who share it and excludes others. In other words, it reflects, creates and maintains a sense of group identity.

Focusing on the effects of creativity on the author/writer, Partington (1998: 124) argues that authors/writers use creative language to 'make highly personalised and context-dependent comments' and to 'impress the reader with the author's ingenuity' (ibid: 142). In her not so dissimilar remark, Renouf (2007: 70) reminds us that creativity is a device 'to convey [...] a sense of authorial learnedness, sophistication, distancing'. On the part of the reader, creative language appeals to them for the pleasure of solving the puzzle after having decoded the novel creation (Veisberg, 2007: 241). This is what Partington (1998) calls the 'collusion' between the paper and the readers, triggering a 'smugness effect' whereby the readers feel flattered because they are intelligent enough to appreciate the wordplay.

Proposing that creativity embraces two components, i.e. pattern-forming and pattern-reforming (see 2.3.1), Carter (2004) distinguishes different functions associated with each type. While pattern-forming creativity aims to create affective convergence and commonality of viewpoint, pattern-reforming creativity draws attention to itself by:

offering some new way of seeing the content of the message; making humorous remarks; underlining what is communicated; expressing a particular attitude, including negative and adversarial attitudes; [...] playing with language form to entertain others (Carter, 2004: 148).

Idiomatic creativity, as a subtype of creative language, is a special tool that creates powerful communicative effects. Idiomatic creativity thus deserves a prominent place in creativity studies in particular and in linguistic research in general.

2.4. Recap of a working framework for idiomatic creativity

From the foregoing discussions of creativity and idiomaticity in the literature, it is now necessary to recap our position on the phenomenon of idiomatic creativity. Lying at the idiomaticity/creativity interface, idiomatic creativity demonstrates a host of interesting characteristics. Idiomatic creativity represents the ‘pattern-reforming’ (Carter, 2004) type of creativity. That is, a use of language is considered creative when it is *new and unexpected* in nature as the result of ‘deliberate and conscious choice to violate normal expectancies of language use’ (Tarone, 2002: 293) or of the ‘bending and breaking the rules of the language’ (Crystal, 1998: 1). ‘Pattern’ as is used in Carter’s definition of creativity embraces the broadest sense of the word, i.e. expected, familiar sound, meaning and usage of a particular linguistic unit, ranging from word to larger chunks. In idioms, the notion ‘pattern’ refers to the familiar arrangement of a particular idiom, including its sound structure, lexical constituents, meanings, usage/distribution, concepts, etc. as recognized by language users and acknowledged as dictionary entries. Idiomatic creativity is understood as any deviation, ‘reforming’, or ‘reshaping’ of such a pattern of an idiom, achieved when an idiom is either formally, semantically, phonologically or pragmatically altered (or all of these together) to create variants to serve specific communicative purposes. Communicability, therefore, imposes a strong constraint on idiom variants, requiring them to be grammatical, semantically compatible and recognizable so that the variants’ meanings and purposes can be communicated. These components of communicability as cognitive constraining factors are reviewed and discussed in chapter 3.

Creativity as a result of pattern re-forming can operate at different linguistic ranks, the three encompassing levels being form, semantics and pragmatics (Cook, 2000), plus wordplay on phonological level. At the *formal level*, creativity can be made available through play with language rules (in this case, the rules of the English language at the graphology, morphology, and syntax levels.) Changes in the formal patterns of an idiom are generally overtly perceivable and readily definable. Changes in other patterns (semantic, pragmatic or phonological) are generally covert and more subtle but by no means less creative. At the *semantic level*, a creative use of language can present itself in the form of newness and unexpectedness of *meaning*. It is necessary to put an emphasis here on the detachment of semantic meanings from context, i.e., we are concerned with the literal meaning of the

sentence in its own right. At the *pragmatic level*, creativity can be attempted through play with *discourse functions* and *context*. When there is a deliberate conflict either between language use and its context, including both context of situation and socio-cultural context, or between language form and its illocutionary forces, creativity might be present. Last but not least, play with the sound characteristics of idioms, at the *phonological level*, particularly to create context-specific puns/wordplay, are highly creative and humorous and will be paid due attention to.

Idiomatic creativity has been suggested to be highly context-specific. Indeed, Carter (2004: 108), using the CANCODE¹, reveals that creativity in spoken discourse is 'almost always contextually embedded in so far as it depends on the social relations which obtain between participants'. However, idiomatic creativity in contexts has not yet been fully explored in existing frameworks. In this thesis, the ultimate goal is to develop a context-based model of idiomatic creativity. In chapter 3, the theoretical background of context is fully reviewed, paving the way for the qualitative analyses of the creative behaviour of idioms in discourse in relation to context (chapter 6).

Since the kind of 'patterns' that idioms possess present the strongest state of institutionalization as compared to other types of patterns (such as semantic prosody, non-idiomatic lexical item), reforming/reshaping them is more difficult and subject to many constraints. Indeed, while creative uses of idioms share 'innovative' features, they are also heavily constrained by features of 'idiomaticity'. For instance, institutionalisation means resistance against pattern-breaking to certain degrees, while non-compositionality means difficulty with formulating new patterns (**kick the blue bucket*). The dual effects of both creativity and idiomaticity on idiomatic creativity make it one of the most fascinating linguistic phenomena to observe and study.

2.5. Conclusions

Chapter 2 has been intended to be purely theoretical, dedicated to the explanations of key terms and concepts and related issues to delineate the theoretical stance that this thesis subscribes to. In the course of this chapter, I have discussed idioms and their defining characteristics in relation to other classes of multiword units within the field of formulaic language, thereby establishing a working boundary between idioms and other multiword units that this thesis adheres to. Being non-compositional and institutionalised word combinations, idioms were traditionally assumed to be fixed and to behave like long words. Some idioms are indeed found to be relatively restricted in the range of alterations they can undergo, if at all. However, evidence from various sources of both written and spoken discourse has shown that many idioms do allow certain degrees of variability via

¹ CANCODE stands for Cambridge and Nottingham Corpus of Discourse in English, a 5 million word corpus of spoken English. See [http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646595/Cambridge-English-Corpus-Cambridge-and-Nottingham-Corpus-of-Discourse-in-English-\(CANCODE\)/?site_locale=en_GB](http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646595/Cambridge-English-Corpus-Cambridge-and-Nottingham-Corpus-of-Discourse-in-English-(CANCODE)/?site_locale=en_GB)

constructional, semantic and pragmatic manipulations to form creative variants that serve special communicative purposes in discourse. Idioms, therefore, are not only characterised by their fixedness to a greater or lesser degree but also by their potential for creativity to a greater or lesser degree. Theories of creativity in general and linguistic creativity in particular have also been reviewed in this chapter to provide the theoretical background for a more fine-grained account of the phenomenon of interest in this thesis, i.e. idiomatic creativity. A working definition and a framework for analysis of idiomatic creativity are devised at the end of the chapter to guide our subsequent analyses of the subject matter.

In the next chapter, more in-depth reviews and discussions of the current theories and hypotheses concerning idiomatic creativity are presented. The particular questions that I concern myself with, and attempt to find answers to, relate to the underlying factors of the differing degrees of variability amongst idioms:

- (1) What facilitates creativity in some idioms and what hinders it in some others?
- (2) Are there limits to how far idiomatic creativity can go?
- (3) How much does context influence idiomatic creativity?

These three questions have been addressed in the literature to certain degrees, but not to satisfactory lengths in my opinion, particularly that relating to the role of context. Chapter 3, as a result, not only reviews the existing theories concerning the facilitating and constraining factors of idiomatic creativity, but also identifies the gaps that this thesis strives to bridge.

CHAPTER 3

THEORIES OF THE DETERMINANTS OF IDIOMATIC CREATIVITY

Before the availability of large bodies of data, particularly electronic corpora, the variability of idioms had largely been underestimated. Idioms had orthodoxly been treated as relatively fixed with minor exceptions that did not deserve to be accounted for. Increasingly, however, the creative behaviour of idioms in discourse has emerged as systematic and rule-governed to certain degrees. Semantic characteristics have often been cited in the literature as one of the most important determinants of idiomatic creativity. Suggestions about the presence of other factors, including cognitive constraints and context, on the other hand, have been tentative. In this chapter, the different theories and hypotheses of the underlying factors of idiomatic creativity are reviewed in detail to pinpoint the place of the thesis in the current debates on the topic.

3.1. Orthodox view of idioms and idiomatic creativity

The fact that idioms are institutionalised, semantically non-compositional, and formally fixed has led to the orthodox view of idioms as being long words which are stored and retrieved as wholes from the mental lexicon (Weinreich, 1969; Fraser, 1970; Katz, 1973), unwittingly causing a reluctance amongst linguists to see the importance of analysing idioms:

The relation between idiomatic and literal meanings is so unsystematic as to deserve no place in the theory (Weinreich, 1969: 76)

Each [collocation] must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts (Palmer, 1933: 4)

In viewing idioms as long words, it is assumed that speakers acquire the figurative meanings of idioms by 'forming arbitrary links' between the figurative and literal meanings of idioms and remembering them 'in a rote manner' or sometimes simply 'inferring the meanings' from context (Gibbs, 1995: 98), so that no analysis of individual component words or grammatical features is required during production and comprehension.

The view of idioms as long words as such rests on number of psycholinguistic studies which ratify, as Glucksberg (1993) calls them, the 'direct look-up models' of idiom processing, according to which idioms are stored as prefabricated lexical units and retrieved directly as wholes from our mental lexicon. However, since idioms have both literal and figurative

meanings, the debates within idioms-as-long-words tradition revolve around the question of how and in what order these meanings are retrieved and processed. Among them, Bobrow and Bell (1973) propose the *idiom-list hypothesis*, which postulates that a potential idiomatic construction is first interpreted literally. If the literal interpretation fits the context, the comprehension process is successful and complete. Otherwise, idiomatic meaning is activated by retrieving it from the mentally represented idiom-list through direct look-up. The idiom-list hypothesis, thus, represents the 'literal meaning first' approach to idiom processing.

Psychological measurements of processing time, however, suggest the opposite route, i.e. 'idiomatic meaning first'. It is found that a linguistic construction, when used idiomatically, requires less processing time than when it is used literally (Ortony et al., 1978; Swinney and Cutler, 1979). The *direct-access hypothesis* (Gibbs, 1980, 1985, 1986), as a result, is formulated, claiming that the idiomatic meaning of an idiom may be directly retrieved from memory prior to the complete processing of literal meaning.

As a compromise between the literal-first and the figurative-first hypotheses, the *lexical representation hypothesis* (Swinney and Cutler, 1979) assumes that the literal and idiomatic meaning are first processed in parallel until the idiomatic meaning is confirmed, at which point the literal meaning is suppressed. In a similar manner, the *configuration hypothesis* (Cacciari and Tabossi, 1988; Tabossi and Zardoni, 1993, 1995) claims that language users try to interpret an idiomatic string according to its literal meaning until the idiom 'key' is recognised, which marks the switch from the literal to the figurative interpretation of an idiom. This hypothesis is based on the assumption that every idiom contains one or more lexical 'keys', which are mental signals that trigger the activation of the idiomatic meaning of a string.

In either of these hypotheses, the figurative meaning of an idiom is believed to be directly retrieved from a list from memory instead of being subject to any type of analysis of its constituents' individual meanings.

Criticisms, however, are levied at the non-compositional view of idioms for its lack of explanatory power with regard to idiom variants, such as adjectival modification (e.g. *leave no legal stone unturned*), quantification (e.g. *touch a couple of nerves*), topicalisation (e.g. *those strings, he wouldn't pull for you*), ellipsis (e.g. *my goose is cooked, but yours isn't*), anaphora (e.g. *we thought tabs were being kept on us, but they weren't*), or passivisation (e.g. *the beans were spilled by John*) (examples by Nunberg et al, 1994: 500-503). According to the authors, these variants can only be explained if parts of idioms be assigned their own interpretations. For example, the verb *spill* in the idiom *spill the beans*, through frequent evocation in the idiomatic context, acquires the idiom-specific figurative meaning of *reveal*, while *the bean* carries the phrase-induced meaning of *information*. As a result, such a variation as *the bean has been spilled* is possible, and so is *all the Ardakkean beans have been spilled*. In order for these constituent

modifications to be possible and for them to be readily understood by both the speaker/writer and the listener/reader, the modified constituents in these cases are argued to possess some autonomous semantic values. This speaks against the non-compositionality view of idioms. Similarly, Langlotz (2006: 24-25) further cites examples of postmodification (*the bull of controversy was grasped firmly by the horns*), the use of determiner and proform (*the Government is willing to grasp nettles that others have shied away from, and I have no doubt we shall grasp this one if we have to*) or comparison (*this way of thinking has to be one of the blinder alleys that we have been led up by psychoanalysis*). In the same line as Nunberg et al. (1994), Langlotz argues that these idiom variations are only possible because the modified constituent in each of case carries an individual figurative meaning.

The non-compositional view of idioms, as a result, is criticised by these linguists as having negative consequences with regard to the explanation of idiom variability. There is a dire need for a new theory of idioms.

3.2. Compositional view of idioms

The linguistic community today has almost unanimously adopted the compositionality view of idioms thanks to the increasing bodies of linguistic and psycholinguistic evidence being accumulated about the possibility to separate components of the figurative meanings of certain idioms via top-down analysis (Nunberg, 1978; Nunberg et al., 1994). While idioms are still considered non-compositional from bottom-up analysis, the figurative meanings of idioms, once known, can be devolved on the constituents of the expressions via top-down semantic re-analysis.

Nunberg et al. (1994), therefore, distinguish between *idiomatic phrases* (non-compositional idioms whose meanings cannot be devolved on their constituents) and *idiomatically combining expressions* (semantically compositional idioms, analysable through top-down analysis, although these might still be unanalysable via bottom-up analysis). An idiomatically combining expression consists of constituents of individual figurative meanings but these meanings are not autonomous; the constituents have to be co-selected within the conventionalised idiomatic context. For example, the figurative meaning of the idiom *pull strings* will not emerge unless they appear together in a specific context that rules out the literal reading of the phrase.

This distinction between idiomatic phrases and idiomatic combining expressions by Nunberg et al. is useful in offering explanations for the syntactic modifications of some idioms. Because they are seen as having internal structures, idiomatically combining expressions can be internally manipulated. Without internal structures, idiomatic phrases like *kick the bucket* do not allow such transformative processes, since the elements *kick* and *bucket* do not have separately assignable meanings.

For analytical purposes, in this thesis, idioms are taken to be potentially decomposable to different degrees. Some idioms are non-decomposable, but some others can be decomposable to certain degrees (fully or partially), so that one or more of their constituents can carry individual figurative meanings. Thanks to this, some idioms can be seen as more transparent than others. The concepts of 'transparency', 'compositionality', 'decomposability' up to this point have intentionally been left vague. In what follows, a more systematic account of the definitions and characteristics of idiomatic transparency as a result of compositionality and other internal characteristics will be provided.

3.3. Idiomatic transparency

There is a certain degree of overlap in the usage of the terms 'transparency', 'compositionality', and 'decomposability' in the field, which will be summarized here in order to indicate some usages of terminology that are divergent from the usage in this thesis. In many studies, 'transparency' is used in the dichotomy with opacity and thus derives its rather simplistic and vague definition from this relationship. That is, idioms are considered transparent if they are literal, i.e. they 'can be interpreted on the basis of their parts' (Fernando, 1996: 60). Idioms can, thus, be ranked on a 'scale of non-literality' from pure idioms, which are non-literal (e.g. *red herring*, *spick and span*), to literal idioms (e.g. *on foot*, *very important person (VIP)*, *opt in favour of*, or *abstain from*) and anything in between.

Other linguists, however, maintain that bottom-up non-literality is essentially one of the defining characteristics of idioms. That is, the meanings of idioms cannot be predicted by adding up the meanings of the constituents (Nunberg et al., 1994; Glucksberg, 2001). Fernando's class of literal idiom as described above is often excluded from the category of idioms. The 'transparency' of an idiom, according to this second school of thought, does not involve bottom-up predictability but comes from top-down re-analysis of the idiom's parts once the meaning is known. Transparency as such is synonymous with analysability as used elsewhere in the literature, a property where 'an idiom has separate, meaningful units contributing to its overall figurative interpretation' (Gibbs and Nayak, 1989: 127).

Transparency in some other studies is understood as metaphorical transparency, which, together with compositionality, makes up the analysability of idioms (Gibbs and Nayak, 1989; Nunberg, 1978; Nunberg et al., 1994). Compositionality in these studies refers to the possibility of breaking an idiom's form down into components which correspond to components of its figurative meaning, although it might not be possible to explain why these components acquire the meanings they do. For example, in the idiom *spill the beans* (meaning 'to divulge information'), *beans* can be understood as 'the information' that is divulged, but it is not clear why *beans* is chosen to mean information. (Note that this definition fits the description of decomposability or analysability elsewhere in the literature). Transparency, on the other hand, refers to the extent to which the literal meaning bears discernible relations to its figurative meaning (Gibbs and Nayak, 1989). For example, with a

metaphorically transparent idiom, there is an overall metaphor that links the literal and the figurative meaning to explain why the idiom means what it does, even though it might not be possible to break the idiom down into meaning components. *Flip your lid*, for instance, can be understood as 'be angry' according to metaphor ANGER IS HEATED FLUID IN A CONTAINER, but it is impossible to decide which part of the idiom actually means angry. Metaphorical transparency as such corresponds to the notion of *motivation* in cognitive-semantic frameworks of idiom variations. Besides metaphor, metonymy, blending, or emblem and knowledge of the world are often cited as other sources of transparency in idioms. The transparency of *saw logs* (meaning to sleep), for instance, is dependent on a person's knowledge of the activity of sawing logs and the resemblance between the sounds produced by the two activities, although there is no decomposition of *saw logs* into components that correspond to the meaning 'to sleep.' With an opaque idiom, the literal and figurative meanings have little to do with each other or can only be seen as connected from a historical point of view that has already become obsolete to modern users. For example, there is no link between the literal and figurative meaning of *kick the bucket*, which is why it is often considered opaque. Although it is sometimes argued that this idiom can be seen as transparent because it is based on an old method of slaughtering pigs, this etymology is largely unavailable to the normal language user.

To further complicate the terminology scenario, Langlotz (2006), on the basis of Geeraerts's (1995) prismatic model of idiom semantics, treats transparency as an umbrella term with four components, including literal compositionality, global motivation, constituentual motivation and figurative-literal analysability. Literal compositionality refers to the well-formedness of the literal forms of idioms, such that *open a can of worms* is well-formed because the phrase complies with the rules of the English grammar, whereas *trip the light fantastic* is not. Global motivation is used to specifically refer to the existence of a metaphorical relationship between the literal and figurative meaning of an idiom but does not entail a one-to-one correspondence between the literal and figurative meaning components (e.g. *flip your lid*). Constituentual motivation, on the other hand, is achieved when constituents of an idiom have semantically autonomous/lexicalised figurative senses that can appear outside of the context of the idiom (e.g. the constituent 'swallow' in *swallow the bitter pill*) or phrased-induced figurative meanings (e.g. 'beans' in *spill the beans*). Last but not least, analysability refers to the top-down analysability of an idiom, achieved via the recognition of direct correspondences between components of the literal meaning and those of the figurative meaning, so that the idiomatic meaning can be devolved on the literal constituents (e.g. *rock the boat* meaning 'spoil a comfortable situation'). Taking all three dimensions (literal compositionality, motivation, analysability) of idiom characteristics into consideration, Langlotz proposes that idioms with all three properties are most transparent, as opposed to opaque ones, which have none.

In sum, different linguists have used different terminologies to describe the internal structuring patterns of idioms. Despite the different names that linguists might have given these phenomena, they stand out as independent aspects of idiomaticity and they are also characteristically distinguished from each other, each addressing one separate facet of the internal structure of idioms, be it formal, semantic or conceptual. The adoption of different terms to call these phenomena does not change their natures, nor does it affect our understanding of idioms and its variable behaviour. In this thesis, ‘transparency’ is used in a similar manner to that in Langlotz’s framework; that is, ‘transparency’ is the encompassing, umbrella term for the openness of idioms to analysis and variation. Transparency is made up of analysability, motivation and literal well-formedness. It is the differences of idiomatic transparency that trigger different forms of idiomatic usage and variability (Langlotz, 2006: 128). The three phenomena of literal well-formedness, motivation, and analysability and the existing theories concerning their contributions to idiomatic creativity in discourse will be discussed in greater detail in the next sections.

3.4. Literal well-formedness and its role in idiomatic creativity

Literal well-formedness refers to the grammaticality of the literal meaning of an idiom. Ill-formed idioms have been identified in the literature to be either syntagmatically or paradigmatically aberrant. That is, their literal meanings cannot be decoded purely compositionally nor encoded freely because their structures break the conventional grammatical rules of English. Čermák (2007) sketches the morphological nature of *irregularly-composed idioms* as following these patterns:

- Idioms contain special unique constituents that do not materialise elsewhere, making up the frozen parts of idioms with highly restricted use: *spick and span*
- Idioms containing unique word combinations not materializing elsewhere, hence appearing as a whole in the idioms only: *go haywire*, *at logger-head*
- Idioms containing constituents made of single morphemes, whose combinatorial capacity outside of the idioms is almost nil: *take umbrage*, *in high dudgeon*, *to and fro*
- Idioms containing constituents which are manifestations of an earlier state of morphology no longer in use in today’s language.

Literal well-formedness can also be the result of semantic incompatibility (Gibbs and Nayak, 1989), so that *crack a joke*, *foot the bill*, *swallow one’s pride* are ill-formed, but *spills the beans*, *kick the bucket* are literally well-formed.

Moon’s (1998) literal well-formed idioms include anomalies on one or more planes (syntax, lexis, or semantics): cranberry morph (e.g. *at someone’s beck and call*), syntax (e.g. *dog eat dog*), word class (e.g. *do the dirty on someone*), violating truth conditions (e.g. *look daggers at someone*), lack of co-referentiality by definite article (e.g. *spill the beans*).

The grammaticality of the literal meaning of idioms is a frequently referenced factor of idiomatic creativity. In Fillmore et al.'s (1988) view, it is because extra-grammatical idioms do not follow normal syntactic patterns (e.g. *all of a sudden*, *by and large*) that they are unproductive. Echoing this view, Moon (1998) did not find in her corpus any cranberry FEIs (fixed expressions and idioms) which undergo transformations such as passives, except the inflection of component verbs. This, according to her, is due to the fact that many of these idioms are unanalysable, which makes it impossible to analogise and perform transformations. Generally, it is argued that the more idiosyncratic a phrase, the more fixed it is, which explains why grammatical idioms are more open to variations than extra-grammatical ones (Fillmore et al., 1988).

On the contrary, in the study by Gibbs and Nayak (1989), ill-formed idioms are actually found to be more productive than well-formed idioms, which contradicts Fillmore et al.'s and Moon's findings. The authors argue that variants of well-formed idioms should be more difficult to be accepted because they are ambiguous – they have perfectly eligible literal readings alongside the figurative ones, which could confuse language users as to which readings should be chosen. Also, some well-formed idioms are opaque and unmotivated (*kick the bucket*), so it is difficult to use them in any variable form. However, Gibbs and Nayak themselves admit that their results are limited to verb phrase idioms and acknowledge the difficulty with which other types of idioms can be perceived as such. Besides, the idioms used in their study (e.g. *crack a joke*, *swallow one's pride*) are all semantically incompatible idioms, which differ significantly from the group of extra-grammatical or anomalous idioms that Fillmore et al. and Moon used in their respective studies. The differences in the definitions of 'ill-formedness' between these studies are potentially the reasons why their results are different.

In short, research in the field has produced contradictory findings with regard to the role of literal well-formedness in idiomatic creativity. As a result, in chapter 5, corpus-based analyses of the relationship between the two features will be carried out to provide an empirical evaluation of such a relationship.

3.5. Motivation and its role in idiomatic creativity

3.5.1. The concept of motivation

Abstracting away from lexico-grammatical and semantic/pragmatic planes to analyse idioms in conceptual terms, cognitive linguists suggest an idiom be regarded not just as an expression that has meaning that is somehow special in relation to the meanings of its constituent parts, but as something that 'arises from our more general knowledge of the world,' which is embodied in our conceptual system (Kosevecz and Szabo, 1996: 330). Idioms (or at least the majority of them) are conceptual, and not linguistic, in nature.

Proceeding from this assumption, cognitive linguists argue that it is possible to rely on knowledge of the world to make sense of idioms to certain extents. The meanings of idioms can, therefore, be seen as *motivated* and not arbitrary (Lakoff, 1987, 1990, 1993). *Motivation* in this sense refers to:

a speaker's ability to make sense of an idiomatic expression by reactivating or remotivating their figurativity, i.e. to understand why the idiom has the idiomatic meaning it has with a view to its literal meaning (Langlotz, 2006: 45).

The important premise underlying the notion of motivation is that the figurative meanings of many idioms are systematically shaped, which means that rules can be teased out with regard to the ways they are shaped and how these patterns systematically govern the behaviour of idioms in discourse. In other words, idioms can be motivated and the degrees of motivation among idioms enable them to be manipulated in discourse to varying degrees.

The sources of motivation (i.e. the tools available to language users to understand why idioms mean what they do) are proposed to be conceptual instead of linguistic, including a host of conceptual structures, from pre-conceptual knowledge (image-schemas) to more complex structures (achieved via conceptual metaphors, conceptual metonymy, etc.). To understand what 'pre-conceptual knowledge' and 'complex conceptual structure' mean, it is important to understand the basic concepts in cognitive linguistics and how they are perceived.

3.5.2. Cognitive linguistics – An overview

The 'cognitive' part of the concept of 'cognitive linguistics' highlights an important commitment of this school of thought; that is, the commitment to providing a characterization of language that accords with what is known about the mind and brain from other disciplines, such as psychology, artificial intelligence, cognitive neuroscience and philosophy. According to Evans and Green (2006), it is this commitment that makes cognitive linguistics cognitive. It entails that linguistic theories cannot that violate known properties of the human cognitive system. Cognitive linguistics, thus, is the branch of linguistics that investigates language on the premise that 'language is based on cognition' (Lakoff, 1987: 291).

More specifically, cognitive linguistics is a school of thought and practice that is concerned with investigating the relationship between human language, the mind and socio-physical experience (Evans and Green, 2006). As a result, besides proposing that language is made meaningful because it is directly tied to meaningful thoughts, cognitive linguists also claim that thoughts are only meaningful because they are directly connected to the world (i.e. linguistic experience is embodied). This is derived from the central assumption that meaning in general is only constructed when cognitive agents interact with their physical

and social world of experience (Lakoff, 1987; Langacker, 1987; Lakoff and Johnson, 1999). Our experience of familiar elements in our environment forms a body of *preconceptual knowledge* that exists prior to conceptual structures and provides the existential grounds for meanings to be constructed. As a result, a full description of language is one that takes into account the relationship between language and the human brain's capacity of processing meaning and the construction of meaning in relation to our embodied experience of the world.

Cognitive linguistics is often referred to as an enterprise due to its diverse range of complementary, overlapping (and sometimes competing) theories (Evans and Green, 2006), although these theories still share the same guiding principles, particularly the cognitive commitment as discussed earlier. (For a detailed discussion of the principles and other aspects of cognitive linguistics, see Evans and Green, 2006). Directly relating to the study of idioms and idiomatic creativity, however, are the theory of image-schemas and the theory of conceptual metaphor, which will be reviewed hereafter.

3.5.3. Image-schemas theory

The theory of image-schemas was first developed by Johnson (1987) on the basis of the embodiment of our experience. It is observed that in our knowledge of the world, there are rudimentary concepts like CONTACT, CONTAINER and BALANCE, which are meaningful because they derive from and are linked to human pre-conceptual experience (Evans & Green, 2006). These image-schematic concepts arise directly from the sensory-perceptual experiences of human beings in their environment. For instance, the CONTAINER image schema is meaningful because containers are meaningful in our everyday experience. As such, image-schemas are abstract kinesthetic patterns that transcend particular modes of perception, including simple force-dynamic, orientational or relational structures that emerge from recurrent patterns, shapes and regularities in our daily activities (e.g. CONTAINER, PATH, BALANCE, UP-DOWN, FRONT-BACK, PART-WHOLE) (Lakoff, 1987: 267).

In short, it is experience, meaningful to us by virtue of our embodiment, that forms the basis of many of our most fundamental concepts. From this fundamental and direct pre-conceptual knowledge of the world, more complex conceptual structures emerge to serve the purpose of shaping our perception of the world. Image-schemas are argued to provide the 'conceptual building blocks' for more complex concepts to be structured (Lakoff, 1987, 1993; Johnson, 1987). For example, the abstract concept of LOVE is structured and understood by virtue of the image-schema CONTAINER (e.g. *fall in love, fall out of love*) via conceptual metaphors (STATE AS CONTAINER-metaphor, so that BEING IN A STATE IS BEING IN A CONTAINER) (Lakoff and Johnson, 1980).

3.5.4. *Conceptual metaphor theory*

As discussed earlier, pre-conceptual structures are resources on which more complex mental networks of world knowledge can be constructed, including idealized cognitive models (ICMs) (Lakoff, 1987), scripts (Schank and Abelson, 1977) or frames (Fillmore, 1975, 1982; Fillmore and Atkins, 1992). Each ICM consists of an ontology (set of discrete entities/elements contained in it, each constitutes an intrinsic basic-level conceptual category) and a structure (coordinated interconnections between these conceptual elements, i.e. the image-schemas). For example, the JOURNEY-ICM has the following conceptual ontology with five elements: (a) a point of departure, (b) a route, (c) a destination, (d) a traveller, and (e) a vehicle, which are arranged by the PATH-schema. More complex and figurative ICMs come into existence by blending two ICMs/domains via specific semantic extension patterns such as metaphor (conceptual metaphor), or metonymy (conceptual metonymy), etc. (Lakoff, 1987; Johnson, 1987; Lakoff and Johnson, 1999), amongst which conceptual metaphors are considered the most prominent.

Conceptual metaphor is, therefore, defined as 'a complex ICM that is cognitively constructed by mapping a concrete source-domain onto a more abstract target-domain' (Lakoff and Johnson 1980: 4). For instance, the conceptual metaphor A LOVE RELATIONSHIP IS A JOURNEY is constructed by mapping the JOURNEY-ICM onto the abstract target-domain DEVELOPMENT OF LOVE RELATIONSHIP. The correspondences between the ontologies and structures of the two domains make it possible to draw metaphorical inferences about the target domain on the basis of the clearly structured and directly comprehensible JOURNEY-ICM. By mapping the two domains, we can easily understand utterances such as *our relationship has hit a dead-end*, *we're at a cross roads*, *we will go our separate ways*, in which lovers are travellers and their relationship is the vehicle to help them reach their goal (destination). These rich conceptual mappings between specific source and target domains are argued to be explicitly represented as part of our conceptual structure and are readily accessible for people in order to arrive at the correct interpretation. Conceptual metaphors are, therefore, a pre-existing, fixed part of our conceptual system of knowledge on the basis of which we can make sense of meanings, particularly of novel, imaginative expressions such as *we are driving on the fast lane of the freeway of love*. In other words, conceptual metaphors are called 'conventional metaphors' which is 'neither part of the grammar of English, nor the English lexicon, but it is part of the conceptual system underlying English' (Lakoff, 1993: 206). Lakoff and Johnson (1980) and Lakoff (1993) indeed contend that our outlook on 'truth' and reality is shaped and defined by conceptual metaphors in their aggregate:

In the intervening years, a huge system of everyday conventional, conceptual metaphors has been discovered. It is a system of metaphor that structures our everyday conceptual system, including most abstract concepts, and that lies behind much of everyday language (Lakoff, 1993: 204).

As a whole, Lakoff and Johnson (1980) posit that underlying most of our fundamental concepts are several kinds of metaphors:

- Orientational metaphors: primarily relating to spatial organization (up/down, in/out, front/back, on/off, near/far, deep/shallow and central/peripheral)
- Ontological metaphors: associating activities, emotions and ideas with entities and substances (most obviously, metaphors involving personification)
- Structural metaphors: overarching metaphors (building on the other two types) which allow us to structure one concept in terms of another (e.g. rational argument is war or time is a resource). This is thought to be the most pervasive type of conceptual metaphors in idiom motivation.

3.5.5. Conceptual metaphors in idioms

With idioms, it is argued that their figurative meanings can be understood (or that idioms can be *motivated*) on the basis of conceptual metaphors (Lakoff, 1987). The meaning of the idiom *spill the beans*, for instance, is motivated by a set of conceptual metaphors THE MIND IS A CONTAINER, and IDEAS ARE ENTITIES. The conceptual (metaphorical) knowledge provides us with a link that makes it possible to map the literal meaning onto the idiomatic meaning where ‘the mind’ is like a container from which information comes out as if beans coming out from a can. The existence of conceptual metaphors allows Kosevecs and Szabo (1996) to argue that it is the conceptual domain (the concept) – and not the individual words themselves – that participates in the process of creating idiomatic expressions. For example, despite the differences between the linguistic arrangements in [1] – [4], they are all related to the concept of *fire*:

- [1] The killing *sparked off* riots in the major cities
- [2] He was *burning the candle at both ends*
- [3] The bank robber *snuffed out* Sam’s life
- [4] The speaker *fanned the flames* of the crowd’s enthusiasm

The conceptual metaphors accounting for [1] – [4] are DANGER IS FIRE, ENTHUSIASM IS FIRE, CONFLICT IS FIRE, and ENERGY IS FUEL FOR FIRE. Many conceptual metaphors such as these are believed to be based on tight, local correlations in experience (Kovecses, 2005), e.g. intensity with heat, purposes with destinations, more with up, intimacy with closeness and so on (Lakoff, 1990, 1993; Kovecses, 1995; Grady, 1997).

Such claims for a conceptual motivation of idioms via conceptual metaphors are supported and extended by psycholinguistic studies by Gibbs (1990, 1992, 1994), Nayak and Gibbs (1990), and Gibbs and O’Brien (1990), who maintain that conceptual metaphors are psychologically real. Gibbs and O’Brien (1990) report the findings from a series of studies

demonstrating that people's mental images for idioms are constrained by different conceptual metaphors. For example, people's images of the idioms *blow your stack*, *flip your lid*, and *hit the ceiling* share similar characteristics such that stacks are blown, lids flipped, and ceilings hit because of internal pressure that causes the involuntary release of some substance upward in a violent manner. The authors argue that the consistency of meanings for different idioms with similar figurative interpretations comes from the constraining influence of conceptual metaphors that provide part of the link between an idiom and its figurative meaning. On the other hand, people's mental images for non-idiomatic phrases, such as *blow your tire*, *flip your hat*, or *hit the wall*, are much more varied because these phrases are not motivated by pre-existing conceptual metaphors. In Nayak and Gibbs' (1990) experiments, similarly, the metaphorical description of emotion concepts like anger (e.g., ANGER IS ANIMAL BEHAVIOUR and ANGER IS HEATED FLUID IN A CONTAINER) is found to facilitate people's speedy understanding of different idiomatic phrases in discourse contexts, such as *bite your head off* and *blow your stack*. Echoing this finding, when Gibbs (1992) asks participants about their understanding of events corresponding to particular source domains in various conceptual metaphors, the participants were remarkably consistent in their responses, showing a shared knowledge of the properties/characteristics of these domains. For instance, the source domain of heated fluid in a container for the metaphor ANGER IS HEATED FLUID IN A CONTAINER is understood as follows: (a) the cause of the explosion is the internal pressure caused by the increase in the heat of the fluid inside the container, (b) the explosion is unintentional because containers and fluid have no intentional agency, and (c) that the explosion occurs in a violent manner. The way participants use this knowledge of the source domain to map onto their conceptualizations of the target domain (anger) was found to be very predictable as well. They inferred *blow your stack*, *flip your lid*, or *hit the ceiling*, as anger being caused by internal pressure, which is unintentional, and done abruptly. These experimental data show how the metaphorical mappings between the source and target domains directly influence people's understanding of idioms.

3.5.6. Conceptual metonymy and other sources of motivation

It is acknowledged that not all idioms are metaphorical (although they definitely prevail), and that they do involve other figurations such as metonymies, hyperboles, etc. (Nunberg et al., 1994). Additional types of conceptual knowledge that are found to be motivating their transparency include conceptual metonymy, blending, emblem, and conventional knowledge (Kovecses and Szabo, 1996; Fauconnier and Turner, 1998; Langlotz, 2006). For example, Langlotz (2006) found that, besides conceptual metaphors, idioms from the word fields of SUCCESS, PROGRESS, and FAILURE in the BNC² World View corpus are also motivated by metonymy, blending and emblems.

² BNC stands for the British National Corpus a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century. See <http://www.natcorp.ox.ac.uk/corpus/index.xml>

According to Lakoff and Johnson (1980), conceptual metonymy is central to human thought and language in the same way conceptual metaphor is, if not even more fundamental to conceptual organization than metaphor (Taylor, 2002; Radden, 2000; Barcelona, 2003). While metaphor is principally a way of conceiving of one thing in terms of another, metonymy has primarily a referential function; that is, it allows us to use one entity to *stand for* another with a more specific focus on certain aspects of what is being referred to. In other words, *metonymic motivation* refers to the recognition of contiguity relationship between the source domain and the target domain, including *adjuncts* (things are found together) or *functional relationships*. For example, in the case of the metonymy THE PART FOR THE WHOLE, amongst the many body parts that can be used to refer to a human, 'good heads' is selected to refer to 'intelligent people'. There is clearly a link between the two concepts, i.e. the particular characteristic of the person (intelligence) which is associated with the head.

Conceptual metonymy (like conceptual metaphor) depends on conventional (cultural, historical, etc.) associations, such as PRODUCT FOR PRODUCER (*he bought a Picasso*), OBJECT USED FOR USER (*the sax has the flu today*), CONTROLLER FOR CONTROLLED (*Napoleon lost at Waterloo*), INSTITUTION FOR PEOPLE RESPONSIBLE (*Exxon has raised its prices again*), THE PLACE FOR THE INSTITUTION (*the White House isn't saying anything*) or THE PLACE FOR THE EVENT (*Watergate changed our politics*) (Lakoff and Johnson, 1980: 38-39).

Emblem, another source of motivation, is sometimes treated as a separate phenomenon, and sometimes subsumed under the broad coverage of metonymic motivation. Emblems are stereo-typical cultural information which can serve as direct bases for idiom motivation, for example WOLF-emblem represents DANGER, EVIL (*a wolf in sheep's clothing*). Besides animals, *emblems/symbols* include objects, places, or events which are used for people, virtue, etc. due to their symbolic significance or eminent characteristics. Some emblems are considered strong, i.e. having well-rooted origins in a culture or community and are therefore well-entrenched and having highly salient cognitive status (e.g. pig for dirtiness), while others are weaker because of their ambiguity and restricted emblematic values (e.g. applicable to a small community or subcultures only) (e.g. goat for perversion or pig for stupidity in Vietnamese culture).

As a result, in a broader context, sources of motivation can include more than just conceptual metaphor and conceptual metonymy. Additional sources of motivation, such as socio-cultural knowledge, are essential for the motivation of idioms to be fully complete:

The conceptual metaphors alone are not sufficient to determine the full motivation of an idiomatic construction. Full conceptual motivation depends on the concrete knowledge associated with the literal scene (Langlotz, 2006: 51).

3.5.7. *The role of motivation in idiomatic creativity*

Systematic idiom variations (as opposed to creative wordplays) are argued to be ‘a reflection of idiom transparency’ (Langlotz, 2006: 289), which in turn is achieved when the speaker is able to re-motivate an idiom in relation to its underlying conceptual bases. In other words, motivation is hypothesised to facilitate the possibility of creating systematic idiom variants in the usage context. In the absence of motivation, the systematic variation potential of an idiom is highly restricted (ibid: 271). In their experiments, Gibbs and Nayak (1989: 125) also find that metaphoric transparency (i.e. motivation in our definition) positively affects how the language user understands and creatively uses idiomatic expressions:

Although semantic decomposition was the best predictor of syntactic productivity, metaphoric transparency was also found to have a strong influence on the syntactic productivity of idioms. [...] Both semantic decomposition and metaphoric transparency emphasize the importance of an idiom’s analysability in accounting for its syntactic productivity

Despite being welcomed and viewed favourably in the literature thanks to its merits in offering an explanation of idiomatic variations more systematic than previous models could, conceptual metaphor theory also faces criticisms for a number of gaps in its framework. Such linguists as Glucksberg et al. (1993), Cacciari and Glucksberg (1995) or Burger (1998a, 1998b, quoted in Langlotz, 2006: 52) contend that although idioms can be (re)motivated, conceptual metaphors do not play a role in this process on the following premises:

- Not all idioms are metaphorical (Burger, 1998b: 30)
- Not all idioms that are metaphorical are structured by general conceptual metaphors. A great number of idioms cannot be described according to systems of conceptual metaphor (Burger, 1998b: 91)
- The actual occurrence of idioms in texts does not suggest that the underlying metaphorical models are used in a consistent way (Burger, 1998b: 36)
- The definition of metaphorical models on the basis of idioms often violates the linguistic data. The metaphorical patterns in idioms are often more specific than the generic models suggest (Burger 1998a: 91)
- The supposed universality of conceptual metaphors neglects the historical and cultural specificity of the phraseological system of a given language (Burger, 1998a: 91)

I also find the claims about the role of motivation in need of further elaboration. First of all, it is true that some motivated idioms might also be analysable, which means they can be flexibly manipulated (to certain extents) in discourse. However, motivation does not necessarily entail analysability; and it is unclear how motivated but unanalysable idioms

(e.g. *flip one's lid, spin the wheels*) might be internally manipulated because their constituents do not carry individual meanings.

Secondly, the existence of conceptual metaphors in our cognitive lexicon is often refuted on the basis that they are not universal (or at least not all of them) but historical and cultural specific (Dobrovolskij and Piirainen, 2005; Kovecses, 2005). While embracing the cognitive theory of metaphor as a valid and useful tool for analysing conventional figurative language, these authors argue that the knowledge of underlying conceptual metaphor is insufficient or not linguistically relevant in all particular cases. Kovecses (2005), for example, notices the inconsistencies in the conceptual metaphorical explanations of some of the concepts. The utterance *our fears are fuelled by acts of terrorism* seems to rely on the metaphor INTENSITY IS FIRE (due to the presence of the concept 'to fuel'). However, the conceptualization of fear is conventionally based on cold, not heat (*he froze at the sight of the killer*). The example then contradicts both the conventional conceptualisation and the physiological embodiment of fear. This is a problem for the cognitive linguistic view of metaphor, which requires that no metaphor contradict the embodiment of metaphorical concepts.

Thirdly, it is suggested that rich mappings between domains are not explicitly represented as part of our conceptual structure, and hence conceptual metaphors are not part of our conceptual knowledge of the world (Glucksberg and McGlone, 1999). This claim is supported by their experiment in which they asked participants to interpret some metaphors rooted in the conceptual metaphor LOVE IS A JOURNEY. With a highly conventional metaphor such as *Our love is a bumpy rollercoaster ride*, the characteristic of the vehicle (rollercoaster ride) as being 'fast, exciting but having ups and downs' will be automatically accessed to make sense of the target concept 'our love'. In case of a non-conventional metaphor such as *Our love is a journey to the bottom of the sea*, the interpretations provided by the subjects point to a tendency of finding the properties of the concrete concepts (e.g. being dangerous and going to kill us/our deepest emotional natures have been revealed) that might plausibly be attributed to the target concept 'our love.' Cross-domain mappings were not found to be explicitly present in the interpretations, which Glucksberg and McGlone take to mean that there is no support for the hypothesis that people automatically retrieve specific source-target domain mappings in order to understand a novel metaphor. Instead, the language user is assumed to have the ability to *infer an attributive category* that the source domain might exemplify and that this attribute is compatible with the intended conceptualisation in the specific context. Glucksberg (1993) proposes that the precise strategy of inferring and determining the meaning of a variant with respect to the meaning of the original idiom comprises the following operations (which are not based on conceptual mappings):

- (1) recognise novel idiom as a variant of a conventional idiom
- (2) retrieve meaning of original idiom
- (3) identify word meanings of both variant and original idioms

- (4) compare the word meanings of the two idiom forms
- (5) identify the relations between those word meanings
- (6) take this relations between word meanings to infer, by analogy, the relations between the meanings of the original and variant idioms

In terms of methodology, Steen (1999) raises questions regarding the use of data in conceptual metaphors research. According to him, it has sometimes remained 'an act of faith' that a particular metaphor in language reflects a particular metaphor in thought. In other words, in his opinion, whether conceptual metaphors are psychologically real or post-hoc theoretical constructs, remains an open question. While acknowledging that the linguistic evidence for the existence of conceptual metaphors in linguistic expressions as provided by Lakoff and Johnson (1980), Johnson (1987), Lakoff (1987), and Turner (1987) is valid, Steen points out that these clear cases serve the purpose of demonstration only. They have not been systematically and exhaustively collected from large stretches of discourse, but have been selected for their persuasive power. Similarly, although the conceptual metaphor approach is judged by Deignan (2005) to be potentially very enlightening as a tool for identifying underlying meaning, she points out the pitfall of overgeneralising on limited linguistic evidence that conceptual metaphor research might encounter. As a result, the arguments for conceptual metaphors in cognitive-linguistic traditions thus far presented cannot be taken as universally applicable:

It presupposes a generally accepted procedure of deriving conceptual metaphors from linguistic metaphors encountered in on-going discourse, and that is currently not available. [...] it is ironic that cognitive linguists are going out of their way to show that linguistic metaphor is fundamentally conceptual, but that in doing so, they have neglected the method for showing how they get from linguistic metaphor to conceptual metaphor in the first place (Steen, 1999: 57-58).

It is, therefore, the purpose of this thesis to use corpus data to highlight aspects of idiomatic creativity that might relate to conceptual metaphors. It should be made clear that this thesis does not aim to validate the existence or otherwise of conceptual metaphors in idiom processing (which requires psycholinguistic experiments) but hopes to find out (1) if there are any significant patterns of conventional metaphorical conceptualization of idioms from top-down analysis, and (2) whether the metaphorical motivation is related to the flexibility of idioms in discourse.

Summary

The concept of 'motivation' is used in the field of cognitive semantics to refer to the possibility of establishing discernable links between the literal meaning of an idiom and its figurative meaning to provide a systematic explanation of the non-arbitrariness of the

figurative meaning of the idiom. Sources of motivation can be attributed to (but not restricted to) conceptual metaphors and conceptual metonymy, general world knowledge and a possible blending of more than one source.

The linguistic community is divided with regard to the role of conceptual metaphors in idiom processing in general and in idiomatic creativity in particular. The most positive assumption is that maintained by Lakoff (1993), that conceptual metaphors are both available and accessible in any context, and thus may serve as the conceptual basis for on-line figurative language comprehension. However, as we have seen, other linguists have doubts about this assumption. Hence, the second school of thoughts posits that conceptual metaphors are available in conceptual-semantic memory, and may be retrieved in certain situations (but not all). In this scenario, conceptual metaphors are not necessary for immediate comprehension, but may be recognized and appreciated in contexts that motivate people to search for an underlying metaphorical theme. This is argued to be the by-product of the context-dependence of conceptual structures. Any given item in memory may be accessible in some circumstances (when people have the time to make considered, deliberate judgments) but not in others (in ongoing speech comprehension and production when people do not have the time for such judgments) (Glucksberg and McGlone, 1999; Nayak and Gibbs, 1990). The least sympathetic view of conceptual metaphors in idiom comprehension is to completely discard the role of conceptual metaphors. As with most words, the comprehension of metaphorical expressions may proceed without recourse to or awareness of their etymological origins. People may be able to appreciate the underlying metaphor when it is pointed out to them, but it need not be explicitly represented as conceptual knowledge.

Adopting a corpus linguistic approach, this thesis does not set out to prove whether or not conceptual metaphors (and other sources of motivation of idioms) are psychologically real, so as to elaborate on the model of the comprehension of idioms as it happens on-line. Instead, with the naturally occurring data from the BNC, the thesis aims to highlight any linguistically significant patterns of conceptual motivation of idioms that appear to have a significant correlation with idiomatic creativity, or the lack thereof, in relation to context.

3.6. Analysability and its role in idiomatic creativity

3.6.1. The concept of analysability

Analysability is a notion referring to the availability of componential meanings in idioms via top-down analysis. While still non-compositional from bottom-up analysis, the idiomatic meanings of some idioms can be devolved on their constituents, allowing for top-down analyses of the idioms. For example, the verb *spill* in the idiom *spill the bean* is considered to have the lexicalised figurative meaning of ‘reveal’ even outside of the idiomatic context (e.g. *come on, spill it, share with us what you know*), and, through frequent evocation in the idiomatic

context, *the bean* carries the phrase-induced meaning of *information*. The idiom, as a result, is viewed as decomposable and hence analysable. Titone and Connine (1994) indeed confirm that a large number of idioms are analysable.

According to Langlotz (2006), the general patterns of figurative-literal analysability are of two types:

- a. *Lexicalised figurative senses*: the constituents have autonomous figurative meanings that are also used outside of the idiom's context
- b. *Phrase-induced figurative senses*: the constituents can be attributed an autonomous figurative sense in the context of the idiomatic construction

Analysability is closely related to, but not synonymous with, motivation in the sense that both notions involve with the top-down reanalysis of idioms to make sense of the figurative meaning in relation to the literal meaning of an idiom, i.e. contributing to the transparency of the idiom. However, the fundamental difference between them is that while analysability is concerned with assigning meanings to idiom constituents (with or without a plausible underlying link for this assignment), motivation explains the underlying links between the literal and figurative meaning of an idiom (with or without the possibility to break the idiom down into components). It is true that analysability might incur on the recognition of the motivated metaphorical correspondences between an idiom's literal and figurative meaning, but not restrictively. In many cases, a correspondence can be made without any motivated link (e.g. constituents' phrase-induced figurative meanings in the context of the idiom only, e.g. 'pickle' in *in a pickle*, 'beans' in *spill the beans*). In other words, analysability might be the result of some degree of motivation, or it could be achieved if a phrase-induced figurative sense is available.

Analysability is argued to be psychologically real according to Gibbs and Nayak's (1989) and Gibbs et al.'s (1989) *decomposition hypothesis*, which states that people do attempt to 'assign independent meanings to its individual parts and quickly recognize how these meaning parts combine to form the overall figurative interpretation of the idiom' (Gibbs et al., 1989: 587). When this cannot be done (as in the case of unanalysable idioms) the reaction time (the processing time spent by subjects to understand an idiom) is found to be longer than in the case of compositional idioms, suggesting that, as part of the comprehension process, people usually try to subject idiomatic constructions to compositional analysis. The impossibility to process some idioms compositionally is thought to be the reason why their processing time is longer.

3.6.2. The role of analysability in idiomatic creativity

It is argued that thanks to its analysability, *spill the beans* ('reveal information') is quite flexible, appearing in such operations as conversion/nominalization (*more beans spill on Ford's*

5.0 V8), passive voice (*the beans (spilled) and a giveaway*), premodification (*Bobby Brown spills the lez-beans*), pre- and postmodification (*they are ready to spill the ugly beans of truth*) etc. This correlation between analysability and idiomatic variability is what Nunberg and his colleagues predict:

We are thus in effect proposing to explain a variety of 'transformational deficiency' of idioms by positing a bifurcation between idiomatic phrasal constructions and idiomatically combining expressions. [...] this approach predicts a strong correlation between semantic analysability and 'transformational productivity' (Nunberg et al., 1994: 508).

Backing Nunberg's claim about the correlation between flexibility and analysability are positive results from various psycholinguistic experiments by Gibbs and colleagues (Gibbs and Gonzales, 1985; Gibbs and Nayak, 1989; Gibbs and O'Brien, 1990; Gibbs et al., 1989). Different groups of native speakers were asked to complete a series of tasks based on various examples of idiomatic expressions, both in their canonical and manipulated forms. The findings show that the decomposability of idioms is not just a theoretical analytic tool utilized by linguists in their research which is alien to the normal language users. On the contrary, people are not only capable of perceiving the different degrees of decomposability of idioms but also able to classify idioms into categories (normally decomposable, abnormally decomposable and non-decomposable idioms). The more speakers are aware of the analysability of an idiom and its separate meaningful units, the more likely that idiom will be viewed as syntactically productive and will be used as such in discourse:

These data suggest that people attempt to do compositional analysis when understanding idiomatic phrases. When an idiom is decomposable, subjects can assign independent meanings to its individual parts and will quickly recognise how these meaningful parts combine to form the overall figurative interpretation of the phrase. [...] the ability to assign the parts of an idiom with independent meanings motivates why many readers view normally decomposable idioms as being syntactically productive and lexically flexible (Gibbs et al., 1989: 587)

These findings demonstrated that the analysability of an idiom is the best predictor of syntactic productivity. The more an idiom's individual components contribute independently to the phrase's overall figurative meaning, the more likely that phrase will be viewed as syntactically flexible (Gibbs and Nayak, 1989: 125)

Admittedly, the view of idiom top-down analysability in processing is helpful in offering explanations of the production and comprehension of idiom variants and idiomatic creativity. However, apart from the work of Gibbs and his colleagues, it is noticed that the psychological relevance of analysability in idiomatic creativity has received very limited

empirical support, and several psycholinguistic studies failed to observe semantic analysability effects in both the recognition and production of idiomatic strings (Cutting and Bock, 1997; Sprenger, Levelt and Kempen, 2006; Tabossi, Fanari and Wolf, 2008; Titone and Connine, 1999). The reason why *the question was popped by John* is acceptable, but *the sack was hit by John* is not, although both idioms are normally decomposable, still needs further explanation outside of the decomposable hypothesis. According to Nunberg (1978) the reason why decomposable idioms behave differently in the passive voice is that there are other factors controlling the variants. Bringing a component of an idiom into focus in a passive construction does not only depend on the meaning of its referent (i.e. if this referent is normally expected to be focused on) but is also constrained by the meaning of other constituents in the idiom (i.e. if they are normally used in passives as well). In the case of *hit the sack*, the ‘hit’ involves the same use of ‘hit’ as in *the paper I dropped hit the floor*, which Nunberg argues to be odd in the passive. On the other hand, the use of ‘hit’ in *hit the jackpot*, for example, is similar to that in *hit the answer*, *hit the lucky number*, which would passivize freely (ibid). Exceptions to passives, as a result, have to be assumed to belong in pragmatic realm, wherein the acceptability/unacceptability of a given passive is ‘due to the fact that the nature of the particular predication is inconsistent with the function associated with the passive construction’ (Nunberg, 1978: 134).

Many authors, such as Schenk (1995), Ackerman and Webehuth (1993), Nicolas (1995) or Abeillé (1995), have either refused to acknowledge idiomatic decomposability altogether, or reported to have found no clear correlation between idiomatic flexibility and semantic analysability in their data. Discussing syntactic operations which are applicable to meaningful expressions in comparison to those applicable to both meaningful and meaningless expressions, using German idioms as examples, Schenk argues that idioms do not allow any of the operations on meaningful expressions such as control or topicalisation. Idioms can only undergo operations typically applicable to meaningless expressions (e.g. raising, verb-second etc.), hence idiom parts have the same distribution as the meaningless existential *there*, or the expletive *it* or the ambient *it*:

Topicalisation:

<i>He believes there to be unicorns</i>	<i>John kicked the bucket</i>
* <i>There he believes to be unicorns</i>	* <i>The bucket John kicked</i>

Control:

* <i>John instructs it to rain</i>	* <i>John instructs the beans to be spilled</i>
* <i>It tries to rain</i>	* <i>The beans tried to be spilled</i>

Schenk took this observation as his argument for the claim that parts of an idiom do not carry meaning (Schenk, 1995: 269). This property, according to Schenk, explains why certain operations, such as internal modification, are not applicable to idioms parts (e.g. **Mary spilled the well-kept beans*). According to him, since idiom parts are meaningless,

modifications of a meaningless expression make no sense because its interpretation will still be meaningless.

Nicolas (1995), by analysing nominal premodification of idiom-internal noun phrases, rejects the view that the variability of these idioms is dependent on the identified meanings of individual constituents. Instead, internal modifications should be regarded as external adverbial modifications of the whole idioms. For example, *make rapid headway* can be understood as *progress rapidly* (manner modification), or *call the political tune* is in fact *dominate, politically speaking* (viewpoint modification). Hence, the internal modifications do not verify the compositional view of idioms but on the contrary can be integrated into a non-compositional view. Nicholas (1995) labels this type of modification 'syntactic modification.' There are a large number of idioms which cannot be interpreted according to this model, for example *raised Labour eyebrow*, which Nicolas calls 'semantic modification' of idioms. He, however, maintains that well-formed idiom modifications can only be syntactic, while semantic modifications are considered the realm of wordplay.

Abeillé (1995), on the other hand, acknowledges compositionality in idioms. However, by observing 2000 French idioms, she claims that there is no clear correlation between semantic decomposability and such syntactic operations as passive or topicalisation or clefting, since opaque idioms behave just as flexible as analyzable ones in her data. Lexical operations such as insertion of quantifiers and relative clauses were found, but Abeillé finds them 'not a clear sign of compositionality' (ibid: 24). Rather, these operations are dependent on the semantics of other constituents within the idioms. Her conclusions, therefore, are open and inconclusive:

I do not mean that no regularities can be found in the syntax of idioms, [...]. Neither do I mean that all idioms should be considered non-compositional. But I did not observe a clear correlation between such syntactic regularities and compositionality (Abeillé, 1995: 24)

Similarly, while acknowledging the role of compositionality (which corresponds to our definition of analysability in this thesis), Glucksberg (1993: 19) is cautious as to how much it determines semantic productivity in idioms:

The more compositional an idiom, the more likely it will be available for variation of one type or another. [...] [however] compositionality is neither a necessary nor a sufficient condition for idiomatic productivity.

By discussing the productive operations of some idioms that he deems non-compositional (e.g. *speak one's mind, two left feet*), Glucksberg contends that semantic productivity may be independent of compositionality. Where idioms are analysable to certain degrees, they can be manipulated in certain ways but not as freely as their non-idiomatic equivalent would be.

In fact, idiomatic manipulations, regardless of the degree of decomposability/analyzability of the idioms, are found to be constrained ‘jointly by the idiom’s idiomatic meaning and the semantics of the words themselves’ (Glucksberg, 1993: 7). For instance, the literal meaning of the constituent *kick* (i.e. an discrete kind of act) in *kick the bucket* dictates that although it is okay to say ‘he’s laying dying all week’ it is impossible to say ‘he’s laying kicking the bucket all week’ (Wasow, Sag and Nunberg, 1983) because if you kick the bucket all week, it means you kick it over and over again, but you cannot die over and over again. This example brings up the constraint exerted on idiom variations by the idiomatic meaning of an idiom as well (which will be discussed in 3.7). By the same constraint of compatibility, even though literally you can ‘kick the bucket silently’, or ‘kick the bucket sharply’, our conventional understanding of the concept ‘to die’ suggests that out of the two phrases, the latter is impossible (you can die silently but not die sharply). Glucksberg, therefore, suggests that general world knowledge and rules of discourse and conversation are also important factors of idiom flexibility, as is decomposability.

Other linguists have also highlighted the empirical fact that non-decomposable/opaque idioms are also capable of variations (Cacciari and Glucksberg, 1991; McGlone et al., 1994; Ackerman and Webelhuth, 1993), for example:

[1] Ned: *By and large, the country is doing quite well economically
and socially these days*

Ted: *By and not-so-large. Have you seen the figures on the
homeless problem?*

[2] Tom: *Did the old man kick the bucket last night?*

Joe: *Nah, he barely nudged it.*

(Cacciari and Glucksberg, 1995: 45)

It is impossible to analyse the idioms *by and large*, *kick the bucket* so as to assign meanings to each lexical word; nevertheless, the idioms can still be creatively altered to bring humorous effects to the otherwise neutral original phrases. This begs a more fine-grained view of idioms in which decomposability is one but not the only explanation to idiom variations. The example suffices to question the absolute validity of ‘analysability’ theory. It is proposed that other factors, including contexts, are at play in the composition and understanding of the meaning of idiom variants.

Last but not least, there are concerns, even amongst advocates of the decomposability hypothesis, over the characterisation of idiom decomposition, since there have not been many formally well-defined frameworks to specify whether a given idiom is semantically decomposable or not (Gibbs and Nayak, 1989). Decomposability, in the eyes of the majority, is a matter of degree (ibid). Clines from the most lexicalized, non-decomposable to the highly decomposable and most syntactically productive have been proposed elsewhere in the literature (Fernando, 1996; Langlotz, 2006). The fact that such an allegedly non-

compositional idiom as *by and large* can be constitutally negated (as in [1] above) suggests that the idiom is not purely or completely non-compositional because the semantics of *large* does bear some functional relation to the idiom's meaning 'generally'. The idiom, in Glucksberg's (1993) opinion, is at least partially compositional/analysable. He even suggests that purely non-compositional idioms may not exist at all. To the extent that a constituent of an idiom may be modified independently of the idiom as a whole, Glucksberg argues that it is compositional. In other words, the relationship between compositionality and idiom productivity is not unidirectional; compositionality can also be viewed as being determined by the productivity of an idiom as well.

Summary

Despite the lack of consensus in the linguistic community regarding the use of terminology, definition and characterisation, the term 'analysability' is adopted to refer to the phenomenon in which an idiom can be reanalysed via top-down analysis into components that correspond to components of its figurative meaning.

The hypothesis in this thesis is that analysability can guarantee certain types of alterations with certain class of idioms, not all of them. This is due to the fact that different syntactic and lexical operations have different purposes and characteristics that might or might not require analysability. Simultaneously, there are constraints on idiomatic creativity in general (see 3.7 below) as well as the effects of context that might extend or restrict the extent of the influences of analysability on specific types of alterations. Since most of the prominent research in this area (as reviewed earlier) has predominantly been carried out using psycholinguistic experiments or introspections, corpus-based analyses in chapter 5 and 6 will provide an empirical angle into the phenomenon of idiom's analysability and the extent of its effects on idiomatic creativity.

3.6.3. *The combined effects of the features*

With recent research findings that confirm the complexity of the phenomenon of idiomatic creativity, it is often acknowledged that neither of the features alone is the necessary nor sufficient condition for the creativity we have observed with idioms. Efforts, therefore, have been made to merge the three clines (literal well-formedness, analysability, motivation) into one 'transparency cline' in order for idioms to be ranked in a way that could explain their creative behaviour. For example, three types of idioms are identified by Nunberg et al. (1994) according to the ways idioms are analysable, which are thought to reflect their flexibility as well. Firstly, *normally decomposable idioms* can be analysed by means of conventions that justify not only the relationship between the literal and the figurative meaning of an idiom, but also the function between the components of the referent (i.e. figurative meaning) and the literal meaning of the idiom (i.e. these idioms are both motivated and analysable). For example, *give up the ghost* (meaning 'to die'), for instance, is

normally decomposable because we can identify death as a two-place relation which holds between people and their spirits. There is a function from ghosts to spirits, and from the act of giving up something to the act of relinquishing one's hold on the spirit. The passive form as in '*once the ghost has been given up, there is nothing medical science can do*' is acceptable because the referent 'the spirit' can be the focus of the new construction. Having the same meaning 'to die', *kick the bucket*, on the other hand, is *non-decomposable*, because it is impossible to find a function between 'kick the bucket' and 'to die' using any convention accessible to the common language user of modern times. Hence, a referent of the noun phrase 'the bucket' cannot be identified and therefore the idiom cannot be passivized. That is, when there is no assurance that the hearer will be able to identify the referents of idioms (as with non-decomposable idioms), it makes no sense for the speaker to modify them. Unanalysability as such, when considered together with the bottom-up opacity, describes the extreme condition of frozenness in some idioms. Last but not least, with *abnormally decomposable idioms*, the decomposability is limited to the possibility of identifying the referents of the idioms as a whole via conventional metaphors, which invoke the relationship between the literal meaning of the idioms and their idiomatic meanings, but there are no identifiable components of the referents to which components of the idioms can refer to (i.e. motivated but unanalysable idioms) (e.g. *flip your lid*). These are not as flexible as normally analysable idioms.

Langlotz's (2006) analysis of the different types of idioms in terms of transparency is more detailed. Idioms with all three properties (literal well-formedness, analysability, global and constituent motivation) are at the top of the cline (most transparent) (e.g. *swallow the bitter pill*, *jump on the bandwagon*), followed by idioms with literal well-formedness, global motivation and analysability (e.g. *grasp the nettle*, *spill the beans*). Considered less transparent than the previous two types are idioms with literal well-formedness and global motivation (e.g. *grit one's teeth*, *spin one's wheels*). The class of opaque idioms are those with only literal well-formedness but none of the types of motivation or analysability (e.g. *kick the bucket*, *bite the dust*). It is quite convincing how idioms should be ordered as such, because it becomes instantly clear how different types of motivation, when interacting with analysability and well-formedness, can help clarify the underlying mechanisms that facilitate the mappings from literal to idiomatic meanings of idioms. Unfortunately, there are other sets of idioms, classed as 'marginal types' by Langlotz, which are not placed in comparison with the aforementioned types, making it difficult to decide where they are placed on the cline. These are idioms with ill-formed literal meanings, either having experientially unrealistic literal meaning (e.g. *cook the books*, *rack one's brain*), possessing only partial well-formedness (e.g. *know one's onion*, *rain cats and dogs*), having idiosyncratic constructions (e.g. *trip the light fantastic*, *come up roses*), containing cranberry morphs (e.g. *blow the gaff*, *cock a snook*), or containing highly specialised/garden-path constituents (e.g. *hide your light under your bushel*, *chomp at the bit*). Many of these are opaque, but many are partially motivated. Creative manipulations of these idioms are therefore found in his data (e.g. *once the gaff about Whisky*

not being suitable for the kitchen was blown (Langlotz, 2006: 132), sometimes more frequently than certain transparent idioms. This raises the question of whether the cline he proposes fulfils its discriminating function. It is easy to find examples to support the extreme cases (familiar transparent and opaque idioms at the two ends of the continuum), but it is the controversial, in-between types that usually cause the most troubles because they usually do not fall neatly into the frameworks as expected. Shunning these controversial data is convenient, but a true description of language will never be reached. It is, therefore, the purpose of this study to use random authentic idioms and idiom variations from real texts and corpora to reveal the true relationship between transparency and flexibility.

Facilitating factors of idiomatic creativity, including literal well-formedness, motivation, and analysability as we have discussed, are still open for further debate due to the inconclusive nature of the results achieved so far. It is, therefore, hoped that the results achieved in this thesis will contribute to the existing body of evidence of the correlation or otherwise between literal well-formedness, analysability, motivation (separately and combined) and idiomatic creativity.

3.7. Cognitive constraints on idiomatic creativity

It has been noticed that while some idioms can be manipulated in certain ways, the manipulations are not always as freely as in literal expressions. This is because, first of all, idioms are more heavily constrained by institutionalization and relative fixedness, which restrict their creative syntactic and lexical alterations. Secondly, although a type of creative language (i.e. breaking with rules and conventions), idiom variants are still linguistic units (i.e. they need to be communicated to the readers), which means they have to conform to certain cognitive principles to be understood and to be able to convey messages. Interpretability and communicability, as such, dictates the maximum range of the violence that idiom variants can have against cognitive rules of language processing.

Linguists have had different views on the constraints on idiomatic creativity, ranging from zero (no constraint at all) (Burger et al., 1982) to seven factors (*recognisability, functionality, compatibility, grammaticality, non-ambiguity, non-conspicuity, and evocational autonomy*) (Langlotz, 2006) with varying degrees of importance attached to each, depending on the specific types of variational strategies. Before deciding on the applicability of some or all of these seven constraints to the current data, let us first take a closer look at each factor.

3.7.1. Functionality

In linguistic creativity in general and in idiomatic creativity in particular, functionality (Langlotz, 2006), or communicative purposes (Glucksberg, 1993), is an important constraint, distinguishing intentional play with language from mistakes or unintentional variations. For Glucksberg, this is *the* most important constraint on idiomatic creativity. The

validity of an idiom variant is verified when it ‘produces a comprehensible difference in interpretation’ compared to the original idiom meaning (Glucksberg, 1993: 21). In other words, a variation of an idiom is deemed acceptable if a reasonable communicative intention/purpose can be inferred by the use of an idiom in some form other than the original. As such, tense inflection in VP idioms is almost universally treated as regular and non-creative due to the lack of purpose, despite obvious lexico-grammatical changes to the canonical forms of the idioms (e.g. *they opened a can of worms*). The inflection is only applied to the head verb of an idiom to convey the tense aspect of the predicate according to the rules of English grammar. Inversions or passives, on the hand, are generally acknowledged as communicating some intention of placing the emphasis on certain parts of the idioms, therefore fulfilling the ‘highlighting’ function as compared to the base-forms (e.g. *out of the woodwork emerges a woman who claims to be the abandoned*). Even more creative are variants achieved via substitutions or insertions with a clear purpose of creating ambiguity via dual referentiality (e.g. *Burma's nuclear bomb alive and ticking*).

It is nevertheless acknowledged that the hearer’s willingness and ability to process a variant to derive weak implicatures and, thus, to recognize the intended purposes, depend on various contextual and cognitive factors (Langlotz, 2006). It follows that communicative purpose is highly subjective. In spoken genres, there are verbal and non-verbal cues in the texts themselves to provide evidence of such purposes. These cues can be the speaker’s intonations, stress patterns or overt verbal lead-ins (e.g. *something funny happened today at work...*) to cue their intention of producing a sarcastic/humorous comment. Cues can also be available from the listener’s laughs, scoffs, facial expressions or overt verbal responses (e.g. *that’s funny*) to show their appreciation/comprehension of the sarcasm/humour, etc. However, in many cases (particularly corpus data) some of these extra details might not be recorded. There is, in that case, no reliable way to measure these special purposes but to depend on the researcher’s judgement about the inferring processes involved in deriving the meanings and purposes of such variants. This is considered a valid method of analysis, since even in the absence of the reception on the part of the listener or the overt signals on the part of the speaker, the intention of special purposes can still be there and need to be detected, no matter how subtle these effects are. In the case of the lack of a communicative purpose, the expression will not be treated as idiomatic creativity:

Rodney: *The secret of my success? Planning. No detail is left unturned, no stone unchecked*

Del: *<confused look on his face, then shook his head and rolled his eyes>*

In this case, the loss of communicative purpose is caused by semantic incompatibility. As a result, subjective as they are, communicative purposes are treated in this thesis as the requirement to ensure that an idiom variant is intentional and conveying a message. This view comes from an empirical fact that some ‘unusual’ uses of idioms might easily be mistaken for idiomatic creativity due to their deviant forms, but these ‘new’ forms might not make sense in the contexts, either due to the speaker’s limited linguistic competence or

genuine mistakes/slips of the tongue. Communicative purpose is believed to be greatly supported and enhanced by contextual knowledge.

3.7.2. *Recognisability*

The next condition a variant is thought to have to satisfy is its *recognisability*. That is, the idiom variant must still be recognizable as a modified form of a canonical idiom instead of a random string or a totally new idiom. Recognisability is perceived as a fundamental variation constraint that ‘defines the maximum degree to which the variant may deviate from the base-form, while still performing its idiomatic function’ (Langlotz, 2006: 216). Glucksberg (1993) also acknowledges that recognizing novel idiom as a variant of a conventional idiom is the first of the many operations a language user must use to determine the meaning of the variant. Glucksberg, however, does not elaborate on how this recognisability is achieved. This gap is (partially) bridged by Langlotz’s (2006) cognitive explanation, in which the following factors are listed as idiomatic recognition triggers:

- (i) Activation probabilities: the activation of an idiom is dependent on the phenomenon of transitional probability - the strong co-activation of its constituents. For example, since *grasp* and *nettle* constitute the complex unit of *grasp the nettle*, the co-occurrence of these two constituents will activate the idiomatic meaning.
- (ii) Salient idiomatic substructures: the presence of unique constituents, including cranberry morphs (*spick and span*) and highly salient lexical items (*tightrope*, *bandwagon*), is recognized as increasing the probability of idiom recognition. Highly salient constructional features (*go places*, *pastures new*) also offer useful leads towards idiom recognition.
- (iii) Conceptual priming: on the basis that an idiom is familiar to the speakers, it can be exploited through a process called ‘parasitic elaboration’ (Langlotz, 2006: 202), i.e. a creative but not independent type of elaboration of the literal scene based on the previous recognition and conventional use of the idiom. The phrase ‘he barely nudged it’ as in the following example would not have been recognised as an idiom variation had it not been preceded by the original idiom itself. The concept ‘kick’ in the first utterance acts as a prime to accommodate the concept ‘nudge’ in the later utterance:

A: *did the old man kick the bucket last night?*

B: *nah, he barely nudged it*

In my opinion, recognisability is indeed the requirement for a creative idiomatic expression to qualify as a valid idiom variant, since it is this recognition of canonical idiom that will trigger the original idiomatic meaning, which can then be interpreted according to the new

arrangement. If a variant strays too far from the canonical idiom to the extent that no connection can be made, the original idiomatic meaning will not be activated, which will render the new expression meaningless, for example:

Conversation 1:

A: The problem is that people don't like to find out that their elected officials are keeping secrets from them. If the leaders we elect are having these wonderful sexual escapades, then they have a duty to share their adventures with the rest of us. In detail.

B: That's why the public turned on Elliott Spitzer and Jim McGreevey--and why they would have turned on Bill Clinton if I hadn't worked so diligently to put his adventures in the public eye where they belong. And I tried to do the same for David Paterson, but he just wouldn't work that blue. I don't know why; it's not like he's Clarence Thomas or Larry Craig. They still have their jobs, because it wasn't about sex they had, just sex they would've had if they had the chance.

Conversation 2:

Ricky: I think Hank is one of the most amazing characters ever ever created because he's got he's got that level of realism where you go oh, god, I feel so sorry for him or what an arse but you you know there's jeffrey tambor really working hard under the water

Garry: jeffrey tambor shockingly came in to audition for the part of Hank and I have the audition tape and it is shocking, because he's exactly in the reading what he was in the show out of the gate I mean he is a brilliant actor

Extended co-texts are provided to show that the underlined phrases appear to be figurative but nonetheless unrecognizable as being connected to any particular idioms (based on the responses by 8 native speakers of English). Hence, their meanings are undetermined.

In the analytical part (chapter 6), recognisability will be further investigated in relation to idiom transparency and contexts to see if these facilitating factors affect the extent of this constraint on idiom variants.

3.7.3. Semantic compatibility and grammaticality

Argued by Langlotz (2006: 221) to be the most important constraint on the creation of systematic variants, *compatibility* requires that 'any formal modifications of the literal scene must be in accordance with the idiomatic meaning'. For examples, if the speaker wants to modify the idiom *kick the bucket* by passivizing it (**the bucket is kicked*) or adding adverbial modification to the verb (**he repeatedly kicks the bucket*), he/she should be aware that these operations are not compatible with the idiomatic meaning of the idiom ('to die'). As a result, such variants do not make sense (see also Newmeyer, 1974). Glucksberg (1993) further shows that, even with decomposable idioms, substitutions of a constituent with a synonym

might still be very tricky and usually do not allow for the full range of synonyms. Consider ‘*crack*’/‘*melt*’/‘*shatter*’ the ice vs. ‘*grind*’/‘*shave*’ the ice (all formally variations of *break the ice*), in which the first set of synonym substitutions are relatively acceptable but not the second. This is because it is possible to *crack* (if the awkward situation is only just warming up) or *melt* (if the change in the social atmosphere was gradual) but not *grind* or *shave* the proverbial ice because the last two substitutions are not compatible with ‘social situation’ scene. This suggests the constraint of semantic compatibility on idiom variants, which dictates that the novel expression can only be recognized as a variant of an idiom if the newly computed form is semantically compatible with the figurative meaning of the canonical idiom.

Besides compatibility, grammaticality is also brought up in early studies on transformational restriction (Burger, 1973; Newmeyer, 1974; Fleischer, 1982). The constraint of *grammaticality* highlights the fact that the formal structure of an idiom variant must be in accord with the grammar of the language (Langlotz, 2006: 222). Even in cases where an operation is compatible with the idiomatic meaning (i.e. semantically compatible), if it renders the idiomatic variant is not grammatically correct, it will not be accepted. For example:

start a project → *the project was started*
get the show on the road → **the show was got on the road*

The constraints of compatibility and grammaticality are argued in this thesis to be justifiable but not yet empirically tested. From a purely theoretical point of view, semanticity and grammaticality are prerequisites not just for idiom variants but for any linguistic expression to make sense in a particular language. An idiom variant, like any stretch of language, expresses a message that needs to be communicated to the reader/listener, which means it needs to be placed in a normal linguistic framework (Quirk et al., 1985). It is, however, unclear whether grammaticality and semanticity is required for all idioms and all types of variants, or if there are exceptions to this constraint. Better still, what if the above example is the exception that is not representative of the true behaviour of idiom variants in discourse? The human brain has long been recognised as capable of accounting for creativity (particularly in poetry and literary texts) that deviates from normal principles of language processing in terms of both grammaticality and semanticity (e.g. *eyes wide shut*, *a bitter warmth*). Particularly, with the appropriate context, ‘unusual’/ ‘anomalous’ uses of language can easily become accepted thanks to the language user’s tolerance towards creativity and novelty. Burger et al. (1982) (quoted in Langlotz, 2006: 215), therefore, claim that the context-specific variability of an idiom is virtually open-ended (i.e. no constraint at all), precisely because any idiom can be subject to any type of alteration, both systematic and creative wordplay.

Whether or not the capability of overriding grammatical and semantic rules of creative language applies to idioms and idiom variants is not yet fully tested. As a result, while recognisability and functionality are treated as required factors of idiom variants,

semanticity and grammaticality will be taken as non-compulsory and subject to further analysis to see whether grammaticality and semanticity are global constraints on idiom variants or local requirements in certain cases. Provided that a variant satisfies the recognisability and functionality requirements, can it be either grammatically anomalous or semantically incompatible (or even both) and still work in the context? Corpus data is used in chapter 6 to find the general pattern of grammatical and semantic constraints on idiom variants in context.

3.7.4. *Non-ambiguity, non-conspicuity and evocational autonomy*

These three constraints are postulated by Langlotz (2006) to especially distinguish systematic variations from idiomatic wordplay. Systematic idiomatic variations in his view must be straight-forward instead of creating stylistic effects via striking deviant, multi-referential expressions. The literal scene of the variant must also not establish direct reference to the target conceptualisation, i.e. must not be context dependent.

Since our treatment of idiomatic creativity is strictly the opposite (i.e. idiomatic creativity at its best is striking, purposeful and creates stylistic effects), these constraints will not be analysed in our data. Recognisability, functionality, compatibility and grammaticality, on the other hand, will be scrutinized in the data in chapter 5 and 6.

Summary

The two constraints on idiom variants are preliminarily argued to include recognisability and functionality. That is, an idiom variant is only valid if (1) it can be recognised as deriving from a canonical idiom, and (2) it produces a comprehensible difference in interpretation so that a clear intention/message can be communicated. The functionality constraint guarantees the intentional nature of the variant, while the recognisability constraint ensures that the canonical idiom is still accessible so that the figurative meaning is activated. Together, they restrict the deviation so that the essence of ‘idiomatic creativity’ is preserved and the variant is perceived as such.

Semantic compatibility and grammaticality are theoretically required as well for an idiom variant, such that the alterations (syntactic or lexical) must be grammatical and compatible with the idiomatic meanings of the constituents and/or the idiomatic meaning of the idiom as a whole.

Although a consensus about what determines idiomatic creativity is more or less achieved with the aforementioned list of the recurring and most fundamental factors, the weight of each factor in each type of alteration and with each class of idiom has not been firmly established. Admittedly, psycholinguistic experiments (and probably follow-up interviews) are needed to identify the patterns of cognitive constraints on variants. Purely linguistic data, on the other hand, can validate the systematicity of the proposed constraints, or the

lack thereof, in authentic discourse. It is postulated that although the proposed constraining factors have their own merits, they vary considerable according to the type of variants as well as the nature of the canonical idioms (whether they are motivated, analysable, well-formed or not). Simultaneously, it is envisaged that the roles of these constraints also vary considerably under the influence of context. These aspects of the relationship between the facilitating and constraining factors of idiomatic creativity and context will be addressed in chapter 5 and 6. Before that, an account of context in idiomatic creativity studies is provided below.

3.8. Contexts and its role in idiomatic creativity

3.8.1. *Theories of context in the literature*

The role and significance of 'context' in pragmatics and discourse analysis cannot be overestimated and has indeed been reiterated in the literature (e.g. Firth, 1957; Halliday, 1978, 1985). Language users as well as analysts have to relate the relevant contexts to make the correct pragmatic interpretations of a particular utterance. To underscore the importance of context in language use, Malinowski (1923: 307) holds that 'a statement, spoken in real life, is never detached from the situation in which it has been uttered... the utterance has no meaning except in the context of situation'. Similarly, Mey (1993: 186) maintains that 'any understanding that linguists can hope to obtain of what goes on between people using language is based, necessarily and uniquely, on a correct understanding of the whole context in which the linguistic interaction takes place.' More specifically, context is argued to be a communicative element that helps with 'narrowing down of the communicative possibilities of the message as it exists in abstraction from context' (Leech, 1990: 60). Context, as a result, is seen as an integrated component of the interpretation process of an utterance (Saed, 1996).

Despite its important role in discourse, whenever 'context' is mentioned, it is implicit which aspect of the broad concept of 'context' is being referred to. This is because 'context' is a complex concept that can mean remarkably different things within different frameworks. Indeed, the role and scope of context vary in different approaches to language use. It has been used to refer to everything from clauses preceding the target linguistic unit within a sentence to multiple socio-cultural issues. According to Auer (1995), in a pre-theoretical, but intuitively plausible way, context is suggested to have five dimensions:

- (i) Linguistic contexts: sometimes called co-texts, also including the intertextual relationship between texts produced on different occasions
- (ii) Non-linguistic sense-data in the surroundings of the linguistic activity: the situation in a physical sense – i.e. physical surroundings of the speech situation, including the 'things' and 'events' in the co-participants' sensual (particularly visual) reach.

- (iii) Features of the social situation: constellation of participants, their social roles and the social activity they are engaged in
- (iv) Features of participants' common background knowledge other than (i)–(iii): e.g. characteristic to a certain profession, a neighborhood, a 'subculture' or knowledge on how to behave properly within a given (ideal) community which is shared by all its members, etc.
- (v) The channel of communication (the medium): mainly speech and writing, but also telephone, telegram, e-mail, automatic answering machines, etc.

Specific accounts of context will include and place the focus on some (or all) of these elements of context while excluding some others as a result of the disciplinary bias and particular goals of the researchers. Two main approaches include:

3.8.1.1. Context as a predetermined aspect of communication

In a large number of theories of context, context is assumed to be 'given' and that propositions are processed against a predetermined contextual background. As such, knowledge of contexts for information processing is a prerequisite. The view of *context as knowledge* is based on the assumption that the use and interpretation of language is guided by the language user's knowledge of the world, including its mental, social and cultural aspects. Broadly speaking, Leech (1983: 13) characterizes context as 'any background knowledge assumed to be shared by speaker and hearer and which contributes to the hearer's interpretation of what speaker means by a given utterance.' Lyons (1977: 574), on the other hand, specifies the type of 'knowledge' in language use to the knowledge of *role and status, location, formality level, the medium (spoken or written), subject matter, and province (or domain)*. Levinson (1983: 13) restricts context to the basic parameters of the context of utterance (*participants' identity, role, location, assumptions about knowledge*), claiming that such an approach to context is 'an introduction to the philosophical-linguistic tradition' rather than 'an exhaustive coverage of all the contextual coordinates of linguistic organization.' The 'ethnography of SPEAKING' model by Hymes (1962) spells out the sixteen components of context, which are grouped into eight groups:

- (i) **Setting and scene:** Setting refers to the time and place of a speech act and, in general, to the 'physical circumstances'. Scene is the 'psychological setting' or 'cultural definition' of a scene, including characteristics such as range of formality and sense of play or seriousness.
- (ii) **Participants:** Speaker and audience.
- (iii) **Ends:** Purposes, goals, and outcomes.
- (iv) **Act Sequence:** Form and order of the event.
- (v) **Key:** Clues that establish the "tone, manner, or spirit" of the speech act.
- (vi) **Instrumentalities:** Forms and styles of speech.

- (vii) Norms: Social rules governing the event and the participants' actions and reaction.
- (viii) Genre: The kind of speech act or event. Different disciplines develop terms for kinds of speech acts, and speech communities sometimes have their own terms for types.

The context-as-knowledge philosophy is markedly reflected in the treatment of context within 'speech act theory' (Austin, 1962; Searle, 1969), Gricean pragmatics (Grice, 1975) or 'interactional sociolinguistics' (Gumperz, 1982, 1992) as well. In Labov's (1972) 'variation analysis', however, context is seen as *situation and text*, with separate components including (1) the social situation (setting and scene), (2) the social identities (gender, age, and ethnicity), and/or (3) the key (formal vs. informal style), in which the first two elaborate on the 'situation' while the third details on the 'text' component of context. In other frameworks (e.g. 'conversation analysis', Garfinkel, 1967; Sacks, 1992), context is viewed as (1) *knowledge*, (2) *situation*, and (3) *text*. In conversation analysis, knowledge cannot be separated from the situation because it is knowledge 'in use' rather than an independent knowledge stored in the brain that is the essence of conversational analysis. Also, each utterance in a sequence is dependent on a prior context and creates context for the next utterance, hence its function as 'co-text' within a broader discourse. An even broader scope of contextual factors includes knowledge, situation, co-text, and others, in which the 'others' component can include various aspects of the language users' *mental*, *social* or *physical world* (Verschueren, 1999: 74-114). While the 'social world' relatively corresponds to the 'situational context' as defined elsewhere in other frameworks, containing social settings, institutions, cultural norms, and values, the 'physical world' focuses on temporal and spatial reference, and the physical properties of language users such as bodily postures, gestures, gaze, sex, physical appearance, etc. Particularly, the mental world activated in language use is argued to contain cognitive and emotive elements such as personality, beliefs, desires, wishes, motivations, and intentions.

In sum, the approaches to context presented above differ in their views of the scope of context and in their foci on different elements of context. This diversity is partially individual (depending on the particular goals of each research/study) and partially disciplinary. However, they all share the same underlying assumption that context is predetermined, intended by the speaker and is required to be correctly retrieved during processing for the successful interpretation of an utterance.

3.8.1.2. *Context as hearer's selection*

Representative of this approach to context is Sperber and Wilson's (1986) theory of relevance, which claims that context for information processing is not a prerequisite, but a consequence of communication. The determination of context, as such, is done by the hearer, and is part of the comprehension process in which context selection occurs at both

environmental and cognitive level. The range of possible contextual information to be drawn upon in the understanding process is delimited only by that which is deemed most relevant to pre-existing assumptions of the hearer, in order to achieve the goal of modifying and improving his/her overall representation of the world (ibid, 1986: 71). The types of information that Sperber and Wilson cite as sources of contextual information are: linguistic and perceptual information, and information in memory.

3.8.1.3. A combined approach

From this rich theoretical background of both speaker's intended context and hearer's selective context, it is conceded that both types of contexts are feasible. It is not impossible to assume that a speaker has a specific context in mind, which is available and predetermined in the situational, textual or cognitive aspects of the communication, when they make an utterance. For successful inferences, the hearer then must be able to utilize that same context. In a particular situation and with a particular topic in focus, it can be assumed that the hearer can focus on the appropriate contexts to identify the speaker's intended message by virtue of a conversational demand that the speaker provides the hearer with. However, as Sperber and Wilson (1986) admit, mutual knowledge is virtually impossible between two individuals due to differences in their cognitive contexts. Hence, there is not always guarantee for such a match in all transactions. The question to consider is how much context, or rather which contextual elements, the hearer (must) either know or retrieve from the universe of contextual factors that is located in their physical, mental, social and interactive environment. There is a real possibility of a rejection of speaker-intended topic and/or context for information processing, in favour of a topic/context that is more 'relevant' to the hearer's prior cognitions or existing representation of the world. Thus, the context can be pre-determined, and also a consequence of the hearer's context selection based on that which is relevant to them.

In this thesis, however, it is the predetermined contexts intended by the speaker that are the objects of investigation. This is due to the fact that the corpus data we use in this thesis is not capable of validating cognitive processes on the part of the hearer when inferring the meanings of variants. What we have available are utterances by the speakers, hence the intended messages embedded in the intended contexts by the speakers. This is not to say that we cannot make suggestions about the inference processes that might be available during idiomatic creativity comprehension; but that those suggestions are tentative.

The notion 'context' as employed in this thesis includes (1) the immediate physical situation of discourse, (2) social/cultural knowledge that is conventional within a speech community, and (3) shared knowledge between speaker/writer and hearer/reader, all of which might be implied or explicitly expressed in (4) discourse co-texts.

Besides its role in discourse in general, the role of context in creative language and in idiomatic creativity in particular has gained increasing recognition amongst linguists. A review of available research in this direction follows in the next section, highlighting the need for a more systematic characterisation of the subject matter via empirical analyses as well as the need to incorporate context in any account of idiomatic creativity in the future.

3.8.2. The role of context in idiomatic creativity

The role and significance of context in the interpretation of creative language in general and of idiomatic creativity in particular have been emphasized in recent literature. Creative behaviour has increasingly been perceived more as co-operative than individual, stressing on the work and play 'with and with respect to others' (Pope, 2005: 65), particularly in conversations and other genres of spoken discourse. This puts the receivers of creativity on an equally important position as the senders. As in the case of literary language, the success of a creative use of language depends on whether listener is willing to see it that way (Carter, 2004; Carter and Nash, 1990). This tolerance is to a considerable extent dependent on the listeners' ability to construct plausible interpretations of speakers' intentions (i.e. to comprehend the communicative purposes underlying a particular creative language). The interpretation process, in its turn, is argued to involve the integration of the linguistic, stipulated and allusional content with the discourse context (Glucksberg, 1993; Keysar and Bly, 1999).

Gildea and Glucksberg (1983), Glucksberg (1993) and Keysar and Bly (1999), indeed, contend that when the figurative meanings are not immediately available, a minimal context can make them so. By conducting a series of experiments using metaphorical sentences, albeit contrived, these authors were able to provide evidence for the fact that even minimally related contexts help the listener/reader to disambiguate the metaphors, making their figurative interpretation feasible. The noun 'question' in *pop the question*, for example, is canonically singular because one usually proposes marriage to only one person at a time, so that the pluralisation of 'question' as in **pop the questions* seems unmotivated in the context of the monogamous society we live in. Nevertheless, in the context of a polygamous community, the variant might make sense (Glucksberg, 1993) (e.g. *The president is also openly in favor of plural marriages, being married to three wives himself. He popped the questions to all three of them on the same day and had a big fat wedding that lasted for a week*).

Focusing on co-text, Gibbs (1980) and Howarth (1996) maintain that it is an important factor in the post-access disambiguation process, providing 'semantic clues for the interpretation of pre-selected idiomatic strings' (Howarth 1996: 54). For example, the expression *he's singing a different tune* can be made explicit in two different contexts, one of which involves the singer Jackson Browne, and the other involves former president of the US, Jimmy Carter (Gibbs, 1980: 150). The correct interpretation of the expression relies on the context expressed in the 'co-text'. Gibbs' experiments also show that while the

conventional use of idiomatic expressions is assumed to reduce dependence on co-text, innovative extensions are more heavily dependent on co-text, requiring extra prompting.

Convincing as these arguments sound, the issue of 'how minimal' the context can be has not been explored to satisfactory depth. What is valuable from these studies, nevertheless, is that context is recognised as playing an essential role in idiomatic creativity. Besides formal linguistic factors (i.e. transparency), general world knowledge, together with rules of discourse and convention seem as important for idiom flexibility and productivity. One of the research questions of the thesis is therefore to explore the extent of the influence of context on different types of idiom variants. It is hypothesized that this role is not equally distributed across all types of idiomatic creativity but varies according to the dependence or otherwise of the particular variant type on contextual information.

Summary

Perhaps the most important result from this body of research is that appropriate contexts facilitate the processing of idiom variants. However, little attention has been paid to the exact relationship between the psycholinguistic observations in the laboratory and the actual patterns of naturally occurring language concerning idiomatic creativity in context in authentic discourse. While the role of contexts in language communication and comprehension in general has been extensively investigated, its role in idiomatic creativity has mainly been suggestive or intuitive. Lost in this debate is the fact that a considerable amount of information has been amassed showing whether and when context facilitates, fails to facilitate, or actually inhibits the uses of certain types of idioms in certain types of transformational operations. Extensive qualitative analyses of corpus data will be carried out in chapter 6 to examine the specific roles of contexts on various types of idiomatic creativity.

Employing an inductive method, this thesis does not set out to rely on a fixed framework of context with predetermined elements. Rather, data will be analyzed so as to highlight the elements of contexts that are salient and need conscious attendance to from the hearer/reader. That is to say, the thesis assumes that any aspect of the physical situation, cognitive state or textual feature of a particular piece of communication is possibly salient, depending on each particular case.

3.9. Conclusions

Empirical research in the field of idiomaticity, particularly studies in corpus linguistic tradition, has convincingly demonstrated that idioms are by no means lexically fixed as held by earlier views of idioms. Instead, idioms are observed to allow certain alterations to their formal and semantic structures to form creative variants to serve specific communicative purposes. In this chapter, the theories concerning what enables such a flexible behaviour of

idioms in discourse are discussed in detail. Indeed, the debates in the literature with regard to the facilitating factors of idiomatic creativity have been prolific and diverse. The most influential and recurring factors have been suggested to include the semantic characteristics of *analysability*, *motivation*, and *literal well-formedness*. That is, the more analysable, motivated and literally well-formed an idiom is, the more likely it is to manipulate it to form creative variants. While this tendency is strong, criticisms have also been levied at the lack of systematic authentic data to back up the hypotheses. Counter-arguments have also been advanced with concrete examples of certain unanalysable, unmotivated and ill-formed idioms being flexibly used in discourse, while some fully analysable, motivated, well-formed idioms are fixed, thus highlighting the inconsistencies of the general trend. It is even suggested that these inconsistencies are not individual exceptions; they are systematic and indicate further dimensions of idiomatic creativity that need attendance to. Possible factors affecting idiomatic creativity besides semantic characteristics are proposed to include *context* (which allows unanalysable, unmotivated and ill-formed idioms to be creatively manipulated), as well as *cognitive constraints* on idiomatic variants (which restrict the flexibility of some of the most analysable, motivated, and well-formed idioms). The validity of such claims, however, needs substantiating using large bodies of authentic data. More importantly, the effect of semantic characteristics, context and cognitive constraints on idiomatic creativity has been investigated separately in individual research instead of collectively in one framework so that their interactions could be accounted for.

Given the current situation, there is a dire need for a two-stage research into idiomatic creativity that covers:

- (1) an empirical examination of the existence or otherwise of a correlation between idiomatic creativity and the degrees of analysability, motivation and literal well-formedness in idioms using naturally occurring data
- (2) an evaluation of the combined effect of the facilitating factors (analysability, motivation and literal well-formedness) with cognitive constraints (recognisability, functionality, grammaticality, and compatibility) in relation to context.

These two issues are addressed in chapter 5 and 6, respectively, in the hope that the patterns emerging from the analyses will reveal the creative potential of idioms in a more accurate manner than if these factors are treated separately.

Before such analyses, however, the data, methodology and analytical tools employed in this thesis are discussed in the next chapter. With the focus on naturally occurring data, the current research is corpus-based and involves both quantitative and qualitative stages of data analyses. The rationale for such an approach is presented, followed by a careful description of the stages of data collection, the analytical processes and their associated difficulties.

CHAPTER 4

DATA DESCRIPTION AND RESEARCH METHODOLOGY

This chapter gives a detailed account of the data sources and research methodology employed in this thesis. Following a short introduction to mixed method research, the unique mixed methodology implemented in this thesis is introduced, which consists of a quantitative, corpus-based hypothesis testing phase, followed by a qualitative corpus-based description of idiomatic creativity. The rationale for the choice of such a mixed methodology is provided, so is that for the choices of specific methods and/or techniques at various stages of data collection and analyses within each phase. The strengths and weaknesses of the mixed method research as adopted in this thesis to face the challenges of idiomatic creativity studies are also discussed to identify room for improvement in future research.

4.1. Mixed method research

In applied linguistics research methodology, the dichotomy of quantitative – qualitative research is one of the best known distinctions, which is based on the nature of the data (numerical versus non-numerical), the method of data collection, and the method of data analysis (statistical vs. non-statistical) as the result of the ideological differences towards research. In essence, in the quest to adopt scientific objectivity into social sciences investigations, the quantitative researcher typically uses predetermined numerical categories to code their variables so that formal statistical analyses can be carried out on large samples, thereby ironing out any individual idiosyncrasies (Dornyei, 2007). In contrast, the qualitative researcher uses flexible verbal categories that are subject to interpretative analyses via the researcher's individual sensitivity. Results from qualitative research 'focus on the unique meaning carried by individual organisms' (ibid: 29). Qualitative research is, therefore, typically used to *explore* a new phenomenon and develop an initial hypothesis, whereas quantitative research can be used to *test* a hypothesis in terms of the breadth of its distribution in the population (McCracken, 1988).

While the differences between the two approaches might appear relatively straightforward, researchers have pointed out that the distinctions are not always clear-cut and consistent, and that they might be seen as forming a continuum instead of extremes (Richards, 2005; Miles and Huberman, 1994; Sandelowski, 2003; Dornyei, 2007). It has even been suggested that, from a pragmatist point of view, it is possible to integrate the two methodologies to 'corroborate (provide convergence in findings), elaborate (provide richness and detail), or initiate (offer new interpretations) findings from the other method' (Rossman and Wilson, 1985: 627). Such a combination of qualitative and quantitative methods within a single

research project has been characteristically described as 'mixed method research' (Dornyei, 2007: 44). The combination can happen either in simple sequential arrangements (i.e. qualitative first followed by quantitative analysis, or vice versa), or in combination at various stages of data collection, analysis, or interpretation.

Naturally, a mixed method is not beneficial in all cases, since the choice of a specific method (either qualitative or quantitative or mixed method) is not arbitrary but heavily depends on the particular research questions or topics. However, it has been argued that 'mixing methods has great potential in most research contexts' (Dornyei, 2007: 30) for the following reasons:

- The mixing of methods is seen as combining the strengths of the two methods while eliminating weaknesses: The simplistic, decontextualised generalizations of quantitative research are neutralized by the depth of the context-specific interpretative results from qualitative research, in the same way that the sampling bias in qualitative research due to unrepresentative, small samples can be improved by large samples and statistical analyses in which idiosyncratic differences are cancelled out.
- The mixed method research offers multi-level analysis of complex issues: While the natures of some research questions or topics are more naturally linked to one of the two methods, it is observed that full understandings of certain complex phenomena can only be gained from converging numeric trends from quantitative data and specific details from qualitative data. As a result, different facets of a phenomenon can be measured, thus producing a fuller picture of an issue.
- The mixed method research has a unique potential to produce improved validity: It is argued that if the same conclusion can be reached about a certain issue via different methods, the convergence offers strong validity evidence by allowing the investigators to assess information from various angles, using various data types and via various data analysis processes.
- The final results produced by combined qualitative and quantitative methods appeals to a wider audience.

Although mixed method research is not without pitfalls (see Hesse-Biber and Leavy, 2006; Maxwell and Loomis, 2003), it is argued that '*principled mixing*' should help ensure the compatibility of mixed method research with different research paradigms and aggregate their strengths. As a result, Johnson and Turner (2003), for instance, maintain that researchers *should* collect multiple data using different strategies, approaches, and methods but bearing in mind the fundamental principle of mixed method research, i.e. the resulting mixture or combination has to result in complementary strengths and non-overlapping weaknesses. Lazaraton (2005: 219) also encourages mixed method research by stating that

he 'would also hope that we would see more studies that combine qualitative and quantitative research methods, since each highlights 'reality' in a different, yet complementary, way'.

Due to the differences between qualitative (inductive) and quantitative (deductive) analysis, researchers are often advised to keep the analyses separate and only mix the results at a later stage (e.g. final interpretation) (Dornyei, 2007). In some cases, however, data can be transformed (i.e. qualitative into quantitative data and vice versa) to allow for mixed analyses (Greene et al., 2005; Miles and Huberman, 1994; Tashakkori and Teddlie, 1998).

4.2. The mixed methods used in the thesis

As is frequently emphasized, the pragmatic foundation of mixing methods is centered around the purpose of the research. With that in mind, it is worth revising the research questions that this thesis seeks answers to and how they could best be achieved. The main research question is:

What are the facilitating and constraining factors of idiomatic creativity in real life discourse and how do they interact?

To answer this question, the following components, each addressing one facet of the phenomenon of idiomatic creativity, need to be resolved:

- [1] Since literal well-formedness, analysability and motivation have been suggested in various psycholinguistic experiments to facilitate idiomatic creativity, it is the goal of thesis to empirically test this hypothesis, i.e. to see if there is a correlation between these factors and idiomatic creativity in authentic English discourse.

The underlying rationale for this research goal relates to the relationship between psycholinguistic experiments and corpus linguistic analyses, which are not to be assumed to produce the same results about linguistic phenomena. Experiments are often carried out to test the psychological reality of corpus linguistic patterns, whereas corpus analyses are needed to validate the robustness of psycholinguistic findings in large bodies of authentic data. As a result, the facilitation (and/or inhibition) effects of transparency as observed in the laboratory might or might not reflect the patterns of idiomatic creativity in authentic discourse, which is to be uncovered with corpus data and analyses (see section 4.3 for rationale and methodology of corpus linguistic approach in this thesis).

To test the hypothesis of the correlation between idiom transparency and idiomatic creativity in authentic discourse, quantitative data and analyses will be necessary. A large sample of idiomatic creativity will be elicited from corpora to ensure the representativeness of the data. Due to the size of the sample (100 cases with 1,300 variants), any individual

idiosyncrasy will be ironed out, therefore guaranteeing that the data pool reflects the commonalities of language in use.

The data collection process that involves the elicitation of idiomatic creativity from corpora itself reflects a mixed methodology. In the absence of an autonomous tool to carry out such a task, I had to resort to a host of techniques, including concordancing different combinations of lexical items and grammatical structures and manually processing concordances to identify creative uses of idioms using informed intuition. (Specific steps in data collection are explained in 4.5.)

The corpus data will be quantitatively analysed in chapter 5 so that any correlation between idiomatic creativity and the transparency of idioms can be empirically validated. Both frequency counts and statistical tests using SPSS Spearman's correlation are used in this chapter (see 4.6 for more details on quantitative methods).

The second component of the research question relates to the search of significant patterns of idiomatic creativity in relation to its facilitating and constraining factors:

- [2] The effects of context on idiomatic creativity have been suggested in various studies but have not been explored to satisfactory depths in naturally occurring data, particularly not in relation to other determinants of idiomatic creativity (transparency, cognitive constraints) to examine the co-effects of these factors on idiomatic creativity in authentic discourse. Hence, the second goal of the thesis is to initiate a new hypothesis of a context-based model of idiomatic creativity that takes into account the internal characteristics of idioms (transparency) and external cognitive constraints of idiomatic creativity.

As can be seen, aspects of the second research goal can only be sketched via qualitative interpretation of significant patterns of corpus data. As a result, qualitative analyses will be carried out with the corpus data pool, supplemented with additional data from the, to explore patterns of idiomatic creativity in each type of transformation (i.e. passive, modification, substitution, etc.) to draw up a pragmatic model of idiomatic creativity under the effects of their determinants in authentic English discourse. Qualitative analyses here mean introspection (guided by rigorous analytical frameworks and informed by existing literature on related subjects), and independent judgment by native speakers (see 4.6 for more detailed discussions on these qualitative techniques).

On balance, due to the two different components of the research goals as stated above, a mixed method research is believed to benefit from both worlds, helping the thesis reach the goals set out for it. By employing the quantitative method where applicable, I was able to (1) test the existing hypotheses of the relationship between the internal characteristics of an idiom and its creative behaviour in authentic discourse, (2) achieve high levels of reliability

and validity, and (3) arrive at conclusions that are objective, and well-defined by minimizing the subjectivity of judgment. The qualitative method, at the same time, is employed to allow for the emergence of new information and interpretations of idiomatic creativity in naturally occurring language, thus providing a thorough understanding of the phenomenon as a more holistic and complex construct that is context-specific.

In 4.5 – 4.6, detailed descriptions of the techniques and methods involving different stages of the research, from data collection to data analysis and interpretation, are given to establish the validity and reliability of the thesis' methodology. Before that, the rationale for the use of corpora in this thesis instead of any other sources of data is explained below.

4.3. The rationale of the use of corpora in the thesis

Within a usage-based model of language, the observation of large amounts of corpus data is viewed as a viable method of grammatical description. The use of corpus data illustrates the focus on 'performance' or 'e-language' (externalised language) (Chomsky, 1965) as it is manifested in corpora. 'Competence' or 'i-language' (internalized language) as it is represented in the speaker's mind will not be the approach in this chapter. In Sinclair's (1991) view, the availability of large-scale corpora has enabled us to identify the most typical patterns amongst the chaos found in real language production without resorting to idealizations (Sinclair, 1991:103). Indeed, as Adolphs (2008) argues, the use of corpus data in pragmatic investigations is particularly relevant and highlights many actual language-in-use features that contradict theoretical descriptions of language which are based on invented examples. Corpus data, therefore, is the way forward into the 'real' competence as opposed to the 'idealized' competence proposed by any linguistic theory.

In the area of multiword units, figurative language and idioms, evidence of language in use produced by corpus-based investigations has been found to contradict unaided intuition about collocations (Sinclair, 1991), metaphors (Deignan, 2005), and FEIs (Moon, 1998). While trained intuition is acknowledged as a valid research method, the reported discrepancies between intuitive claims and corpus evidence in many cases call for corpus-based investigations of language as it happens in real life.

Besides the advantage of being truthful representations of the language landscape, corpora have enabled objective information about language in use that is not readily intuitively available to the native speaker, such as probabilities, frequency counts, etc. (Leech, 2001). As a result, Leech maintains that probabilistic language processing systems produce a more accurate result, even on unrestricted input data, in a way that 'outperforms most rule-driven language modeling systems' (ibid: 17). Cognitive linguists might argue that a machine for manipulating probabilities cannot represent a model of human linguistic knowledge. In other words, the probabilities provided by computer programs, due to their lack of human knowledge and cognitive abilities, can achieve a limited resemblance to human ability.

However, corpus linguists have never claimed to be able to characterize the psychology and cognitive aspects of language. What is achieved is a model of human linguistic *behaviour* – although behaviour can tentatively be argued to also provide some insights into the psychology and cognitive processes of language, on the grounds that behaviour reflects psychology to certain extents.

By using corpus data in this thesis, I hope to bridge the gap between findings from psycholinguistic experiments and intuition about idiomatic creativity and the actual behaviour of this phenomenon in authentic discourse. It is believed that corpus data can always offer more in-depth understandings of the subject matter via exceptions and special individual cases that fall outside the patterns, thus helping elaborate on the general trends. The corpus linguistic approach in this thesis is another attempt to collect linguistic facts from which underlying systems concerning idiomatic creativity can be traced.

Summary

With the goal of sketching a context-based model of idiomatic creativity in authentic English, corpora are the ideal sources of data. In order to uncover underlying facilitating factors of idiomatic creativity as well as its possible constraints, it is essential that a large collection of variants of idioms be accessible for analytical purposes. In many studies, idiom variants (e.g. passives) have been conveniently made up by the linguists to illustrate points of arguments about possibilities and restrictions on particular operations on idioms. While this technique has certain merits, it fails to capture the more random, creative, context-dependent variants that corpora can supply (particularly wordplays with idioms). The corpus linguistic approach, therefore, is adopted in this thesis to ensure the authenticity as well as the variety of the data, which in turn ensures the true-to-life analyses of a context-based model of idiomatic creativity in authentic English.

There are different ways a corpus can be of use in a study, depending primarily on the specific goals of the research. Leech (2001:14) summarizes the roles of corpora as ranging from being ‘simple tools’ for sorting and counting data (while all the serious data analysis is performed by the human investigator) to being ‘the main agent implementing the discovery procedure’ (while the human analyst provides no linguistic insight, just programming ability). The next section will describe how corpora are used in general and in this thesis in particular.

4.4. The role of corpora and the human analyst in the thesis

To emphasize the various ways corpora can be used in linguistic research, Leech (2001: 18) details four different paradigms of research balance between the role of the human analyst and the computer:

- (1) data retrieval model: machine provides the data in a convenient form, human analyses the data
- (2) symbiotic model: machine presents the data in (partially) analysed form, human iteratively improves the analytic system
- (3) self-organising model: machine analyses the data and iteratively improves its analytic system, human provides the parameters of the analytic system and the software
- (4) discovery procedure model: machine analyses the data using its own categories of analysis, derived by clustering techniques based on data, human provides the software

A different but not unrelated dichotomy is also used to describe the ways a corpus linguistic analysis is carried out; that is, corpus-based vs. corpus-driven approaches. Other terms used including corpus-informed (Carter and McCarthy 2006: 11-12), corpus-assisted (Stubbs, 1996; Partington, 2006). By elaborating on the distinctions and relationship between both approaches, it is then possible to justify the approach within which this the current thesis can be positioned.

Corpus-based research, in essence, uses corpora as a resource from which to 'expound on, or exemplify, existing theories, that is theories which were not necessarily derived with initial reference to a corpus' (Tognini-Bonelli, 2001: 74). The approach lends itself to application in many linguistic fields. Corpus-based study is a valuable tool which allows researchers to validate and elaborate on theoretical claims.

Corpus-driven research, on the other hand, is inductive, attempting to approach data with no explicit preconceptions as to what it will find, hence allowing patterns to emerge from the data itself, producing descriptions wholly consistent with corpus data in so much as they are informed wholly by it (ibid). Corpus-driven research arises from the observation that traditional grammatical categories do not always fit with naturally-occurring data, so that new descriptions of language are required from authentic language. Corpus-driven research usually relies on frequency in indicating significance, as well as a focus on lexico-grammatical patterning, with insights emerging through concordances, keywords, word-frequency lists, and clusters (Sinclair, 1991, 2004).

Both approaches rest largely (though in different ways) on frequency. While the corpus-based approach describes language through investigating feature frequencies and distributions (though these are not autonomously generated by computer programs but done manually by the analyst), the corpus-driven approach uses computer-generated frequencies (rather than existing theoretical categories). It is sometimes argued that the two approaches do not necessarily have to be mutually exclusive. Indeed, McEnery et al. (2006: 11) and McEnery and Gabrielatos (2006: 36) suggest that the boundary is 'fuzzy' or

'overstated' and that their relationship is best represented as a cline, since many research can be seen as both corpus-based and corpus-driven.

Within this context, the current thesis is best described as corpus-based. In the first part of the thesis, i.e. the hypothesis testing using quantitative methods, corpus data is used as raw material that can be quantitatively described and fed into statistical tests so that hypothesis can be validated or nullified. The second part of the thesis, in exploring the creative uses of idioms in authentic discourse, uses corpus data to characterize idioms according to a predetermined set of parameters (context, transparency, cognitive constraints). Most of the qualitative research is done using introspection and inter-raters. The methodology of the thesis illustrates the 'mixed methods research' approach, in which the corpus is regarded as '*one of the linguist tools, to be used together with introspection and elicitation techniques*' (Johansson, 2001: 313), since '*wise linguists, like experienced craftsmen, sharpen their tools and recognize their appropriate uses*' (ibid). If we have to further elaborate the role of corpora in this thesis using Leech's (2001) system, I would say it is the first level, i.e. the data retrieval model in which machine provides the data in a convenient form, human analyses the data.

Admittedly, corpora do not have all the answers to creativity. Socio-cultural or cognitive information about creativity cannot usually be directly elicited from corpora or by corpus techniques. However, the speed and the level of sophistication with which corpora annotation is evolving today mean that more layers of social and cultural information are being added, which will arguably bridge these gaps in the near future.

4.5. Step-by-step techniques in data collection

Generally, corpora are praised for their capability of performing endlessly repetitive tasks faster and more accurately than any manual method does. In this thesis, however, economy is not the strongest value of corpora, because computer programs to extract and analyse creative language in general and idiomatic creativity in general have not yet been fully developed. At the moment, advances in the field of corpus linguistics have mainly been associated with patterns and trends. Research studies that have been carried out in the field of creativity and idiomatic creativity have been mostly limited to the data retrieval model, i.e. corpora provide the sources of authentic uses of language from which the human analyst manually retrieves creative linguistic uses via a combination of searches. In the absence of a fully developed computer program that can autonomously carry out this task given the canonical idioms, the current thesis is no exception to this trend. In what follows, the techniques involving the elicitation of creative uses of idioms in corpora are explained step-by-step.

4.5.1. Data collection methods and associated difficulties

In creativity studies, particularly idiomatic creativity, the roles of computer tools have unfortunately been limited with regard to the collection of data. Defined as ‘a large and principled collection of natural texts’ (Biber, Conrad and Reppen, 1998: 12), ‘to *represent*, as far as possible, a language or language variety as a source of data for linguistic research’ (Sinclair, 2004 – emphasis added), a corpus is characterized by its truthful representation of the naturally-occurring discourse as it is produced in everyday life activities. It entails a generally accepted fact that at the heart of corpus research lie *patterns of language* and *recurrences of linguistic items*:

Corpus linguists [...] have typically seen their goal as the uncovering of recurrent patterns in the language, [...] with probability of co-occurrence, [...] with fluency in language rather than creativity (Hoey, 2005: 152)

This bias towards patterns in corpus linguistics can be attributed to the popular belief that uniqueness ‘cannot be observed with certainty in a corpus, because uniqueness in a corpus does not entail uniqueness in a language’ (Sinclair, 2004). Since creativity is essentially characterised by ‘newness’ and ‘unexpectedness,’ or, in other words, ‘uniqueness,’ it appears that corpora and creativity cannot benefit each other. Linguistically-variable phenomena determined by meaning (instead of surface lexis) (e.g. semantic prosody, creativity, metaphors) are not appreciated by computers (Groom, 2006; Deignan, 2005) and thus cannot automatically be identified and retrieved.

In the specific case of idiomatic creativity, Philip (2006) admits that ‘there is a degree of incompatibility between a methodology which favours the identification of recurrent patterns, and the search for non-recurring variants’. In fact, one of the most commonly cited drawbacks of corpus related research into phraseological variation is that ‘you find what you look for: search tools will only match the pattern sought’ (Moon, 1996: 252). As a result, the location, or rather the discovery, of non-canonical forms in corpora is still generally considered to be a matter of good fortune (Moon, 1998: 51).

A data collection technique proposed by Deignan (2005) in her quest to identify metaphors in corpora will be replicated, since it fits the purpose of the current thesis of locating creative uses of idioms in corpora. That is, to use a smaller corpus, which can be read in its entirety, and working from it into the larger corpus using the identified patterns found in the small corpus. The benefits of this technique are manifold. First of all, it reduces the intuitive processes involved as well as the reliance on previous research results, hence an improved objectivity and validity of the results. Second, as Deignan herself admits, there are added benefits of using small corpora, which ‘usually provide more detailed information about context, so adds to the richness of interpretation’ (2005: 93). Last but not least, the use of a smaller corpus to identify the initial set of idiomatic creativity before they are

concordanced in larger corpora undeniably helps avoid the issue of ‘convenience sampling’ (Miles and Huberman, 1994; Read and Nation, 2004), which is often criticised as easily leading to biased, unrepresentative data.

As a result, a small corpus is compiled specifically for the purposes of this thesis, comprising of a range of authentic texts, thus ensuring the randomness of the selection as well as the authenticity and representativeness of the data. Idioms and idiomatic creativity will be located in this corpus manually before a manageable sample can be chosen and further searched for in large corpora. The sample is chosen in such a way that it reflects the composition of the population it is taken from. This method of sampling is called *stratified sampling*. Stratification is the process of dividing members of the population into homogeneous subgroups before sampling. During sampling, proportionate allocation uses a sampling fraction in each of the strata that is proportional to that of the total population. For instance, if the total number of idioms in the corpus consists of 60% in the speech stratum and 40% in the writing stratum, then the relative size of the two samples (54 spoken idioms, 46 written idioms) should reflect this proportion. A stratified survey could thus claim to be more representative of the population than a survey of simple random sampling or disproportional systematic sampling.

4.5.2. *Compilation of a small corpus of spoken and written texts*

The selection of spoken and written texts from which the initial idiom set is identified is of various genres in both the ‘popular’ and ‘serious’ texts types (Semino and Short, 2004: 22). In Carter’s (2004) mapping of creativity in everyday talk, popular types would include ‘socialising’ or ‘intimate’ contexts while serious genres generally materialize in ‘transactional’ or ‘professional’ contexts. The rationale behind the choice of both ‘popular’ genres and ‘serious’ genres lies with the goals of the thesis, i.e. to have a set of data that is large enough to contain adequate examples as well as encompassing enough to be representative of everyday language use. While ‘popular’ genres (comedy shows, TV entertainment interviews) are marked with ‘entertainment’ and/or ‘intimate’ elements that generally promise more creative usage of language than transactional, serious genres (tourist commentaries, academic discussions) (Carter, 2004), the addition of the serious genres adds to the data set their own characteristic idiomatic uses. As mentioned earlier, since the idiom data set is meant to be representative of authentic English discourse, it is felt that a balanced composition of both popular and serious texts will provide for a non-discriminatory set of idioms that is valuable for analysis.

4.5.2.1. *Spoken texts*

Spoken texts constitute a cline ranging from naturally occurring speeches to near naturally occurring speeches to scripted, written-to-be-spoken texts, covering a wide range of genres, including tourist commentaries, TV interviews, comedy shows, movie scripts and academic

discussions during PhD supervision sessions. Totalling nearly 100,000 words, the texts provide a valuable source of data into spoken idiomatic creativity which have been largely under-researched in the literature. Specific descriptions of spoken data used in this thesis include:

- 3 complete academic discussions in the form of supervision sessions, totalling approximately 30,000 words. These are live, naturally occurring discourse in a more or less formal manner and having academic topics.
- 4 pieces of live tourist commentaries, recorded on 4 different tour buses in London, totalling 2 and half hours talking time (approximately 16,000 words). These are on-the-spot, naturally occurring speeches. It might be argued that parts of the commentaries might not be truly online naturally occurring speech but rather rehearsed due to the fact that the guides might have given a commentary to a particular attraction more than once. However, there is no clear evidence of which parts might have been 'memorised' and which part have not. Hence, the pieces are treated as 'near naturally occurring discourse' due to the possible elements of rehearsedness.
- 3 complete television interviews, totalling nearly 3 hours talking time (approximately 28,000 words), all directed by contemporary British comedian Ricky Gervais. The interviews are original, unedited, and verbally unscripted as claimed by the producers. The prominent feature that these interviews share is that the roles of interviewer/interviewees are often reversed where the guests would ask the host questions about his career. As a result, the interviews resemble friendly chats between colleagues more than formal interviews. There were awkward moments, allegedly due to the unedited and unscripted nature of the interviews (e.g. overlaps, latching, facial expressions and long pauses and gazes and fake laughs, etc.). As a result, these interviews represent 'naturally occurring speech'.
- 1 complete comedy movie script, *For Your Consideration*, which is 1 hour 30 minutes in length (approximately 14,000 words). The comedy (starred and directed by Christopher Guest) is verbally improvised. In other words, the uses of linguistic expressions in general and idiomatic expressions in particular, (if any, either in canonical forms or creatively) are random, unplanned and on the spot, thus comparable to the 3 interviews described above in terms of near natural speech quality. Similar to the tourist commentaries, certain language use in this movie, although unscripted, might be subject to rehearsedness due to the repetition of the process of filming each scene. Language in this movie, as a result, is classed as 'near natural'.
- 1 complete season of the comedy show *The Office* (starred and directed by Ricky Gervais). This is strictly scripted, hence having the cognitive processing of written discourse despite the spoken forms/characteristics of 'written to be spoken' discourse.

In terms of spontaneity/scriptedness, some speeches can be heavily scripted (hence the status of written-to-be-spoken speech) while others can be completely spontaneous and in-between types naturally occurring data with an added awareness of an audience).

4.5.2.2. Written texts

The written component also totals approximately 100,000 words in both popular (short stories, entertainment news) and serious genres (travel guides, political satires, and academic essays). Break-down of the written data is as follows:

- A collection of short stories (14,000 words) from online magazines, by various contributing authors. Each story is between 1,500 – 2,000 words, telling a story about Christmas.
- Entertainment/celebrity news from e-magazines (20,000 words), containing miscellaneous entertainment news (fashion, cinema, celebrity, etc.) and ranging from 200-600 words each.
- Web-based travel guides to various tourist attractions (16,000 words) from four websites, including Thomson, Expedia, Thomas Cook and Co-op Travel.
- Political satires (24,000 words) from online newspaper columns. Each piece is by a different writer and of a different topic but they are all politically-related. Only about 500 words in length each, these satires are short and witty.
- Academic essays/reports (13,000 words) by university students of an undergraduate course at the University of Nottingham and Nottingham Trent University. Topics include various aspects in the field of economics and finance.

These different data types are not separated, but treated as part of a common linguistic databank. Collectively, in terms of variety, the data are rich, varied, representative (to a certain extent) of authentic English discourse of different genres, in different contexts.

The spoken and written texts in this small corpus are read through so that uses of idioms can be located. Even though reading through the corpus to locate idiomatic creativity is a qualitative process that relies on the researcher's intuition, it is strictly guided and informed by an extensive understanding of what 'idioms' and 'idiom variants' are (see chapter 2).

4.5.3. Manual identification of idioms in the self-compiled corpus

It has been brought to the attention of the linguistic community that when idioms are involved, intuitions are thought to be very often resorted to in a systematic, conscious manner so that an expression is judged as idiomatic if it 'sounds right', 'regularly considered by a language community as being a unit' (Moon, 1997: 44). It is true that in many cases, it is fairly easy to identify very familiar idioms, particularly if they are used in their canonical forms. However, in other cases (unfamiliar idioms or creative variants) intuitions might not

be extremely reliable. To many linguists, 'human intuition about language is highly specific and not at all a good guide to what actually happens when the same people actually use the language' (Sinclair, 1991: 4). As a result, to ensure that my intuitions will not affect the process, all figurative uses of language (i.e. non-literal expressions) are marked and checked for their idiomatic status using The Free Dictionary website by Farlex at <http://idioms.thefreedictionary.com/> which is an online database compiled from the Cambridge International Dictionary of Idioms (containing 7,000 entries), the Cambridge Dictionary of American Idioms (5,000 entries) and the McGraw-Hill Dictionary of American Idioms and Phrasal Verbs (24,000 entries). The reason why this website is chosen is because it is probably the world's largest corpus-based online database of idiomatic expressions currently available. The two Cambridge publications are based on the Cambridge International Corpus (CIC) of more than 1 billion words of written and spoken texts. The texts come from novels, newspapers, non-fiction books, websites, junk mail, TV and radio programs, recordings of people's everyday conversations, etc. on a wide range of topics. The dictionaries as a result show a strong commitment to the authenticity of language, offering examples of the uses of idioms in real life discourse. Although the McGraw-Hill volume does not use corpus information, it takes advantage of the large database of idioms and phrases from a selection of previous McGraw-Hill's publications, which makes it one of the most comprehensive of its kind. The fact that the website is easy-to-navigate with a user-friendly format makes searching for idioms quick as a wink, an obvious advantage over the manual method of consulting printed dictionaries.

More importantly, the definition of 'idiom' in the Cambridge dictionaries also coincides with ours in this thesis, having fixedness/institutionalisation, compositeness and non-compositionality as their criteria:

An idiom is a phrase whose meaning is different from the meanings of each word considered separately. These phrases have a fixed form – they usually cannot be changed – and they are often informal, but they can also be slang, rude slang, or even slightly formal. [...] Not all fixed expressions are idioms. For example, *close your eyes* is a common fixed phrase, but it is not an idiom because each word in it is used in its standard meaning. The phrase *keep your shirt on* is an idiom, however, because the phrase does not mean "do not take off your shirt" – it means "stay calm" (Cambridge International Dictionary of Idioms, 1998: vi)

As suggested by Deignan (2005), the idioms and their patterns of creative uses as identified in the small corpus can then be individually searched for in larger corpora to build up a rich data pool of idiomatic creativity which serve to facilitate answers to our research questions about the correlation between idiomatic creativity and the internal characteristics of idioms (motivation, analysability and composition), as well as the roles of constraining factors (recognisability, compatibility, functionality and grammaticality) and contexts on idiomatic creativity.

4.5.4. Selection of idiom sample set via stratified sampling

The total number of idiomatic uses found in spoken texts is 208, while that in written texts is 141 (see appendix 1 and 2 for detailed lists of these idioms). To analyse the variable behaviour of all 349 idioms in corpora is impossible, since the efforts associated with finding variants of one idiom alone is considerable (as we shall see in 4.5.5). As a result, out of the 349 idioms in both spoken and written texts, a selection of 100 idioms is singled out as the *actual sample* that will be subject to further corpus analysis. The sample is chosen using stratified sampling technique, so that the sample reflects the composition of the original 349 idioms in various respects (phrase structure, semantic characteristics, etc.) (see appendix 3 for a complete list of the idiom set):

Structure:

	<i>VP</i>	<i>NP</i>	<i>PrepP</i>	<i>AdjP</i>	<i>Clause</i>	<i>Undefined</i>
349 idioms	197	51	56	23	18	4
100 idioms	59	16	14	5	4	2

Mode:

	<i>Idioms from spoken texts</i>	<i>Idioms from written texts</i>
349 idioms	208	141
100 idioms	54	46

Motivation:

	<i>Full</i>	<i>Constituential</i>	<i>Global</i>	<i>Partial constituential</i>	<i>Partial global</i>	<i>Unmotivated</i>
349 idioms	129	31	84	41	9	55
100 idioms	37	8	27	9	2	17

Analysability:

	<i>Full</i>	<i>Partially analysable</i>	<i>Unanalysable</i>
349 idioms	167	52	130
100 idioms	56	15	29

Literal well-formedness:

	<i>Well-formed</i>	<i>Ill-formed</i>
349 idioms	189	160
100 idioms	57	43

Length of phrase:

<i>Number of content words</i>	<i>1 word</i>	<i>2 words</i>	<i>3 words</i>	<i>4 words</i>
349 idioms	25	219	93	12
100 idioms	16	59	21	4

Complexity of phrase

<i>Number of phrases</i>	<i>1 phrase</i>	<i>2 phrases</i>	<i>3 phrases</i>	<i>4 phrases</i>
349 idioms	2	179	145	23
100 idioms	2	50	41	7

4.5.5. *Eliciting idiom variants from corpora*

Three different corpora are used as rich sources of idiomatic usage in real-life discourse, including MICASE³, CANCODE, and most importantly BNC, with the purpose of creating an extensive pool of data for analyses regarding the different ways idioms are manipulated in discourse and about the constraint(s), if any, on such manipulation. Our decision to use data from the BNC was motivated by a number of factors. First, the BNC is a sizeable corpus of current British English. It contains a 100 million word collection of samples of written and spoken language from a wide range of domains and genres from the later part of the 20th century. The wide range of domains and genres is desirable to produce data that will be relatively domain-independent and useful for a number of application scenarios. Second, the BNC XML Edition is already annotated with rich layers of information, e.g. automatically assigned part-of-speech tag, which makes it easier to extract idiom variants. The CANCODE is smaller and contains 5 million words of spoken texts. It is used to supplement the spoken part of BNC (which is significantly smaller than the written part) to find more uses of creative idiomaticity in spoken discourse. Last but not least, the MICASE is consulted due to the presence of a body of academic discourse in the data.

Unfortunately none of the corpora comes with semantic or contextual annotations, hence certain difficulties with the extraction of idiom variants. Indeed, technical aspects of the process of elicitation have proved to be famous cans of worms in creativity studies. In the field of formulaic language and collocations, a number of statistical methods have been developed for the identification of collocations and the description of their frequencies in corpora, among which z-score, MI score, t-score, and chi-square are the most popular. Unfortunately, none of these methods are suitable for the identification of idioms and idiom variants, and here are the reasons why. Essentially, these statistical methods are all based on computer-generated calculations, although each with a different formula due to their different assumptions and hypotheses, designed to establish whether a particular word combination is a collocation (i.e. if it is significant enough to earn the status of collocation). These methods are hypothesis testing methods of identifying collocations, testing the null-hypothesis that words appear together no more frequently than we would expect by chance alone. They can, therefore, be seen as formalisations of Hoey's definition of collocations as sequences of words 'appear with greater than random probability in its (textual) context' (Hoey, 1991: 7). However, as we know from the previous chapter, idioms are not only defined on the basis of frequency of co-occurrence, but more heavily on semantic criteria. In fact, even though highly institutionalized, the frequency with which idioms occur in speech and writing is quite low, sometimes extremely low. As a result, none of the statistical methods can effectively extract idioms and idiom variants from texts and corpora.

³ MICASE stands for Michigan Corpus of Academic Spoken English, a 1.8 million words of transcribed speech from the University of Michigan. See <http://micase.elicorpora.info/>

Alternatively, and most frequently, linguistic items/units can be extracted from corpora using built-in corpus search/query tools, which yields neat and user-friendly concordances containing the searched items. Unfortunately, we do not know exactly what we are looking for, which makes it impossible to decide on a sensible set of search units to submit. Since this is a study of idiomatic creativity, we expect the idiom constituents to be varied in some ways other than pertaining to the base-forms, but there is no way to ‘guess’ or ‘control’ the variations. To capture this variety, instead of entering the whole idiomatic phrases into the search/query box (which will give us all *invariable* uses of the idioms) a number of tailored searches are carried out for each idiom base-form in order that the maximum number of variants is located.

To increase the success rates of locating idiom variants and reduce the labour, the searches are tailored to each particular type of idiom and each specific type of variant instead of generically applied across the board. The underlying rationale originates from observations about the two factors that might affect the outcomes of corpus searches, i.e. the type of idiom and the type of variant we wish to find. First of all, it is observed that each type of variants, due to the differences in the ways they change the lexico-grammatical arrangements of canonical idioms, requires different extraction techniques. Existing literature as well as patterns found in the small corpus of real texts suggests the following types of modifications to the canonical arrangements of an idiom:

- [1] Morphological changes to constituents: found in such types of operations as changing numbers (pluralisation), nominalization. The original construction and all the constituents are preserved, except that one or more constituents can have different inflectional forms.

Corpus searches for these variants are most straightforward, because in the search query, constituents are treated as lemmas instead of words. As such, a search for two lemmas ‘sell’ + ‘soul’ will produce variants containing ‘sell’, ‘sells’, ‘selling’, ‘sold’ and ‘soul’, ‘souls’. As a result, no special technique is required here.

- [2] Altered word order, with or without morphological changes to constituents: involves passives, grammatical highlighting (via cleft structures, inversions, fronting, etc.). All the constituents are preserved, but the order between them is changed using different constructions.

In corpus query, it is possible to search for any combination of two lemmas without the need to specify the constructional relationship between them. As a result, this type of variant can be found by searching for the constituents but in a reverse order. For instance, instead of searching for the lemmas ‘ring’ + ‘bell’, we search for ‘bell’ + ‘ring’, which will then give us variants such as *a bell rang somewhere in D'Arcy's subconscious* or *a dim distant bell ringing in his mind*.

- [3] Insertions of lexical items: includes insertions of articles, adjective phrases, noun phrases, prepositional phrases etc. into the existing arrangement of an idiom to produce various types of premodifications and post modifications. The most prominent characteristic of this type of variant is that all the constituents are present as well, just like in type 1 and 2, except for insertions in between them.

To successfully locate insertions, it is important to expand the gap between the search items (i.e. the constituents) so that any possible insertions can be found. As a result, by allowing a span of at least 5 tokens between ‘keep’ and ‘lid’, we can successfully identify insertions such as *he keeps a very tight lid on his own private life*.

- [4] Truncation of idiom/allusion: containing variants in which parts of the canonical form is missing.
- [5] Substitution of one or more of the constituents of the canonical idiom: One or all of the constituents are replaced. The construction might also be altered according to the new lexical items.

Type 4 and 5 are discussed together because, despite the different ways they change a canonical idiom, the problems ensuing their extraction from corpora are similar. With type 4, since parts of the idiom are missing, which, in extreme cases of allusion, can mean that only one constituent is used in the variant. To extract these variants is difficult, since we have to assume the worst (i.e. only one constituent is present in the variant) by conducting searches using only one lemma, which often yields large, unmanageable results, particularly if the lemma is a core lexical item with high frequency in the corpora. This is also the problem that is encountered with extracting substitutions, type 5. Since the canonical lexico-grammatical arrangement is replaced with a new arrangement, and we do not know in advance what it is, the amount of noise, and as a result labour, involved in the extraction process is sometimes prohibitive.

Techniques are available to deal with the issues to some extents (see below) but they do not guarantee a very high success rate for all idioms. This is because the difficulties with which variants of an idiom are extracted also depends on the idiosyncrasies and/or the complexity of the idiom’s lexico-grammatical structure. Variants of an idiom with cranberry morphs are easier to locate than those of an idiom whose constituents are all core vocabulary, in the same way that variants of idioms with only 2 content words are easier to be extracted than those of idioms with more than 3 content words, and so on. The specific issues with idiom’s characteristics and the techniques to overcome them are discussed below.

4.5.5.1. Search by one lemma

First of all, variants of idioms containing one or more cranberry constituent(s) or rare/non-core words (*in the **offing**, **spick** and **span***) are the rather more straightforward to extract.

The search for the idiosyncratic lemma alone is able to retrieve most of the types of variants, except those with the substitution of the cranberry morph itself. For example, the search for the lemma ‘spick/spic’ shows the substitutions of the second constituent (‘span’), the omission of ‘span’ altogether, the pluralisation of ‘spick’, and the nominalization of the whole phrase as in [1] – [6]:

- [1] *Come clean! With the Spick and Spam set, there's no excuse for dirt in your runtime*
- [2] *Spick and Clean – Speciality chemicals and formulators*
- [3] *Spick and Sparkle is a family run domestic and commercial cleaning business*
- [4] *Spicks and spectacular*
- [5] *His buttoned flies, however spick - and they were freshly laundered - might carry some old aroma all the same*
- [6] *The reliever of a swindler to a study placebo affected in a such insurance, interpreted as a vision of spick-and-spanness*

4.5.5.2. Search by two or more lemmas

As we have seen, searches with only one control node word are preferred due to their generous scopes and maximized possibilities of finding variations, particularly substitutions. Unfortunately, unless constituents are rather rare (non-core) in the English language, such a search will produce a very large number of hits (in many cases it is a matter of hundreds of pages of concordances), most of which contain non-idiomatic uses of the words, which is practically impossible to process due to the amount of ‘noise’ and the labour involved in manually eliminating the ‘noise’. Attempts to narrow down such searches are, therefore, crucial, while there is still a need to keep them as open as possible to ensure the maximum number of variations. Conducting searches using a combination of two or more lemmas (as we have seen in the case of *spick and span* substitution) is an alternative.

These searches, by combining two or more lemmas while allowing a gap between them (a span, usually set at 5 tokens to the left or right), are capable of identifying many types of variants. For example, the search for the lemma ‘worm’ in the context of the lemma ‘can’ within 5 tokens to its left yields passive, allusion, pluralisation, postmodification, insertion, substitution as in [9] – [14] respectively:

- [7] *this can of worms has been opened up*
- [8] *it's a bit of a can of worms that to be honest*
- [9] *although cans of worms keep arising*
- [10] *they are terrified of the can of worms they might find*
- [11] *it opens the Oval office's can of worms*
- [12] *we are taking the lid off a can of worms here*

4.5.5.3. Search by construction pattern

The construction of the canonical idiom can be taken into consideration when searching for substituted variants. This is done via the use of wild cards with control on their parts-of-speech. For example, to find out whether 'ring' can be substituted by other verbs in the idiom *ring a bell*, a search for the node 'bell' within the context of a verb to its left will bring up all patterns of *V + bell*. Then a manual scan of the concordances is needed to decide if any of the other verbs were indeed idiomatic substitutions (e.g. *Hainan - Does this name **strike** a bell?*) or literal expressions (*there was an old lady who swallowed a bell*). Unfortunately, this search will not identify passives or inversions in which the construction pattern of *V + Obj* is reversed.

Second, sometimes it is necessary to narrow the span between the wild card and the node word down to less than 5 tokens if the construction of the idiom suggests a strong tendency for this. For example, a search for variants of *not bat an eye/eyelid* using the *V + 'eye'* with a span of 5 tokens between them still gives us 945 pages. Reducing the word span between the node word and the wild card from 5 down to 2-3 greatly bring the search results down to a manageable number while still allowing a certain amount of gap for insertions in between them (if any) to be detected. Although the gap is not as wide as the 5 token span, it greatly reduces the amount of work involving reading concordances. This is a compromise the researcher has to make with the limited search tools at his disposal.

4.5.5.4. Search by related lemmas

In some cases, the large number of concordances requires further filters. For example, to find substitution variants of *keep the lid on something*, we can apply the aforementioned techniques of searching for the *V + 'lid'* construction with control on the span. However, the concordances produced by the search with a span of 3 tokens still contain 29 pages, most of which are literal uses. Instead of manually checking all the concordances, the researcher can choose to check only the lines containing verbs which are semantically related to the substituted verb (in this case, 'keep'). It has been observed in the literature (and will be elaborated in later chapters) that substitutes usually hold a semantic relationship to the original constituents. For example, 'strike' is more likely to be substituting 'ring' in *ring a bell* because of a near synonymous relationship between them in this context. A built-in tool in corpora, called WordSketch, is particularly useful for filtering out such variants. WordSketch, besides providing information about the most frequent combinations a word enters into, can cluster semantically related words (synonyms, antonyms, same word field, etc.) can under the same headings:

Table 4.1. WordSketch list of collocates of ‘lid’ in the BNC (V + Object ‘lid’)

object_of	506	3.9
prise	6	7.85
flip	7	7.78
lift 67	82	7.7
raise 15		
screw	10	7.63
unscrew	4	7.6
slam 9	13	6.98
bang 4		
close 28	64	5.77
shut 8 open 28		
lower	7	5.67
snap	4	5.22
scal	3	5.22
replace 17	33	4.89
remove 16		
blow	5	4.57
secure	6	4.37
keep 35	121	4.03
put 39 take 26 hold 5 get 12 leave 4		

In the list of verbs that form significant patterns with ‘lid’ as presented in the Word Sketch table, we can choose to check only the concordances for the verbs which are semantically related to ‘keep’ (i.e. ‘put’, ‘take’, ‘hold’, ‘get’, ‘leave’ according to WordSketch), thus avoiding wasting time on unlikely verbs such as ‘need’ or ‘want’. By checking these related verbs we find variations via substitution such as [13] – 15]:

- [13] *There is no rift, it's time to put the lid on all the talk*
- [14] *A major piece of research that takes the lid off the French marital laundry basket*
- [15] *They could at least hold the lid on the Indo-Chinese kettle for the predictable [...] future*

I personally do check other items outside of the suggested cluster by the corpus, but only when a sensible relation can be observed between them and the original constituent. For example, in the table, the cluster is suggested by WordSketch to include ‘keep’, ‘put’, ‘take’, ‘hold’, ‘get’, and ‘leave’, but I also check ‘secure’, ‘screw’, ‘unscrew’, ‘close’, ‘open’, ‘slam’, ‘blow’, ‘lift’, which I feel are also related in some way. The extra variants found with these verbs ([16] – [20]) apparently enrich the list of variations:

- [16] *We must screw the lid on spending as tightly as possible*
- [17] *The UK slump has slammed the lid on margins*
- [18] *That opens the lid on buried secrets*
- [19] *They would blow the lid on the hundreds of millions of pounds of defence equipment*
- [20] *Her True Stories first lifted the lid on marriage problems*

Apparently, this technique of checking semantically related items proves to reduce the amount of manual editing significantly. However, by checking only the *expected* patterns (i.e. holding a semantic relationship with the original constituents) we have in effect eliminated any chance to spot ‘creative’, ‘unexpected’, ‘irregular’ variants. As a result, the checking of

unrelated items in the list might offer more interesting examples of substitution but it means a lot more work and time invested, so a line has to be drawn somewhere within the time scale allowed of the thesis.

Unfortunately, phonological relatedness has not been made available for corpus tools, so variants such as [21] – [22] can only be identified via the use of wild cards in place of the substituted constituent. The number of concordances might be manageable for some idioms (e.g. *kiss and ** yields 124 concordances, which is perfectly within the processing limits) and it might still be very high for some others depending on the frequency of the other constituent, so again a compromise has to be made with regards to how much time and efforts the researcher can invest in data collection.

[21] *Don't kiss and swell*

[22] *Spoilt four choice*

4.5.5.5. *Limitations of the current tools and techniques*

All of these searches (with or without some types of control, depending on each idiom) involve manual scanning of the search results (concordance lines) to eliminate the ‘noise’ (i.e. non-idiomatic uses of the same lexical constituents) and to find the meaningful, purposeful variants.

The use of the built-in search function so far described works under the assumption that at least one original constituent is preserved in each variant, because minimally one node word has to be entered into each query. In reality, however, there are variants in which all the original constituents are replaced:

[23] *I believe the appropriate metaphor here involves a river of excrement and a native American water vessel without any means of propulsion*

With the currently available tools, it is impossible to locate such a variant. This variant of *up shit creek without a paddle* was indeed encountered in a random manner outside corpus data.

Summary

There is no single route that can ensure a successful extraction process, but a combination of the described techniques will maximize the success rates while minimizing the manual efforts required. There are advantages and disadvantages with the types of searches described here. It is worth bearing in mind that while one type is better designed to work with a particular type of idiom and to find a particular type of variation, others will work better with other idioms and find other types of variation. It is indeed a time-consuming and laborious process which involves a series of successive searches per idiom in order to find its variations, although compromises are unavoidable with the current tools.

4.6. Data analysis procedures

To recap, as detailed in 4.3, there are two big components of the thesis' goal (which will be addressed in chapter 5 and 6 respectively):

- [1] Empirically testing the hypotheses about the relationship between literal well-formedness, analysability, motivation and idiomatic creativity.
- [2] Searching for significant patterns in idiomatic creativity in relation to context and other facilitating and constraining factors, including the internal characteristics of idioms and cognitive constraints.

It follows that data analyses to answer each question will differ significantly, with the former being quantitative using statistical tests, whereas the latter qualitative with the employment of mainly introspection and inter-raters where necessary. This section will elaborate on the uses of specific methods in each stage.

4.6.1. Quantitative analyses of the correlation between idiom transparency and idiomatic creativity

4.6.1.1. Variables

If we treat each idiom as one 'case' then its internal characteristics (motivation, analysability, literal well-formedness) can be considered independent variables (just like a person's gender or race), which are primarily categorical. However, since linguists have been proposing viewing these characteristics as cline (e.g. from fully motivated to unmotivated idioms), they can be treated as ordinal data to which numerical values can be assigned to indicate ranks. The decisions concerning the degrees of motivation, analysability and well-formedness of each idiom (which also entails the assignment of ranks to each idiom) is not volitional but strictly controlled by a rigorous theoretical framework derived from previous research in the literature (see chapter 3).

An idiom's creative behaviour, on the other hand, is the dependent variable that is interval, i.e. the values are at an equal distance from each other, so that the differences in values indicate the degree/size of the variable. In this thesis, it is proposed that statistically an idiom's creative behaviour is measured by two components, i.e. its frequency of use (how often the idiom is used in variant forms) and range of variants (how wide the range of variants is). The two concepts are distinguished on the basis of performance/competence, such that 'frequency of use' (FOU) is closely related to performance and signifies 'productivity', whereas 'range of variants' (ROV) is associated with competence and signifies 'flexibility'. FOU, in essence, is the frequency with which a particular idiom is actually used in creative/variant forms in corpus data. ROV, on the other hand, is reflected in the potential/ability by idioms to be open for different types of manipulation in different

contexts. Internal flexibility is usually assumed to entail external discursual productivity, hence the interchangeable use of the two terms in many studies. However, such a correlation, if any, will be reassessed using corpus data and statistical tests.

FOU is determined by straightforward frequency counts of the number of creative variants per idiom in the data (or number of *variant tokens*). To determine ROV, on the other hand, involves categorizing variants into types and determining the number of *variant types* an idiom can allow. The higher or lower number of variant tokens and variant types found of is taken as statistical representation of the actual productivity and flexibility of an idiom. Such an approach is based on the assumption popular in usage-based linguistic tradition that frequency of actual usage reflects the linguistic categories that speakers have in their minds (Bybee and Hopper, 2001: 3).

Frequency counts of variant tokens and variant types can be argued to be highly relative or even somewhat deceptive, since corpora, even the most comprehensive ones to date, are still not capable of reflecting the diversity, creativity, and instability of the real life discourse, particularly in spoken genres. However, since the idioms are retrieved from the same database, the frequency counts of their variants and types of variants can provide a comparative account of their productivity. Moreover, even though they will never be the exact record of real life discourse, electronic corpora of the sizes and designs today are thought to provide a trustworthy representation of the prominent patterns of actual language usage. Frequency counts of creative tokens of idioms in corpora can, therefore, reveal the degrees of relative productivity and flexibility amongst idioms.

Frequency counts of variant tokens and variant types in this thesis are carried out manually via reading concordances line by line and selecting idiom variants from the search results, and are to be distinguished from the ‘traditional’ raw frequency method used in phraseology, executed via computer-generated counts of the number of times a particular collocation occurs. Raw frequency obtained via computer searches and automatic counts is usually criticized for their lack of sophistication by letting through a large number of ‘noise’ due to co-incidence by chance instead of valid sequences whose words stand in particularly interesting relationship to each other. This method also runs the risks of missing out on significant collocations whose constituent words are individually rare in the language. However, raw frequency counts in this thesis will not have the same issues because they are not computer-generated but manually executed via a selection of searches and manual scanning of the concordance lines to find legitimate variants of idioms.

4.6.1.2. Statistical tests: Non-parametric tests

After both the independent and dependent variables can be determined for each idiom, the values will then be entered into SPSS for statistical analyses. The use of SPSS statistics for non-human cases (in our case, idioms) is not traditionally popular, but has been attempted

before. Most notably, Biber (1988, 2003), Biber and Finegan (1994, 1997) use factor analysis of corpus data to analyse the co-occurrence of a large number of linguistic features. The same philosophy is applied here to use statistics to analyze the correlation of different idiom internal characteristics with idiomatic creativity in corpus data.

To adopt the best tests for the data, it is first of all important to check the normality of the data distribution. In table 4.2 below, it can be seen that the significance values *p* are smaller than 0.05 for both variant token and variant type in both tests (Kolmogorov-Smirnov and Shapiro-Wilk), which means that that the distributions of both sets of data are significantly different from the normal distribution (see also histograms in figure 4.1).

Table 4.2. Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Variant token	.249	100	.000	.675	100	.000
Variant type	.161	100	.000	.915	100	.000
a. Lilliefors Significance Correction						

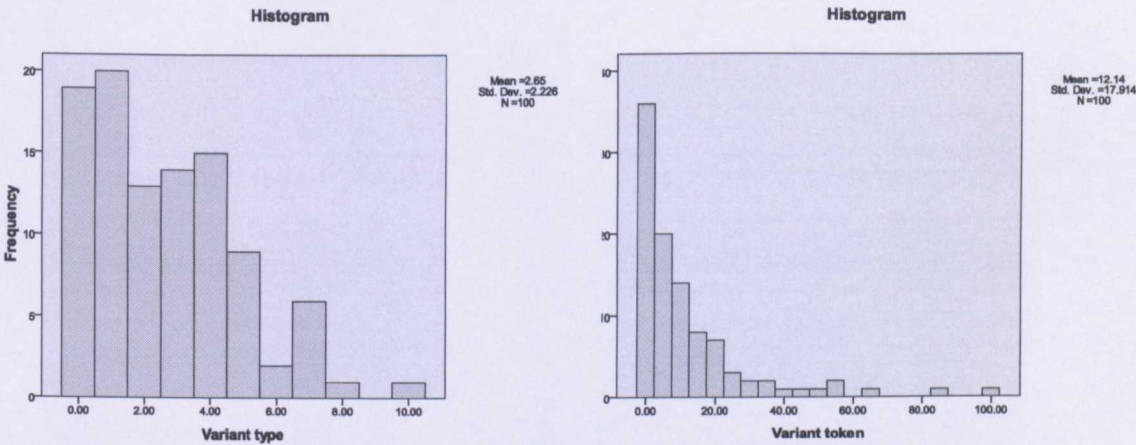


Figure 4.1. Histograms of the normality of distribution of variables

Since the data is found to be skewed instead of normally distributed, parametric tests will not be appropriate. To test the differences between different groups of idioms in terms of creativity, the Mann-Whitney and Kruskal-Wallis tests are used. These are non-parametric analogs to the independent samples t-test and the one-way ANOVA. They can be used when we do not assume that the dependent variable is a normally distributed interval variable (which is our case). The results from both Mann-Whitney and Kruskal-Wallis test can tell us whether or not different groups of idioms (in terms of motivation, analysability, and well-formedness) significantly differ in terms of creativity (productivity and flexibility).

The non-parametric *Spearman's rank order correlation* test will also be used where appropriate to test the correlations between sets of variables. A Spearman correlation is particularly useful when one or both of the variables are not assumed to be normally distributed and interval

(but are assumed to be ordinal). The values of the variables are converted in ranks and then correlated.

The use of scatterplots to map the distributions of the variables in each case will also be used to support the statistical tests. Visual representation of the data in the scatterplots is usually found to ease reading and help illustrate the results in a more reader-friendly manner. Correlations may be positive (rising), negative (falling), or null (uncorrelated).

By adopting the ‘meaning in numbers’ approach in this section of the thesis, it is hoped that systematicity is added to the data analysis phase, thereby assisting the researcher’s subjective interpretation and eliminating individual variability.

4.6.2. Qualitative characterisation of idiomatic creativity in relation to its facilitating and constraining factors in context

Besides the quantitative analyses about the correlation of the internal semantic characteristics of idioms and their creative usage in discourse, qualitative analyses will also be carried out with the corpus data pool, supplemented with additional data from the internet, to explore patterns of idiomatic creativity in each type of transformation (i.e. passive, modification, substitution, etc.) to draw up a context-based model of idiomatic creativity under the effects of their internal characteristics and external factors in authentic English discourse.

4.6.2.1. Extra data

As Dornyei (2007) observes, due to the iterative, nonlinear pattern of qualitative research, it is not unusual for extra data to be required even at the later stages of the analytical process, which is the case with this particular thesis. During this qualitative analysis stage, additional sources of data are exploited here to enrich the data pool with the types of variants that are not found in corpora. Indeed, that a variant is not found in corpora does not mean it will not appear in a different text. Extra variant types will therefore be added to the data to show the wide range of variants an idiom *can* have despite the fact that these types might not be found in corpora. These extra variant types are collected from the Internet via Google tool as well as from TV shows, movies, newspapers, websites, etc., i.e. in a random manner.

The internet with its rich source of authentic texts covering a wide range of genres, content domains and written text types have been proved to be beneficial to linguistic research. Fletcher (2004, 2007) compiled a sample corpus from a selection of web texts (5.25 million tokens) and compared it with the written texts from the British National Corpus (BNC) to assess their similarity. Both corpora are found quite similar, but web pages seem to provide more representative language samples, hence arguably more useful than off-the-shelf corpora for special information needs.

4.6.2.2. Generic analytical process

The main qualitative method used in this part of the thesis is qualitative in the most generic way using 'coding', 'memoing' or 'data display' (Dornyei, 2004) that partially rely on informed intuition. In favour of the intuitive approach to linguistic research, it is usually argued that the subjective and reflective involvement of the researcher in the analysis as well as the fluidness and creativity that he/she brings to the analytical process allows new theories to emerge freely (ibid). It is generally agreed that intuitions have made considerable contributions to linguistic research, enabling us to reach outside the scope of a purely corpus driven approach (Labov, 1975; Newmeyer, 2003; Sampson, 1975). However, Miles and Humberman (1994) strongly advise the use of accepted analytical conventions, thus achieving 'rigorous flexibility' by applying procedures and frameworks in a way that they assist the generation of new insights without limiting the results. Indeed, to be able to achieve a balance between 'structure' and 'freedom' of the analytical process is not easy. However, this thesis strives for such a balance by using intuition that is guided by rigorous analytical frameworks and informed by existing literature on related subjects.

4.6.2.3. Inter-raters

When the researcher is a non-linguist or a non-native, his/her intuition is judged as easily becoming unreliable due to a host of external factors (Botha, 1981: 304). These factors might include (but are not restricted to) linguistic background, knowledge of the subject matter, over exposure to a particular dialect, etc., all of which can possibly affect the researcher's judgments.

Besides, experimenter's bias, i.e. a tendency to deviate towards what is expected by the rater, might occur even with native speaker researchers. When interpreting and presenting the results, there may be variations as to which is acceptable and which is significant.

As a result, at this analytical phase, 3 native speakers are used to provide judgments on unclear cases to offset any possible bias caused by the added fact that I am a non-native speaker. Inter-rater reliability is not carried out via statistic methods (correlation coefficient tests, etc.) because the use of independent raters is few and far between throughout various stages, only when potential bias is anticipated. Instead, the simple joint probability agreement method is used, in which raters agree with the official rating of a performance, or they might agree with each other about an alternative.

4.7. Conclusions

Amongst many benefits of corpora, this thesis mainly capitalizes on the size and breadth of the corpus data to build a rich data pool in the hope to gain in-depth insights into the description of idiomatic creativity that might have escaped intuition. In general, the role of

corpora in this thesis is as the data source, and the human analyst is responsible for the analysis of this data.

To be specific, the first phase of the analytical process is to use statistical tests on corpus data to validate existing hypotheses of the correlation between the semantic characteristics of an idiom and its creative behaviour. The use of corpora for the purpose of hypothesis testing as such is what Adolphs (2008) and Biber (1988) believe to be one of the main contributions of corpus linguistics to language variation studies, i.e. to exploit the huge amount of data at corpus linguists' disposal to explore and extend observations made in pre-corpus research.

The second phase of analysis, i.e. qualitative observations of the interaction of facilitating and constraining factors of idiomatic creativity under the influence of context, is intended to flesh out the quantitative results that the first phase arrives at. Interpretations of corpus data to tease out meaningful patterns in this phase are argued to complement the decontextualized claims by statistical tests, thus bringing us a fuller picture of the underlying factors of idiomatic creativity.

The methodology employed in this thesis, as a result, can be seen as a mixed method approach to idiomatic creativity. It draws on various resources and uses various techniques at different phases of the analytical process. Quantitative analyses of corpus data are capable of shedding light on unclear/controversial cases as well as providing robust unbiased statistics of language in use, while qualitative analyses bring more sophistication and precision to the frequency data and raw quantitative descriptions.

The first phase of analysis, i.e. the quantitative description of data and statistical testing of hypotheses, is carried out in the next chapter. To be specific, the next chapter reports on:

- (1) the creative behaviour of idioms in the data in terms of distribution, frequency and range.
- (2) the statistical tests of the correlation between the semantic characteristics of an idiom (analysability, motivation, and literal well-formedness) and its creative behaviour in discourse (demonstrated in the variant type and variant token counts in the data)
- (3) the discussions of such statistical results.

CHAPTER 5

EMPIRICAL EVALUATION OF THE CORRELATION BETWEEN THE INTERNAL CHARACTERISTICS OF IDIOMS AND IDIOMATIC CREATIVITY

This chapter is dedicated to an empirical evaluation of the existing hypotheses in the literature about the relationship between the internal semantic characteristics of idioms and idiomatic creativity in discourse. Semantic factors (literal well-formedness, motivation, and analysability) have been suggested in various psycholinguistic and cognitive-linguistic studies to have strong impacts on the creative behaviour of idioms. In this chapter, authentic data will be used to test these hypotheses and validate the existence or otherwise of such correlations in real-life discourse.

5.1. General description of idiomatic creativity in the self-compiled corpus

In the approximately 200,000 words of data in the self-collected corpus, 349 cases of idiomatic expressions were found (0.17%), among which 239 were non-creative (i.e. they are either used in their canonical forms or with tense inflections, which are argued to be non-creative – see chapter 2). That means only 110 uses of idioms were found to be creative (0.06%). Statistically, this result is considered very low. For example, the frequency of FEIs is estimated by Moon (1998) to be roughly 4-5% in her corpus. Creative language in general is thought to have higher frequencies as well (Carter, 2004). Idiomatic creativity, as found in this thesis, is rather rare.

5.2. General frequency distribution of idiomatic creativity in the BNC, CANCODE and MICASE

The break-down of the frequency distribution of the variants of 100 idioms (table 5.2) shows that some idioms are completely invariable (at least in the corpora consulted), while others have higher types and tokens of variants. To be specific, 19% of the idioms have 0 variant, while a massive 81% of the idioms have at least one variant.

The pattern in table 5.2 also demonstrates that idioms in general can be productive and flexible, but rather moderately. The frequency of variant tokens and variant types cluster more around the low end than the high end of the creativity cline. Indeed, only 10% of the idioms have a very high number of variant types (more than 5 types) or very high number of variant tokens (more than 30 tokens). The majority cluster around the marks of 1-5 variant types and 1-25 variant tokens. There are idioms which score high on both token and type, but these are rare and not typical of the data pool.

Table 5.1: Frequency distribution of the creative types and tokens of variants in the three corpora

	Frequency band (variant tokens)	Proportion of idiom	Frequency (variant types)	Proportion of idiom
	0	19%	0	19%
	1-5	32%	1	20%
	6-10	16%	2	13%
	11-15	10%	3	14%
	16-20	5%	4	15%
	21-25	6%	5	9%
	26-30	2%	6	2%
	31-35	1%	7	6%
	36-40	1%	8	1%
	41-45	2%	9	0%
	46-50	0%	10	1%
	51-55	2%		
	56-60	1%		
	61-65	0%		
	66-70	1%		
	71-75	0%		
	76-80	0%		
	81-85	0%		
	86-90	1%		
	91-95	0%		
	96-100	1%		
Total	1214 tokens			

Despite relatively significant raw numbers, the frequency of the creative variants of the 100 idioms in relation to the corpora sizes (more than 100 million words) is less impressive (table 5.2). They occur less than 1 time per million words in the three corpora, with the majority clustering around the 0 – 0.2 time per million words (88%). Only 12% of the idioms have the creative usage frequency of 0.2 – 1 time per million words. The average creative variant in our corpora, therefore, is 0.1 times per million words, which is very low and might be considered insignificant (Biber et al., 1999; Biber and Conrad, 1999).

Table 5.2: Frequencies of creative variants in the three corpora

Frequency band (token per million words)	Proportion of idiom
0	19%
0.01 – 0.23	69%
0.24 – 1	12%
> 1	0%
Average	0.1 token per million words

The frequency counts of the invariable tokens against the variable tokens of the 100 idioms in the three corpora show that an average idiom is used in its canonical form roughly 3.3 times more often than in creative forms (table 5.3). In other words, there is still strong tendency amongst language users to use idioms in their original forms. Creative manipulations of idioms, although less frequent than canonical uses, are noticeable and an empirical evidence of the flexibility of idioms in discourse against the claims about the fixedness of idioms.

Table 5.3: Invariable versus creative usage of 100 idioms in three corpora

Invariant tokens	Creative tokens
3962	1214

Individually, there are idioms with high invariable tendency (having higher invariant tokens than creative tokens) (69%) and there are idioms with high creative tendency (having higher creative tokens than invariant tokens) (31%). This result partly echoes Philip’s (2006) who claims that creative idiom variants can sometimes outnumber canonical idioms in real life usage. The list of the idioms which are more often used in creative forms is shown in table 5.4.

The most common denominator of these creative idioms is motivation. These idioms are all motivated to various degrees, including full motivation, constituent motivation or global motivation. There is only one exception, which is *a fine/pretty kettle of fish*. This idiom, although unmotivated, is analysable.

In this set of creative idioms, analysability and well-formedness seem to be less reliable predictors, since a number of idioms, despite being unanalysable and/or ill-formed, still have creative variants (*crawl out of the woodwork*, *a wolf whistle*, *pat someone on the back*, *ram something down one’s throat*, *hook line and sinker*). A more detailed analysis of the correlation between motivation, analysability and literal well-formedness will be carried out in later sections.

Table 5.4: List of idioms with creative tendencies

Idiom	Invariant tokens	Creative tokens	Creative types
<i>a fine/pretty kettle of fish</i>	2	9	1
<i>come/crawl out of the woodwork</i>	0	2	1
<i>out of one’s hands</i>	0	4	1
<i>a wolf whistle</i>	7	9	1
<i>stop (someone/something) in their tracks</i>	10	16	2
<i>(fall for something) hook line and sinker</i>	6	9	2
<i>out of bounds</i>	0	4	2
<i>throw the baby out with the bathwater</i>	2	7	3
<i>see which way the wind blows</i>	1	30	4
<i>back (someone) into a corner</i>	5	22	4
<i>pat (someone) on the back</i>	37	51	4
<i>put the brakes on (something)</i>	19	57	5
<i>a babe in the woods</i>	1	9	5
<i>turn back the clock</i>	23	86	5
<i>make a splash</i>	26	66	5
<i>ram (something) down one’s throat</i>	13	20	6
<i>in the spotlight</i>	37	55	6
<i>keep a lid on (something)</i>	7	19	7
<i>touch/hit a raw nerve</i>	9	42	8
<i>open a can of worms</i>	5	28	10

Summary

The frequency of occurrence of creative variants of idioms in the corpora is low, particularly in comparison to other linguistic features. This is thought to be the lateral effects of the low frequency of idioms in discourse in general. The fact is that language users do not use idioms as often as they use literal expressions, and when they do, they tend to use idioms in their canonical forms more than in creative forms. This definitely brings about a state of affairs in which the general average frequency of idiomatic creativity is low.

However, out of the 100 idioms investigated, 81% have at least one variant token, and as many as 48% have more than 2 variants types. This finding is significantly higher than Moon's (1998) report of the creativity of FEIs in her corpus (40% of FEIs have at least 1 variant). The results are a positive proof of the variability of idioms in discourse, with much higher counts of both variant tokens and variant types than previously estimated. Rare as they may seem, a vast majority of them are not fixed. They are used variably in different communicative events to serve special communicative purposes. This variability is observed to differ from idiom to idiom, which can be the effects of different factors, including literal well-formedness, motivation and analysability. The hypotheses that this chapter aims to empirically test are:

- (1) Is there a significant correlation between idiomatic creativity and literal well-formedness?
- (2) Is there a significant correlation between idiomatic creativity and motivation?
- (3) Is there a significant correlation between idiomatic creativity and analysability?
- (4) Is there a significant correlation between idiomatic creativity and the transparency of an idiom (determined by the combination of all three factors)?

5.3. Literal well-formedness and idiomatic creativity

5.3.1. An analytical framework of literal well-formedness

Different authors have had different views regarding which types of anomaly should be treated as ill-formed (see chapter 3). In this study, a broad approach is taken, i.e. treating all types of anomalies (grammatical, semantic, and pragmatic) as blocking normal compositional processes, contributing to the difficulties with which the literal meanings are understood. However, how much this well-formedness/ill-formedness affects creative usage of idioms in discourse is open for further investigation, which is carried out hereafter.

For analytical purposes, idioms are categorised into two groups (well-formed and ill-formed idioms). Ill-formed idioms include the following types of anomaly:

- (i) Syntactic anomaly: odd phrase structures, ellipsis, inflections, articles, transitivity patterns, etc. that deviate from usual syntactic behaviour: *none the wiser, fall for something hook line and sinker, the long and the short of it*.

These correspond to Fillmore et al.'s (1988) class of 'extra-grammatical idioms'. Fraser (1970) notices that these never or only rarely undergo any transformations due to the impossibility to analyse and analogise them.

- (ii) Lexical anomaly: cranberry morphs, archaic/rare words, garden-path constituents, words that only appear as part of the idiom: *spick and span, live in cloud-cuckoo land, have had one's chips, chomp at the bit*.

In these idioms, one or more constituents do not have meanings outside of the context of the idioms (*spick, cloud-cuckoo*). Alternatively, the constituents might have specialised meanings which are not generally known by many speakers (even though they might be homographic with other independent items) (*chips, chomp*). The inability to activate the concrete meanings for these constituents apparently impedes the understanding of the literal meanings of these idioms. Fraser (1970) and Moon (1998) both report negative findings about their transformations.

- (iii) Semantic anomaly: violation of truth conditions, contradictions, tautologies, etc.: *make one's blood boil, lose your mind/nerve*

Syntagmatically, these idioms are correct, but they are anomalous paradigmatically (Moon, 1998) because they are literally impossible, i.e. violating the truth conditions in semantics.

- (iv) Pragmatic anomaly: lack of clear co-referentiality by determiners/pro-forms in co-texts: *lose it, bite the bullet*.

The pronoun 'it' in *lose it* is not supported by any further information within the idiom for us to know what 'it' refers to. The idiom, as a result, is thought to display a type of ill-formedness (Moon, 1998). Similarly, the use of definite article 'the' in *bite the bullet* suggests some kind of ill-formedness (we do not know which specific 'bullet' the definite article refers to), hence enforcing the retrieval of the article directly from the institutionalized meaning of the idioms without any possibility to creatively manipulate them (Moon, 1998).

Care should be taken when one tries to determine the ill-formedness of idioms in this category. First of all, such idioms as *see which way the wind blows, out of the blue, pat someone on the back, out of the starting gate, roll out the red carpet, plant the seeds of something*, have their definite NPs made explicit within the idioms themselves by their unique referents (*the wind, the blue, the back of a person*), pre-modification (by adjectives), or post-modification (by *of*-construction). Also, there are idioms with special noun phrases which are considered definite in reference because they are inferable from institution (*turn back the clock*), or by having pragmatically

presupposed referents from normal expectations/ shared knowledge, such that we normally expect these places or things to be found/self-referential in a particular environment (*around the corner, a fly on the wall, come out of the woodwork, turn on the waterworks*) (using the same principle as *the beach, the zoo, the post office, the bank*). As a result, the uses of definite NPs in these idioms are still considered grammar conforming and easily understood. On the contrary, some idioms are found with definite nominal groups which do not have automatic independent definite references as such. Their definite references are only inferable from the contexts in which they occur (*bury the hatchet, spill the beans*). As a result, the participants of the conversations/readers of the texts need to be aware of the existence of the appropriate contexts (hostile relationship between two parties, secrets to be revealed) and/or preceding co-texts that make explicit the noun phrases.

Last but not least, some idioms with definite noun phrases seem to violate all the aforementioned rules of determiner selection: *kick the bucket, bite the dust, chew the fat, shoot the breeze*. They cannot be interpreted either literally or metaphorically, hence appearing non-denoting. According to Fellbaum (1993), it is this class of idioms that are semantically opaque and syntactically frozen due to the emptiness of the nouns.

To single out grammatical anomalous idioms, the rules of the English Grammar as detailed in Quirk et al. (1985) are used, complemented with features of dialectal varieties and non-standard, journalistic genres as described in Freeborn (1995). These provide us with guidelines of the grammatically 'correct' phrases in English, any deviation from which should be treated as anomalous. Similarly, taking away the differences between different theories, the core of traditional semantic rules of English, including *lexical semantics* (Cruse, 1986), *truth conditions, presuppositions, entailments* (Davidson, 1967), and pragmatic's notion of *reference* are used as standards against which idiomatic expressions are analysed.

As Moon (1998) admits, it is sometimes ambiguous whether an expression is literally well-formed or not, even when clear criteria are used. A phrase like *time flies* can be judged as 'impossible, violating truth conditions' if 'fly' is interpreted as 'move through the air on wings'. The phrase, however, becomes possible if 'fly' is interpreted as 'move fast.' Apparently, dictionary entries are usually listed with several meanings, including literal, figurative and slang meaning, making it difficult to class idioms as being literally well-formed or otherwise. However, in this study, it is the literal meanings of the constituents that will be used in the composition of literal meanings of the idioms. Figurative meanings and slang meanings, it is argued, are also derivative in some ways, although they might have become institutionalized in the speech community and earned their places in dictionaries. As such, the type of semantic anomaly in *time flies* is treated as ill-formed.

With these analytical tools in hand, it is possible to categorise the 100 idioms into two groups of well-formed and ill-formed idioms (see appendix 3), from which statistical analyses can be carried out.

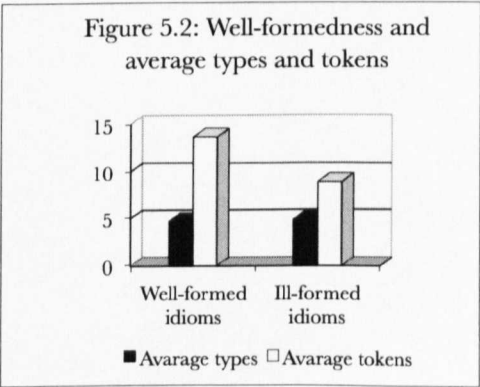
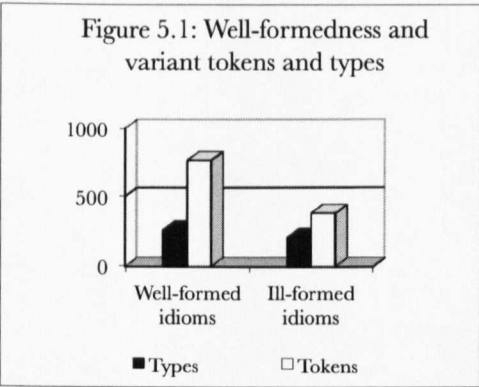
5.3.2. Charts and discussions

5.3.2.1. Is there a correlation between well-formedness and idiomatic creativity?

By comparing two groups of idioms, i.e. well-formed and ill-formed, it is hoped that the question of whether the literal well-formedness of an idiom entails its higher creative usage in discourse can be answered to certain extents.

The descriptive analysis (table 5.5) shows that, in pure numbers, well-formed idioms are more flexible and productive, with the average token of 15 and average type of 3 per idiom, compared to 8 and 2 for ill-formed idioms, respectively. Figure 5.1 and 5.2 show the relative distribution of well-formed and ill-formed idioms along two clines, i.e. variant token (the x-axis) and variant type (the y-axis). (See also scatterplot 1 in appendix 4)

Table 5.5: Descriptive analysis of well-formedness and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Well-formed	56	15.3393	18.17719	2.42903	10.4714	20.2072	.00	86.00
	Ill-formed	44	8.0682	16.90933	2.54918	2.9273	13.2091	.00	99.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Well-formed	56	3.1964	2.26772	.30304	2.5891	3.8037	.00	10.00
	Ill-formed	44	1.9545	1.98781	.29967	1.3502	2.5589	.00	8.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00



In short, there are differences between the two groups of idioms in raw frequencies. The question is if the differences are statistically significant. The Mann-Whitney test (table 5.6) reveals that the differences between well-formed and ill-formed idioms in terms of variant token ($z = -3.149, p = 0.02$) and variant type ($z = -2.81, p = 0.04$) are statistically significant. In other words, the tests show that well-formed idioms have higher variant tokens as well as higher variant types than ill-formed idioms. This means that, in the current data, people appear to creatively manipulate well-formed idioms more often than they do ill-formed idioms. Also, people can create more creative types of variants with well-formed idioms than they do with ill-formed idioms.

Table 5.6: Mann-Whitney test for well-formedness and variant token and variant type		
	Variant token	Variant type
Mann-Whitney U	780.500	823.500
Wilcoxon W	1770.500	1813.500
Z	-3.149	-2.871
Asymp. Sig. (2-tailed)	.002	.004
a. Grouping Variable: Literal well-formedness		

The hypothesis about the correlation between well-formedness and idiomatic creativity is tested using the Spearman correlation test (table 5.7). The test shows that literal well-formedness is significantly correlated with variant token ($r_s = 0.277$, $p = 0.05$) and variant type ($r_s = 0.306$, $p = 0.02$). In other words, literally well-formed idioms are significantly more productive and more flexible than ill-formed idioms. However, the correlations are rather weak with the values of r in each case being closer to 0 than to 1. This means that the hypothesis about the correlation between literal well-formedness and idiomatic creativity is empirically validated, but it is a weak correlation.

Table 5.7: Spearman correlation between well-formedness and idiomatic creativity					
			Literal well-formedness	Variant type	Variant token
Spearman's rho	Literal well-formedness	Correlation Coefficient	1.000	.277**	.306**
		Sig. (2-tailed)	.	.005	.002
		N	99	99	99
	Variant type	Correlation Coefficient	.277**	1.000	.796**
		Sig. (2-tailed)	.005	.	.000
		N	99	100	100
	Variant token	Correlation Coefficient	.306**	.796**	1.000
		Sig. (2-tailed)	.002	.000	.
		N	99	100	100
**, Correlation is significant at the 0.01 level (2-tailed).					

This finding goes against the result from Gibbs and Nayak's (1989: 212) psycholinguistic study, which 'rules out the simple alternative that the syntactic productivity of idioms is simply a matter of literal well-formedness'. According to the authors, their subjects respond better to ill-formed idioms than to well-formed ones due to the semantic composition of the particular idioms used in the test. Even though ill-formed, *pop the question* is rated as normally decomposable (i.e. analysable), while the well-formed idiom *kick the bucket* is non-decomposable (i.e. unanalysable). As such, well-formedness alone cannot determine flexibility and productivity but works together with other semantic characteristics of idioms. In this thesis, well-formed idioms are found to be more creative than ill-formed ones. An explanation for the conflicting results between the two studies would pertain to their respective scales. It is possible that on a small scale (16 subjects) and with a small list of idioms (20 idioms), the selection of the particular idioms plays an important part, since these idioms might have different motivation and analysability that, due to the small scale of the study, greatly affect the outcomes. In this thesis, on the other hand, with a sample of 100

idioms and 5176 idiom variants from three corpora of over 100 million words, any human manipulation or subjectivity is effectively eliminated. As such, on a larger scale, it can be imagined that well-formed idioms are more productive and flexible than ill-formed idioms.

5.3.2.2. *Does type of anomaly among ill-formed idioms correlate with idiomatic creativity?*

Since idioms can be ill-formed as the result of different types of anomaly (grammatical, lexis, semantic, or pragmatic), our question is whether the different types of anomaly cause idioms to have different creative usage? In other words, is one type of anomaly more restraining than another in terms of idiomatic creativity?

Ill-formed idioms in the data are grouped into four subcategories, i.e. syntactic, lexical, semantic, and pragmatic anomaly. First, looking at the descriptive analysis (table 5.8), it can be seen that the groups have rather similar mean variant types (around 2 types per each group). In terms of variant token, semantic anomaly has the highest mean token (12 tokens per idiom), whereas syntactically and pragmatically anomaly have the lowest (4 tokens per idiom).

Table 5.8: Descriptive analysis of types of anomaly and idiomatic creativity									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Syntactic	7	4.4286	3.59894	1.36027	1.1001	7.7570	1.00	10.00
	Semantic	20	12.7000	23.95193	5.35581	1.4902	23.9098	.00	99.00
	Lexis	6	7.0000	9.25203	3.77712	-2.7094	16.7094	.00	24.00
	Pragmatic	11	4.5455	6.63873	2.00165	.0855	9.0054	.00	18.00
	Total	44	8.5682	17.05040	2.57044	3.3844	13.7520	.00	99.00
Variant type	Syntactic	7	2.1429	.69007	.26082	1.5047	2.7811	1.00	3.00
	Semantic	20	1.9500	2.11449	.47281	.9604	2.9396	.00	8.00
	Lexis	6	2.0000	2.60768	1.06458	-.7366	4.7366	.00	7.00
	Pragmatic	11	2.1818	1.88776	.56918	.9136	3.4500	.00	5.00
	Total	44	2.0455	1.91633	.28890	1.4628	2.6281	.00	8.00

Scatterplot 2 (see appendix 4) shows the detailed distribution of ill-formed idioms along the variant type and variant token parameters. It becomes clear that the high average token for the group of ‘semantic anomaly’ (12 tokens) is thanks to three high frequency idioms (between the 50 and 100 mark). The rest of the data clusters around the 0-30 mark. Along the ‘type’ axis, the groups of ‘semantic anomaly’ and ‘lexical anomaly’ share quite a similar range of variant type (0-8 and 0-7 respectively). Syntactic anomaly has the lowest range of type (1-3).

Again, it is important to emphasize that the differences in terms of pure number and relative distribution as revealed in table 5.8 and scatterplot 2 are not statistically significant as revealed by the Kruskal-Wallis test (table 5.9). As shown in table 5.9, with $p>0.05$ for both variant token and variant type, the null hypothesis cannot be rejected. We have to

assume that all four groups have similar distribution in terms of flexibility and productivity. In short, types of anomaly do not seem to affect the frequency and range of creative variants among idioms.

Table 5.9: Kruskal-Wallis test for ill-formed idioms and variant token and variant type		
	Variant token	Varian type
Chi-Square	1.532	1.607
Df	3	3
Asymp. Sig.	.675	.658
a. Kruskal Wallis Test		
b. Grouping Variable: Ill-formed		

5.3.2.3. *Does structural complexity correlate with idiomatic creativity?*

Since the idioms in the data pool are highly heterogeneous in terms of phrase types, lengths, and levels of complexity, the question raised here is whether structural complexity plays a role in idiomatic creativity. This is based on the anticipation that if an idiom is composed of more components (i.e. more content words, or more embedded phrases), it can possibly accommodate more types of alterations (insertions into different positions, substitutions of different constituents, reductions of different parts, etc.), which also leads to higher number of tokens. An idiom with a simpler structure, on the other hand, might not have enough material to work with to create many variants. The hypothesis, therefore, states that the more complex the structure, the more productive and flexible the idiom is.

The determination of the number of content words within each idiom is carried out on the theoretical division between function words and content/lexical words in the lexicon. By definition, function words are words that have little lexical meanings or have ambiguous meanings, but instead serve to express grammatical relationships with other words within a sentence (including all closed-classed words such as prepositions, pronouns, auxiliary verbs, conjunctions, grammatical articles or particles). Content words (or open class words or lexical words), on the other hand, are those that have meanings, including nouns, verbs, adjectives, and most adverbs.

The hypothesis about the correlation between content words and idiomatic creativity is roughly based on the observation that content words are the items getting manipulated most often (insertion of pre/post modification, substitution, truncation, coordination, etc.). As a result, the more content words there are in an idiom, the easier it is to manipulate it, resulting in more types and tokens of variants. Apparently this observation is not absolute, because function words such as articles or prepositions have been found to be substituted or pre-modified too (*here at Jordans we think inside the box, advertisers forced to think way outside the box*), but these seem to happen less frequently than with content words.

A summary of the mean token and mean type of each group of idioms (table 5.10) shows that idioms with 2 content words seem to be the most productive (nearly 14 tokens per idiom), while idioms with 3 content words are the most flexible (more than 3 types per idiom). Idioms with one content word score the lowest on both counts.

Table 5.10: Descriptive analysis of number of content words and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	1.00	14	9.5000	16.15192	4.31678	.1742	18.8258	.00	55.00
	2.00	61	13.3770	19.91412	2.54974	8.2768	18.4773	.00	99.00
	3.00	21	10.3333	13.88284	3.02948	4.0139	16.6527	.00	51.00
	4.00	4	12.0000	12.51666	6.25833	-7.9168	31.9168	1.00	30.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	1.00	14	2.0000	1.83973	.49169	.9378	3.0622	.00	6.00
	2.00	61	2.5738	2.12492	.27207	2.0296	3.1180	.00	7.00
	3.00	21	3.3810	2.76543	.60347	2.1221	4.6398	.00	10.00
	4.00	4	2.2500	1.25831	.62915	.2478	4.2522	1.00	4.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00

Scatterplot 3 (see appendix 4) shows the patterns amongst the groups in terms of distribution instead of mean frequencies. It can be seen that idioms with one content word seem to cluster around the lower ends of both axes (0-10 tokens and 0-4 types per idiom). While idioms with 2 content words dominate in productivity, idioms with 3 content words do so in flexibility. However, all four groups have some low frequency items and some high frequency items, which prompts the question of whether the differences are statistically significant.

The Kruskal-Wallis test (table 5.11) reveals that the differences between the groups of idioms are not significant in both variant token ($p = 0.695$) and variant type ($p = 0.517$). Reflecting this insignificant difference, the Spearman correlation test results (table 5.12) show that the number of content words of an idiom is unrelated to its number of variant tokens and variant types in discourse. The p values in both cases (0.31 and 0.86) indicate that the correlation is not significant.

Table 5.11: Kruskal-Wallis test for number of content words and variant token and variant type		
	Variant token	Variant type
Chi-Square	1.445	2.274
Df	3	3
Asymp. Sig.	.695	.517
a. Kruskal Wallis Test		
b. Grouping Variable: Number of content words		

Table 5.12: Spearman correlation between number of content words and variant token and variant type					
			Number of content word	Variant token	Variant type
Spearman's rho	Number of content word	Correlation Coefficient	1.000	.103	.173
		Sig. (2-tailed)	.	.310	.086
		N	100	100	100
	Variant token	Correlation Coefficient	.103	1.000	.796**
		Sig. (2-tailed)	.310	.	.000
		N	100	100	100
	Variant type	Correlation Coefficient	.173	.796**	1.000
		Sig. (2-tailed)	.086	.000	.
		N	100	100	100
**. Correlation is significant at the 0.01 level (2-tailed).					

The complexity of an idiom is also reflected in the number of phrases contained in that particular idiom. Granted that the number of content words usually coincides with the number of phrases (*spill the beans, kick the bucket*) but they are two distinctive features of phrases. Two-phrased idioms (V + NP, Prep + NP, etc.) are observed to be most popular, and hence most extensively researched in the literature. In this thesis, a comparative analysis will be conducted between them to highlight the role (if any) of structural complexity in idiomatic creativity. The premise, similarly to previous hypothesis, is that the more ‘raw materials’ there are in an idiom (more phrases) the more ways there are to change it lexico-grammatically (insertions in different positions, reduction of different components, substitution of different phrases, etc.). The hypothesis, therefore, states that the more phrases an idiom is made of, the more productive and flexible it is.

A summary of the mean frequencies of the groups is provided in table 5.13, showing that 1-phrase idioms are the least productive and also the least flexible while those with two or more phrases have more creative tendencies. 3-phrase idioms seem to be most creatively used in discourse, having both the highest mean variant token and mean variant type.

		95% Confidence Interval for Mean							
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min	Max
Variant token	1.00	2	5.0000	7.07107	5.00000	-58.5310	68.5310	.00	10.00
	2.00	50	12.2000	17.70852	2.50436	7.1673	17.2327	.00	86.00
	3.00	41	13.1220	19.67765	3.07313	6.9109	19.3330	.00	99.00
	4.00	7	8.0000	10.03328	3.79222	-1.2792	17.2792	1.00	30.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	1.00	2	1.5000	2.12132	1.50000	-17.5593	20.5593	.00	3.00
	2.00	50	2.6200	2.29365	.32437	1.9682	3.2718	.00	10.00
	3.00	41	2.7317	2.33478	.36463	1.9948	3.4687	.00	8.00
	4.00	7	2.7143	1.11270	.42056	1.6852	3.7434	1.00	4.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00

The relationship between the groups can also be seen in scatterplot 4 (appendix 4). It is clear from the chart that 2-phrase and 3-phrase idioms are the most creative groups in both types and tokens of variants. It is, however, not clear whether a pattern can be found with 1-phrase and 4-phrase idioms. The Kruskal-Wallis test results (table 5.14) reveal that the four groups of idioms (discriminated by the number of phrases within their constructions) are not significantly different from each other in variant token ($p = 0.933$) and variant type ($p = 0.833$).

Table 5.14: Kruskal-Wallis test for number of phrases and variant token and variant type		
	Variant token	Variant type
Chi-Square	.436	.869
df	3	3
Asymp. Sig.	.933	.833
a. Kruskal Wallis Test		
b. Grouping Variable: Number of phrase		

The Spearman correlation test (table 5.15) confirms that the correlations between number of phrases and variant token and variant type are not statistically significant ($p = 0.930$ and $p = 0.539$ respectively). In other words, the number of phrases in an idiom is unrelated to its variant tokens and types in discourse. The differences that we can observe from raw frequencies, as a result, can only be assumed to be possibly meaningful in a different set of data, which will be the topic for future research.

Table 5.15: Spearman correlation between number of phrases and variant token and variant type					
			Number of phrase	Variant type	Variant token
Spearman's rho	Number of phrase	Correlation Coefficient	1.000	.062	-.009
		Sig. (2-tailed)	.	.539	.930
		N	100	100	100
	Variant type	Correlation Coefficient	.062	1.000	.796**
		Sig. (2-tailed)	.539	.	.000
		N	100	100	100
	Variant token	Correlation Coefficient	-.009	.796**	1.000
		Sig. (2-tailed)	.930	.000	.
		N	100	100	100

**.

5.3.3. Conclusions

To sum up, the following results are found for the relationship between literal well-formedness and idiomatic creativity:

- (i) In pure number, the average well-formed idiom is both more productive and more flexible than an ill-formed idiom, evidenced by the higher mean variant token and mean variant type.

- (ii) Statistically, the differences are significant, i.e. well-formed idioms are found to have higher number of tokens and higher number of types of variants in discourse. However, the correlation in each case is rather weak.
- (iii) There is also a significant correlation between the productivity and flexibility of idioms, i.e. the large number of variant tokens entails a large number of variant types and vice versa.
- (iv) Amongst ill-formed idioms, there is no significant difference between the types of anomaly (syntactic, lexical, semantic, and pragmatic anomaly) in terms of variant token and variant type. That is, no single type of anomaly is found to be more restraining than another in terms of idiomatic creativity.
- (v) There is no correlation between structural complexity and idiomatic creativity, either. The number of content words and the number of phrases within an idiom does not affect how it is creatively used in discourse.

All in all, the most important finding is that there is a significant correlation (albeit weak) between literal well-formedness and idiomatic creativity. This is a proof of the strong preference amongst language users for the creative use of well-formed idioms over ill-formed idioms.

5.4. Motivation and idiomatic creativity

5.4.1. An analytical framework of motivation

The motivation of idioms (Lakoff, 1987; Geeraerts, 1995; Kovecses and Szabo, 1996; Langlotz, 2006) characterises the semantic associations between two readings of the same expression, contributing to the understanding of the semantic extensions and/or schemata that links the original (literal) meaning of an expression to its transferred (figurative) reading. Via top-down interpretative processes, a speaker can reactivate, or remotivate the figurativity an idiom to make sense of its idiomatic meaning with a view to its literal meaning. In other words, motivation is generally understood as the possibility of an interpretation of an idiom that makes sense of its use. This is in full accord with evidence from psycholinguistic studies about the ability to decompose and analysable idioms by language users. The meanings of idioms can, therefore, be seen as motivated and not arbitrary (Lakoff, 1987, 1993; Gibbs, 1990, 1994).

The motivation of an idiom can be on a global or constituential scale. If an idiom is *globally motivated*, it can be motivated as a whole by relating the literal scene to the idiomatic meaning (e.g. *grit one's teeth*, *spin one's wheel*, *go round in circles*). On the other hand, an idiom is seen as *constituentially motivated* if its constituents have autonomous figurative sense or acquire

phrase-induced sense in the context of the idiomatic construction (e.g. *swallow the bitter pill*, *jump on the bandwagon*).

On the basis of both global and constituent motivation, the degrees of motivation of idioms can be seen as constituting a cline as follows:

- 1 = Unmotivated
- 2 = Partially globally motivated
- 3 = Partially constituentally motivated
- 4 = Globally motivated
- 5 = Constituentally motivated
- 6 = Fully motivated (with both global and constituent motivation)

The ordering of the types of motivation into a cline as such is based on Langlotz's (2006) typology. While it is clear that fully motivated idioms and unmotivated idioms should be at two ends of the continuum, the placement of the in-between types are theoretical rather than empirically tested. Constituent motivation is argued to render the idioms more transparent than global motivation because the lexicalised figurative senses of the constituents, such as *swallow* (meaning 'accept/bear') and *bitter pill* ('unpleasant fact') in *swallow the bitter pill* are also used outside of the context of the idiom, making it easy to understand the figurative meaning of the phrase as a whole. Global motivation, on the other hand, is not as readily meaningful. For example, the metonymy involving the idiom *go round in circles* entails a literal reading of 'stagnation in moment', which is then metaphorised to denote the more abstract idiomatic meaning. Since there is no direct constituent mapping, the activation of the idiomatic meaning depends on the language user's capability of making sense of the metonymic extension and metaphorical transfer between the literal meaning and the idiomatic meaning. Hence, the idiomatic meaning is less readily available as with the case of constituent motivation. Again, this is only a theoretical argument, and psychological further experiments will have to be carried out to test this hypothesis. In this thesis, this theoretical order of degree of motivation is assumed to be correct for analytical purposes.

In terms of the underlying sources of motivation, there is little evidence of how each individual speaker processes a particular idiom (Dobrovol'skij and Piirainen, 2005), which makes it impossible to speculate the actual processes/sources employed by an individual speaker in each case. Psycholinguistic experiments have to some extent revealed some trends, mostly towards the validity of conceptual motivation of idioms via conceptual metaphor and conceptual metonymy, although the results are still very much inconclusive (see chapter 3). In the current thesis, the underlying mechanisms of idiomatic motivation are treated as an open group of sources of information that promote a better understanding of *why* idioms mean what they do. In other words, my treatment of the phenomenon is not restricted within conceptual metaphor frameworks but inclined towards a more holistic

approach in which motivation is an effect of various *interpretation/inferential strategies* amongst which conceptual metaphor might be one contributing factor. Other types of motivation used in this thesis include symbols/emblems, knowledge of the etymology of the idioms, and general socio-cultural world knowledge. In this section, through analyses of linguistic data from corpora, the empirical relationship between the motivation of an idiom and its creative uses in real life discourse will be illuminated.

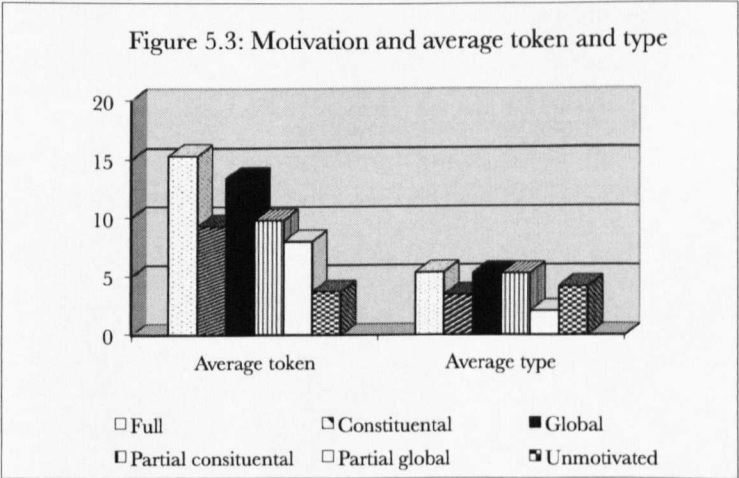
5.4.2. Charts and discussions

5.4.2.1. Does motivation correlate with idiomatic creativity?

By grouping idioms into two categories, i.e. motivated (regardless of the degree or type of motivation) and unmotivated idioms, a descriptive analysis of the productivity and flexibility of each group is portrayed in table 5.16.

In both aspects (productivity and flexibility), motivated idioms are found to outperform unmotivated idioms, which is made explicit by the higher mean variant token (14 tokens per motivated idiom compared to 3.5 tokens per unmotivated idiom) and higher mean variant type (3 types per motivated idiom and 1.3 types per unmotivated idiom). The dominance of motivated idioms is apparently most significant in variant token. This suggests a preference amongst the speaker/writer in using motivated idioms in creative forms in discourse.

Table 5.16: Descriptive analysis of motivation and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Motivated	83	13.9036	19.06319	2.09246	9.7411	18.0662	.00	99.00
	Unmotivated	17	3.5294	5.31646	1.28943	.7959	6.2629	.00	19.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Motivated	83	2.9157	2.25888	.24794	2.4224	3.4089	.00	10.00
	Unmotivated	17	1.3529	1.53872	.37319	.5618	2.1441	.00	4.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00



Statistically, the differences in pure numbers between the two groups of idioms are found to be significant by the Mann-Whitney test (table 5.17). The difference between motivated and unmotivated idioms in terms of variant token is found to be significant at $z = -3.019$ ($p = 0.03$) and in terms of variant type at $z = -2.759$ ($p = 0.06$). Thus, motivated idioms have significantly higher frequency as well as wider ranger of creativity than unmotivated idioms. In the sample, as a result, it seems that if people can perceive an idiom as ‘making sense’ (motivated), they tend to use it more often in creative forms than they do an unmotivated idiom, which does not make sense to them. At the same time, with well-formed idioms, people use more manipulation strategies to create a wider range of creative variants than they do with ill-formed idioms.

Table 5.17: Mann-Whitney test for motivation and variant token and variant type		
	Variant token	Variant type
Mann-Whitney U	378.000	408.500
Wilcoxon W	531.000	561.500
Z	-3.019	-2.759
Asymp. Sig. (2-tailed)	.003	.006
a. Grouping Variable: Availability of motivation		

The Spearman correlation test (table 5.18) confirms that the differences between the groups as found by the Mann-Whiney test also mean that there is a significant correlation between motivation and idiomatic creativity. Motivation is statistically validated as correlated with variant token at $r_s = 0.303$ (rather weak) ($p = 0.02$) and with variant type at $r_s = 0.77$ ($p = 0.05$). It can therefore be claimed that the general hypothesis is substantiated; that is, motivation is significantly related to creativity usage of idioms in the data.

Table 5.18: Spearman correlation between motivation and variant token and variant type					
Spearman's rho	Availability of motivation	Correlation Coefficient	Availability of motivation	Variant token	Variant type
		Sig. (2-tailed)	.	.002	.005
		N	100	100	100
	Variant token	Correlation Coefficient	.303**	1.000	.796**
		Sig. (2-tailed)	.002	.	.000
		N	100	100	100
	Variant type	Correlation Coefficient	.277**	.796**	1.000
		Sig. (2-tailed)	.005	.000	.
		N	100	100	100
	**. Correlation is significant at the 0.01 level (2-tailed).				

5.4.2.2. *Does the degree of motivation correlate with idiomatic creativity?*

There is an undeniable fact that the group of ‘motivated idioms’ is highly heterogeneous. By grouping idioms according to their degrees of motivation (from idioms with full to partial motivation to unmotivated idioms) we can further test the effects of the degrees of

motivation on idiomatic creativity. The hypothesis is that the more motivated an idiom is, the more creative its usage is.

According to the mean analysis in table 5.19, it is immediately apparent that, in general, fully motivated idioms have the highest degree of creativity, boasting both the highest mean variant token (16 tokens per idiom) and highest mean variant type (more than 3 types per idiom). Unmotivated idioms, as expected, have the lowest mean variant token (3.5 token per idiom).

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Unmotivated	17	3.5294	5.31646	1.28943	.7959	6.2629	.00	19.00
	Partial global motivation	2	8.0000	11.31371	8.00000	-93.6496	109.6496	.00	16.00
	Partial constituentual motivation	9	9.7778	13.82731	4.60910	-.8508	20.4064	1.00	45.00
	Global motivation	27	13.7778	15.71215	3.02381	7.5623	19.9933	.00	66.00
	Constituentual motivation	8	9.3750	10.01338	3.54027	1.0036	17.7464	.00	30.00
	Full motivation	37	16.2973	23.81744	3.91556	8.3562	24.2384	.00	99.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Unmotivated	17	1.3529	1.53872	.37319	.5618	2.1441	.00	4.00
	Partial global motivation	2	1.0000	1.41421	1.00000	-11.7062	13.7062	.00	2.00
	Partial constituentual motivation	9	2.7778	1.64148	.54716	1.5160	4.0395	1.00	5.00
	Global motivation	27	3.0000	2.05688	.39585	2.1863	3.8137	.00	7.00
	Constituentual motivation	8	1.6250	1.50594	.53243	.3660	2.8840	.00	4.00
	Full motivation	37	3.2703	2.59967	.42738	2.4035	4.1370	.00	10.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00

Scatterplot 5 (appendix 4) shows the distributions of idioms by degrees of motivation and their relative variant tokens and types in discourse. A pattern can be detected in variant token, in which the more motivated idiom groups have the higher average token than the next group. Variant type can be seen as forming a partial pattern in which the order between some groups falling out of the pattern expected by the hypothesis.

The Spearman correlation test (table 5.20) reveals a significant correlation between the degrees of motivation and variant token ($r_s = 0.202$ and $p < 0.05$) as well as between the degree of motivations and variant type ($r_s = 0.226$ and $p < 0.05$). As a whole, this means that the degrees of motivation are related to the degrees of creative usage of idioms in discourse. The more motivated an idiom, the more often it is used in a variant form, and the more ways there are to manipulate it. However, the correlations are both weak, since the correlation coefficient in each case is much closer to 0 than to 1.

Table 5.20: Spearman correlation between the degree of motivation and variant token and variant type					
			Degree of motivation	Variant token	Variant type
Spearman's rho	Degree of motivation	Correlation Coefficient	1.000	.202*	.226*
		Sig. (2-tailed)	.	.044	.024
		N	100	100	100
	Variant token	Correlation Coefficient	.202*	1.000	.796**
		Sig. (2-tailed)	.044	.	.000
		N	100	100	100
	Variant type	Correlation Coefficient	.226*	.796**	1.000
		Sig. (2-tailed)	.024	.000	.
		N	100	100	100
*. Correlation is significant at the 0.05 level (2-tailed).					
**. Correlation is significant at the 0.01 level (2-tailed).					

5.4.2.3. *Do conceptual metaphors affect idiomatic creativity among motivated idioms?*

One of the most debated questions in the literature involves the role of conceptual metaphors in both the understanding of original idioms and the creation and interpretation of idiom variants. Attempts to probe into this particular can of worms have traditionally been in the form of psycholinguistic experiments, in which the comprehension and interpretation processes can more or less be monitored. With corpus data, the question that might possibly be answered would involve the correlation or otherwise of conceptual metaphor motivation with flexibility and productivity in idioms.

By dividing the 100 idioms into two big groups of metaphorically motivated idioms and non-metaphorically motivated idioms, analyses can be carried out to find the correlation (if any). It is duly noted, however, that within the group of metaphorically motivated idioms, there might be other types of motivation working in combination with conceptual metaphors (indeed, there are not a lot of idioms with single motivation). The group of non-metaphorically motivated idioms includes those which are motivated via conceptual metonymies, emblems, lexicalised figurative meanings, etymology or knowledge of the world.

According to mean frequencies (table 5.21), it is found that non-metaphorically motivated idioms have slightly higher mean variant type (3 versus 2.8 types per idiom), while metaphorically motivated idioms have slightly higher mean variant token (14 versus 12 tokens per idiom). In other words, both groups seem to have the same creativity potential. Scatterplot 6 (appendix 4) shows the relative idiom-by-idiom distribution of the two groups of idioms, which helps elaborate on the actual range of productivity and flexibility by each individual idiom in each group. It is clear that some of the metaphorically motivated idioms have the highest variant token and variant type. However, there are some which have lowest scores in both aspects, bringing their average productivity and flexibility scores down to approximately the same level as non-metaphorically motivated idioms.

Table 5.21: Descriptive analysis of metaphorical motivation and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Non-metaphorical motivation	14	12.0000	14.05484	3.75631	3.8850	20.1150	1.00	51.00
	Metaphorical motivation	69	14.2899	19.98904	2.40640	9.4880	19.0917	.00	99.00
	Total	83	13.9036	19.06319	2.09246	9.7411	18.0662	.00	99.00
Variant type	Non-metaphorical motivation	14	3.0000	1.75412	.46881	1.9872	4.0128	1.00	7.00
	Metaphorical motivation	69	2.8986	2.35862	.28394	2.3319	3.4652	.00	10.00
	Total	83	2.9157	2.25888	.24794	2.4224	3.4089	.00	10.00

The statistical significance between the groups is tested in the Mann-Whitney test (table 5.22). The results reveal that the differences between the groups are found to be insignificant ($p = 0.667$ for variant token and $p = 0.756$ for variant type). In other words, the fact that an idiom is metaphorically motivated, or not, does not have an effect on how creatively it is used in discourse.

Table 5.22: Mann-Whitney test for metaphorical motivation and variant type and variant token		
	Variant type	Variant token
Mann-Whitney U	448.000	457.500
Wilcoxon W	2863.000	2872.500
Z	-.430	-.311
Asymp. Sig. (2-tailed)	.667	.756
a. Grouping Variable: Motivation by metaphor		

5.4.2.4. Does the complexity of motivation affect idiomatic creativity?

It has been noticed in the literature that there are idioms with only one source of motivation, while there are others with a combination of different sources of motivation. It might be hypothesized that, compared to idioms with single motivation, idioms with complex motivation are likely to be more productive and flexible, since there are more sources of information available to motivate these idioms, which means there are more chances of these idioms appearing more transparent to the language users. With idioms with only one source of motivation, if the motivation is undetected, they will falsely be regarded as opaque, entailing invariable usage.

The mean types and tokens per idiom in each group are presented below in table 5.23, which speaks against our hypothesis in many ways. First of all, in both respects, idioms with single motivation are actually more productive (17.5 variants per idiom) and more flexible (3.3 types per idiom) compared to idioms with complex motivation (12.5 and 2.7 respectively). The difference is more remarkable in variant tokens than in variant type.

Table 5.23: Descriptive analysis of complexity of motivation and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Single motivation	23	17.5652	18.37445	3.83134	9.6195	25.5109	.00	66.00
	Complex motivation	60	12.5000	19.28598	2.48981	7.5179	17.4821	.00	99.00
	Total	83	13.9036	19.06319	2.09246	9.7411	18.0662	.00	99.00
Variant type	Single motivation	23	3.3913	1.80250	.37585	2.6118	4.1708	.00	7.00
	Complex motivation	60	2.7333	2.39962	.30979	2.1134	3.3532	.00	10.00
	Total	83	2.9157	2.25888	.24794	2.4224	3.4089	.00	10.00

The differences in pure numbers and distribution are significant, as revealed by the Mann-Whitney test (table 5.24) with regard to variant token (at $z = -1.964$ with $p = 0.05$), which means people use single motivation idioms in creative forms more often than they do complex motivation idioms. However, the two groups are not significantly different in terms of variant type. In other words, the number of types of creative variants people create is similar for both groups of idioms.

Table 5.24: Mann-Whitney test for the complexity of motivation and variant type and variant token		
	Variant token	Variant type
Mann-Whitney U	497.500	534.500
Wilcoxon W	2327.500	2364.500
Z	-1.964	-1.599
Asymp. Sig. (2-tailed)	.050	.110
a. Grouping Variable: Sources of motivation		

The Spearman correlation test (table 5.25) confirms that the complexity of motivation is significantly correlated with variant token (at $r_s = -0.217$ with $p < 0.05$) but not with variant type ($p = 0.11$).

Table 5.25: Spearman correlation between the complexity of motivation and variant token and variant type					
			Complexity of motivation	Variant token	Variant type
Spearman's rho	Complexity of motivation	Correlation Coefficient	1.000	-.217*	-.177
		Sig. (2-tailed)	.	.049	.110
		N	83	83	83
	Variant token	Correlation Coefficient	-.217*	1.000	.748**
		Sig. (2-tailed)	.049	.	.000
		N	83	83	83
	Variant type	Correlation Coefficient	-.177	.748**	1.000
		Sig. (2-tailed)	.110	.000	.
		N	83	83	83
*. Correlation is significant at the 0.05 level (2-tailed).					
**. Correlation is significant at the 0.01 level (2-tailed).					

The negative correlation means that the less complex the motivation, the more tokens are found for the idiom. Both groups of idioms appear to have similar range of type of creative variants. The reason why the multiple sources of motivation do not seem to affect variant type might be due to the fact that they do not render an idiom more transparent than a single source motivation during comprehension and manipulation. Further psycholinguistic experiments will need to be carried out to flesh out this hypothesis.

5.4.3. Conclusions

Findings regarding the relationship between motivation and productivity and flexibility can be summarised as follows:

- (i) Motivation and idiomatic creativity are found to be significantly correlated, in which motivated idioms are more productive and more flexible than unmotivated idioms.
- (ii) A significant correlation between the degree of motivation and creativity is also found, such that the more motivated an idiom (e.g. fully motivated, constitutently motivated), the more variant token and variant type will be found for that idiom.
- (iii) No significant correlation is found between the availability of metaphorical motivation and creativity. The types and tokens of idiom variants do not depend on whether the motivations of an idiom are based on conceptual metaphors or not.
- (iv) There is a significant negative correlation between the complexity of the motivation of idioms and variant token, which suggests that idioms with single motivation have higher number of variant tokens than those with complex motivation. Variant type, on the other hand, is not correlated with structural complexity.

The most important finding is that there is a significant correlation between motivation and creativity as well as between the degree of motivation and creativity. As such, the hypothesis about a possible correlation between motivation and creativity in idioms can be regarded as substantiated. In other words, these results confirm psycholinguistic findings regarding the facilitating effects of motivation on idiom comprehension and idiom manipulation.

It is acknowledged that, unlike well-formedness, motivation cannot be one hundred per cent accurately and objectively determined because the precise course of applying different sources of knowledge (conceptual metaphor, conceptual metonymy, etymology, general knowledge, etc.) to motivate a particular idiom is highly dependent on each language user. In reality, a fully motivated idiom might still remain rather opaque to certain people,

leading to limited creative uses of said idiom by said people. However, this is also where corpus data has a clear advantage over psycholinguistic experiments, in that, considering the size of the BNC with its multimillion entries from different genres by a heterogeneous random selection of speakers/writers, the average results should still reflect an unbiased, non-manipulated picture of the relationship between motivation and creativity.

5.5. Analysability and idiomatic creativity

5.5.1. *An analytical framework of analysability*

Analysability refers to the interpretative process through which a 'one-to-one correspondence between the parts of the semantic value of the expression as a whole and the meanings of the constituent parts of the expression can be detected' (Geeraerts, 1995: 60). It is sometimes referred to as the 'isomorphism' of an idiom.

Unlike motivation, which helps language users to understand *why* idioms mean what they do, analysability shows the possibility to break the idioms down into analysable, meaningful components even though there might not be a plausible explanation for such a break-down. For instance, 'beans' in *spill the beans the book* can be understood as 'information' (hence an analysable relation) this correspondence is not motivated because there is no explanation as to why 'beans' means what it does. In other words, motivation and analysability are two very closely related, and often mutually facilitating, aspects of idiomaticity, but they are definitely two distinctive phenomena. Thanks to analysability, unmotivated idioms such as *by the book* can still be creatively manipulated (e.g. *doing business by the (good) book*).

The hypothesis to be tested here is that the more analysable an idiom is, the more productive and flexible it is in discourse. The foundation of such a hypothesis is based on the assumption that there is an inclination towards not using a particular component in creative ways if it does not have an analysable meaning. Linguists who have suggested a plausible link between analysability and idiomatic creativity include, among others, Nunberg (1978), Nunberg et al. (1994), Gibbs (1990, 1994), Gibbs and Nayak (1989), Gibbs et al. (1989) (see chapter 3):

If a component part of an expression is opaque in relation to the meaning of the whole expression there will be a tendency against using it productively in that expression. Thus 'red' in *red herring* will not be substituted by any other colour since the significance of redness in the expression is entirely opaque (Hudson, 1998: 35)

To empirically validate the existence or otherwise of such a correlation between analysability and idiomatic creativity, the 100 control idioms in the data pool are categorised into three groups for further statistical analyses:

1. *Fully analysable*: there is absolute symmetry between the literal and figurative meaning so that components can be detected and linked between the two meanings
2. *Partially analysable*: one or more of the components of the literal meaning can be linked to components of the figurative meaning but there are still unanalysable parts
3. *Unanalysable*: no correspondence can be found between components of the literal meaning and those of the figurative meaning.

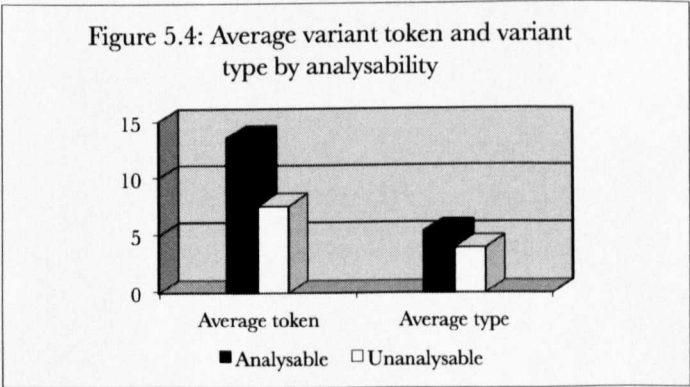
5.5.2. Charts and discussions

5.5.2.1. Is there a correlation between analysability and idiomatic creativity?

First of all, if we group the idioms into two categories of ‘analysable’ (including both fully and partially analysable idioms) and ‘unanalysable’, the hypothesis to be tested is that: there is a correlation between creativity and analysability.

It appears that, in pure numbers (table 5.26), as portrayed in figure 5.4), idioms with analysability (regardless of the degree of full or partial analysability) are more productive and more flexible than idioms without analysability. In terms of productivity, the dissimilarity between two groups can be clearly observed (more than 14 for analysable idioms versus nearly 7 for unanalysable idioms.) In terms of flexibility, the difference between two groups is less marked, with analysable idioms having only 1 type more than unanalysable idioms (3 versus 2 types per idiom).

Table 5.26: Descriptive analysis of analysability and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Unanalysable	29	6.8621	8.83469	1.64056	3.5015	10.2226	.00	37.00
	Analysable	71	14.2958	20.15894	2.39243	9.5242	19.0673	.00	99.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Unanalysable	29	1.9310	1.66757	.30966	1.2967	2.5653	.00	6.00
	Analysable	71	2.9437	2.36575	.28076	2.3837	3.5036	.00	10.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00



Such a pattern, however, is found to be statistically insignificant by the Mann-Whitney test (table 5.27). With $p > 0.05$ for both variant type and variant token, the groups of analysable and unanalysable idioms have to be assumed to have similar frequency and range of creative variants. It seems that possibility of breaking the idioms down into analysable parts does not affect people's creative usage of idioms in discourse.

Table 5.27: Mann-Whitney test for analysability and variant token and variant type		
	Variant token	Variant type
Mann-Whitney U	798.000	782.000
Wilcoxon W	1233.000	1217.000
Z	-1.766	-1.903
Asymp. Sig. (2-tailed)	.077	.057
a. Grouping Variable: Availability of analysability		

Similarly, no significant correlation between analysability and idiomatic creativity was found by the Spearman correlation test (table 5.28). The availability of analysability is not correlated to either the productivity (variant token) or flexibility (variant type) of idioms in discourse. It is postulated that this lack of correlation might be due to the low frequencies of some analysable idioms, leading to low counts of both token and type.

Table 5.28: Spearman correlation between analysability and variant token and variant type					
			Availability of analysability	Variant token	Variant type
Spearman's rho	Availability of analysability	Correlation Coefficient	1.000	.178	.191
		Sig. (2-tailed)	.	.077	.057
		N	100	100	100
	Variant token	Correlation Coefficient	.178	1.000	.796**
		Sig. (2-tailed)	.077	.	.000
		N	100	100	100
	Variant type	Correlation Coefficient	.191	.796**	1.000
		Sig. (2-tailed)	.057	.000	.
		N	100	100	100
*. Correlation is significant at the 0.05 level (2-tailed).					
**. Correlation is significant at the 0.01 level (2-tailed).					

5.5.2.2. *Is there a correlation between the degree of analysability and idiomatic creativity?*

Another hypothesis is formulated relating to the degree of analysability amongst idioms and their creativity, such that the more analysable the idioms, the more creative usage they have in discourse. By grouping the 100 idioms by the degree of analysability (fully analysable, partially analysable and unanalysable), we can examine the relationship of these three groups in terms of creativity. If the hypothesis holds, in pure number, fully analysable idioms should have the most number of creative variant tokens and the widest range of creative variant types, followed by partially analysable idioms, while unanalysable idioms fall behind.

The mean variant type and mean variant token per idiom in each group can be seen in table 5.29 below. It is immediately apparent from the table that fully and partially analysable idioms outperform unanalysable idioms in both respects. The pattern can be seen as supporting the hypothesis, with fully analysable idioms having the highest means (14.3 tokens and 3 types per idiom), followed closely by partially analysable idioms (14.2 tokens and 2.6 types per idiom) and, then unanalysable idioms (nearly 7 tokens and 2 types per idiom). That is, the more analysable an idiom, the more creative its behaviour is in discourse.

Table 5.29: Descriptive analysis of degrees of analysability and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Unanalysable	29	6.8621	8.83469	1.64056	3.5015	10.2226	.00	37.00
	Partially analysable	15	14.2667	18.50663	4.77839	4.0180	24.5153	.00	51.00
	Fully analysable	56	14.3036	20.73725	2.77113	8.7501	19.8570	.00	99.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Unanalysable	29	1.9310	1.66757	.30966	1.2967	2.5653	.00	6.00
	Partially analysable	15	2.6000	2.16465	.55891	1.4013	3.7987	.00	8.00
	Fully analysable	56	3.0357	2.42685	.32430	2.3858	3.6856	.00	10.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00

In terms of statistical significance, however, the differences between the four groups are not found to be significant according to the Kruskal-Wallis test (table 5.30). With $p > 0.05$ for both variant token and variant type, the null hypothesis about the similar distribution of the four groups of idioms cannot be rejected. That is, the groups have to be assumed to be similar in terms of variant token and variant type. In other words, people seem to use idioms of all degrees of analysability equally in terms of creativity.

Table 5.30: Kruskal-Wallis test for the degrees of analysability and variant token and variant type		
	Variant token	Variant type
Mann-Whitney U	798.000	782.000
Wilcoxon W	1233.000	1217.000
Z	-1.766	-1.903
Asymp. Sig. (2-tailed)	.077	.057
a. Grouping Variable: Availability of analysability		

The Spearman correlation test (table 5.31) does not find a significant correlation between the degrees of analysability and idiomatic creativity, either ($p = 0.98$ for variant token and $p = 0.53$ for variant type). This result, as well as the Kruskal-Wallis test result above, is contradictory to suggestions made in the literature, particularly those followed psycholinguistic experiments (see chapter 3), about the role of analysability in idiomatic creativity.

Table 5.31: Spearman correlation between the degrees of analysability and variant token and variant type					
Spearman's rho	Degree of analysability	Correlation Coefficient	Degree of analysability	Variant token	Variant type
			1.000	.167	.194
			Sig. (2-tailed)	.098	.053
	Variant token	Correlation Coefficient	N	100	100
			.167	1.000	.796**
			Sig. (2-tailed)	.098	.000
	Variant type	Correlation Coefficient	N	100	100
			.194	.796**	1.000
			Sig. (2-tailed)	.053	.000
			N	100	100

** . Correlation is significant at the 0.01 level (2-tailed).

5.5.3. Conclusions

Frequency counts and statistical tests on corpus data suggest the following trends about the relationship between analysability and productivity and flexibility in idioms:

- (i) In pure numbers, analysability does seem to produce a pattern in terms of both productivity and flexibility. Analysable idioms have both higher mean token and mean type than unanalysable idioms.
- (ii) Also in pure number, the degrees of analysability do result in differences in terms of productivity and flexibility. Fully analysable idioms are found to have higher mean token and mean type, followed closely by partially analysable idioms, while unanalysable idioms are far behind in both regards.
- (iii) The differences, however, are not statistically significant. In other words, we cannot say with confidence that people use analysable idioms more often in creative forms than they do unanalysable idioms
- (iv) No correlation was found between analysability and creative usage of idioms in discourse.

The missing of a correlation between analysability and idiomatic creativity is counter-intuitive and discordant with most of the claims in the literature about the role of analysability in idiomatic creativity. This might be due to a number of factors. First of all, analysability is a concept that is sometimes criticised for its dependence on the particular evocation of the figurative meaning (Burger, 1998a, in Langlotz, 2006). For example, *make one's blood boil*, if explained as 'make someone angry' will not have a clear correspondence between the literal arrangement and the figurative reading. However, if it is explained as 'make someone's temper rise' then there is a possible one to one correspondence between the components of the literal and the figurative meaning, in which 'make' is used literally, while 'blood' can be roughly understood as 'temperature' and 'boil' means 'rise'. The idiom

can now be viewed as analysable. This issue highlights the subjectivity of the concept of analysability. It is highly dependent on the language user's ability and willingness to see the links and correspondences within an idiom. If a speaker only sees part of the analysability of an idiom, he/she is most likely to focus on manipulating that part and leaves the rest of the phrase intact. Hence, it seems to boil down to the each individual's ability to assign phase-induced meaning to constituents. This subjectivity is allegedly eliminated when the database is large enough, which is the case of this thesis. Further investigations into the technical aspects of statistical tests might explain the missing of a significant correlation between analysability and idiomatic creativity, even though raw frequency counts strongly suggest such a correlation.

5.6. Transparency and idiomatic creativity

Since the three factors of well-formedness, motivation, and analysability jointly shape the transparency of idioms, it is desirable that they be analysed in combination as well as separately. There have been efforts in the literature to combine the three parameters into one transparency cline in the quest for better insights into the mechanisms of idiomatic creativity (see chapter 3). If our dependent variables (variant token and variant type) were normally distributed, or if the residuals from a regression using these variables would be normally distributed, we could test the combined effects of well-formedness, motivation, and analysability on variant token and variant type using multiple regression tests. Unfortunately, our data is not normally distributed. As a result, we need to try to combine the three factors into one cline of transparency before further tests can be run. There is a certain degree of difficulty with the process (see chapter 3), since there are medial types that are problematic to place on the transparency cline. The first approach is, therefore, to consider these factors as equally contributing to the transparency of idioms so that an order can be established based on the presence of one or more factors in an idiom:

1. *Fully opaque idioms*: idioms without any of the factor, i.e. ill-formed, unmotivated, unanalysable
2. *Partially opaque idioms*: idioms with only one out of three factors, including: (a) ill-formed, unmotivated, analysable idioms; (b) ill-formed, motivated, unanalysable idioms; and (c) well-formed, unmotivated, unanalysable idioms
3. *Partially transparent idioms*: idioms with two out of three factors, including: (a) ill-formed, motivated, analysable idioms; (b) well-formed, unmotivated, analysable idioms; and (c) well-formed, motivated, unanalysable idioms
4. *Fully transparent idioms*: idioms with all three factors, i.e. well-formed, motivated and analysable idioms

In pure descriptive terms (table 5.32), we can see a clear pattern in both variant token and variant type, such that fully transparent idioms have the highest means (16.5 tokens and 3 types per idiom), followed by partially transparent idioms (12.5 tokens and 2.5 types per

idiom), then partially opaque idioms (5 tokens and 2 types per idiom) and, expectedly, fully opaque idioms bottom the table with less than 1 token and 0.5 type per idiom.

Table 5.32: Descriptive analysis of degrees of transparency and variant token and variant type									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Fully opaque	9	.8889	1.36423	.45474	-.1597	1.9375	.00	4.00
	Partially opaque	12	4.8333	4.36585	1.26031	2.0594	7.6073	.00	12.00
	Partially transparent	38	12.4211	18.35186	2.97707	6.3889	18.4532	.00	99.00
	Fully transparent	41	16.4878	20.32624	3.17443	10.0721	22.9036	.00	86.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Fully opaque	9	.5556	.72648	.24216	-.0029	1.1140	.00	2.00
	Partially opaque	12	1.9167	1.44338	.41667	.9996	2.8337	.00	4.00
	Partially transparent	38	2.6842	2.08087	.33756	2.0002	3.3682	.00	8.00
	Fully transparent	41	3.2927	2.45198	.38293	2.5187	4.0666	.00	10.00

The neat pattern as found in raw frequencies is found to be significant in statistic terms using the Kruskal-Wallis test (table 5.33). The four groups are determined to be significantly different from each other in terms of variant token ($p = 0.02$) as well as in terms of variant type ($p = 0.03$).

Table 5.33: Kruskal-Wallis test for degrees of transparency and variant token and variant type		
	Variant token	Variant type
Chi-Square	15.042	13.783
df	3	3
Asymp. Sig.	.002	.003
a. Kruskal Wallis Test		
b. Grouping Variable: Combination of three factors		

Unfortunately, post-hoc test to the Kruskal-Wallis test to find out which groups are significant different from each other are not available from SPSS. As a result, a series of Mann-Whitney tests are carried out on each pair (see appendix 5), which show that the groups which are found to be different from each other are:

- Group 1 – group 2
- Group 1 – group 3
- Group 1 – group 4

In other words, group 1 (fully opaque) is significantly different from other groups, which confirms claims elsewhere in the literature about the blocking of creative operations in opaque idioms. However, there is no difference found between groups 2 (partially opaque), 3 (partially transparent), and 4 (fully transparent), which rejects the hypothesis about the influence of the degrees of transparency on idiomatic creativity. The most we can draw

from the results is that as long as idioms are transparent to a certain extent, they are more creatively used than fully opaque idioms.

This partial result is reflected in the significant correlation between the degrees of transparency and idiomatic creativity by the Spearman correlation test (table 5.34). The degrees of transparency and variant token are correlated at $r_s = 0.325$ with the significance of $p = 0.01$, which is remarkably similar to the correlation between degree of transparency and variant type ($r_s = 0.316$, $p = 0.01$). In other words, the more transparent an idiom, the more frequent it is used in creative forms in discourse and the wider range of creative variant types it can accommodate. The correlations, it is noted, are not strong, allowing for exceptions to the trends (i.e. the tie between medial types).

Table 5.34: Spearman correlation of degrees of transparency and variant token and variant type					
Spearman's rho	Combination of three factors	Correlation Coefficient	Combination of three factors	Variant token	Variant type
		Sig. (2-tailed)	1.000	.325**	.316**
		N	100	100	100
	Variant token	Correlation Coefficient	.325**	1.000	.796**
		Sig. (2-tailed)	.001	.	.000
		N	100	100	100
	Variant type	Correlation Coefficient	.316**	.796**	1.000
		Sig. (2-tailed)	.001	.000	.
		N	100	100	100
**. Correlation is significant at the 0.01 level (2-tailed).					

Criticism, however, might be levelled at the simplistic computation of transparency as employed in previous tests. It might be argued that the three factors (well-formedness, motivation and analysability) do not contribute to the transparency of idioms in an equal manner. For example, despite both having only one factor, the group of ‘ill-formed, unmotivated, but analysability’ idioms has to be treated as more transparent than ‘well-formed but unmotivated, unanalysable’ idioms, because well-formedness does not contribute as much to transparency as motivation and analysability do. As a result, a more refined framework of idiom transparency is required.

A second approach is, therefore, taken in which the four groups of idioms are further categorised, resulting in 8 groups of idioms with differing degrees of transparency. However, the groups are treated as categorical instead of constituting a cline:

- Group 1: well-formed, motivated, analysable
- Group 2: ill-formed, motivated, analysable
- Group 3: well-formed, unmotivated, analysable
- Group 4: ill-formed, unmotivated, analysable
- Group 5: well-formed, motivated, unanalysable

- Group 6: ill-formed, motivated, unanalysable
- Group 7: well-formed, unmotivated, unanalysable
- Group 8: ill-formed, unmotivated, unanalysable

Mean comparisons of variant token and variant type of each group are presented below in table 5.35. The group with the highest average token is group 1 (well-formed, motivated, analysable), whereas the group with the highest average type is group 3 (well-formed, unmotivated, analysable). Having the lowest means on both variant type and variant token is group 8 (ill-formed, unmotivated, unanalysable).

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Variant token	Group 1	41	16.4878	20.32624	3.17443	10.0721	22.9036	.00	86.00
	Group 2	24	12.5417	21.88503	4.46726	3.3004	21.7829	.00	99.00
	Group 3	2	11.5000	10.60660	7.50000	-83.7965	106.7965	4.00	19.00
	Group 4	4	3.7500	4.50000	2.25000	-3.4105	10.9105	.00	9.00
	Group 5	12	12.3333	10.99862	3.17503	5.3451	19.3215	.00	37.00
	Group 6	6	4.8333	4.11906	1.68160	.5106	9.1560	.00	10.00
	Group 7	2	7.0000	7.07107	5.00000	-56.5310	70.5310	2.00	12.00
	Group 8	9	.8889	1.36423	.45474	-.1597	1.9375	.00	4.00
	Total	100	12.1400	17.91395	1.79139	8.5855	15.6945	.00	99.00
Variant type	Group 1	41	3.2927	2.45198	.38293	2.5187	4.0666	.00	10.00
	Group 2	24	2.5417	2.26465	.46227	1.5854	3.4979	.00	8.00
	Group 3	2	4.0000	.00000	.00000	4.0000	4.0000	4.00	4.00
	Group 4	4	1.2500	1.89297	.94648	-1.7621	4.2621	.00	4.00
	Group 5	12	2.7500	1.86474	.53831	1.5652	3.9348	.00	6.00
	Group 6	6	2.1667	1.32916	.54263	.7718	3.5615	.00	4.00
	Group 7	2	2.5000	.70711	.50000	-3.8531	8.8531	2.00	3.00
	Group 8	9	.5556	.72648	.24216	-.0029	1.1140	.00	2.00
	Total	100	2.6500	2.22645	.22264	2.2082	3.0918	.00	10.00

The Kruskal-Wallis test results show that the groups are indeed significantly different both in terms of variant token ($p = 0.020$) and variant type ($p = 0.022$ (table 5.36).

Table 5.36: Kruskal-Wallis test for groups of transparency and variant token and variant type		
	Variant token	Variant type
Chi-Square	16.634	16.313
df	7	7
Asymp. Sig.	.020	.022
a. Kruskal-Wallis Test		
b. Grouping Variable: Combination of characteristics		

Further Mann-Whitney tests (see appendix 6) find the specific groups that are significantly different from each other, including:

- Group 1 – group 8
- Group 2 – group 8

Group 3 – group 8
Group 5 – group 8
Group 6 – group 8
Group 7 – group 8

That is, group 8 (ill-formed, unmotivated, unanalysable) (i.e. without any transparency facilitating factor) has significantly lower variant type and variant token than any other group that has at least one facilitating factor. Other groups (2, 3, 4, 5, 6, and 7) are not different from each other. These results mirror the previous test results of transparency using four groups of idioms (fully opaque, partially opaque, partially transparent and fully transparent). As a result, we claim with a certain degree of confidence that if an idiom is fully opaque (without literal well-formedness, motivation and analysability), it is used less creatively than other idioms that do have some transparency.

Summary

The determination of the degrees of transparency of idioms on the basis of the combination of well-formedness, motivation and analysability has been known in the literature to pose a serious challenge to researchers, particularly those who seek a rigid framework for quantitative analyses such as the current thesis. Applying two different approaches to idiom transparency to the analytical process, we achieve one significant result. The group of idioms that is least used in creative form is fully opaque idioms, which is significantly different from other semi or fully transparent idioms. Other groups cannot be claimed with confidence to differ from each other in terms of creative tokens or creative types. In other words, regardless of how we compute transparency (as a cline or as groups), only a partial correlation can be found between transparency and idiomatic creativity. That is, if an idiom is completely opaque, it has least creative usage in discourse compared to other types of idioms. People, therefore, appear to avoid using opaque idioms in creative ways.

It is noticed that, despite the statistics, there are exceptions to this general claim about transparency and idiomatic creativity. For example, the opaque idiom *red herring* is found to have variants where substitutions are made of the constituent 'red' (which is considered semantically empty within the idiom):

- [1] *Green herring*
- [2] *The laptop herring*

Of course, without supporting co-texts and contexts, the substitutions are not automatically meaningful. However, they become clear and meaningful when the full contexts are exposed. In the conference paper title [1], the subtitle *resisting the cooptation of environmental ideologies and anti-systemic movements* makes explicit the meaning of the adjective 'green', which is conventionally associated with environmentalism. Thus, *green herring* in this context refers

to what is being mobilised to make the speaker or organization appear to hold themselves accountable to the moral requirements of environmental problems, while distracting from the truth, e.g. exploitation of nature or of human resources for maximization of profits. Similarly, *the laptop herring* was created in which 'laptop' is used to address the topic of the article (organizations banning the use laptops at meetings, believing they lower concentration levels). The creative variant, as a result, communicates a clever critic, i.e. blaming laptops is just a diversion; the actual problem is that people do not understand the value of the meeting).

Admittedly, such coinages as these are highly sophisticated, but the point is that it is not impossible to substitute 'red' with other lexical items even though this constituent is empty and the idiom is regarded as opaque. With the right context, a lot of the 'unorthodox' manipulations with opaque idioms become meaningful, and are often highly creative, humorous, and/or emphatic.

In short, these examples of creative uses of unmotivated unanalysable idioms highlight the exceptions to the rules, particularly in the right contexts. It is the goal of the next chapter to qualitatively look at each variant, particularly the exceptions to the rules, to sketch a framework based on theories of context that can explain the productivity and flexibility of idioms.

5.7. Conclusions

This chapter aims to use corpus data to test the hypotheses (mostly borne out of psycholinguistic experiments and cognitive frameworks) of the effects of the semantic characteristics of an idiom on its creative usage in real life discourse.

It is found that when considered separately literal well-formedness and motivation have significant correlations with creativity, but analysability does not. This means that in this particular data pool, people tend to use motivated, well-formed idioms in creative forms more often than they do unmotivated, ill-formed idioms. The range of variant types that motivated, well-formed idioms can accommodate is also significantly better than that by unmotivated, ill-formed idioms. Analysability, however, is not found to have a significant effect on idioms in terms of creative usage. The frequency and range of variant types by both analysable and unanalysable idioms do not appear to be significantly dissimilar.

The combined effect of the three factors of well-formedness, motivation, and analysability, which make up the transparency of an idiom, however, is less clear. Fully opaque (ill-formed, unmotivated, and unanalysable) idioms are used least often in creative forms and have the narrowest range of variant types in comparison to other groups of idioms that have some degrees of transparency. The differences between partially transparent and fully transparent idioms, however, are not statistically significant.

In short, the hypotheses concerning the effect of semantic characteristics on idiomatic creativity are only partially validated in the current data pool. While parts of the hypotheses hold true, other parts do not, which suggests interfering forces that have not been accounted for in the data. These forces have been suggested elsewhere in the literature to include cognitive constraints and context. It is because of these factors that, despite the strong trends, semantic characteristics do not always determine idiomatic creativity, particularly when context-specific wordplay and puns are created for specific communicative purposes.

The next chapter addresses the necessity for a careful qualitative re-examination of idiom variants so that individual features and exceptions can be accounted for. Using a rich data pool of corpus and internet data, I will examine the creative behaviour of idioms under the combined effect of semantic characteristics, cognitive constraints and context, with an emphasis on the interaction of these factors. To be specific, the next chapter covers the followings:

- (1) the definition and analytical framework for the concept of *context dependency* amongst idiom variants
- (2) the qualitative analyses of context-dependent variants and context-independent variants to identify the most salient features in terms of lexico-grammatical adaptations in each group
- (3) the assessment of the effect of semantic characteristics, cognitive constraints and context on each subgroup of variant

It is postulated that the relationships between the factors are complex, in which the dominance of one factor over other, if any, changes according to the specific variant type in a specific context. The analyses of different types of idiom variants within a context-based framework as such is believed to reflect the true behaviour of the determinants of idiomatic creativity, thus providing a truthful representation of the creativity of idioms in real-life discourse.

CHAPTER 6

A CONTEXT-BASED MODEL OF IDIOMATIC CREATIVITY - POSSIBILITIES AND CONSTRAINTS

The role of semantic characteristics in creating more opportunities for creativity among idioms is undeniable, as confirmed in chapter 5. However, semantic characteristics alone cannot fully describe the phenomenon of idiomatic creativity in real life discourse due to the presence of other interfering forces. First of all, besides grammatical and semantic constraints, which are prerequisites for any linguistic unit above the word level, idiomatic variants are also subject to such cognitive constraints as recognisability and functionality, as we have discussed in chapter 3, which restrict the possibilities to derive certain types of variants from certain idioms despite the transparency of their semantic structures. Secondly, contextual elements might also further restrict idiomatic creativity in certain cases while facilitating it in others, depending on the particular communicative purposes of the variants. In other words, I believe that any conclusion about idiomatic creativity that is based solely on the semantic characteristics of idioms is flawed and partial. As a result, in this chapter, a corpus-based description of the transformational potential of idioms in relation to specific constraints on each type of transformation in specific context is given. The qualitative analyses in this chapter will sketch a picture of the actual possibilities and constraints on idiomatic creativity in discourse under the influence of context.

6.1. A context-based model of idiomatic creativity

With the view of looking upon the construction of an idiom variant as both constructing a grammatical structure and constructing a message in context, the analytical model of idiom variation in this chapter is fundamentally context-based, achieved via the segmentation of the material into pragmatically effective units and the analysis of these units via the appropriate lexical and grammatical devices in context so as to reveal the embedded messages.

It is argued that when we construct a message, it is a courtesy to the reader/hearer as well as a convenience for the writer/speaker, to provide the point of the message (new information, or focus) with enough context (given information, or theme) for this point to be both clearly identified and unambiguously understood (Quirk et al., 1985: 1360). Simultaneously, we will want the construction of the message to be placed in a normal linguistic framework (ibid). This view of meaning-in-use is in harmony with the view of language investigation in pragmatic traditions, i.e. 'the study of those relations between

language and context that are grammaticalised, or encoded in the structure of a language' (Levinson, 1983: 9).

The relationship between grammar and context has been the topic of interest in the linguistic community since the beginning of the pragmatics tradition. Generally speaking, the grammatical rules of a language, according to which a message is required to be coded to ensure its communicability, are seen as superseded by appropriate contexts. Quirk et al. (1985: 1325) indeed acknowledge that 'special contexts create their own usage where the general rules of the language may be waived.' Similarly, it is noticed by Nunberg (1978) that any unfamiliar expression can be understood if the context is made informative enough. This, however, does not mean that it is possible to waive *all* grammatical rules at any one time; the sanction of unfamiliar, ungrammatical expressions via context is case sensitive and subjective. In the case of idioms, which are characterized by institutionalization and liability to cognitive constraints (see chapter 3), the questions we concern ourselves with are (1) how do these restraining factors counteract with context, and (2) to what extent can context overwrite them. Glucksberg (1993) rightly comments that the determinants of idiom flexibility/productivity are complex and feature a combination of both the internal semantics of the idiom and the pragmatics of the discourse context. In this thesis, context is postulated to be the overarching facilitating factor of idiomatic creativity, capable of overwriting certain grammatical rules and/or cognitive constraints in certain types of operations to create more creative possibilities than had been acknowledged before with frameworks based solely on the internal semantic characteristics of idioms. This power of context over semantic and cognitive constraints, however, should not be overestimated and needs to be put into perspective in relation to other elements. This is the ultimate claim in this thesis.

As Lyons (1977) reminds us, natural languages are primarily designed, so to speak, for use in face-to-face interaction, and thus there are limits to the extent to which they can be analyzed without taking this [i.e. contextual information] into account. Arguably, the most obvious way in which context affects the interpretation of utterances is through the phenomena of 'reference' and 'deixis' (see Levinson, 1983; Lyons, 1977). For example, the demonstrator 'this' in *this car is nice* depends for its interpretation on the immediate context of speaking (i.e. where the interlocutors are and which car is being singled out), without which we cannot claim to understand the meaning of the utterance. The context-dependency of deixis, in other words, is reflected in the fact that the meaning of a deictic expression is not predetermined but is 'a variable or place-holder for some particular entity given by the context' (Levinson, 1983: 54). On a broader scale, the contrast between context-dependent and context-independent modes of reference is not restrictively observed in the deixis/non-deixis dichotomy but in many other kinds of referential expressions that rely on contextual information as well (ibid: 62). Indeed, many traditionally non-deictic

linguistic units have been found in the idiomatic creativity data to have context-dependent reference, as we will see later on.

By analysing idiom variants in the data according to their context-dependency, this chapter hopes to:

- (1) tease out the patterns of idiomatic creativity strategies on the basis of context
- (2) examine the role of context in overwriting grammatical and cognitive constraints as well as in maximizing the possibilities of idiomatic creativity irrespective of their internal characteristics, and
- (3) observe the failure of context in facilitating certain types of alterations, thereby highlighting the constraining factors that cannot be overwritten

The ultimate goal is to sketch a context-based model of idiomatic creativity that is capable of reflecting the joint roles of context, the internal semantic characteristics of idioms, cognitive constraints and grammatical rules in the facilitation and restriction of idiomatic creativity.

6.2. An analytical framework of idiom variants

6.2.1. Context-based classification of idiom variants

It should be emphasized here that no utterance is completely context-independent per se, because, as we have argued earlier, language is essentially meaning in context. As with idiom variants, knowledge of the meaning of the canonical idiom from which the variants are derived (i.e. intertextual context) is required first and foremost. If the hearer does not know what *open a can of worms* means, he/she will not be able to interpret the meaning of the variant *open a can of constitutional worms* either. Simultaneously, contextual information (physical or textual) is usually required for the figurative meanings of idioms to be activated, particularly those which are literally well-formed. For example, *dark horse* could be understood literally as a description of a real horse, but in the context of someone with unexpected potential to win a competition, its figurative meaning is activated. In other words, no idiom variant can be assumed to be completely independent of context as such, because background knowledge of the meaning of the canonical idiom as well as supporting context for this meaning to be activated is a prerequisite before any interpretative/inferential process can take place to make sense of the variants.

However, assuming that there is enough contextual information for the language user to recognize and understand the meaning of the canonical idiom, by context-independent variants, we mean that no extra context is required to process the variants. Upon the activation of the idiomatic meaning of the original canonical idiom, the variants can be

readily understood using inferential skills without the need for extra context. Consider these examples:

[1] *So I think we are skating on increasingly thin ice*

[2] *This is throwing the baby into the bathwater*

Variants [1] and [2] exemplify a host of context-independent substitutions in the data (which we will discuss in the later half of the chapter), in which the substitutes bear organic relationships to the original constituents so that the meanings of the variants can be worked out via general interpretative strategies. In the case of variant [1], it is considered context-independent because the language user, upon the activation of the idiomatic meaning of *skating on thin ice* ('be in a dangerous situation'), can use interpretative skills to analyse the differences between the canonical idiom and the variant in terms of form, thereby inferring the differences in terms of meaning. Here, it is the knowledge of the analysability of the idiom and the phrase-induced meaning of 'thin' ('dangerous'), so that 'increasingly thin' can be interpreted as 'becoming more and more dangerous'. In other words, the variant is meaningful on account of the knowledge of the canonical idiom and the process of analogy. Similarly, in variant [2], *throw the baby out with the bathwater* is analysable with two important meaning components, i.e. *the baby* ('the good thing') and *the bathwater* ('the bad thing'), which are linked by the relationship 'throw out with' ('remove together'). This relationship, however, is replaced with its directional opposite 'throw into', which describes a reverse relationship. The variant's meaning, therefore, can be understood as 'mixing together' the good thing and the bad thing.

Variants [3] – [4] below, on the other hand, are context-dependent, i.e. their meanings cannot be worked out via interpretative strategies due to the lack of underlying motivations between the substitutes and the original constituents. Instead, their meanings are a function of the original meanings of the canonical idioms and the meanings of the substitutes in contexts.

[3] *Here at Jordans we think inside the box*

[4] *Shetlanders roll out the straw carpet*

Without context, the meaning of variant [3] appears to be inferable on the basis of the sense relation between the original constituent ('outside') and the substitute ('inside'), which are directional antonyms. With *outside the box* meaning 'creatively', the opposite *inside the box* can be understood as 'uncreatively'. The problem is that it seems strange that anyone wants to claim that they are uncreative (which is generally considered a negative trait). The context, indeed, clears up the intention behind the reversed meaning. Being an advert for Jordans, the cereal brand, *inside the box* provides an allusion to 'inside the Jordans cereal boxes', i.e. Jordans cares about what is inside their cereal box, proving quality food instead of fancy packages. The dual meaning and the creative play with the idiom is interesting and

captivating and can only be fully activated with contextual knowledge. Variant [4], in the same manner, is ambiguous even after the figurative meaning of *roll out the red carpet* ('welcome with respect') is activated, partly because 'carpet' does not have a clear phrase-induced figurative meaning, partly because 'straw' seems incompatible with the idiomatic meaning of the idiom. Both analysability and incompatibility appear to block the interpretation process of this variant. However, when extra context is made explicit (that is, the invention of a Shetland crofter: using straw mats to mop up oil slicks as they come ashore, which is welcome and applied to parts of the region), the meaning of the variant becomes clear. 'Straw carpet' here is not meant to be interpreted figuratively in relation to 'red carpet', but is merely an allusion to a piece of contextual information.

In short, context-independent variants of an idiom, as we use in this chapter, are motivated variants whose interpretations are readily available through general interpretative processes once the figurative meaning of the canonical idiom is activated so that no extra context is required. Context-dependent variants, on the other hand, are unmotivated and rely on knowledge of specific aspects of the contexts (physical, socio-cultural, textual), without which their meanings cannot be disambiguated even when the canonical idiom is recognized and understood.

6.2.2. *Lexico-grammatical classification of idiom variants*

Since it is envisaged that the role of context, as well as the effects of well-formedness, motivation and analysability, varies according to the types of manipulation, idiom variants are grouped, on a lower level, according to their lexico-grammatical alterations to the base-forms, including:

1. *Rearrangement*: By rearrangement we mean a number of lexico-grammatical operations through which variants of an idiom can be created such that the message remains the same, but established arrangement of the idiom (including its constituents and the grammatical structures connecting them) is reconfigured. Essentially, no new information is introduced into the original message of the idiom (as with insertions and substitutions), while no information is lost either (as with ellipsis or allusions). The message is preserved but the idiom is expressed in a novel lexico-grammatical arrangements. In Quirk's et al (1985: 1355) words: 'the truth value is unchanged but the presentation is different.' The differences in presentation might be executed via a number of lexical and grammatical operations in which either new lexical items or new structures, or both, are used, including *passivisation*, *grammatical highlighting*, *nominalisation*, and *rankshift*. Verb inflection is not treated as a type of creative variant due to the systematic and automatic nature of the operation and the lack of stylistic purposes. Admittedly, some types of systematic syntactic reconfiguration do not score high on the creativity ladder either; however, there are clear communicative intentions (e.g. placing emphasis) with these operations, while there is no such communicative purpose with verb inflection.

2. *Extension*: The extension of an idiom's message is different from rearrangement in that the message of the idiom is extended via the introduction of new information. This group involves a number of processes by which the established configuration of an idiom gains additional information via the insertion or substitution of lexical items.

Specific types of extension include *modification*, *substitution*, and *coordination*. With modifications, we can distinguish *adjectival premodification*, *nominal premodification*, *adverb premodification*, and *prepositional postmodification*). Substitutions, on the other hand, are analyzed according to the semantic and phonetic relations between the substitute and the substituted (*substitution by synonyms, antonyms, homophones, homographs, etc.* as well as *substitution by free lexical items*). Co-ordinations do not include further classes due to the small number of variants in the data, hence the lack of material for further classification.

3. *Reduction*: The reduction of an idiom's message is the process of removing certain components of the message while only retaining the important information that needs to be conveyed. On the surface (syntactic) level, this involves certain constituents of the original arrangements being omitted. The result is a reduced, shorter version of the original message. This is argued elsewhere to be possible via the power of allusion. When a component of the message is realised by a salient lexical item(s) and/or salient structure, the hearer/reader is able to recognise it to understand the intended allusion.

This attention to greater detail in this framework is not only desirable but also required for subsequent analyses in this chapter in meticulously examining the role of constraining and facilitating factors on idiomatic creativity.

6.3. Context-dependent idiom variants

First and foremost, context-dependent idiom variants are found to be ubiquitous in our data, featuring a wide range of alteration strategies and covering a large number of idioms. Out of the 100 idioms, 52 were found to have at least one context-dependent variant. Examples include:

- [1] *Did Activision hit a raw Niro?*
- [2] *How to eat humble hummus?*
- [3] *Can Jobs think outside the pretty box?*
- [4] *It's raining cats, dogs and lizards too*
- [5] *Strike before the iron freezes*

It is not difficult to recognise these variants as deriving from *hit a raw nerve*, *eat humble pie*, *think outside the box*, *rain cats and dogs*, and *strike while the iron's hot*. Nevertheless, even after the figurative meanings of these idioms are activated, it is still ambiguous what the intended meanings of the adaptations in [1] – [5] are due to the lack of contextual information about what the substitutions/insertions refer to. Extra context in these cases is needed for the

disambiguation of the variants. In [1], the context is that the producer of Guitar Hero 5, Activision, uses rock band Nirvana's image without permission, causing frustration amongst Nirvana's fans. The homophone 'Nirv', short for Nirvana, is used in place of 'nerve' to create a playful adaptation of the idiom while referring to the context-specific incident. Similarly, the meanings of [2] – [4] become transparent in the context of

- *eat humble hummus*: a recipe for the Middle Eastern paste 'hummus'
- *pretty box*: Steve Jobs, chief executive of Apple Inc, is sometimes criticised for his aesthetic philosophy that is style over substance, or 'form fetish'
- *rain cats, dogs and lizards*: the pet competition for kids, including a wide variety of species from chickens, dogs, cats, rabbits to spiders, lizards.

Variant [5] represents a group of variants which, on the surface, looks uncreative and unmotivated due to the synonym substitution. The variant can be mistaken as just another wording for the idiom *strike while the iron's hot*. However, in the context of this expression being an advert for a weapon shop, the play upon the polysemy of the verb 'freeze' ('the degree of cold', and 'to stop suddenly and remain motionless as when someone is held at gunpoint') renders the variant humorous and alludes to the context of weaponry. This example serves to show that even when a variant appears banal, contextual information can reveal its true stylistic value as being creative, humorous adaptation of a canonical idiom.

While these variants share the characteristics of being context-dependent, they are achieved via different of lexico-grammatical operations. Here, we have substitution by homophone (*nerve* – *nirv*), substitution by a lexical item of the same semantic field (*pie* and *hummus* are both types of baked goods), adjectival modification (*pretty*), coordination (*and lizards*), and literal paraphrases (*before the iron freezes*).

In the data, substitution and insertion are the two most popular strategies to create context-dependent variants; however, other operations are found in the data too. In what follows, these subtypes of context-dependent variants are examined in detail to highlight the role of context in creating favourable conditions for the understanding of 'unusual', incompatible and ill-formed variants, particularly those deriving from unmotivated, unanalysable idioms.

6.3.1. Context-dependent substitutions

Substitution as a strategy to create idiom variants is one of the most popular as well as problematic strategies due to its possible negative interference with the recognisability constraint. The question that is often asked with substitution is to what extent the constituents of an idiom can be substituted without losing its status and meaning? With the assumption that idioms are conceptually (not lexically) stable (see chapter 3), each idiom is seen as having a salient conceptual core that serves the purpose of activating the figurative meaning of the idiom even when the original constituents are replaced by other lexical

items. In this section, the extent to which this salient conceptual core contributes to the recognisability of substitution variants will be scrutinised.

In the data, context-dependent substitutions are found to be executed via semantically related words (synonyms, antonyms, items from the same word field), phonetically related words (homophones, homographs, polysemic words), and even unrelated lexical items. Examples of these types include:

Substitution by phonologically related items:

- [1] *Think outside the socks. With Father's Day fast approaching, if you need reminding, you may once again be wondering what to buy the Dad who really doesn't need any more pairs of socks*
- [2] *Kiss and Patel. There's more to Patel than just being one of Britain's most common surnames - it's a club too, in which members like to marry others with the same last name*

Substitutions by semantically related items:

- [3] *Green herring: Resisting the cooptation of environmental ideologies and anti-systemic movements*
- [4] *A push to add another 50 per cent to the ethanol content of some automobile fuel has opened a barrel of worms.*

Substitution by free items:

- [5] *Open up a can of spam. A federal law intended to curb the scourge of junk e-mail appears to have had little effect so far in discouraging spammers from deluging inboxes.*
- [6] *Wasps, and hornets in particular, are major predators of bees and the largest ones can make even the giant bees look puny.[...] The giant bees might seem easy pickings for a hungry hornet, but they have a trick up their abdomens.*

These substitutions share the characteristic of being context-specific, i.e. referring to particular events and dependent on knowledge of such events for their interpretations. In context-dependent substitutions such as these examples, the constraints of compatibility and grammaticality as well as the negative impacts of opacity are lifted, as long as the context is made informative enough to ensure recognisability and functionality. In other words, incompatible, ungrammatical substitutions can be encountered, even in ill-formed, unmotivated, and unanalysable idioms.

6.3.1.1. Incompatible context-dependent substitutions

The compatibility constraint on idiom variants, as discussed in detail in chapter 3, requires that any formal modification be semantically compatible with the figurative meaning of the original idiom. For example, *silently kick the bucket* is a valid variant but *repeatedly kick the bucket* is not, because general knowledge of the world tells us that that we can die silently but we cannot die repeatedly. It will be shown in this section that while the compatibility constraint

is observed in some cases, it is relaxed in others, particularly in the case of context-dependent substitutions if appropriate contexts are provided. Examples in the data include:

- [7] *A little mishap while I was building my own CAT scanner.... In fact, I was briefly able to see the inside of my sister's guinea pig, Snowball, before he caught fire. It led to an interesting expression in our house: 'not a snowball's chance in a CAT scanner'.*
- [8] *The Whigs shed some 'Dark' on the 40 Watt. Tonight, rock band The Whigs will be playing their new album 'In the dark' on Athens's most exciting stage, the 40 Watt.*
- [9] *Dead and kicking. A horror about a girl who discovers unexpected things, including an angry ghost who wants her to go away. And there are even more secrets beneath the surface.*

These variants are characterised by the *incompatibility* between the substitutes and the figurative meanings of the canonical idioms. For example, the canonical idiom *not a snowball's chance in hell* derives its figurative meaning from the opposition between 'snowball' and 'hell' (hence the meaning of 'the impossibility of something happening'). By substituting 'hell' with 'a CAT scanner' in [7], the underlying motivation for the canonical idiom is removed and replaced with an incompatible and not readily meaningful relationship between 'snowball' and 'a CAT scanner'. This compatibility is often argued in the literature to be unacceptable (see chapter 3), but context, as we see here, makes explicit the context-specific referents of 'Snowball' and 'CAT scanner' (the name of a cat that was killed in a DIY CAT-scanner), which renders the variant meaningful, purposeful and witty. In other words, despite the incompatibility and the lack of a perceivable connection between the variant and the base-form, this example showcases the role of context in facilitating context-dependent substitutions.

Perhaps the most extreme case of incompatibility amongst context-dependent substitutions can be seen in [9], which is nonsensical and paradoxical within itself on both literal and figurative level. Literally, you cannot be kicking if you are dead, so 'dead' and 'kicking' can not be used to describe the same subject. Figuratively, as the opposite of *alive and kicking*, the variant *dead and kicking* can be interpreted as 'dull/lifeless and active', which is also illogical. This meaninglessness is generally regarded as unacceptable according to the compatibility constraint (the same way that *kick the bucket repeatedly* is disqualified as an idiom variant). However, in the context of a ghost story, it becomes clear that the main characters of the story are ghosts who are taking revenge. Hence, they can actually be described as both literally dead, but still figuratively kicking (active) because they still come back and haunt people instead of resting in peace.

Appropriate context, as we have seen, play the decisive role in turning the 'unusual' and 'unacceptable' substitutions (according to standard rules of compatibility) into creative, valid variants by making explicit the intended purposes of the substitutions – that is, alluding to a specific aspect of the specific context in each case.

6.3.1.2. Ill-formed context-dependent substitutions

Despite claims that idiom variants have to be grammatically correct, it is found that context-dependent substitutions can be ill-formed on the literal level, resulting in literally ill-formed variants such as:

- [10] *Sight for snore eyes: Vision fixed in bed. Short-sighted people can now get 20/20 vision with a revolutionary contact lens which corrects their eyesight while they sleep.*
- [11] *Indiana Pols forced to eat humble pi. Egged on by an amateur mathematician, the Indiana General Assembly almost passes a bill adopting 3.2 as the exact value of pi (or π). Only the intervention of a Purdue University mathematician who happens to be visiting the legislature prevents the bill from becoming law, saving the most acute political embarrassment.*
- [12] *Take with a grain of Sultanov - The 'expert' on the Jordan bombs, Sultanov, believes King Abdullah ('who, by the way is half-English') may be behind the bombing 'to strengthen his personal authority.'*
- [13] *West Hollywood to roll out the 'dead carpet' for 500000 revelers at the 22nd annual Halloween Costume Carnival.*

In these examples, the substitutions are ungrammatical via the use of semantically non-compatible lexical items as pre-modifier (*snore eyes*, *dead carpet*), object (*eat pi*), or postmodifier of a NP (*a grain of Sultanov*). These are all ill-formed modifications according to standard rules of English about the semantic agreement between constituents of a phrase or clause. In other words, these are ill-formed by the violation of truth conditions in semantics. However, with the help of contexts, it becomes clear that the substitute *snore* in [10] refers to the night-time during which we snore and during which the vision-correct lenses work. No figurative meaning is attributed to 'snore', nor is it to 'pi', 'Sultanov' or 'dead' in [11] – [13]. In *eat humble pi*, the topic of the mathematical value 'pi' is blended into the idiomatic meaning of the idiom (to reluctantly apologise) to create the variant with the intended meaning of 'Indiana politicians are forced to admit having made mistakes and apologise about the pi bill incident.' Similarly, 'Sultanov' in [12] refers to a specific person whose words we should take with a grain of salt, and 'dead' in [13] to the Halloween characters. These substitutions, as a result, act as deictic expressions to specific aspects of the contexts. These examples serve to prove that with the right context, the grammatical rules can be waived in the case of context-dependent substitutions.

6.3.1.3. Context-dependent substitutions in unmotivated/unanalysable idioms

It is widely believed that unmotivated and unanalysable (i.e. fully opaque) idioms cannot accommodate substitutions because, arguably, if people cannot make sense of the idioms, they are unable to manipulate them. For example, 'red' in *red herring* cannot be substituted because it is semantically empty (Hudson, 1998) (see chapter 3). From a cognitive

perspective, it is argued that opaque idioms cannot undergo substitutions, because their meanings can only be activated if all the constituents are present (Langlotz, 2006). In this thesis, it is acknowledged that on a large scale people tend to use transparent idioms in creative forms more often than they do opaque ones (as revealed in quantitative analyses in chapter 5), but this does not mean that unmotivated, unanalysable idioms cannot be manipulated via substitutions. Examples in the data include:

- [14] *Mine's a pint: the car in front is a green herring. This is what bothers me about hybrid cars – they are helping to propagate the totally misguided notion that burning petrol and diesel is a bad thing, that the emissions muck up the climate, drown polar bears and give people diseases etc etc etc.*
- [15] *Spick and spam - there's no excuse for dirt in your runtime*
- [16] *The long and the short of Steve Poizner. When Whitmaniacs tell you that Poizner is 'underwhelming' or 'unimpressive', you have to wonder if what they really mean is: he's short.*
- [17] *A fine cauldron of fish.*

All these idioms are unmotivated and analysable, but variants are still possible via the substitution of one of their constituents to achieve special communicative purposes. With 'green' being conventionally associated with environmentalism, in the context of environmental issues, the meaning of 'green' and 'red herring' can be co-activated and fused to yield a creative phrase 'green herring' in [14] to refer to any false claim to protect the environment while concealing the true exploitation of nature.

It is also interesting to see the fully opaque idiom *spick and span* in the clever wordplay in [15] against the claim that unanalysable, unmotivated idioms cannot be creatively manipulated. 'Spam' in this case is used to introduce the topic of the utterance (junk emails). By fusing the meanings of *spick and span* (brand new and clean) and 'spam', the variant *spick and spam* is a clever pun to refer to the software that cleans up spam in mailboxes.

The substitution of 'kettle' with 'cauldron' in [17] might appear to be a simple unmotivated substitution via near synonyms, except that 'cauldron' is actually an intentional choice of word that serves special stylistic effects. *A fine cauldron of fish* is the title of a story of gods and other fictional creatures on the Isle of Man. The use of 'cauldron' in this case refers to a real cauldron in the context of the story (a cauldron in the middle of a sea goddess' cave, which has to be filled with a mortal's blood). Besides the exploitation of the ambiguity of the phrase 'kettle of fish' in the context of underwater gods and monsters (which are, essentially, fish), the use of 'cauldron' is intentional and enhances the cohesion between the title and the content of the story.

It is shown via these examples that idioms do not have to be motivated and analysable for substitution. The deciding factor here is context, which can sanction variants that might appear nonsensical on the surface due to the lack of motivation and analysability. In all these cases, contexts are indispensable to the identification of the correct referents of the substitutions, hence their unmissable role in the successful interpretations of the variants.

6.3.1.4. *Recognisability constraint on context-dependent substitutions*

Despite the incompatibility in many cases on both literal and figurative level, the variants as we have analysed can still be recognised as deriving from certain canonical idioms instead of being completely new or random strings of words. Recognition in each case is achieved via one or more of the followings:

- Salient lexical items that are still present (*spick, herring*)
- Combination of lexical items that are still present (*not a snowball's chance in..., barking up the wrong..., eat humble..., sight for ... eyes*)
- Salient constructions (*alive and V-ing, kiss and V-ending in the sound /el/*).
- Phonologically related items (*sight for snore eyes, spick and spam, eat humble pi*)
- Semantically related items (*a fine cauldron of fish, a fine kettle of piranhas*)

In the case of a salient constituent or a salient structure, the presence of this element alone is enough to trigger the activation of the idiomatic meaning of the phrase (while the rest of the constituents can be altered). In the cases of non-salient constituents/structures, it is the co-activation of two or more of the remaining constituents that is responsible for the recognition of the variants. If these two conditions cannot be satisfied, the recognisability of variants can still be guaranteed if one or more constituents are substituted with phonologically and semantically related lexical items. This is argued to be the effects of priming.

To explain the effects of priming in cognitive semantic terms, it is postulated that all concepts are linked to related concepts within the same schema or context. That is, each concept is primed with a series of other connected concepts so that when the target concept is encountered, these conceptual associates are activated as well to different degrees (i.e. *conceptual priming*), reflecting their similarity to the activated node (Eysenck and Kean, 2000; Gabrieli, 1998; Graf et al., 1984; Vaidya et al., 1995). For example, prior study of *chair* makes that word more likely to come to mind as a response to *table*. This is thanks to the mind's unique way of organize concepts into categories, by attributes, by prototypes, by schema or by general knowledge of their shared features. In the case of idioms, the original constituents can be seen as primes, in relation to which targets (i.e. new substitute words) can be activated. For example, when the variant *dead and kicking* is heard, the canonical idiom *alive and kicking* is also activated thanks to the presence of the constituent 'kicking' and the conceptual relatedness between 'alive' and 'dead'. The probability of *alive and kicking*

being activated as a response to *dead and kicking* is a lot higher than that of, say, *running and kicking the ball in the field* thanks to the semantic relationship between the prime 'alive' and the target 'dead', not to mention the facilitation of the context.

Similarly, various studies on auditory word recognition suggest that the processing of a target word is influenced by the prior presentation of a phonologically related prime, i.e. the phenomenon of *phonological priming* (c.f. Dumay et al., 2001; Monsell and Hirsh, 1998; Norris et al., 2002; Radeau et al., 1995, 1998; Slowiaczek et al., 2000). For example, when the target word 'snore' is encountered in the context of 'a sight for ... eyes', the prime word 'sore' comes to mind as a response to 'snore' on the basis of phonological similarity, as well as the co-activation of other constituents (sight, for, eyes). Phonological priming dictates that the probability of 'sore' being activated as a response to 'snore' in this case is much higher than that of, for example, *a sight for the tourists' eyes*. Such processes of priming, apparently, are dependent on the speaker's familiarity with the original idiom.

Recognisability, evidently, is crucial in substitutions, for the lack thereof will result in indeterminate expressions such as these:

- [18] *It's good to know I'll never have to pull another set of balls like that again.*
- [19] *But you you know there's Jeffrey Tambor - really working hard under the water.*
- [20] *That's why the public turned on Elliott Spitzer and Jim McGreevey - and why they would have turned on Bill Clinton if I hadn't worked so diligently to put his adventures in the public eye where they belong. And I tried to do the same for David Paterson, but he just wouldn't work that blue.*

It is difficult to determine from which canonical idioms are variants [18] – [20] derived, or even whether they are idiomatic variations at all. The figurative meanings of these expressions cannot be retrieved with more certainty than pure speculation. This is because the literal and (possible) figurative meanings of each expression are at a 'horse race' but there are no salient concepts or structures to activate the figurative readings. The results are ambiguous utterances. One of the observations of the study is, therefore, that in the extreme cases of substitutions where the conceptual cores are disturbed to the extent that the idioms cannot be recognised, if the contexts are not strong enough to compensate for this disturbance, the variants will not make sense to the reader/listener.

In conclusion, the recognisability of context-dependent substitution variants can be seen as the effects of the conceptual cores being evoked via one or more of the following cognitive processes: (1) activation of salient concepts or structures, (2) co-activation of less salient concepts/structures, and (3) phonological and conceptual priming, all of which are placed in context.

6.3.1.5. *Functionality constraint on context-dependent substitutions*

Whether an adaptation of a particular idiom fulfils a specific communicative purpose in a given communicative situation is perhaps the most important governing factor in idiomatic creativity in general and in context-dependent substitutions in particular. Grammatical and semantic constraints, as we have seen in variants [7] – [13], can, with the appropriate contexts, be bypassed because the variants can be recognized as conveying intentional messages. Unusual variants derived from unmotivated, unanalysable idioms as in [14] – [17] are also accepted thanks to their clear communicative purposes in contexts. In other words, despite the opacity of the idioms and the incompatibility, ungrammaticality of the substitutions in many cases, as long as they can be seen as serving particular functions in the context of discourse, unusual variants are tolerated. The reverse is also true. Without a clear purpose, variants will be treated as uncreative or unintentional usage of idioms, such as:

- [21] *Take me, for example, I'm a minimalist. This suggests, if you return into my home, you might say one thing like "Hmmm...Returning out of the woodwork. Funny issue is, I still remember the day one where one of my colleagues, Pierrette, said this very issue to me.*
- [22] *Did Niccolo Paganini soled his soul to the devil? It's obvious that few people who read this article might think that at the great composer and Italian violinist as a person who gained his talent as a result of a pact to the devil. Moreover, even from his birth moment, in a poor salesman and no luck family, his mother had a premonition dream, in which, her future son will be world's greatest violinist.*

Since contexts could not help justify the choices of 'return' and 'sole' in the two expressions, the substitutions are considered meaningless in contexts. Considering the spelling and grammatical errors in [22], it is confirmed that the adaptation is an error due to a poor command of English rather than a creative use of idiom.

This observation about contexts and accepting unusual variants raises the issue of the listener/reader's tolerance towards idiomatic creativity (Carter, 2004) in which the listener/reader plays a vital role in the success of a creative language use. It really depends on whether they are willing to see certain creative usage as creative, or just plain bad English, since some uses of language might be more controversial than others. Consider the following contrived examples:

- [23] A: *what are Rafael and Ali doing out there, shooting the breeze?*
B: *shooting the storm more like, you know with the fight over job re-assignment and all*
A: *oh yeah I forgot*
- [24] *In a fine 'jackpot' of fish (and chips) – the casino is struggling.*

In [23], the replacement of the noun *breeze* by *storm* suggests a heated and possibly loud conversation as opposed to the calm and peaceful nature of *breeze* as in 'light and aimless

conversation, just chit-chat'. Although it is semantically empty, the constituent 'breeze' can still be manipulated to code a different meaning. In [25], the play with the binominal *fish and chips* and the polysemic word *chips* ('potato chips' and 'casino chips') create a pun that fits the context of casinos, and showing a high degree of creativity, recognizable and fits the context of casino perfectly. This interpretation, however, depends on the tolerance on the part of the listener/reader and their willingness to see it as a creative manipulation of the original idiom. Nevertheless, with the availability of context, they become much more relevant, intentional and meaningful than otherwise. This is in full accordance with Glucksberg's (1993) proposal that any variation of an idiom can be deemed acceptable if a reasonable communicative intention/purpose can be inferred in a specific context.

The functions of context-dependent substitutions can be observed in the data to include (1) creating cohesion with co-texts, (2) alluding to specific aspects of contexts, and (3) introducing topics, which are context-specific. These purposes are not acquired via perceivable links between the substitutes and the original constituents (because there is none), but achieved via direct references to the contexts.

In short, context-dependent substitutions refer to contextual elements and, in turn, depend for their interpretations on contexts. The variants can be incompatible or ill-formed, or they can be derived from unmotivated, unanalysable idioms, as long as they are recognizable in the specific contexts and express clear messages, they are considered intentional and valid.

Summary

Being one of the most productive idiomatic variational strategies in the data, substitution features a wide range of variants, the most notable of which are substitution by phonologically related words, substitution by semantically related words, and substitution by context-specific words. Substitutions which are context-dependent, as we have analysed in this section, are those in which the substitutes refer to specific aspects of context and depend on context for their successful interpretations. Analyses of individual cases yield an important insight into the interplay between facilitating and constraining factors in context-dependent substitutions. That is, context emerges as the overarching factor in (1) facilitating the recognisability and functionality of unusual variants, and (2) lifting the constraints of grammaticality, compatibility and opacity on idiom variants.

In fact, majority of the variants in this group of context-dependent substitutions are ill-formed on the literal meaning level and/or semantically incompatible on the figurative meaning level. Nevertheless, in the right contexts, the variants can still be understood and accepted.

Context-dependent substitutions are accounted for in both unmotivated and unanalysable idioms. Arguably, this is due to the fact that the purposes of context-dependent substitutions

are to directly refer to aspects of the contexts instead of manipulating the internal structure or semantic configuration of the canonical idioms. Thanks to this deictic function, a context-dependent substitution only depends on the availability of context for its functionality to be communicated, not the motivation or analysability of the idioms.

Since the possibilities with substitution are virtually endless, the most important question regarding context-dependent substitution is ‘where is the limit?’ Recognisability constraint is, therefore, crucial, requiring that the canonical idioms be recognized so that their idiomatic meanings can be activated for further interpretations. Recognisability, as such, is an organic component of functionality, both of which ensure that the idiom adaptations are intentional and meaningful in context. The recorded functions of context-dependent substitution include (1) allusion to context-specific entities, (2) context-specific topic-indicator, and (3) cohesive devices. The discoursal effects of such functions are found to be striking and highly creative.

All in all, the claim in this section is that context has the power to overrule the constraints of grammaticality and compatibility as well as to relieve effects of opacity on context-dependent substitutions. However, the variants are required to satisfy the recognisability and functionality constraint to qualify as intentional idiomatic creativity. This recognisability and functionality is again argued to be dependent on the availability of appropriate context.

6.3.2. Context-dependent premodifications

In theory, the complexity of a noun phrase can be extended exponentially via two processes, premodification and postmodification. Premodifying items can include lexical and grammatical items of a wide range and indefinite complexity and interrelationship. In our data, adjectival phrases and nominal phrases are the major types of premodifying items in context-dependent variants. Postmodification, on the other hand, is realized by either a finite or a non-finite clause, or a prepositional phrase. It is observed in the data that variants created via postmodification are rarely context-dependent. Discussions in this section, as a result, focus on premodification instead, leaving postmodification until the latter half of the chapter, amongst context-independent variants. Examples of context-dependent premodification are:

Adjective phrases as premodifiers:

- [1] *Does it ring a sobering bell for you? Bing drinking is never a good idea.*
- [2] *A babe in the genetic woods – who returns to write a book in which strange genes appear in improbable places.*

Noun phrases as premodifiers:

- [3] *Galaxy tell Real Beckham's ship has sailed. Real Madrid have been rebuffed by David Beckham and LA Galaxy after making an approach to buy the midfielder out of his*

contract with the US club. Real were told yesterday there was no clause in Beckham's deal that would allow him to stay on in Madrid.

- [4] *Joseph Dalli, a Republican running the Assembly 93rd district against Mike Spano, was assaulted last night by Spano's friend and campaign worker, Julio Sophia. Of course, once the feds open one small can of Spano worms, the big can will be all too irresistible.*

It is generally claimed that only motivated and analysable idioms can have systematic modifications on their constituents (Langlotz, 2006; also see chapter 3). This is because modification is hypothesized to be only possible if the speaker is able to establish a systematic, directly meaningful and functional relationship between the modifier and the modified head noun. In opaque idioms, since the constituents do not have individual figurative meanings, their modifications will not make sense. While this is true with systematic (i.e. context-independent) modifications, it will be shown in this section that this hypothesis does not hold with context-dependent modifications. The element of being context-dependent indeed allows this type of variant to occur with unmotivated and unanalysable idioms as well as to elude the constraints of compatibility and grammaticality, in the same manner that context-dependent substitutions do.

6.3.2.1. *Incompatible context-dependent premodifications*

Although a large number of premodifications are literally well-formed, they are not compatible with the figurative meanings of the canonical idioms, for example:

- [5] *Stock market analysis at DeepMarket - Think outside the black box. Most people, including experts and even sometimes this blogger, use what I call the Black Box Theory of Stock Market Returns*
- [6] *Roll out the (code) red carpet. NRCC dep. exec. dir. Johnny DeStefano penned a memo 3/8 a.m. to GOP candidates under the header, "Code Red: How Republicans Can Upend the Obama-Pelosi Health Care Agenda."*
- [7] *Our leaders are in carbon-cloud cuckoo land. For a perfect example of what is meant by 'gesture politics' - an empty pledge given solely for effect, which the politician has no hope of honouring - one could not do better than this week's commitment by the G8 leaders to fight climate change by halving their emissions of carbon dioxide by 2050*

The common characteristic these variants share is their empty figurative meanings. There is no underlying motivating mechanism (e.g. conceptual metaphor, knowledge of the world, etc.) to help establish a meaning correspondence between *the box* (loosely assigned the figurative meaning of 'limits' or 'conventions') and *the black box* in [5]. The adjective 'black' in this case, although literally meaningful, is not compatible with the meaning of 'think creatively'. However, in the context of finance and trading, the premodifier 'black' is understood as an allusion to a stock trading model called Black Box. Similarly, although

‘code red’ in [6] and ‘carbon’ in [7] are incompatible with the figurative meaning of *roll out the red carpet* (‘welcome with respect’) and *live in cloud-cuckoo land* (be naïve and/or unrealistic), respectively, their intended meanings are made explicit in the co-texts. That is, ‘code red’ is used to introduce and relate to the Obama-Pelosi health care agenda, while ‘carbon’ refers to the issue of carbon dioxide emissions.

In other words, many of the adjectival premodifications in the data are found to be incompatible with the figurative meanings of the canonical idioms, so that no meaning connection can be made between the variants and the canonical idioms. This is because the premodifications are not used for bona fide modifying purposes. Instead, they work as indicators of the topics of discussion, which are made explicit in contexts.

6.3.2.2. *Ill-formed context-dependent premodifications*

Premodification by adjectives is rarely subject to constraints (Quirk et al., 1985: 1323), except the minimal restriction on permanence/temporariness. That is, according to the standard rules of English, the function of adjectival premodifiers in NPs is to describe permanent characteristics of the head nouns (ibid), which means that adjectives which are strongly identified with temporary status cannot appear in premodifying position (e.g. **an afraid man*, **the ready man*). The possibility of modification by an adjective, therefore, depends on the potentiality of the adjective to indicate a permanent characteristic feature in the context.

The same restriction is found with nominal premodifiers, such that nouns describing temporary characteristics of the head nouns cannot appear in premodifying position (e.g. **the corner man*), in which case a prepositional postmodification should be used instead (*the man in the corner*). Additionally, premodifying NPs usually have reduced explicitness in comparison to other types of modification (e.g. *the church house* is less explicit than *the house beyond the church*) to the extent that they can become unclear (*?a stream tree*), in which case the expression is unacceptable. Explicitness and permanence, as a result, are the two constraints on the semantic agreement between the premodifying items and their head nouns.

In the case of context-dependent premodification in idioms, these rules are largely non-applicable. Many variants in the data do not observe these constraints of explicitness and permanence:

- [8] *Russia throwing its gassy weight around, again. Is anyone really surprised that Russia is stopping natural gas shipments to Ukraine and the EU? They did the same thing three years ago, in a cold spell*
- [9] *A babe in the genetic woods – who returns to write a book in which strange genes appear in improbable places*
- [10] *Eating accessibility humble pie: We all make mistakes. Right? Particularly when it comes to accessibility. Often in the rush to ready a site for launch, we forget to check the*

details that can make a world of difference. That's what I did when I launched the latest For A Beautiful Web.

- [11] A: *No USB port is a deal killer for me. I understand Apple's need to control the experience and USB being an open standard opens a big can of compatibility worms*

To introduce the topics of gas supply, genetics, internet accessibility for websites, and USB port compatibility in computers, the speaker/writer uses the corresponding adjectives 'gassy' and 'genetic' as premodifiers to the head nouns in [8] and [9] and nominal premodifiers 'accessibility' and 'compatibility' in [10] and [11]. However, it can be observed that all four cases of premodifications are not literally well-formed due to the non-congruence between the adjectival/nominal premodifiers and their respective head nouns, resulting in a lack of explicitness in the expressions ('gassy weight', 'genetic woods', 'accessibility humble pie,' and 'compatibility worms'). However, this ambiguity on the literal meaning level is resolved when contexts are made available, so that these modifications are understood as modifications on the figurative meaning level instead (exert your power in gas supply, an inexperienced person in the field of genetics, a mistake concerning accessibility, and a problem with compatibility, respectively).

In short, the possibility of modifying the head nouns on the figurative meaning level overrules the principles of explicitness and permanence as explained earlier. Where an adjectival/nominal premodification is seen as incompatible/implicit, it can still be interpreted as a reference to a context-specific topic or information. Context, as such, again proves to be the facilitating force behind the endorsement of ill-formed (and incompatible) modifications that do not make sense via traditional interpretation strategies.

6.3.2.3. *Context-dependent premodifications in unmotivated/unanalysable idioms*

It is acknowledged in this thesis that unmotivated, unanalysable (i.e. fully opaque) idioms do not have as many context-dependent premodifications as transparent idioms do. However, the number of variants derived from opaque idioms via context-dependent premodifications is by no means scarce nor trivial. In fact, due to their opacity and relatively high degree of fixedness, opaque idiom variants are more striking and attention-grabbing:

- [12] *It's another African team who bite the African dust*
- [13] *A fine kettle of Pfizer fish. Did Pfizer, the world's largest pharmaceutical company, cross the line on the eve of last month's crucial test trial involving its controversial anti-seizure medication Neurontin?*
- [14] *Home Team Fitness is the husband and wife team of Chad and Cat Smith that has really turned Hagerstown, Maryland on its fitness ear*
- [15] *Take it with a grain of iodised salt. Nutrition programme manager David Roberts says iodine is an essential nutrient for growth and development, however iodine deficiency is re-emerging as a population-wide health problem*

Since the modified nouns do not have individual figurative meanings, it is not possible to interpret these variants using general interpretative strategies. The premodifying adjectives and nouns in these cases are not used to describe the head nouns but to refer to context-specific topics, places, and organizations. In [12], for example, *bite the dust* (to die) is found to have an incompatible premodification on the figurative meaning level ('African' is not semantically compatible with 'to die'). However, in the context of a football tournament that was held in Africa, in which one African team had just lost their last game and was therefore out of the tournament, the variant can be understood as 'another African team had just ended their journey in the tournament in Africa.' The phrase 'African dust' is used to refer to Africa via the ENTITY FOR LOCATION-metonymy. The premodifier, as a result, provides an allusion to a specific location.

In the same manner, the premodifier 'Pfizer' in [13] refers to a specific organization – a pharmaceutical company. Even though 'fish' in *a fine kettle of fish* does not have an individual figurative meaning, 'Pfizer' can still be used in the premodifying role because it is not meant to modify the noun but refers to 'the Pfizer scandal'. It is noteworthy that when the referents are widely known (i.e. part of general knowledge, as 'African' in [12]), extra context might not be needed, but the variant is still treated as context-dependent because it requires the language user to know the concept 'Africa'. In the case of lesser-known referents (such as 'Pfizer' in [13]) co-texts are clearly helpful in introducing the background knowledge on which the variant is based.

Variants [14] – [15] are slightly different in that their premodifiers are not proper nouns of specific locations or organizations, but they function in very similar ways, i.e. introducing the context-specific topics of the texts instead of modifying the head nouns in the standard sense. 'Fitness' in [14] refers to the fitness programmes that are aimed at putting people in shape, and 'iodised' in [15] to the issue of iodine deficiency amongst the population.

In other words, that fact that these idioms are unanalysable does not hinder their capability of allowing context-dependent premodifications of their constituents, as long as sufficient context is provided to disambiguate the intended meanings of each variant in each communicative situation.

It is acknowledged that not all of the opaque idioms are found to have context-dependent premodification variants in the data, for instance *in the offing*, *off to the races*, *not five a shit/hoot*, or *take a piss*. However, this is argued to be the limitation in terms of the breadth of data, because the principle as we have examined can be applied to any opaque idiom to create context-dependent premodification variants, provided that enough context is provided to sanction their usage.

6.3.2.4. *Recognisability constraint on context-dependent premodifications*

Recognisability is secured across the whole range of adjectival and nominal premodification thanks to the fact that no component of the idiom's message is lost or replaced; nor are they rearranged in any way that might hinder recognition. In all cases, the insertions of premodifying items do not destroy the integrity of the conceptual core of each idiom; they only add extra information to enrich this core. As a result, the constraint of recognisability is always satisfied.

6.3.2.5. *Functionality constraint on context-dependent premodifications*

The functions of context-dependent premodifications can be observed to roughly coincide with those of substitutions, i.e. (1) alluding to specific aspects of contexts, and (2) introducing topics, which are context-specific. As we have discussed with substitutions, functionality is key in facilitating incompatible, ill-formed premodifications, especially when they occur in unmotivated, unanalysable idioms. As long as a meaningful purpose can be inferred in context, the idiomatic expression is considered a valid variant instead of being erroneous. In the absence of clear intentions behind the modifying items, idiomatic expressions such as [16] – [17] are discarded:

[16] *A little lemony garden bird told me a secret*

[17] *I'll read this about 50 more times then I'll cry my Munchin eyes out, ok?*

Unlike variants such as *a sexy little gay bird told me* or *a little Las Vegas bird told me*, which are compatible modifications, the modification in [16] is incompatible with the figurative meaning of the idiom (? 'a little lemony garden person who reveals a secret'). 'Munchin' as the premodifier for 'eyes' in [17] is incompatible as well. As we have learned from previous analyses, usually when an 'unusual' expression is encountered, background contextual information will play the part of the disambiguating factor that delineates the intended meaning of the expression, hence justifying its 'unusualness' and sanctioning their usage. However, in these two cases, there are no contexts to offer any explanation about the choices of premodifiers, resulting in a failure to grasp the intended messages in these expressions. Without a communicative purpose, these expressions, although easily recognized as variations of *a little bird told me* and *cry one's eyes out* on the formal level, cannot be treated as valid idiomatic variants.

Summary

Adjectival and nominal premodification are the two most popular types of context-dependent modification in the data with the most number of variants. Like transparent idioms, unmotivated and unanalysable idioms are also found to have context-dependent premodifications. This fact undermines the claims in the literature about the lack of creativity among opaque idioms. Opaque idioms are found to be capable of being creatively

manipulated via context-dependent premodification, provided that appropriate contexts are available to guarantee recognisability and functionality.

Context-dependent premodifications can be both incompatible on figurative meaning level and ill-formed on literal meaning level. Although ambiguous, and ill-formed, these variants are not treated as 'wrong' but tolerated if they fit the contexts and serve special communicative purposes. This tolerance of incongruent, implicit adjectival/nominal premodifiers in idioms is thanks to language users' willingness to see past grammatical and semantic idiosyncrasies for the sake of idiomatic readings. The interpretations of such premodifications are taken beyond their literalness to derive relevant context-specific meanings of the variants.

In premodification, thanks to the fact that no components of an idiom's message is lost (as in the case of allusions or truncations), replaced (as in substitutions) or rearranged (as in structural manipulations) in any way that destroys the integrity of the conceptual core of the idiom, the recognisability constraint is always met. Meaningfulness/functionality, on the other hand, might be hard to detect if the context is missing, result in a failure to see an idiomatic expression as a valid variant.

The claim regarding the interplay between facilitating and constraining factors in context-dependent premodification is similar to that made about context-dependent substitution. That is, context can overrule the constraints of compatibility and grammaticality as well as relieve the negative effects of opacity upon context-dependent premodification. The constraints of recognisability and functionality, however, must be observed for the variants to be valid. These two aspects, particularly functionality, in turn, depend on the availability of context to be successfully achieved.

6.3.3. *Minor type of context-dependent variants: Coordination*

Besides the two major types of substitution and modification, a minor type of context-dependent variants is found in the data: coordination. By 'minor' I do not mean that this type is less creative or less striking in terms of discoursal effects, but only that the frequency of occurrence of this type of variant is low in the present data. However, it is their humorous effects and their occurrences within opaque idioms that earn them a quick mention here, so that future research might follow to uncover their true potential.

In terms of form, coordination involves the linking of units of the same rank at the same level of constituent structure, marked by overt signals of coordination (*and, or, but*) or by implicit tone units boundary (in speech) or by punctuation (in writing). Coordination can link clauses or clause constituents (subject, predicate, complement, object, modifier, etc.) The conjoins in each coordinated construction are required to be parallel to one another in meaning, function and also generally in form (Quirk et al., 1985). However, this

requirement is observed with varying degrees of strictness, particularly when an expression is used for humorous or rhetorical effects (e.g. *she made up her mind and then her face*, which is based on an incongruous association of two homonymous idioms).

Although the context-dependent variants created via coordination are creative and humorous, there are not enough examples in the data to draw any sensible conclusion about their behaviour in discourse in general. The variants that do occur pertain to a few idioms only (*rain cats and dogs*, *think outside the box*, *give a hoot*), for example:

- [1] *Think and play outside the box!* (context: the launch of a new computer game)
- [2] *Give a hoot (or a roar or a warble)* (context: a website where people can download animal calls ringtones to help raise awareness of extinct species)
- [3] *It's raining cats, dogs and bears at Christie's* (context: Christie's summer Teddy Bears sale of cuddly animals, including cats, dogs, rabbits, bears and other furry things)
- [4] *It's not raining cats and dogs, it's raining birds - but why?* (context: reports of 75 starlings falling out of the sky within a week)
- [5] *Raining cats, dogs and Memorex Digital Audio Players* (context: Memorex announced new players in the burgeoning portable music player market)

The parallelism between conjoins is maintained in these variants on the literal meaning level. In terms of figurative meaning, on the other hand, the coordinated NPs in [2] – [5] are incompatible with the respective canonical idiom ('care or roar or warble', 'rain heavily and bears,' etc.). This incompatibility is probably due to the fact that 'a hoot' and 'cats and dogs' in these idioms do not have phrase-induced meanings, so they cannot be extended via coordination.

With regards to interpretation, the coordinated phrases/clauses are not attributed figurative meanings but relate to specific referents in the specific contexts of discourse. For example, [1] can be understood literally as 'think and play creatively'; however, the intended meaning of the variant is to allude to the fun and creative features of a new computer game, hence 'play' in the variant is specified as 'game playing' instead of having the generic meaning of 'activity'. Similarly, in [2], the figurative meaning of *give a hoot* (care) is maintained while the literal meaning of 'hoot' (the cry of an owl) is exploited in the coordination with 'roar' and 'warble' to introduce the topic of animal call ringtones. As such, the variant creates a dual effect to send the message of animal rescue and to introduce the website's features.

Due to the intentional referential purpose, it is not possible to understand the true intentions behind the coordination in each case if contextual information is not provided. It is noteworthy that context-dependent coordination occurs in both transparent idioms (*think outside the box*) and opaque idioms (*raining cats and dogs*, *give a hoot*).

In short, in the small number of variants with context-dependent coordination, the following observations can be recorded:

- (i) Some context-dependent co-ordinations are compatible and well-formed and some are incompatible and ill-formed. The constraint of compatibility and grammaticality, as a result, can be seen as waivable when sufficient context is made available to specify the communicative purposes of 'unusual' expressions, thereby sanctioning their usage.
- (ii) Both motivated idioms and unmotivated/unanalysable idioms are open to this type of context-dependent coordination without any differences in the ease with which each group can be interpreted. Due to the context-dependency of this type of variant, all the coordinated phrases are to be interpreted in context as (1) introducing a topic, or (2) referring to an aspect of context.

These observations are in line with those found with substitutions and modifications.

6.3.4. *Conclusions: Context-dependent variants*

As a whole, the findings about context-dependent variants in this chapter dispute the claims in the literature about the lexical fixedness in opaque idioms, contradict the hypotheses about the constraining factors of grammaticality and compatibility on idiom variants, but substantiate the theories of the roles of context and recognisability and functionality as the *sine qua non*s of the valid idiom variant.

The common characteristics among context-dependent variants are, first and foremost, their context dependency, and their functions as deictic expressions in discourse. Thanks to these two characteristics, idiom variants created via the strategies of context-dependent substitution/modification or coordination are not restricted to transparent idioms and are not required to observe the rules of grammaticality and/or compatibility, provided that the context is made informative enough to facilitate their recognisability and functionality.

Recognisability is the only aspect that is type-specific amongst context-dependent variants. That is, only substitution variants are potentially ambiguous due to the lack of recognisability, which means that care should be taken to ensure that this constraint is satisfied for the variants to be valid. Modification and coordination, on the other hand, are non-disruptive strategies, hence the guarantee of recognisability across all of their variants.

These findings about context-dependent variants will be compared against those about context-independent variants, which will be analysed hereafter, to highlight the similarities and differences between the two types of variants, thereby formulating a comprehensive context-based model of idiomatic creativity with regards to the interplay between context and other facilitating and restraining factors in each type.

6.4. Context-independent variants

Context independent variants of an idiom, as I use the term in this thesis, are explicit variants whose interpretations can be derived from inferential/ interpretative processes once the figurative meaning of the canonical idiom is activated so that no extra context is required, for example:

- [1] *You're striking while the iron is cold*
- [2] *Striking while the iron is lukewarm is sometimes the best strategy in philosophy.*
- [3] *The critics of Spinning were swallowing large slices of humble pie*
- [4] *Those Right-Ons who had written off the Labour Party for ever now began to chew humble pie*

To understand [1] – [2], it is also important to know the phrase-induced meaning of ‘hot’ (timely) in the idiom *strike while the iron’s hot* (‘act immediately when the situation is opportune’). Cognitive linguists might want to argue that the conceptual metaphor URGENCY/IMMEDIATENESS IS HEAT is the underlying motivational force for this idiom (e.g. *the hot freight must be delivered tomorrow*). Since ‘hot’ is mapped onto the domain of ‘promptness’ according to the metaphor, ‘lukewarm’ (less hot) is mapped onto the lesser degree of promptness (act after a certain amount of time has passed), whereas ‘cold’ (the opposite of ‘hot’) maps onto the opposite meaning of ‘promptness’ (i.e. the delay/lateness in response). Other linguists would argue that the conceptual metaphor is not necessarily activated during comprehension; the knowledge of the meaning of the idiom alone is enough to assign individual figurative meaning to the constituent ‘hot’. The sense relation between ‘hot’, ‘cold’ and ‘lukewarm’ can then be used to infer the meanings of [1] and [2]. Either way, it can be seen that the meanings of [1] and [2] can be inferred on the basis of the knowledge of the figurative meaning of the canonical idiom via analogy/conceptual metaphor without the need for extra context.

Similarly, the meanings and stylistic effects in [3] - [4] are achieved via the differences in connotation between the substitutes (‘swallow’, ‘chew’) and the original constituent ‘eat’ in *eat humble pie*. While they are near synonyms, ‘swallow’ (as in *swallow a bitter pill*) and ‘chew’ (as in *chew a tough steak*) has an autonomous connotation of involuntary or unpleasant consumption due to an aversion to the substance consumed. Variants [3] and [4], therefore, indicate a degree of reluctance in admitting your faults than that evoked by the original idiom. Possible conceptual motivations for this idiom and its variants include: DEALING WITH AN ISSUE IS CONSUMPTION metaphor, and MANNER OF DEALING WITH AN ISSUE IS MANNER OF CONSUMPTION metaphor. Again, extra contextual information is not required for the successful interpretations of these variants.

The foregoing discussion highlights the following important aspects of context-independent variants: (1) the routes by which language users make sense of an idiom variant, and (2) the

interplay between the facilitating and constraining factors in the absence of context. These aspects will be examined in this part of the thesis to highlight the similarities and differences between context-independent and context-dependent variants in terms of possibilities and constraints, so that the roles of context in idiomatic creativity in general can be fully captured.

6.4.1. Understanding the meanings of context-independent variants

Up until now, we have argued that unlike context-dependent variants, which are interpreted via contextual information, the meanings of context-independent variants are inferred via interpretative processes. However, it has not been explained what we mean by 'inferential/interpretative processes'. In the literature, there has been little agreement on how this inferential process in idiom variant comprehension looks, although suggestions have been made. One school of thought is the *conceptual metaphor theory* (Lakoff and Johnson, 1990; Lakoff, 1993), which posits that conceptual metaphors serve as the conceptual bases for on-line figurative language comprehension, including idioms and idiom variants (see full review of this theory in chapter 3). The second approach is to see the processing of idiom variants as the result of *inferential/interpretative strategies* on the basis of the comparative similarities and differences between the variants and the base-forms via *analogy* (Glucksberg, 1993; Glucksberg and McGlone, 1999; Keysar and Bly, 1999) (also see chapter 3).

In this chapter, it is argued that there are merits to each approach, but also that a middle ground is beneficial in explaining the wide range of context-independent variants that we have in the data. That is, conceptual metaphors may be retrieved in certain situations, but not all, because not all idioms are metaphorical and not all idiom variants are metaphorically created. This suggestion is, however, a tentative one, since psycholinguistic experiments are the desiderata for this kind of an insight into human linguistic processing. Corpus data that we have here can only claim to point out the trends that might guide such experiments. For example:

- [5] *Bots popping out of the woodwork*
- [6] *60s radicals suddenly tumbling out of the woodwork*
- [7] *Racism makes absolutely NO SENSE to most people, but evidently there's some still slithering out of the woodwork since President Obama was elected.*
- [8] *Every puny author who ever got his ass kicked by yours truly scuttles out of the woodwork*
- [9] *An unexpected opponent poked out of the woodwork as an option for Shane*
- [10] *Villa supporters leaping out of the woodwork*
- [11] *Of course, the usual crazy villains seep out of the woodwork in search of their target*
- [12] *These creatures have crept their way out of the woodwork to explore Christchurch Art Gallery*

The intended meanings of these variants rely on the different connotations of the near synonyms of ‘come/crawl’ in *come/crawl out of the woodwork* (‘to appear suddenly and unexpectedly, especially to do something unpleasant’). The canonical idiom can be argued to be motivated via the WOODWORK-emblem (‘place to hide unpleasant secrets’) (e.g. *a skeleton in the cupboard, come out of the closet*). Also, general knowledge tells us that most of the things that come out of the woodwork are verminous insects or other invertebrates which are considered unpleasant.

All the verb substitutes in these variants denote the act of exiting a dark, obscure hiding place into the light, hence the shared meaning of ‘emerging from a secrete place’ with the overall negative connotation attached to the subject (being unpleasant). Be that as it may, the verbs differ in literal meanings with regards to manner (speed, strength, etc.). With ‘come’ as the neutral base-form of the movement, the different aspects of these substitutes describe the different manners with which the subjects appears, adding colors and dimensions to the descriptions in the variants. Especially, the last variant, [12], does not only provide elaboration of the action but also creates cohesion within the text. The substitute ‘creep’ is congruous with the subject ‘creatures’ whose signature/characteristic manner of movement is ‘to creep’. The stylistic effect is, hence, increased cohesion within the text.

In the data, the verb ‘come’ is also substituted by verbs that have different transitivity, so that the variants can benefit from additional highlighting effects when the subjects do not voluntarily fulfil the action of ‘coming out’ but passively affected by other agents, as in [13] – [17] below:

- [13] *In June, Microsoft lit a fire under Cisco and has squeezed the less-talked-about vendors out of the woodwork and into the limelight*
- [14] *Uchitel, 34, was the first Woods woman dragged out of the woodwork last fall*
- [15] *Coaxing ghost social entrepreneurs out of the woodwork*
- [16] *Michigan Secretary of State job draws candidates out of the woodwork*
- [17] *March Madness office pools luring sociopaths out of the woodwork*

By analogy, it is possible to understand the meanings of these substitutions without the need to refer to context or conceptual motivation (conceptual metaphor, conceptual metonymy, etc). The role of analogy, however, cannot be claimed across all idiom variants. In the set of examples below, conceptual metaphors appear to the conceivable source of explanation for these four variants of *think outside the box*:

- [18] *Way to think outside but press right against the box!*
- [19] *Advertisers forced to think way outside the box*
- [20] *The designers there think so far outside the box that very little is squared*
- [21] *I don't just think outside the box, I tear it up.*

The underlying motivation of the canonical idiom *think outside the box*, from which [18] – [21] are derived, can be argued as involving the BEING CONTROLLED IS BEING INSIDE A CONTAINER metaphor (e.g. *push the envelop, feel suffocated under surveillance*). As such, the idiom can be seen as analysable with the phrase-induced meaning of ‘restrictions/limits’ attached to the constituent ‘box’ and the meaning of ‘breaking away’ attached to the constituent ‘outside’ in the context of the idiom. With this analysability, variants [18] – [20] are also motivated via the LINEAR SCALES ARE PATHS metaphor, which underlies the inserted adverbial phrase in each case. The LINEAR SCALES ARE PATHS metaphor is one of the basic metaphors, which maps the starting point of the path to the bottom of the scale and the distance travelled onto quantity (Ortony, 1993). From the metaphor, other mappings can be made to express the relationship between the path travelled and the degree on the scale, so that LONGER PATH IS HIGHER DEGREE. The expression *press right against the box* in [18] indicates a short path, while *way outside the box* in [19] and *so far outside the box* in [20] indicate longer paths. As a result, the meaning of [18] can be understood as ‘low degree of breaking away from restrictions’ (not very creative) while [19] and [20] are intended to mean ‘higher degree of breaking away from restrictions’ (i.e. very creative).

The conceptual motivation behind the interpretation of [21] is not the LINEAR SCALES ARE PATHS metaphor, but the ABSTRACT CONCEPTS ARE PHYSICAL ENTITIES metaphor, according to which ‘restrictions’/‘limits’ can be torn up. The act of *tearing up the box* is symbolic of removing the constraints/restrictions, which implies the highest degree of creativity, because the constraints do not even exist anymore.

Indeed, there is a certain degree of systematicity amongst idiom variants concerning the possibility of conceptual metaphors and analogy processing routes. The problem, however, is that controversial data is not scarce, which means that we cannot hypothesize with any confidence the dominance of one route over the other. It is also observed that the comprehension process, either via conceptual metaphors or analogy, differs from variant to variant according to the intended meanings of the variants and the variational strategies chosen to code these meanings. For example, as we have seen earlier, variants [18] – [21] are conceivably motivated by conceptual metaphors, but the specific metaphor in each case depends on the intended meanings of the variants. Variant [22] below, on the other hand, although derived from the same idiom as [18] – [21] (*think outside the box*), is not even metaphorically motivated but relies on general knowledge of the world and analogy for its interpretation. To understand variant [22], it is helpful to know that a box is essentially square and it is one of the basic shapes (hence the slang meaning ‘being square’ = being conventional, boring) (e.g. *he’s a bit square*). The hexagon is less basic, more unusual, and thus can be understood as ‘more creative, more interesting’.

- [22] *‘Think outside the box’, they say. ‘Think outside the hexagon’, I say. Dare to be different? Dare to be odd! That’s right - throw moderation out the stained-glass window because it’s the extremes that are the most memorable*

The claim made in this section, as a result, is that a combined approach is arguably beneficial in accounting for the meanings of individual context-independent variants. It seems sensible to take an open approach in which both conceptual metaphors and analogous inferences are equally possible, but subjective and case-sensitive. That is, each variant and each type of modification depends on a different set of motivation mechanisms and/or inferential processes that might even overlap. As will become clear later in this chapter, operations such as substitution and constructional adaptation are seen as more potentially reliable on analogy, while modification appears to depend on complex conceptual motivations.

In the subsequent sections, these specific types of variational strategy are scrutinised to flesh out the role of motivation in context-independent variants in relation to other facilitating and constraining factors.

6.4.2. Context-independent substitutions

The lexico-grammatical operation to create a context-independent substitution variant is via constituent substitution by synonyms, antonyms, or gradable lexical items that are semantically related to the original constituent in some ways, for example:

Substitution by synonyms and near synonyms:

- [1] *She had definitely rubbed a raw nerve*
- [2] *The UK slump has slammed the lid on margins*

Substitutions by antonyms and gradable opposite items:

- [3] *A: Did you do something behind my back?*
B: I'm not sure how to unring that bell
- [25] *You really need brass ovaries to get up that pole and dance - it's harder than it looks*

Substitutions by gradable items with comparable size/weight, shape, state:

- [4] *A: Come on, at your age, relationships aren't serious, they're just you know puppy love*
B: No no this wasn't just puppy love, this was dog love
- [5] *Skating on thinning ice*
- [6] *Well now that's a different erm barrel of worms isn't it really?*
- [7] *Thus the NEB's reassurances were taken with a strong dose of salt*

Substitution by topic-introducing items:

- [8] *Alive and shopping*
- [9] *Trading on thin ice*
- [10] *So you're ready to grasp life by the horns now*

Although one or more constituents of these idioms are replaced with new items and no extended co-texts are provided to specify the contexts, variants [1] – [10] above can still be recognised and understood. This is the fundamental distinction between context-independent and context-dependent substitution. This difference entails a host of specific

discrepancies between the two types of variants in terms of possibilities and constraints, which will be discussed subsequently.

6.4.2.1. Context-independent substitutions in transparent and opaque idioms

Substitutions are found to frequently occur with analysable components of idioms, possibly due to the readily meaningful connections made in such variants, for example:

- [11] *Skating on cracked ice*
- [12] *Skating on thinning ice*
- [13] *Skating on thicker ice*
- [14] *Skating on bumpy ice*

These variants exemplify a host of variants in which the states of being of the subjects/object NPs are adapted to create meaningful and comparable variants. The idiom *walk/skate on thin ice* is postulated to be motivated and analysable according to the following conceptual metaphors:

ACTIVITY IS BEING ON A PATH metaphor (e.g. *we are on the way to a break-through in medicine*)

SUCCESSFUL DEVELOPMENT IS PHYSICAL BALANCE/STABILITY metaphor (e.g. *his organization is going through some ups and downs, he learns how to walk a tight rope*)

BASIS OF DEVELOPMENT IS A PATH → QUALITY OF BASIS OF DEVELOPMENT IS QUALITY OF PATH metaphor (e.g. *it has been a long and bumpy journey for his company to get to where it is today*)

INSECURITY IS WATER metaphor (e.g. *they're drowning in debt, he's scared to dip his to into the water*)

General knowledge: thin ice over water is dangerous to walk on (because it might crack and break and the traveller would be drowned in the cold water underneath)

Walking/skating on thin ice, according to these metaphors, can be understood as to try to progress (by walking towards a destination), but the basis of this progress is dangerous and capable of terminating the progress (thin ice). The constituent 'thin' has an individual figurative meaning of 'dangerous/risky' in this context. When 'thin' is replaced with 'cracked', 'thinning', 'thicker' and 'bumpy' in [11] – [14], the 'quality of the path' (i.e. the quality of development) in these variants changed in comparison to that of the canonical idiom. First of all, 'cracked' and 'thinning' in [11] – [12] indicate a change in the quality of the path in terms of *breakability* in the direction of worsening ('already cracked'/'thinning' ice = more dangerous to walk on than the neutral state of 'thin' ice). The figurative meanings of the two variants are those of more dangerous situations. On the contrary, the substitute 'thicker' in [13] indicates a change in quality of the path in the direction of improving (less dangerous to walk on). Its figurative meaning is therefore an improvement in the situation (*Swedish banks: skating on thicker ice. We believe that there are clear reasons to favour the*

more positive view of Swedish banks put forward by S&P. First, the macro-economic scenario is improving.) Last but not least, the meaning of ‘bumpy ice’ in [14] is not based on the *breakability* aspect of the quality of the path but the quality of the surface of the path. General knowledge dictates that smooth-surfaced paths are easier to move on than bumpy ones, which means that ‘bumpy ice’ indicates obstacles which hinder movements along the path (HINDRANCE TO PROGRESS IS OBSTACLE ON THE PATH metaphor). Hence, the derived figurative meaning of the variant is ‘more difficult situation to move forward because of obstacles’ (not danger).

In other words, it is generally easy and straightforward to interpret the meanings of context-independent substitutions in motivated and analysable idioms, because the underlying motivations of the idioms are clear and available for inferences to be based upon. The question is, can the same be said about unmotivated, unanalysable idioms? Without underlying motivation, constituent’s meaning, or extra context, how can a context-independent substitution be explained?

In the data, context-independent substitutions are found to occur with unmotivated, unanalysable idioms as well, for example:

- [15] *A: Come on, at your age, relationships aren’t serious, they’re just you know puppy love*
B: No no this wasn’t just puppy love, this was dog love
- [16] *A: it’s snowing!*
B: yup, snow, rain, you’ll get everything today, soon you’ll get cats and dogs
A: yeah (laughs)
B: raining cats and dogs
A: it’s only raining kittens and puppies right now
- [17] *Taken with a fistful of salt*
- [18] *A: I don’t want to leave you high and dry*
B: oh I’ve never been lower or wetter
- [19] *A: It’s rather out of the blue don’t you think?*
B: This is not out of the blue. It’s smack-dab in the middle of the blue
- [20] *Spick and clean*

It seems that, despite the lack of motivation and analysability, variants of opaque idioms can still be created via certain strategies. The first strategy is the analogy in terms of size, which relies on the conceptual mapping between the domain of SIZE and the abstract domain of DEGREE OF INTENSITY. Indeed, the primary metaphor LINEAR SCALE IS SIZE includes a number of derived metaphors (IMPORTANCE IS SIZE, INTENSITY IS SIZE, etc.) which share the specific mappings of the lower end of the scale of something (e.g. importance, intensity) to the small size and the higher end of the scale to the big size. The comparison between *puppy love* and *dog love* in [15] is enabled in a way that the analogy in size between a puppy (small) and a dog (big) can be mapped onto the intensity of the love (‘mild crush in

young people' becomes 'serious grown up feelings'). The same principle can be applied to variants [16] – [17], so that the comparable size between 'cats and dogs' (big) – 'kittens and puppies' (small), and 'fistful' (big) – 'pinch/grain' (small) can be mapped onto the comparable of degree of intensity of the rain ('rain heavily' becomes 'rain lightly') and the degree of caution ('with caution' becomes 'with much more caution'), respectively.

The second strategy among opaque idiom context-independent substitution is the use of antonyms and words of opposite meanings to suggest opposition. It might seem counterintuitive on the surface that [18] is possible, since *leave someone high and dry* is unmotivated and unanalysable. However, 'high and dry' as a whole roughly translates into 'alone/ helpless', (even though there is no underlying motivation for this correspondence). As a result, variant [18] is created using antonyms 'lower and wetter' to utilise the opposition with *high and dry* as a whole, which enables the derivation of the meaning of the variant ('leave someone alone/helpless' becomes 'not alone or helpless'). The same strategy is found in [19], in which the variant *smack-dab in the middle of the blue* can be interpreted on the basis of the opposite relationship between 'in the middle of' and 'out of', so that the base-form *out of the blue* ('without warnings'/'suddenly') is changed into 'with warnings' or 'not suddenly'. With everything else being equal, *in the middle of the blue* is in its totality the opposite of *out of the blue*, hence the variant's meaning can be derived as the negation of the base-form's meaning. In other words, this type of negation is possible because the substitution of 'out of' with 'in the middle of' negates the whole clause instead of just one meaning component, in the same way that the negation of the verbs also brings about negation of the clauses (*ring/unring a bell*).

Last but not least, the third strategy is less striking in terms of meaning; that is, substitution by synonyms, as in [20]. Note that the synonymous relationship between the substitute 'clean' in this variant is not with the literal or figurative meaning of the original constituent 'span' (because *spick and span* is highly opaque without constituent meanings) but it is with the figurative meaning of the idiom as a whole ('clean, tidy'). *Spick and clean* might be just a simple variant without any meaning adaptation, but it shows another strategy of manipulating opaque idioms. Furthermore, with the general assumption that it is highly fixed, the variant still appears interesting (much more so than the original *spick and span*).

What is important in the preceding discussion is that context-independent substitutions do not have to rely on transparency and analysability, as long as there are still discernible connections between the variants and the original canonical idioms that allow comparisons can be made (in these cases, in terms of size, opposition and similarity). Due to the opacity of the idioms, however, these strategies have to induce changes in the meanings of the idioms as wholes instead of locally.

6.4.2.2. Grammaticality and compatibility constraints on context-independent substitutions

There is no incompatible, ill-formed substitution among context-independent variants, arguably because these variants can only be context-independent if they are compatible and grammatical. In other words, among others, the constraints of compatibility and grammaticality are the prerequisites for context-independent substitution. Indeed, an incompatible and ill-formed variant is usually only meaningful under the facilitating power of context for the purpose of referring to a contextual aspect of utterance. In the case of a context-independent variant, on the other hand, there is no extra context to sanction unusual usage, and the purpose is to induce meaning adaptation via lexical and grammatical devices. As a result, they are required to be compatible and well-formed for their meanings to be inferred on the basis of linguistic and cognitive processes.

For example, in the variant *she rubbed a raw nerve*, both the substitute ‘rub’ and the original verb ‘touch’ denoting the act of ‘making physical contact’. Hence, the figurative meaning of the variant can be estimated to be similar to that of the base-form idiom *touch a raw nerve* (upset someone/cause emotional reaction). However, while ‘touch’ denotes ‘a gentle, slight tap using hand or finger’, ‘rub’ expresses the kind of touch that applies pressure and force, hence a difference in terms of vigour/intensity between the two. Conceptual metaphors are envisaged to include EMOTIONAL EFFECT IS PHYSICAL CONTACT metaphor (e.g. *the movie touched him deeply*) and QUALITY OF EMOTION EFFECT IS QUALITY OF CONTACT metaphor (e.g. *his mother’s death hit him hard*)

The difference between two verbs in terms of vigour can be mapped onto the difference in terms of *emotional effects*. The derived meaning of the variant is, therefore, to ‘cause stronger emotional reaction/really upset someone’. Such an inferential process is only possible because the substitute ‘rub’ is compatible and well-formed in the idiom, which allows perceivable links to be formed between the variant and its base-form. On the basis of these links, the intended meaning of the variant can be interpreted. Other verb substitutes in the data include ‘*tweak*’/‘*strike*’/‘*hit*’/‘*pluck*’ a raw nerve, which uphold the same principle.

Consider the expression *dance a raw nerve*, which is incompatible and destroys the underlying motivation of the original idiom. Although it may be recognized as coming from *touch a raw nerve* due to the salience and co-activation of ‘raw’ and ‘nerve’, its meaning is not inferable because of the incompatibility of the verb and the figurative meaning of the idiom. The only way to understand this expression is if the context is revealed, i.e. a dance competition in which a couple was accused of copying the steps from another couple, causing frustration among fellow competitors and the audience. However, by referring to contextual information as such, the variant enters the realm of context-dependent substitution as we have discussed in the first half of the chapter. Context-independent substitution, unfortunately, cannot accommodate incompatible, ill-formed, context-specific variants.

In other words, context-independent substitutions are both compatible and well-formed to allow for interpretations via either conceptual metaphors or general interpretations via analogy. This characteristic sets context-independent substitutions apart from their context-dependent counterpart and could be argued to be an inherent feature of context-independent variants in general (and not just substitutions) as we shall see later.

6.4.2.3. *Recognisability and functionality constraints on context-independent substitutions*

The recognisability of context-independent substitutions is argued to be similar to that of context-dependent substitutions, since they both involve replacing meaning components with new information. The facilitating factors of recognisability in these cases of substitutions, in other words, include (1) salience of constituents/structures, (2) co-activation of non-salient elements, and (3) conceptual and phonological priming, all of which to be placed in context.

In terms of functions, the comparativeness in terms of meanings between the substitutes and original items are most frequently utilized for *intensification* or *elaboration* purposes. A small proportion of the data include *topic-introducing* substitutions as well. This is another fundamental difference between context-independent substitutions (which induce meaning adaptations for intensification/elaboration purposes) and context-dependent substitutions, whose function is primarily referential.

As with other types of variants, context-independent substitutions are required to have clear functionality in order for the variants to be accepted. Without supporting context, such meaningless (hence not having clear communicative purposes) substitutions as **sleep a raw nerve*, **leave someone high and white* will not be tolerated by language users.

The facilitating factor of functionality is not context (as in the case of context-dependent substitutions) but the interpretability of the variants according to general interpretative strategies (most notably analogy) and/or conceptual motivations of the idioms.

Summary

As suggested by the name of this type of variant, context does not have a role in context-independent substitutions. As a result, none of the constraints can be lifted (as in the case of context-dependent variants), which means that context-independent substitutions are liable to the constraints of compatibility and grammaticality as well as recognisability and functionality.

Interestingly, even without context, opaque idioms are still found to have variants via context-independent substitutions, which is a powerful proof of the flexibility of opaque idioms, challenging the popular belief that they are fixed and non-creative. The key here, it is believed, is to recognize the strategies that can be applied to them. Admittedly, the range

of strategies and operations that opaque idioms can undergo is more restricted than that of transparent idioms due to the lack of constituent meanings. Nevertheless, they are by no means fixed; and when they do appear in creative forms, the effects are striking.

The role of conceptual metaphors to make sense of context-independent substitutions is observed to be more pronounced in comparable substitutions than in the cases of synonymous/antonymous substitutions (which can be directly inferred via analogy thanks to the similarity or opposition between the items). With comparable items, however, because there is a *degree* of differences, it is important to understand the underlying conceptual motivations of the original idiom as well as those of the variational strategies in order for the variants to be interpreted accordingly. Apparently, psycholinguistic experiments are needed to probe into the precise processing routes by which language users arrive at their own interpretations of the variants. Corpus data as we have here can only provide tentative suggestions that serve as premises for such experiments in the future.

6.4.3. Context-independent modifications

Modifications to noun phrases within idioms that are context-independent include both premodification and postmodification of various types, for example:

Adjectival premodification:

- [1] *AT&T has just opened perhaps the most vindictive, messy can of worms it could have possibly found*
- [2] *A fetid kettle of rotting, decomposing fish, that is.*
- [3] *Sounds like Ryan really hit the controversial nerve here by actually naming a few names*

Nominal premodification:

- [4] *I eat humble ignorance pie*
- [5] *An open (data) can of worms*
- [6] *Start the conversation ball rolling*

Insertion of determinatives:

- [7] *He welcomes the opportunity to set several records straight*
- [8] *They had walked on some thin ice before, never like this*
- [9] *The hassle of having to learn all the ropes*

Postmodification:

- [10] *A fly on de_(sic) wall of China*
- [11] *Mr. Duncan sets out on a journey that registers on the odometer but fails to touch the raw nerve of the West.*

6.4.3.1. Context-independent modifications in opaque idioms

With the flexibility of opaque idioms so far discovered, it is not surprising to see in the data the number of opaque idioms undergoing context-independent modification to create highly creative variants. Even though the modified nouns do not have individual figurative

meanings, it is still possible to interpret the modifications via a combination of inferential strategies. Consider the following examples:

- [12] *He was on the brink of rallying back from obscurity to really turn Hollywood on its hypocritical ear.*
- [13] *There is one thing that you will want to keep in untutored mind: most internet users do not take pleasure in looking at videos that appear as spam.*
- [14] *The State of California murdered a four-time Nobel Peace Prize nominee in calculated, ritualized, cold blood.*
- [15] *These stories of players being offered money to lose games is a different kettle of footballing fish*

The meanings of the adjectival modifications in these cases do not apply to the modified nouns but are transferred to other nouns in the utterances. In [12] for example, although ‘hypocritical’ appears to modify the head noun ‘ear’, in terms of meaning, it actually modifies ‘Hollywood’ who supposedly has a hypocritical ear. The transfer of meaning relies on the ORGANIZATIONS ARE PEOPLE conceptual metonymy (so that Hollywood can be perceived as a person with a hypocritical ear) and the BODY PARTS FOR PERSON metonymy (which enables the meaning of ‘Hollywood is hypocritical’). Variant [13] features the same strategy, in which ‘untutored mind’ is transferred via the BODY PARTS FOR PERSON metonymy to mean the person is untutored.

Variants [14] and [15] represents a slightly different strategy in which the premodifying items ‘calculated, ritualized, cold’ and ‘footballing’ modify the meaning of the idioms as wholes, not the constituent ‘blood’ or ‘fish’ individually, since they do not have individual figurative meaning. As such, the metaphorical ‘murder’ of the four-time Nobel Peace Prize nominee is described as calculated, ritualized, and in cold blood in [14]. The adjectival premodifier in this case are parallel to the idiom *in cold blood* as a whole, in the same way that the premodifier in [15] does to introduce the topic of ‘football scandals’.

In short, while it is true that transparency does allow a wider range of variational strategies to be employed to create a wider range of variants, opaque idioms are not completely fixed. Specific to each idiom, context-independent modification variants are still possible if the right techniques are used. These are found in the current data to include (1) modifying the idiom as a whole, or (2) using ‘transferred’ modification via metonymies, both to fulfil the function of giving descriptions or introducing the topics of the utterances.

6.4.3.2. *Grammaticality and compatibility constraints on context-independent modifications*

For the same reason that context-independent substitutions are subject to grammaticality and compatibility constraints, context-independent modifications are also bound by this rule, except for some minor distinctions. That is, the premodifying items and their modified nouns might be compatible on the literal or figurative meaning level, depending on the

intended purposes of the variants (to give literal descriptions of the head nouns, to introduce the topics of the utterance, etc.) For example, the premodifying adjectival phrase ‘greasy, squirming’ in [16] is compatible with the head noun ‘worms’ on the literal meaning level, but not on the figurative meaning level (*‘greasy, squirming problems’). However, the variant is still readily meaningful because it evokes a rich figurative meaning in which the quality of the worms (‘greasy, squirming’) is mapped onto the quality of the problem (‘unpleasant’).

[16] *You can open up a can of greasy, squirming worms if you begin to study negative political advertising.*

[17] *That would open up an entire can of constitutional worms.*

On the other hand, ‘constitutional’ in [17] is incompatible with ‘worms’ on the literal meaning level, but it is on the figurative meaning level (problems with constitution). In the data, both types of compatibility are equally popular, suggesting an extension to the general constraint of compatibility that accepts both compatibility on the figurative meaning level and literal meaning level, as long as their meanings can be inferred and understood.

Grammaticality and compatibility, therefore, in the case of context-independent modifications, are flexible and can refer to the grammaticality and compatibility of either the literal or the figurative meaning of a variant depending on the purpose of the utterance. That is, if the purpose is to modify the literal meaning of the original constituent, then literal grammaticality and compatibility must be achieved. On the contrary, an intention of modifying the figurative meaning of the original constituent would necessitate figurative grammaticality and compatibility.

6.4.3.3. *Recognisability and functionality constraints on context-independent modifications*

The recognisability of pre/postmodification variants is argued to be guaranteed across the whole range, context-dependent or context-independent, thanks to the preservation of the conceptual cores of the idioms. That is, the original constituents are not manipulated or replaced, and the original constructions are only marginally disrupted by the insertions of the modifiers, which means little effect on the integrity of the conceptual cores of the idioms. As a result, pre/postmodified idiom variants can be always recognised as deriving from certain canonical idioms, provided that the language user is familiar with such idioms to start with.

In the data, context-independent modifications have a wide range of specific communicative functions, including (1) topic-introducing, (2) descriptive, and (3) intensifying/elaborative. Out of the three, the first two are more straightforward in terms of interpretation and arguably unlikely to be subject to conceptual metaphors and other conceptual motivation processing routes. Instead, upon the activation of the idiomatic

meanings of the idioms as wholes and the constituent phrase-induced meanings (if available), the modifying items are understood as specifying these meanings (in the cases of topic introduction or figurative modification in [18] – [20]) or the literal meanings of the constituents (in the case of literal modification in [21]) according to the rules of premodification in English:

- [18] *Each freshman representative or senator needs time to learn the legislative ropes*
- [19] *Earth Day finds itself in a political pickle*
- [20] *AT&T has just opened perhaps the most vindictive, messy can of worms it could have possibly found*
- [21] *Kasabian, after all, had a few tricks up their embroidered sleeves*

The last function, i.e. elaborative/intensifying modification, is more sophisticated and depends on complex systems of underlying motivations, without which the intended communicative purposes is unlikely to be understood:

- [22] *The Beauty Myth casts bright light on the ugly facts about woman's current situation and social identity*
- [23] *These studies shed a dim light, and do not illuminate any important scenes, nor indicate others which were before unknown*
- [24] *The top executive who is constantly looking ahead rather than backwards will rarely steer his company into a tight corner*

These variants feature compatible adjectival modifications which are embedded into the literal meaning of the idioms to evoke richer figurative meanings. 'Bright light' in [22] intensifies the phrase-induced figurative meaning of 'light' (explanation) to mean 'very clear explanation', whereas 'dim light' in [23] is downgrades this meaning into 'poor/sufficient explanation'. The bases for such successful interpretations of the two variants are postulated to be two conceptual metaphors, including UNDERSTANDING IS SEEING metaphor (e.g. *I see*), and QUALITY OF UNDERSTANDING IS QUALITY OF SEEING metaphor (e.g. *keep someone in the dark*).

In a similar manner, the modification in [24] intensifies the quality of the situation, so that being in a 'corner' (disadvantageous position) becomes 'tight corner' (very/extremely disadvantageous position). The motivation for the processing of the variant's meaning is associated with BOXING script, in which to be in a corner is bad due to the restricted room for movement. Conceptual motivation includes BEING IN A STATE IS BEING IN A LOCATION metaphor (e.g. *she is in a bad place right now*), and QUALITY OF THE SITUATION IS QUALITY OF THE LOCATION metaphor (e.g. *in the lion's den*).

Functionality, on the other hand, might not be recognised if the idiomatic expressions are poorly coded, e.g. violating the constraints of compatibility and/or grammaticality, in which case they will not be tolerated. Being context-independent, with the purpose of

coding the idiomatic meaning adaptations along normal linguistic parameters, context-independent modifications are required to be compatible, well-formed and motivated (i.e. interpretable on the basis of meaningful links between the variants and the original idioms.)

In other words, thanks to the availability of conceptual backings and/or other interpretation strategies, the meanings of many variants can be inferred without the need of contexts. Admittedly, some types of variants do not appear to necessitate complex motivating bases because their interpretations are immediately available once the meanings of the original idioms are activated (e.g. topic-introducing, descriptive modifications). As a result, while there is evidence of conceptual metaphors in language use, whether these large-scale schemas are psychologically real conceptual systems, or post-hoc theoretical constructs, remains an open question. From our linguistic data analyses in this section, conceptual metaphors do seem to have an explanatory power in meaning adaptation processes (intensification, elaboration), regardless of whether they are on-line or post-hoc.

Summary

Context-independent modifications to idioms, as they appear in the data, are amongst the more flexible and productive types of variants, possibly due in part to the leniency of the compatibility and grammaticality constraints on them. Depending on specific communicative purposes, a modification can be well-formed and compatible, or ill-formed and incompatible, as long as it is recognised as fulfilling either of the following functions: (1) topic-introduction, (2) description, and (3) intensification/elaboration. While the last function is executed via a number of meaning adaptation strategies (which are motivated by conceptual metaphors and/or other conceptual motivation sources), the first two are genuine modifications on either the literal or figurative meanings of the modified constituents, interpretable once the meanings of the idioms and their constituents are activated.

It has been argued in the literature that the opacity of some idioms blocks their creative behaviour in discourse, particularly with regards to modification, because modification requires the idioms to be analysable into meaningful components for the modifications to make sense (see chapter 3). While this hypothesis is true to certain extents, evidenced by the higher frequency of modifications among transparent idioms, it is not complete. It is in fact not impossible for opaque idioms to be creatively manipulated. It seems to boil down to the choice of appropriate strategies to code specific meanings that can be accommodated by opaque idioms, taking into account the lack of motivation and analysability among them. In the data, the type of context-independent modifications that are sanctioned to occur with opaque idioms include (1) modifications of the meanings of idioms as wholes, (2) transferred modifications, both of which are for the purposes of providing literal descriptions of the NPs and/or introducing the topics of discussions.

Last but not least, while the constraint of recognisability is non-applicable to modification, functionality is a prerequisite (just as with other types of variants), so that unintentional/erroneous expressions are filtered out from the class of idiomatic creativity.

Lexical adaptation processes as we have examined so far (substitution, modification) are characteristically divergent from the next group of variants, constructional adaption, in terms of possibilities and constraints, which is I argue to be the result of the differences in terms of the weight of each facilitating and constraining factors on them. The frequency of constructional adaptation variants in the data is very low (approximately 9% of the data), but it is believed that the analyses of the infrequent types of constructional adaption, although statistically insignificant, are still beneficial to our full understanding of idiomatic creativity and the positive and negative forces upon it. As a result, the subsequent section will zoom in on the processes of constructional adaption to highlight another aspect of idiomatic creativity that is more restricted and rule-abiding.

6.4.4. *Minor types of context-independent variants: Constructional adaptations*

Many grammatical structures do not just carry information, they are associated with *information focus*, i.e. they enable the user to select (within limits) which element of the sentence to be highlighted (Quirk et al, 1985: 89) while suppressing those less informally important elements. Different grammatical structures, therefore, serve the communicative purposes of emphasizing different important piece of information in a sentence.

Constructional adaptation as a variational strategy in idiomatic creativity includes a range of distinctive syntactic operations, each with its own strict grammatical rules and restrictions, including passive voice, pluralisation, grammatical highlighting, nominalization, and rankshift, which will be discussed in turn hereafter. Due to the low frequency of these types of variants in the data, their analyses will not be as detailed as previous types. However, the findings are found to be consistent enough to draw tentative conclusion about their behaviour and the contrast with previously discussed lexical adaptation processes.

6.4.4.1. *Passivisation*

Passive variants of verb phrase idioms is amongst the first, and possibly the most extensively researched, type of transformations by idioms, notably in early studies by Nunberg and his colleagues (Nunberg 1978, Nunberg et al. 1994). The fact that verb phrase idioms have passive variants has been the first proof of the productivity and analysability of idioms. Although this correlation has yet to be confirmed, a causal link is certainly observable. Despite their popularity in earlier days, it is recently observed, especially with the help of corpus data, that this type of variant does not occur very often in discourse. In our data, it

makes up approximately 5.5% of all the variants, which is very low. Some examples of the passivisation of verb phrase idioms include:

- [1] *The baby was thrown out with the bathwater*
- [2] *Here's a bud that needs nipping*
- [3] *Should they succeed, the record will require considerable straightening*
- [4] *Lips lined- so are his pockets*

The low frequency of the passive voice is argued to be partly due to the strict grammatical and semantic constraints on this type of transformation, which either block the possibility of passives completely or strongly discourage language users to use certain idioms in the passive. The restrictions for the passivisation of an active clause, according to Quirk et al. (1985), are summarized to include (1) verb constraint: the idiom has to have at least one verb phrase within it, and the meaning of the head verb has to be compatible with the passive, (2) object constraint: this verb has to be transitive, taking a phrasal object.

These conditions mean that the passive voice is blocked for verbs that only take active forms (e.g. stative verbs of 'being' and 'having' like *have a cow*/**a cow is had*; *get the chop*/**the chop is gotten*, *not have a snowball's chance in hell*/**a snowball's chance is not had in hell*). Intransitive verbs (*to boot*, *kiss and tell*) or verbs followed by adjuncts instead of objects (*keep in mind*, *live in cloud-cuckoo land*) are also not accepted in the passive voice. As a result, in the case of a strong inclination towards the passive voice to achieve foregrounding effects, these unacceptable verbs will have to be replaced by another verb which can appear in passive form, albeit entailing a certain degree of meaning alteration to the original meaning, for example:

- [5] Alan: *I think she wants to have sex with me*
 Charlie: *hang on a second* (opens newspaper and flips through the pages) *no, no snowball reported in hell, no sighting of flying pigs*

In fact, rule-breaking passive variants are not found in the data. Structurally, 100% of the passive voice variants of idioms strictly observe the rules of grammar to form syntactically well-formed passives such as in the above examples.

Besides grammatical rules, there are the constraints of *recognisability and functionality/meaningfulness* that all idiom variants must satisfy to ensure that the canonical idioms are recognized and their idiomatic meaning activated so that the variants can be interpreted accordingly. It is found that many factors can interfere with these two constraints, namely opacity, non-salience, and convention, for example:

- [6] **She started screaming and it was lost*
- [7] **Thirty three, or thirty five? A bell is rung by thirty five*
- [8] **To top off the horrible experience, there were a couple of women crying hysterically outside. A scene was made by them and it attracted a lot of unwanted attention*

In [6], it is the semantic emptiness of the constituent 'it' that blocks the passive voice transformation. The opacity of the idiom is the result of the lack of co-referentiality between the pronoun 'it' and its referent, i.e. it is not known what 'it' refers to.

Secondly, when idioms are literally well-formed and their constituents are idiomatically non-salient, their recognisability is compromised. In the case of *lose it*, for example, both constituents are core items with high frequency in the language (*lose* BNC_{freq} = 26605, *it* BNC_{freq} > 1 mil), which are linked by a non-salient structure, which means that only when they appear in the fixed order (*lose + it*) in the context of 'someone getting angry' can the meaning of the idiom be activated. Otherwise, it is almost certain that the phrase will be understood literally or unrecognized (e.g. **certain people who've had fame, and then lost it; *she started screaming and that was when he relinquished it*).

The third constraint reveals itself in variants such as [7] and [8]; that is, convention. In theory, *ring a bell* and *make a scene* should qualify to appear in the passive voice because they are both analysable and recognizable if the contexts are made informative enough. However, in reality, the five native speakers of English I interviewed reported that although they can recognize the idioms and the intentions behind the passives, the wordings of the idioms as such sound weird and mechanical. As a result, *conventionality* prevents certain idioms to appear in passive voice, even when the conditions are met and the right contexts are provided. However, it is worth realizing that conventionality and creativity tolerance are subjective and depend on a language users' openness to creative usage of language, as Carter (2004) mentions. Until large-scale interviews/survey questionnaires are carried out to investigate people's attitude towards 'unusual' passive idiom variants, conventionality remains a situational instead of global constraint.

Although only 1 opaque idiom was found in the data to have passive variants, it nevertheless undermines the hypothesis about the obstruction of the passive voice by opacity:

[9] *Prince William: 'I get the piss taken out of me for liking Linkin Park'*

In short, there are many grammatical and semantic rules that strictly govern the usage of the passive voice among idioms. Additionally, idiom variants are required to satisfy the recognisability and functionality constraints, which further restrict the possibility of passivising idioms. Conventionality is suggested to also exercise a restraining influence on idiomatic passives, which could be the topic for further research in the future.

6.4.4.2. Pluralisation

The insertion of quantifiers and/or the change in the morphological arrangements of the NPs, thereby switching them between singular and plural forms, are mostly executed on the analysable constituents of idioms ([1] – [4]), but not restrictively as can be seen in [5] – [6]:

- [1] *Several cans of worms open here*
- [2] *We speak of the 'sacrifices' we make for one another - that notion rings many bells*
- [3] *He touches a few raw nerves*
- [4] *Minton began increasingly to make scenes and when drunk would physically attack Norman*
- [5] *For the rest, the plan contains two stumbling blocks*
- [6] *They are the dark horses in the competition this year*

The reason why pluralisation is popular with analysable idioms is because, upon knowing the meanings of the idioms and the phrase-induced figurative meanings of the constituents, the language user can interpret the plural NPs in the variants according to these figurative meanings (e.g. 'several cans of worms' = 'several problems', 'many bells' = 'many memories', etc.), as long as these figurative constituent meanings are compatible with plural meaning (countable). Unanalysable idioms, on the other hand, having no individual constituent meaning, can only be pluralised as wholes, as in [5] – [6], as long as their meanings are compatible with plurals. In other words, both analysable and unanalysable idioms can be pluralised, albeit via different routes.

Recognisability amongst pluralisation variants is not an issue, because the original constituents and constructions are preserved, so their co-activations are enough to trigger the idiomatic meanings. Functionality, on the other hand, is only guaranteed when the requirements are met. In the case of the absence of analysability and/or the incompatibility of the NPs' with plural meaning, functionality is at best speculated and the expressions are often regarded as poorly coded instead of creative:

- [7] *There's no way I'm not doing it by the books. No Can't afford to do that*
- [8] *The houses were abolished by the authority while the rain has been falling cat and dog. As a result, the victims were facing acute difficulties in the rainy condition*
- [9] *Later my silly wee sister has put her feet down and refuses to let me near her Power Pack*
- [10] *Nutcase bike helmets...spoilt for choices!*

6.4.4.3. Grammatical highlighting

Fronting and *cleft sentences* are grammatical devices to highlight parts of a clause by placing important elements at the initial position of the sentence or in marked focus constructions (Quirk et al., 1985):

- [1] *Then, seemingly out of the woodwork, emerges a woman who claims to be the abandoned*
- [2] *It is fairly thin ice on which they are skating*
- [3] *He never made it to the fight. It was nerve that he lost, not his way*
- [4] *What the boys never lost was nerve*
- [5] *Humble pie is definitely a dish I should eat on a regular basis*

Grammatically speaking a fronted item is frequently an entire sentence element (subject, verb, object, complement or adverbial) but at times it can be a part of such a clause element (e.g. the postmodifier of a complement, etc.) (Quirk et al., 1985: 1377-1378). With cleft sentences, only clause elements (subject, direct object, adverbial of time and adverbial of position, object complement) can be placed in focus. Analysability, as a result, is essential to ensure the semantic completeness of the highlighted elements to enable their movement to different positions. Unanalysable idioms seem to be blocked from this operation:

**The piss is what you shouldn't take*

**It is a cow that you should not have*

**A cow you should not have*

There are also a number of specific grammatical restrictions on the structures of fronting and cleft sentences, including (1) the polarity constraint (i.e. not all fronting allow negation of the verb) (**out of the woodwork didn't come dark family history*), (2) transitive verbs constraint (**down my throat forced that idea by Tom*), (3) complex auxiliaries constraint (**on thin ice could have skated the young couple*), (4) semantic constraint on the verb and the fronted elements (*in cloud-cuckoo land lives a blond haired girl* but not **outside the box think the team*) (Quirk et al., 1985). Grammatical constraints, as such, are very specific in each case of highlighting. As a result, in reality, only a very small number of variants are found created highlighting devices.

6.4.4.4. Rankshift

Introduced by Halliday (1961), downward rankshift happens when 'an item normally having the function of rank x loses these functions on taking over those of rank y' (Halliday, 1966: 114). Rankshift, therefore, involves a change in the external functions of a symbolic unit, such that a unit of a given rank is 'demoted in size and reclassified as a unit of a lower rank, as a result of which it takes on the grammatical and semantic properties inherent to the lower rank unit' (Gregor, 1997: 128).

Rankshift in idioms is as rare as highlighting in the data, comprising approximately 1% of the data. Examples include:

- [1] *If you have a big kiss and tell story, our publicists will negotiate directly with the newspapers*
- [2] *Turn back the clock beauty: make-up tricks for grown-ups*
- [3] *They adopted 'a fly on the wall' role*
- [4] *The 'hook-line-and-sinker-buyers' will be responsible for any shortfall*
- [5] *All-you-can-eat humble pie buffet*
- [6] *This is my give-a-shit meter*
- [7] *What is this, users-crawl-out-of-the-woodwork day?*

In all of these variants, different phrases (VPs, NPs, AdvPs) and clauses have been rankshifted down to become parts of clause elements only (in these cases, premodifications in NPs). Stylistically, this type of variant is highly appreciated, particularly in creative genres where originality is valued (e.g. *their day-after-day complaints*, *a come-and-fight-me attitude*) (Quirk et al., 1985). Grammatically, the rankshifts of phrasal and clausal idioms into premodifying positions are sanctioned for use by all idioms given the semantically congruent head nouns (*a see-which-way-the-wind-blows kind of love*/**a see-which-way-the-wind-blows murder*). Semantic congruence, as such, ensures the functionality/meaningfulness of the variants. The negative effect of opacity is non-existent in this type of variant (e.g. *my give-a-shit meter*, *one of her have-a-cow moments*, *his dislike of the everyone-kicks-the-bucket ending of the story*).

In short, semantic congruence to ensure functionality is the only prerequisite of rankshift. Despite this flexibility and leniency, there are not many rankshift variants in the data, arguably due to the fact that they are a more wordy and creative type of arrangement that are only used in poetic/literary texts. Further investigation might follow to flesh out this observation.

6.4.4.5. Nominalisation

Nominalisation is a process that involves the functional reclassification of a non-nominal into a nominal unit. Syntactically, nominalisation is defined in traditional grammar as a noun phrase that has a systematic correspondence with a clausal predication which includes a head noun morphologically related to a corresponding verb or adjective, i.e. a deverbal or deadjectival noun (Quirk et al., 1985: 1288). For example, the noun phrase *refusal to help* corresponds to *he refuses to help*, in which the head noun *refusal* is morphologically related to the verb *refuse*.

Nominalisation can be achieved with or without morphological transformation of the head verb or adjective (Kolln, 1998: 63). In the data, nominalisation is also rare, taking up only 1.5% of the variants, including the following types:

Forming verbal nouns:

- [1] *Listening to him is like hitting a frightened raw nerve inside yourself over and over*
- [2] *The bosses require an immediate straightening of the record.*
- [3] *The opening of a scandalous can of worms*
- [4] *She didn't show any emotion, not even a batting of an eyelid*

Using deverbal nouns:

- [5] *It's a walk on thin ice to capture media about film production*
- [6] *It was a withering critique of the West's loss of nerve*
- [7] *There is still a massive gap to rebuild to call that a corner turn*
- [8] *Well-lined pockets and no ties. They're freelance, they're ambitious*

End-of-phrase suffix

- [9] *A vision of spick-and-spanness*

Conversion

- [10] *Women must learn that no good comes from a kiss and tell*

Examples of nominalisation in the data reveal a certain degree of creativity on the part of the speaker/writer when applying liberal interpretation to grammatical constraints, employing 'unusual' structures or morphological arrangements to create new NP idioms at times. For example, *spick-and-span* is treated as one word serving as the root word for the nominalization via suffix *spick-and-spanness* in [9]. Despite the lack of further evidence in the data, this principle can be envisaged to be applicable to other idioms provided that suitable suffixes are used to code relevant meanings, and that the reader is willing to accept it (e.g. *?the blatant kiss-and-tellability of her revelation is shocking*).

Another unusual type of nominalization is the conversion of VP idiom *kiss and tell* into NP without morphological changes as in [10]. All other variants, on the other hand, are rule-abiding. Indeed, there are specific grammatical constraints and discourse principles regarding weight and structural complexity that, when combined with recognisability constraint and opacity, restrict the possibility of nominalizing a large number of idioms:

**Chop getting is everyone's nightmare*

**I don't like your cow possession*

**His loss of it during the presentation ruined his chance*

**The rain of cats and dogs yesterday was unexpected*

Summary

Arguably, the low frequency of constructional adaptation variants in the data is due to the adverse effects of grammatical rules, opacity, recognisability, functionality and convention while at the same time other facilitating factors (motivation, analysability, context) are not powerful enough to relieve any of these effects. As a result, it is more difficult to create a variant via constructional adaptation than it is via other lexical processes (e.g. substitution, modification). This is an important distinction between constructional and lexical adaption variational strategies.

A question arises, however, concerning the tolerance among readers towards unusual, and otherwise unaccepted, constructional adaptations given the right contexts, such as:

?We didn't expect him to do so well in this competition. Let me tell you, dark is what this horse is.

?I have been very wrong lately with my strategy. Eating so many humble pies in a month is not a good thing is it?

?A bit drenched because of the rain! it's cat and dog rain, u know! I bought a new umbrella, coz I didn't bring mine'.

?Villagers are protesting against the plan to put a huge picture of Homer in their beloved fields. However, the government thinks their having of a cow over Homer is ridiculous

Overall, ambiguity due to the non-salience of the constituents (which results in the failure to meet recognisability constraint) is an important factor on nominalization. Taken out of context, these expressions are ambiguous due to the possibility of the expressions being literal. With context provided, the expressions are more or less meaningful but still depend on the subjective judgement of the reader. Particularly, the use of suffixes at the end of complex phrasal idioms to form nominal units seems a lot more unnatural and unacceptable than in other cases:

?Tom's full-of-himselfness is irritating

?New candidates' crawl-out-of-the-woodwork-ation changed the election picture

While enthusiasts can argue that these can still be argued to be creative variants, the absence of this type of nominalisation in the corpus data and on the internet suggests a strong aversion to such transformations. Tolerance of creativity, as a result, needs to be further investigated in future studies so that models of idiomatic creativity in particular and creative language in general can be fully formulated.

6.4.5. Minor types of context-independent variants: Conceptual adaptations

Some additional types of low frequency variants are found in the data (allusion, coordination, idiom blending), and it is unclear why they are unpopular. Unlike constructional adaptations, which are strictly constrained by many rules and principles, allusion, coordination and idiom blending are highly creative and striking and marginally constrained by recognisability and functionality only. The only explanation that this thesis would attempt to account for this unpopularity of these types of variants is either (1) a subjective choice among language users, or (2) the limitation of the current data that fails to capture these types, both of which needs to be tested with further studies in the future.

6.4.5.1. Coordination

This group of coordination extends the canonical idioms without the need for context thanks to their explicit meanings:

- [1] *In fact radio is alive, kicking and expanding, and there are thousands of hours of radio going out*
- [2] *Bobby Lavender had, as they say in LA, 'dressed the dress, walked the walk and talked the talk ' of the black urban gangster*
- [3] *Think and profit outside the box, but how?*

- [4] *It is raining cats, dogs, and just about everything else outside right now*
- [5] *Anyone have a grain of salt... or a whole bag.*
- [6] *Traditions are great, but there comes a time when one must be willing to throw the baby, the bathwater, and tub out and start over*

There are two distinct functions of coordination idiom variants in the data, including (1) description of concepts/actions/states of affairs, as in [1] – [3], and (2) intensification/elaboration, as in [4] – [6].

While the first function is seen as straightforward and immediately explicit from the meanings of the idioms and the meanings of the coordinates, the second function requires inferring skills for the variants' meanings to be correctly interpreted. For instance, the coordinate in [4] is not meant to be interpreted as 'rain heavily and just everything else outside' (which is incompatible), but since *rain cats and dogs* means 'rain heavily', *rain cats and dogs and everything else* is an intensification in terms of size/quantity, which is understood as intensification in terms of degree ('rain very heavily') via the conceptual metaphors LINEAR SCALES ARE QUANTITY. The same metaphor is applicable to [5], which can be interpreted as an intensification of the idiom *with a grain/pinch of salt* to create the intended meaning of 'be extra/very cautious'.

There seems to be no conceptual metaphors behind the interpretation of [6], but thanks to the analysability of the idiom and the general knowledge of the world associated with the images evoked by the idiom, we can understand the intended meaning of the variant, i.e. throw out the good thing, the bad thing, and everything associated with them, and start over.

There are minor grammatical semantic restrictions regarding the order of the conjoins in a coordinated clause, (e.g. the conjoins are parallel in terms of grammatical role and meaning, and the second conjoin is usually a chronological sequent or expressing consequence/result of the first, etc.) (Quirk et al., 1985: 971). However, these restrictions are more suggestive than regulatory. In the data, all variants follow these rules to create well-formed coordination variants. There is no incompatible coordination in the data.

Coordination can occur within unanalysable, unmotivated idioms, as with the examples [2], [4] and [5] above. Even though the idioms do not have constituent meanings or underlying motivations that render them transparent, it is still possible to infer their meanings via conceptual metaphors and/or general knowledge to understand their intended communicative purposes.

6.4.5.2. Allusion

In theory, allusion can be achieved via the use of salient lexical items or salient structures. In the data, the following types emerge:

Salient constituents:

- [1] *Humble pie is gonna be added to the menu*

Non-salient constituents in appropriate contexts:

- [2] *It's a pickle Terry! Women!*

- [3] *Our 1,000 subscribers will be happy to let these priceless pearls be passed on to your millions of readers*

Combination of (potentially non-salient) constituents as allusions

- [4] *Finance Bill: Baby-less bathwater from Congress*

- [5] *Ooh, good cookies, maybe these will take the taste of foot off my mouth*

- [6] *If you see a bull looking for directions to the china shop, tell him it's too late, Barbara's beaten him to it.*

Combination of salient concepts as allusions:

- [7] *I believe the appropriate metaphor here involves a river of excrement and a native American water vessels without any means of propulsion*

- [8] A: *what if I get him to promise to behave?*

B: *then we would let him back in the team and while we're doing that you can take an aerial tour of LA on your pigs*

The use of a noun phrase as an allusion to the original idiom ([1]) is the most popular type of allusion in the data. Moreover, the noun phrase used as allusion in this example is the salient (key) constituent within a familiar idiom, which means it has the power to trigger the full recovery of the original idiom. In the case of less salient constituents ([2] – [3]), appropriate contexts significantly increases the chance of the idiomatic meaning activation. The constituent *pearl* in the idiom *pearl of wisdom* in [3] is non-salient, but it is successfully exploited given the right context of shared knowledge between readers, hence narrowing down the interpretation and suppressing the literal reading of 'piece of jewellery'. As a result, context is given maximum credit in helping with recognition of non-salient allusions.

The use of a combination of constituents as allusions such as [4] – [6] (instead of a single one) is another useful type of allusion when the constituents within the idioms are not salient enough and the context is not sufficiently provided, in which case the co-occurrence of two or more constituents greatly enhance the recognisability of the variants).

Last but not least, in the extreme case of allusion, in which all the original lexical items and constructional features are replaced, the co-occurrence of two or more salient concepts is found to be capable of activating the figurative meaning of the idiom, as found in [7] – [8]. Even though the whole idiom *up shit creek without a paddle* is paraphrased in [7], the concepts of 'shit creek' (realised by the phrase 'river of excrement') and 'without a paddle' (realised by the phrase 'water vessel without any means of propulsion') can be recognised and their co-activation is salient enough to activate the figurative meaning of the idiom in the context of somebody in trouble without solution. The variant, therefore, is an extreme type of

allusion, whose meaning is triggered not by a salient constituent or salient structure but by salient *concepts* or *schemas* in context.

The strictest constraint on allusions, as a result, is recognisability, the lack thereof proves to result in ambiguous, awkward and unrecognizable allusions such as:

- **That's a big cow!* (idiom: *have a cow*)
- **Have you heard of hand?* (idiom: *force one's hand*)
- **Do you expect the races in the near future?* (idiom: *off to the races*)
- **That would learn the* (idiom: *learn the ropes*)
- **They are having a* (idiom: *have a cow*)

6.4.5.3. *Idiom blending*

Although striking and creative, this type of variant is not very popular in the data. In fact, the only examples of mixed idioms include:

- [1] *Another nice kettle of fish you pickled me in*
- [2] *they could at least hold the lid on the Indo-Chinese kettle for the predictable [...] future*
- [3] *A silver pocket lining - Many a dark cloud can provide a silver lined pocket and, let's face it, our industry is better equipped than most to calculate on challenging times with innovative thinking*
- [4] *Most people at the newspaper think of this job as a career black hole, but the editor wants Holly to be less aggressive than her predecessors and build a better relationship with the local force. In the editor's words, "Eat humble tart. Don't upset the apple pie."*

These are highly unpredictable and one-off examples of idiomatic creativity. It is postulated (although very tentatively due to the low number of variants) that the constraints on idiom blending will be minimal, including the recognisability and functionality of the variants in the communicative contexts.

6.4.6. **Conclusions: Context-independent variants**

As suggested by their name, context-independent variants do not rely on context for the success of their interpretations. Instead, a language user with a sufficient knowledge of the grammatical/semantic rules of English and familiarity with the figurativity of English idioms should be able to infer the meanings of these variants. This context-independence entails a number of important characteristics that stand in stark contrast with those of context-dependent variants that we discussed earlier in the chapter.

One of the most prominent characteristics of these variants is their liability to grammaticality and compatibility, which grows directly out of the necessity to ensure the interpretability and communicability of context-independent variants. Thanks to their well-

formedness and compatible meanings, these variants can be interpreted without the need for extra context. Admittedly, grammaticality is observed with varying degrees of strictness according to the specific types of variants. That is, while the rules are most rigid and unyielding with constructional adaptations (passive, fronting, cleft sentence, nominalisation), they are more lenient with lexical adaptations (substitution, modification, allusion, coordination, blending).

Compatibility, similarly, is observed to be more lax with lexical adaptations, in which case compatibility can be on either the literal or figurative meaning level, depending on the specific purposes (topic introduction, literal description, figurative description, intensification, etc.) In constructional adaptations, on the other hand, compatibility is a prerequisite. For example, pluralisation can only be done if the figurative meaning of the constituent is compatible with the plural sense; or the passive voice can only be employed if the figurative meaning of the idiom is compatible with the passive.

In other words, the constraints of grammaticality and compatibility, although required with context-independent variants, can take different forms and have different degrees of strictness according to the types of variant.

The constraint of recognisability is more or less guaranteed in non-disruptive types of variant such as modification, rankshift, coordination or pluralisation. In these types variants, all the original constituents and structures are preserved, which theoretically guarantees recognisability even in the absence of context (e.g. *eat compatibility humble pie, up-shit-creek without-a-paddle situation, rain cats and dogs and lizards too, touch a few raw nerves*, etc.). It is acknowledged, however, that some idioms have equally strong literal readings, which necessitates appropriate contexts to terminate the 'horse race' between the literal and figurative readings of the strings and support the choice of either meaning in the particular contexts (e.g. *this name rings a loud bell, *teach your dog to ring a bell to indicate it needs out*, etc.).

In other types of variants (e.g. substitution, allusion), the conceptual cores of the idioms are interrupted by the removal and replacement of the original constituents/structures with new elements, which compromises the salience of the strings and inhibits the activation of their idiomatic meanings. As a result, the variants created via these strategies need to satisfy the 'recognisability' constraint by either (1) retaining salient constituents, or (2) ensuring the co-occurrence of two or more constituents or concepts, or (3) relying on the effects of priming.

In other words, although in theory the preservation of the conceptual cores should guarantee recognisability. In reality, ambiguity is still present with non-salient, literally well-formed idioms, in which case context is crucial in disambiguating them.

All of the aforementioned constraints (grammaticality, compatibility, recognisability) are to ensure that an adaptation of a particular idiom in a specific communicative situation is functional, i.e. meaningful and fulfilling a specific communicative purpose. The successful interpretation of a context-independent variant is markedly different from that of a context-dependent variant. In the latter case, as we have discussed earlier, it is via direct reference to the context of utterance that the meaning of a variant is understood. Context-independent variants, on the other hand, feature various meaning adaptation operations that can be decoded via inferential processes. It is found that conceptual motivation (including conceptual metaphors, conceptual metonymies, emblems) and analogy are two of the most important inferential techniques, most notably in the cases of intensification/elaboration by substitution, modification and coordination, in which the figurative meanings of the original idioms are elaborated along conceptually motivated parameters. When the same operations (substitution, modification and coordination) are used for the purpose of topic-introduction and description, the variants are potentially explicit via direct composition of the meanings of the constituents according to the composition rules of English, just as in the cases of passive voice, rankshift, nominalisation, fronting and cleft sentence. The conclusion here is that there are different inferential strategies that a language user can potentially employ in specific cases. Overgeneralisations about the role of conceptual metaphors appear counter-factual according to the current data.

The adverse effects of opacity, argued in many previous studies to block idiomatic creativity, are observed to be offset in certain types of variants, including substitution, modification, rankshift, nominalisation, allusion. In other operations, such as passivisation, fronting, cleft sentence, on the other hand, opaque idiom variants are not found, arguably because these types of manipulation require that the constituents have individual meanings. The conclusion concerning opacity, therefore, is that it is only detrimental to certain types of variants, while in other cases opaque idioms can have as much creative potential as transparent ones.

Last but not least, conventionality is found in the data to be an extra constraint on idiom variants. Some idiomatic expressions, although appearing eligible on the surface, are not widely accepted due to their 'unnatural' wording (e.g. **a scene is made by them*). Some other expressions are meaningful because of their conventionalized usage despite the conflict with common inferential strategies (e.g. *she's my left hand and he's my right hand*, *I can't run this office without them*). This factor has not been extensively researched in the literature and has only been marginally been discussed here due to the limited data. Future research could expand on this aspect of idiomatic creativity to expand our understanding of the phenomenon.

In short, the degrees of rigour of grammaticality, compatibility, recognisability, functionality, conventionality, and the negative effects of opacity are mirrored in the creative behaviour of idioms in different context-specific types of variants.

On balance, with both context-dependent and context-independent variants, the functional profile of the creative behaviour of idioms in use can be established on the basis of the context-specific lexico-grammatical types of variants. Each type has its own functions and requirements which determines the relative roles of context, transparency, cognitive constraints, and phraseological constraints within it. The incorporation of context into the traditional lexico-grammatical classification of idiom variants offers new insights into the relationship between lexico-grammar and context and the extent to which context modifies cognitive and phraseological constraints on idiom variants in specific communicative situations. The suggestion, as the result, is that a model of idiomatic creativity should at least show discrimination between different types of variants in terms of both context-dependency and lexico-grammatical alterations, taking into account the changing roles of the facilitating and restraining factors in each type. An attempt at such a model is presented below.

6.5. Proposal of a context-based model of idiomatic creativity

The foregoing discussions highlight an important factor of idiomatic creativity that has not been systematically characterised in the literature – context. Although acknowledged to be a powerful force in the understanding of figurative language in general, context in idiomatic creativity in particular has largely been under-researched. Apart from some fragmented comments in several studies, context-specific variants have traditionally been treated as wordplays and marginalized in mainstream frameworks. The proposed framework in this thesis goes against this trend by fleshing out the role of context in idiomatic creativity in relation to other highly recognised facilitating and constraining factors.

Context is the overarching factor in all types of unmotivated, context-dependent variants. These types of variant are characterised by their referential communicative purposes and the dependence on context for the successful interpretation of the variants as such. That is, they refer to specific aspects of the contexts of discourse that can only be correctly understood if this contextual information is made explicit in the discourse co-texts. Due to this context-dependency, the interplay between facilitating and constraining factors in these variants is context-dominant. Context can overrule the constraints of compatibility and grammaticality as well as relieve the negative effects of opacity. In other words, context-dependent variants are not restricted to transparent idioms and are not required to observe the rules of grammaticality and/or compatibility, provided that the context is made informative enough in each case to facilitate the variant's recognisability and functionality. These two aspects, in turn, depend on the availability of context to be successfully achieved. As such, the role of context in context-dependent variants is unassailable.

In context-independent variants, the role of context changes to being a facilitator of recognisability and functionality of idiom variants while none of the other constraints can be overruled. Particularly, in context-independent constructional adaptations, the cognitive

constraints of recognisability, functionality, grammaticality and compatibility as well as the constraint of conventionality are all exerting equal power, causing difficulties to the manipulation of an idiom according to these types. In context-independent lexical adaptations (substitution, modification, coordination, allusion, blending), compatibility, grammaticality and conventionality are secondary and can be flexible if recognisability and functionality are achieved.

In other words, the creativity of idioms is best described in terms of context-specific lexico-grammatical types with reference to the specific communicative purposes of each type. Indeed, each type of variant, created via a particular lexico-grammatical variation operation for a specific communicative purpose, has its own set of determinants in which the dominance of one determinant over others determines the range of idiom variants available to it.

Such a context-based model of idiomatic creativity does not only have a powerful descriptive strength, it also helps predict the maximum potential for creativity by different groups of idioms. As a whole, the model supports the general claims in the literature about the linear relationship between transparency and idiomatic creativity, i.e. transparent idioms *can* have the widest range of variants, from context-dependent types to context-independent types. However, the creative behaviour of opaque idioms is predicted by the current context-based model to be much greater than had previously been acknowledged. Opaque idioms are found to be able to accommodate all types of context-dependent variants and most types of context-independent variants as well, except for constructional adaptations.

On balance, a context-based model of idiomatic creativity as such is the ultimate proposal of the thesis, which is believed to offer both descriptive and predictive power to characterise the phenomenon of idiomatic creativity in use.

6.6. Conclusions

This chapter deals exclusively with qualitative analyses of the idiom variants found in corpora and on the internet to identify significant patterns among them that might help pinpoint the underlying factors of idiomatic creativity and their interactions in relation to context.

In general, the data shows a struggle between the three most important aspects of idiomatic creativity: creativity, idiomaticity, and communicability. Characterised by creativity/novelty, idiomatic creativity breaks away from expected patterns and norms to create new context-specific instantiations of an idiom, sometimes using ungrammatical and incompatible lexical/grammatical devices, etc. Idiomatic creativity, thus, represents the 'organised violence against cognitive processes' as the Russian formalists put it. At the same

time, such a novelty is restricted to certain extents by idiomaticity and communicability. Indeed, idioms are institutionalised phrases that have become familiar and are not easily accepted if used in different forms. Simultaneously, the novelty needs to be communicated to a reader/listener, hence its liability to general cognitive principles to ensure interpretability and communicability. The struggle between these three components renders the phenomenon of idiomatic creativity complex and difficult to capture.

In this chapter, idiom variants in the data are analysed into two groups according to their dependency on context (i.e. context dependent and context independent idiom variants) and then within them subgroups of variants with regard to the type lexico-grammatical transformation. Such a classification shows clear patterns in terms of the creative behaviour of idioms, the constraints on each type and the role of context in facilitating creativity and relieving such constraints.

In context-dependent variants, context is seen as the most powerful factor that can support recognisability and justify functionality even when they are not immediately clear. Context maximises the creative potential of idioms, creating windows for opaque idioms to be manipulated as well as sanctioning 'unusual' variants, thanks to the power of context in facilitating recognisability and functionality in ways that no other facilitating factors (e.g. motivation, analysability) can do. As a result, the differences between transparent and opaque groups of idioms in terms of creative behaviour are insignificant.

In context-independent variants, however, the role of context is minimised while that of semantic transparency and cognitive constraints is emphasised. The motivation, analysability and well-formedness of an idiom evidently generally allow more room for creativity, although this creativity is still bound by one or more of the cognitive constraints depending on the type of transformations. The prevalence of transparent idioms over opaque idioms in terms of context independent creativity partially confirms the hypotheses in the literature concerning the role of semantic characteristics in idiomatic creativity.

By characterising idiomatic creativity according to the degrees of dependency on contextual information by different types of idiom variants, the creative potential of idioms can be elaborated with greater detail and accuracy, which arguably truthfully reflects the highly complex nature of the phenomenon of idiomatic creativity in use. A context-based model of idiomatic creativity is, therefore, proposed in this chapter.

In terms of methodology, the eclectic and flexible attempt to be driven as much as possible by qualitative consideration of the corpus data while still maintaining a structure based on existing literature and quantitative analysis fills a gap in the language variation literature, particularly with the incorporation of context into the framework. The result is a context-based model of idiomatic creativity using corpus data that benefits from both qualitative and quantitative methods.

CHAPTER 7

CONCLUSIONS AND OUTLOOK

7.1. Summary

This thesis has been concerned with the characterisation of idiomatic creativity from a pragmatic perspective, which builds around the concept of context and the interplay between context and other recognized facilitating and constraining factors of idiomatic creativity. It is a response to the recognition of the complexity of the nature of idiomatic creativity and the heterogeneity of its set of determinants, which involve the internal semantic characteristics of idioms (motivation, analysability, literal well-formedness), cognitive constraints (recognisability, functionality, compatibility, grammaticality) and conventionality constraints. The effects of all of these factors are argued to be mitigated by context to varying degrees in different types of variants with different communicative purposes. Ultimately, the thesis proposes a model of idiomatic creativity that describes and predicts in relation to authentic English discourse the full creative potential of idioms under the effect of both context and other facilitating and constraining factors.

Being two highly indeterminate and notoriously difficult-to-define phenomena, idiomaticity and creativity are discussed in chapter 2 to lay the foundation for the delineation of the concept of *idiomatic creativity*, the phenomenon of interest in this thesis. The hypotheses in the literature about the effects of the internal characteristics of idioms (literal well-formedness, motivation and analysability) and cognitive constraints (recognisability, functionality, compatibility, grammaticality) on the creative behaviour of idioms are reviewed in detail in chapter 3 to provide theoretical support for the analyses in the second half of the thesis as well as to pinpoint the gap in the literature where this thesis fits. That is, there is a need for a model of idiomatic creativity that (1) outlines the interplay between its determinants instead of treating them in separation, so that a full description of idiomatic creativity in use can be achieved, (2) takes into account the roles of context as the facilitator of idiomatic creativity in specific communicative situations, and (3) includes all types of idiomatic creativity in the model, particularly the context-dependent types that have traditionally been treated as wordplays and excluded from systematic analyses. The context-based model in this thesis claims to have done this.

Before the empirical testing of the hypotheses and the qualitative analyses of corpus data in chapter 5 and 6, respectively, a description of the data sources and research methodology is given in chapter 4. Following the principles of a corpus-based analysis of language, the thesis makes use of the rich sources of data from the BNC, CANCODE and MICASE to probe into the idiomatic can of worms using both quantitative and qualitative analytical

methods in the hope that their combined effects will help illuminate various aspects of the phenomenon of idiomatic creativity.

To be specific, quantitative analyses are carried out in chapter 5 to test the hypotheses about the roles of transparency in idiomatic creativity. Raw frequency counts as well as statistical procedures are performed to reveal the relationship between the two constructs. In pure number, the dominance of well-formed, motivated and analysable idioms in terms of creativity is confirmed. Statistically speaking, however, only the effects of literal well-formedness and motivation are found to be significant. Analysability, surprisingly, is found to be unrelated to the degree of creativity among idioms according to statistical tests.

Chapter 6 addresses the main question about the possibilities and constraints on idiomatic creativity: whether the creative potential of idioms are globally restricted by opacity, functionality, recognisability, grammaticality, and compatibility, or whether their adverse effects are case-specific and can potentially be alleviated under special circumstances by sufficient context. Our conclusion will be that the effects of the internal semantic characteristics of idioms and the cognitive constraints on idiom variants are context-specific and variant-type-sensitive.

To model the phenomenon of idiomatic creativity according to pragmatic principles of utterances, which view meanings as essentially context-embedded, idioms are described as constructing a message in context via lexical and grammatical devices. That is, certain types of variational operations are more flexible and minimally restricted (e.g. modification, substitution) while some others are subject to rigid rules (e.g. constructional adaptation). In most cases, the availability of context can lift certain constraints to maximize the potential for creativity by idioms. The model of idiomatic creativity proposed, therefore, is a systematic context-based model of idiomatic creativity, which constitutes an integrated and theoretically motivated attempt to reassess the possibilities and constraints on the creative behaviour of idioms. In doing so, the model paid tribute to both the heterogeneous nature of idioms, the power of context, and the diverse communicative qualities of different types of idiom variants.

7.2. Contributions of the thesis

This investigation into idiomatic creativity contributes to the existing knowledge of the field in three ways: firstly, it produces a comprehensive description of idiomatic creativity as a context-specific creative phenomenon as well as phraseological units; secondly, it empirically confirms the hypotheses in the literature about the relationship between the internal semantic characteristics and the creative behavior of idioms in discourse; and thirdly, it submits useful suggestions for the development of the methodology used in corpus investigation of idiomatic creativity. The first two relate to the content insights that this

thesis offers via its quantitative and qualitative analyses of the data, while the last is associated with the strength in research methodology that the thesis prides itself on.

7.2.1. Content insights

First of all, the generous take on idiomatic creativity in this thesis extends the existing theories of the phenomenon over to the realm of wordplay, which has traditionally been ignored in the literature due to its indeterminate nature. Arguing that they represent the essence of creativity (reflected in their unexpected, ill-formed and incompatible forms as well as their surprising and striking meanings), this thesis includes puns/wordplays (context-dependent variants) into the idiomatic creativity data pool for a comprehensive model of the phenomenon. Far from being random, these variants show patterns in terms of forms and functions, and, together with systematic context-independent variants, they complete our understanding of idiomatic creativity. The model of idiomatic creativity developed in thesis is, therefore, more comprehensive than has previously been attempted.

As a whole, the thesis substantiates the following assumptions about idiomatic creativity:

- (i) The transparency of idioms is related to their creative usage in discourse.
- (ii) It is rather simplistic to view idiomatic creativity only in relation to its transparency. Besides transparency, the determinants of idiomatic creativity also include cognitive constraints (recognisability, functionality, compatibility, grammaticality) and phraseological constraints (conventionality), all of which exist in an interplay with one another as well as with context.
- (iii) Context is seen as the overarching facilitating factor that can maximize the creative usage of idioms in discourse by lifting the constraints of compatibility and grammaticality, supporting recognisability and justifying functionality while relieving the adverse effects of opacity and conventionality.
- (iv) The degree with which the determinants of idiomatic creativity can be infiltrated by context is variant-type-sensitive. At one end (context-dependent variants), the role of context is absolute (i.e. it is capable of all of the above). At the other end (context-independent constructional adaptations), context facilitates the recognisability and functionality of idiom variants but none of the other constraints can be overruled.
- (v) Conceptual metaphors are one of the sources of motivation that a language user might employ to infer the meaning of a context-independent variant. Other sources include context, conceptual metonymy, analogy, knowledge of the world, knowledge of the etymology of the idioms, etc., depending on the specific cases.

- (vi) Each idiom is perceived as having a stable conceptual core which triggers the activation of the idiomatic meaning of the particular idiom. 'Stable' in this sense does not mean that the core has to remain strictly invariable but that changes are required to have motivated links to the original concept, so that the original concept can still be accessed. Such links are argued to be established by the cognitive power of conceptual salience, conceptual co-activation and cognitive priming.

All in all, context, the internal semantic characteristics of idioms, cognitive constraints and conventionality constraint do not apply to all idiom variants in the same way. Each type of variants has its own set of determinants with one or more of these factors and varying degrees of importance attached to them. As such, successful interpretations of some types are based on their well-formedness according to the constraints, and of some others on the degree of support they receive from context. These types are not the traditional lexico-grammatical types, but context-specific with reference to the particular communicative purposes of the expressions.

7.2.2. *Methodological insights*

The contributions made to corpus methodology by the thesis are twofold, including (1) its role as an example of a corpus-based investigation that reflects the tenet of corpus linguistics with a mixed research methodology of both quantitative-qualitative values, and (2) its discussion of issues relating to corpus design that will benefit idiomatic creativity studies. Taken together, these aspects pave the way for further corpus investigation into idiomatic creativity and other forms of creative language.

7.2.2.1. *Methodological strength of the thesis*

First of all, I argue that in terms of methodology, the thesis meets the usual kind of rigour that is required for applied linguistic research so that tentative generalizations can be made with certain degrees of confidence. The thesis is placed within the context of existing studies in the field to ensure continuity and solid theoretical foundation; at the same time, it pushes the field forward in a direction that has not been extensively researched before, i.e. a context-based characterization of idiomatic creativity.

The proportionate stratified sampling method of data collection employed in this thesis effectively minimises the issues of biased, convenient sampling. The idiom sample set of 100 cases was not arbitrarily chosen from dictionaries or an existing list, but identified in real texts. The text collections (200,000 words) from which the idiom set was drawn were chosen at random and comprise equal proportions of written and spoken texts across both serious/academic genres and entertainment genres, with the goal that they represent, as close as possible, the spoken and written language use in general. Apparently, the texts

chosen cannot be claimed to be a completely true representation of all the texts that exist, but then this certainty can never be guaranteed in applied linguistics. As a result, with all the care that has gone into data collection, it is argued that the representativeness of the data can be claimed with relative confidence.

The creative uses of the 100 idioms are systematically searched for in three corpora: BNC, CANCODE and MICASE. Given that BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, it is generally acknowledged to be a trusted database that is representative of modern English. The CANCODE and MICASE corpora are also consulted to increase the proportion of spoken data.

The random but principled selection of the idiom sample in real texts, the size and representativeness of the corpora and the naturally occurring nature of corpus data are valuable in supporting the claims made in this thesis and set it apart from other studies that either use conveniently chosen samples, limited data or contrived examples which are not naturally occurring. The corpus data indeed offers robust and objective results/observations of our linguistic phenomenon of interest.

Another strength that this thesis demonstrates is its sophisticated handling of corpus statistics via quantitative processes to test existing hypotheses about the determinants of idiomatic creativity. The testing of hypotheses on corpus data is carried out using both frequency counts and statistical tests. While frequency counts provide invaluable proof of prominent trends and differences, they are often questioned with regard to the significance of such trends/differences; hence the addition of the statistical tests to observe significance. In turn, statistical tests are not without pitfalls, e.g. they are prone to errors if there are some issues with data noise or representativeness, etc. As a result, frequency counts and statistical tests are used to supplement each other, so that a reliable conclusion can be made about the hypotheses.

Last but not least, the detailed qualitative analyses of individual instances of idiomatic creativity in corpus data in chapter 6 offer a robust description of the phenomenon, including its general trends, functions, distributions and manifestations in real life usage under the influences of context and other governing factors. This holistic corpus-based approach has brought valuable insights into the nature and discoursal behaviour of idiomatic creativity that has not been made available before.

The outcome of such a combination of detailed qualitative and systematic quantitative analyses of corpus data, therefore, renders the research findings highly robust, reliable and at the same time replicable, offering a comprehensive outlook on the discoursal behaviour of the phenomenon of idiomatic creativity.

7.2.2.2. *Suggestions about corpus design*

The difficulties with which idiom variants are retrieved from corpora using the current tools and the techniques to overcome these difficulties to maximize the success rates of the retrieval process have been presented in chapter 4. What makes idiomatic creativity such a challenging data set to retrieve is that it is highly varied and unpredictable, including unconventional usage of language, and not having been fully characterised. As a result, based on our extensive description of various types of idiomatic creativity in real life discourse, it is realized that patterns start to emerge that could be used as the guidelines for the future developments in corpus design so as to facilitate idiomatic creativity studies. Suggestions include:

(i) Semantic annotation and idiomatic creativity

The stability of the idiomatic conceptual cores and their role in triggering the idiomatic meaning activation as well as in facilitating extreme cases of manipulation have been substantiated via empirical analyses of corpus data in chapter 6. As a result, it is suggested that corpora could be developed towards having semantic annotation, i.e. being tagged into semantic categories on the basis of their sense relations, including synonyms, antonyms, hypernyms and hyponyms (which are used in the electronic lexical database WordNet) (see Fellbaum, 1998). In effect, such a task has been proved possible by such a corpus tool as UCREL Semantic Analysis System (USAS), developed at Lancaster University. The software offers automatic semantic annotation of English texts where each content word in the texts is assigned a value within 21 primary semantic fields, which are then further subdivided into 232 categories (see Wilson and Thomas, 1997; Rayson et al., 2004). F1 for example is the category of FOOD. The level of sophistication of these categories might still need further evaluation, but the principles can be envisaged to be applicable to a corpus so that each word can be tagged with semantic information as well as the currently available grammatical information.

When semantic features become available in corpus data, corpus tools/software can then allow words that are semantically related to be included in a query. For example, a search for *eat humble pie* will also yield such variants as *swallow/chew/taste/stomach humble pie* if a 'verb synonym' relation is selected. As such, the probability of identifying creative variants of the idiom would significantly increase, hence easing the laborious process of performing repeated individual searches for each entry.

(ii) Phonological annotation and idiomatic creativity

The large amount of variants that are created via punning (using homophones, homographs, polysemic words, etc.) pinpoints another possible development in corpus

design that will help address this type of idiomatic creativity, i.e. phonological annotation of the data.

It is suggested that corpora be annotated with phonological information, so that homophones, rhymes, assonance, alliteration, consonance, etc. could be grouped into a coherent system. For instance, such words as *day*, *pain*, *whay*, *rein* will be grouped together under [æɪ] sound, and so on, using the IPA for English as the tagset. In effect, automatic phonemic annotation of words has been attempted on small data set using different grapheme-to-phoneme conversion techniques (see Bosch & Daelemans, 1993; Divay & Vitale, 1997; Auran et al, 2004). As a result, the task of tagging corpora with phonemic information is feasible, which, when completed, will allow variants such as *kiss and sell*, *dis and tell*, *kill and tell*, *kiss and swell* to be retrieved for the 'kiss and tell' query. Such a multiple retrieval process would significantly reduce the time and efforts put into individual searches, especially when we do not know exactly what we are looking for.

(iii) Figurative annotation and idiomatic creativity

The retrieval of idiom variants (and other types of figurative language) is envisaged to benefit from a special type of tagging that can identify figurative uses of language in texts, such that *dark horse* as a literal description of an animal is ignored (e.g. *you can understand that light coloured breeches make it easier for the judge to see where the rider's legs are against a dark horse*) but *dark horse* as an idiom is tagged (e.g. *Peter is always a dark horse in any championship*) to distinguish figurative uses of a string from the literal ones. As such, searches for idiomatic variants can be designated to apply to a specific part of the corpus only (i.e. the figurative language part), thereby reducing the number of searches as well as the number of concordances for each query.

Admittedly, the annotation of a corpus, especially a large commercial one as the BNC, is a mammoth task that might involve reading through the corpus. Suggestions made in this section have been intentionally kept on the superficial conceptual level, while sparing the technical details of the actual tasks to experts in the field. It might take a while before these features become available for the researcher who is interested in idiomatic creativity, but when they do, he/she is better equipped to carry out the task of retrieving creative uses of idioms in discourse.

7.3. Limitations of the thesis and suggestions for improvements

Having highlighted the insights gained from a context-based approach to idiomatic creativity, it is now necessary to acknowledge the limitations involving the progress of this thesis so that implications for future studies for the improvement of the thesis can be pinpointed.

7.3.1. Data assumptions

The works that inspire me to delve deeper into the complex relationship between idiomatic creativity and its determinants include mainly a body of psycholinguistic and cognitive semantic studies which postulate a connection between the possibility of understanding and analysing certain idioms and our capacity to manipulate them in speech and writing. To empirically evaluate this claim by using corpus data to analyse the correlation between the internal semantic characteristics of idioms and their creative types and tokens (chapter 5), the thesis assumes that the different ways people analyse idioms is reflected in the ways they manipulate them. That is, if a speaker sees an idiom as analysable, he/she is likely to use it more creatively than otherwise. As such, if an idiom is analysable, its usage in corpora (i.e. variant types/tokens) is expected to be an indicator of this analysability. In reality, however, an analysable idiom might have low creative usage due to a number of other factors (context, familiarity, data collection, personal preferences, etc.) As a result, one might argue that the use of decontextualized corpus data for statistical tests to evaluate the connection between idiomatic creativity and the internal semantic characteristics of idioms will not yield a truthful representation of the relationship between them.

Acknowledging the potential noise in the data, it is argued that corpus data as used in this thesis (BNC, CANCODE, MICASE), as well as the large number of cases (100 idioms), should be able to cancel out individual idiosyncrasies, so that noticeable trends of the relationship between transparency and creativity would emerge. Exceptions and special cases, which are ignored in statistical procedures, will be analysed qualitatively in chapter 6 to flesh out the model of the determinants of idiomatic creativity.

To further confirm the findings in this chapter, duplicate studies can be carried out with a different set of idioms in order to find out if the trends revealed in this chapter can be directly transferred to other idioms. On the basis of the present evidence, one might predict that they do.

7.3.2. Data size

Although the three corpora (BNC, CANCODE, MICASE) make a large data base, idioms and idiomatic creativity tend to occur relatively infrequently. This could be argued to be a weakness of the current thesis. However, the total number of creative variants found is approximately 1,300, which is acceptable. Acknowledging the issue with data size, internet data is incorporated into the qualitative analysis part of the thesis to increase the richness of the database. Nevertheless, it is also acknowledged that, due to the low frequency of idiomatic creativity in discourse in general, the issue of data size will not be resolved until more extensive sources of data become available. As it stands at the moment, the three corpora I use (of which BNC is one of the biggest commercial corpora available) and the supplement internet data is the best option.

Future corpus-based research on idiomatic creativity therefore could attempt to incorporate more extensive databases (preferably idiom-prone genres) to improve on the size issues to substantiate the insights of the present thesis.

7.3.3. *Data representativeness*

One could also argue that the 100 idiom sample might not be representative of idioms in general, nor the collection of texts from which they are extracted is representative of all the genres there is in English. It entails that the findings here are only true with the present set of data. While this is true, it can be said about any applied linguistic study or any social science investigation in general in which analyses are carried out on samples instead of the entire population. In the current thesis, attention has been paid to ensuring the representativeness of the sample via (1) the random selection of a wide range of texts from which idioms are identified and (2) the 'proportionate stratified sampling' method via which the idiom sample is chosen. The combination of these two factors effectively minimizes the chance of 'bias' that might occur with convenient sampling technique.

To improve data representativeness and to test the validity of the claims made in this thesis, future research could replicate the sampling process on a different population (i.e. different set of texts) to see if the results are different. Idiosyncratic features might still emerge but general trends and tendencies as revealed in this thesis will be expected to be similar.

7.3.4. *Data balance*

Despite its size, the BNC is seriously unbalanced, in that the spoken section of the BNC is relatively small (10% of the corpus) compared to the written counterpart. This is why the CANCODE and MICASE corpora were also consulted to increase the proportion of spoken data. Admittedly, these two spoken corpora are not big enough to create a balanced database, but compromises have to be made here due to technical difficulties concerning spoken data collection and the limited availability of spoken corpora. Besides, the findings in this thesis do not distinguish between spoken and written creativity, so the skewed ratio can be ignored.

It will be an improvement to the validity of the claims in this thesis if spoken data became available so that future research could be carried out on a more balanced data. Additionally, such a speech-writing balanced data can be used in investigations into the distribution of spoken and written idiomatic creativity, which will be of tremendous value to our current understanding of the phenomenon.

7.3.5. *Individual bias and mistakes*

The use of a mixture of corpus and internet data in this thesis, although beneficial in many ways, as I have discussed in chapter 4, is not free from limitations. The wide range of

sources from which the data was collected, while capable of capturing the heterogeneity of the phenomenon of idiomatic creativity, cannot distinguish between different groups of language users in terms of mother-tongue, location, educational background, and so on, all of which contribute to how they use idioms in context. Particularly, the differences in terms of idiomatic usage between different L1 users are sometimes discussed in studies that focus on the notion of 'speech community', claiming that the institutionalization of idioms is specific to the speech community they come from. It follows that the creative usage of idioms can only be correctly measured and described if the data is collected from within their particular speech community (in our case, British English idioms and the native speakers of British English speech community). As such, the heterogeneity of corpus and internet data, which include users from all over the world, is seen as working against this selection, thereby leading to potential concerns about the validity of any claim about idiomatic creativity made on this data pool.

While this criticism has its merits, I believe that the use of corpus and internet data does not necessarily mean unreliable analytical values. Admittedly, a more fine-grained selection of L1 users helps set the social boundaries within which the use of British English idioms can be described and accounted for, as opposed to different groups of L2 users from different backgrounds. However, this has not been my desire since the beginning of the thesis. Instead, I am interested in understanding the state of affairs in a wider setting, with the definition of 'speech community' being the community of English speaking individuals, regardless of their L1. In fact, the definition of 'speech community' in the literature has never been specific about the L1 requirement, for instance:

Early attempts to grapple with the speech community date back to the Prague School notions of *Sprechbund* or 'speech bond', focused on 'shared ways of speaking which go beyond language boundaries', and *Sprachbund* or 'language bond', involving "relatedness at the level of linguistic form" (Romaine 1994: 23, in Patrick, 1998)

A speech community made up of individuals who regard themselves as speaking the same language; it need have no other defining attributes (Corder, 1973: 53)

The speech community is defined by ... participation in a set of shared norms ... [which] may be observed in overt types of evaluative behavior, and [in] the uniformity of abstract patterns of variation" (Labov 1972: 120-1)

As such, on the basis that when a non-native speaker learns English, he/she also learns and acquires the same set of norms regarding the use of English as native speakers do, he/she can be considered a part of the English speech community, although the level of proficiency might differ significantly from user to user. Differing levels of proficiency, strictly speaking, can also be observed within different groups of L1 users as well depending on the

individual's educational background, field of interest, etc. As a result, linguists with more radical points of views such as Hudson (1996) or Duranti (1997) even argue that language is located in the individual rather than at the community level, thus rejecting the concept of 'speech community' all together, because 'our socio-linguistic world is not organised in terms of objective 'speech communities' (Hudson, 1996: 29). In the case of idiomatic creativity in particular and language production in general, the role of individual speaker perceptions is paramount and likely to differ from others, there can be no objective basis to any delimitation of a speech community (ibid).

Consequently, I did not take L1 as the criterion to define a specific speech community for analysis in this thesis but focus on 'the micro level of the individual person and the individual linguistic item' (ibid: 229). The corpus and internet data in this thesis, hence, does not have to be restricted to a group of native speakers but represents 'the product of the communicative activities engaged in by a given group of people' (Duranti, 1997: 82) who speaks English, regardless of their native tongue or any other attributes.

This said, comparative analyses between L1 and L2 users in terms of idiomatic usage are by no means undesirable and could become a productive direction in idiomatic creativity studies in the future. The individual and subjective usage of idioms as the result of differing personal backgrounds (L1, education, environment, etc.) can be further investigated in the future using psycholinguistic experiments and questionnaires/interviews to probe into the links between people's backgrounds and the way they manipulate it in context-specific communicative situations.

Due to this subjective nature of language use in general and creative idiom usage in particular, bias might incur during the analytical process of the highly subjective characteristics of the transparency of each particular idiom. Since transparency depends on the ways each speaker understands, stores, and processes idioms in his/her mind, it is expected to vary from individual to individual. In this thesis, idiom transparency has been determined from *my* cognitive representation of the idioms, but this does not necessarily coincide with that of the next speaker. For example, I claim that *force one's hand* is motivated and analysable via the understanding of the CARD GAME script, but if it is not perceived as such by other speakers, their usage of the idiom will be different. As a result, there is a certain amount of bias in this thesis due to the subjectivity of the features analysed.

The individual and subjective usage of idioms as such can be further investigated in the future using psycholinguistic experiments to probe into the link between the way people understand an idiom and the way they manipulate it in context-specific communicative situations.

Last but not least, the literature on idiomaticity, creativity and corpus development in language investigations, on which this thesis is based, is enormous. Out of necessity, I limit

myself to the most important and relevant works, and might therefore unintentionally miss some of the more up-to-date publications in the fields. This can only be improved with time as I progress along my career.

7.4. Related research possibilities

7.4.1. *Teaching idioms and idiomatic creativity*

The importance of teaching and learning idioms and idiomatic creativity has been emphasized in applied linguistic studies and pedagogical literature. For example, it is argued that the ability to master the meanings and usage of idioms (and their variants) has been one of the important requirements for efficient language users, such that 'people are not considered competent speakers of a language until they master the various cliché, idiomatic expressions' (Gibbs, 1995: 97). Idioms in particular and fixed expressions in general are therefore 'not only for efficient language use, they're essential for appropriate language use' (Schmitt, 2004: 10). Idiomatic creativity, as a type of language play, is another important aspect of English language acquisition that needs to be mastered. Being a counterbalancing force to the frequency-driven expectancies in language processing, linguistic creativity helps preventing fossilization, promoting permeability and openness to change in interlanguage (Tarone, 2002: 296). Also, as Cook (2000), Tarone (2000), and Sullivan (2000) point out, language play does not only encourage students to expand their vocabulary but also provides situations of authentic language use and can be seen as linguistic preparation for conversations outside the classroom. Language play, thus, is seen as a pedagogic tool that is intrinsically motivating and facilitates L2 learning (Cekaite and Aronsson, 2005) as well as L1 acquisition (Crystal, 1998) and thus should be brought into schools as soon as possible. Strongly criticizing current children's reading schemes for mistakenly leading children to conformity rather than creativity in language use, Crystal (*ibid*) argues that language play is one of the right directions for education which should be made a new movement in literacy policy.

Despite the importance of incorporating idioms and idiomatic creativity into the classroom, throughout the decades, idioms have consistently been mentioned in pedagogical studies as one of the most challenging items to teach and learn. In reality, they are also largely ignored in language curricular (O'Keeffe, McCarthy and Carter, 2007: 80). Nattinger and Deccarico (1992) for example, stress the importance of incorporating *lexical phrases* in the teaching of English but did not have a separate treatment of idioms from other fixed expressions. Further research in the direction of finding the ways to bring idioms and idiomatic creativity into school programmes will inevitably be beneficial for students in the long run. As an example of how to incorporate idioms into classroom activities to raise students' awareness and appreciation of this phenomenon, Steward (2005) suggests using a list of creative newspaper headlines containing aspects of idiomaticity or culture-specific references that are not readily accessible to non-native speakers. Students are then required

to use the BNC to work out the wordplay based on any departure from usual patterns shown in concordances. (See also Cook, 2000). Similarly, applying findings from corpus analyses to language teaching, O’Keeffe, McCarthy and Carter (2007) do not only offer practical guidance to help learners locate idioms in corpora, discuss the functions and meanings of idioms, but they also emphasize the importance of providing the contexts in which idioms typically occur to motivate learners because contexts contain clues to the meanings of idioms and idiom variations if any. Decontextualized, contrived examples of idioms should therefore be avoided.

The challenge of teaching them to native as well as non-native speaker of English deserves more extensive research in the field. Idiomaticity and idiomatic creativity and their place in education constitute an interesting research direction that should be further explored in future research.

7.4.2. *Translating idioms*

Research into the translation of idioms from one language to another has been few and far between, but attempted nevertheless (see Glucksberg, 2001). It has traditionally been acknowledged that translating idioms is notoriously difficult, mainly due to cultural differences surrounding the first language and the target language. For example, the French phrase *casser sa pipe* (to break one’s pipe) is equivalent to the English *kick the bucket*. Engel (1996), however, found that not all idioms are equally difficult to deal with. Transparent, analysable idioms are generally better and faster learned by second language learners than either opaque or oligosemic idioms.

Cross-cultural studies of idioms will illuminate cultural aspects of idioms, idiom variations and their processing routes, from which translation studies will undoubtedly benefit. Kovecses’ (2005) investigation of the cultural variations of metaphors includes a section of analyses of cross-cultural idioms, which serves as an excellent departure point for further research in this direction.

7.4.3. *Understanding socio-cultural movements with idioms*

The social significance of formulaic language is pursued in depth by Kuiper (2004) on the use of routine formulas before during and after China’s Cultural Revolution (Kuiper, 2004: 45-6). It was found that formulaic sequences after the Revolution were either ‘proscribed or altered to represent the new order’, reflecting the ideological and political changes during the Revolution. Kuiper suggests that there is a strong trend for young people to aim for, amongst many other things, conformity in terms of linguistic performance with their peers; and this was exploited for socio-political ends: ‘formulaic speech is not only sensitive to socio-cultural change but also easily manipulated by the powerful for socio-political ends’.

7.4.4. Idiomatic functions and their lexico-grammatical, phonological and nonverbal features

It has been noticed that the choice of use of idiomatic expressions over their literal neutral alternatives expresses more intense and more evaluative versions of the literal statements (Carter, 2004: 132; Strassler, 1982; Moon, 1998; McCarthy and Carter, 1994). Pragmatically, the interpersonal functions realised by idioms in particular and formulaic language in general include maintaining social interaction, discourse organization, and information transaction in an efficient way (Schmitt and Carter, 2004). Thanks to the set of conventionalized phatic phrases which are non-threatening and standardized, the issues of politeness and face can be addressed more effectively, allowing conversations to flow more easily, particularly under special occasions (expressing condolences, handling embarrassing situations, etc.) Since they are considered 'tools which individuals employ in order to relate to others in an accepted way' (Coulmas, 1981: 2) they help to form and maintain social bonds and the sense of belonging to the same speech community and culture. It is suggested that the specific function a formulaic sequence fulfils is potentially realized in different ways using distinctively different lexical and grammatical devices, accompanied by distinctive non-verbal behaviour. Further research in this direction has indeed started to gain popularity in recent years and should be further pursued.

Similarly, the distinctive phonological features associated with a particular function represent a real and interesting field of research. Certain idiomatic expressions function like discourse markers in speech and writing - they are used to summarise, close or side shift a topic (Holt and Drew, 1995, 2005; Drew and Holt, 1998), signalling problem-solution patterns, negotiating roles and identities and referring to speaker stance (Simpson, 2004; Koester, 2006). In other words, idiomatic expressions signal the talks/conversations for improved comprehension. Moreover, the use of idiomatic expressions in spoken language is found to create a higher degree of informality/intimacy (Carter, 2004). It is imagined that different functions might have their phonological features, i.e. speed, intonation, pauses, etc. This information would be beneficial to both speaker/writer and listener/reader during the production and reception process of a particular idiomatic expression for a particular function.

All in all

Although idioms make up only one, and not even the largest, part of the spectrum of fixed expressions, they have important roles and interesting behaviour in discourse. The study of idiomatic creativity in this thesis has revealed the flexibility of idioms in authentic English language usage. This flexibility is context-specific, and constrained to certain extents by different idiom internal and external factors depending on the specific communicative purposes of a particular use. It is hoped that the insights from such a context-based model

of idiomatic creativity inspires further research in the field towards a fuller context-oriented understanding of the phenomenon.

APPENDIX 1

Idiomatic uses in written texts
<p>Travel guide (16,084 words)</p> <ol style="list-style-type: none"> 1. <u>stop life in its speedy tracks</u> 2. a popular magnet for the island's <u>night owls</u>. 3. why not climb aboard a boat and <u>set sail</u> on a romantic dinner cruise 4. the playing field next to the shores is regularly <u>alive and kicking</u> with Bajan cricket 5. It doesn't solely[sic] cater for the <u>on-the-go</u> crowd 6. <u>set your sights</u> on the nearby Chattel House 7. Holidays in Hastings are a <u>far cry</u> from the <u>hustle and bustle</u> of the daily grind 8. Holidays in St James are <u>the cream of the crop</u> 9. Your[sic] <u>spoilt for choice</u> when it comes to choosing somewhere to stay 10. <u>fall for beach-beautiful Rockley – hook line and sinker</u>. 11. <u>put pen to paper</u> 12. your daily routine will be <u>turned on its head</u> 13. it'll <u>whip your breath away</u> 14. should you <u>loose[sic] your nerve</u>. 15. something to <u>keep in mind</u> when booking your Bangkok holiday 16. it's the same <u>hustle and bustle</u> of the city streets 17. haggling is <u>de rigueur</u> 18. everyone wants a <u>piece of the action</u> 19. enjoying a <u>new lease of life</u> with a wave of bars 20. Recent rocketing property prices <u>have breathed new life into</u> the area 21. Scarperia's <u>claim to fame</u> is the bone-handled pocket knives 22. <u>the rest, as they say, is history</u>. 23. stepping in out of <u>the bustle</u> of the market area 24. they are little more than a <u>shell of their former glory</u> 25. portraits of Cuba <u>tell a more complete tale</u> of the largest island in the Caribbean 26. but the Golden Age <u>came to an end</u> with the arrival of the Panama disease
<p>Satires (24,161 words)</p> <ol style="list-style-type: none"> 1. Amy Winchouse <u>fits like a glove</u> 2. the dude was <u>crying like a baby</u> 3. when he was still a <u>bona fide</u> black man 4. he can <u>turn back the clocks</u> 5. <u>carpet has been firmly swished</u> from underneath them 6. mortgaged <u>up to the hilt</u> 7. never <u>sold his soul</u> for humanitarian ends. 8. murdering them <u>in cold blood</u> 9. did not <u>sell his soul</u> in order to accommodate humanity. 10. I am <u>above the law</u> and I got away with it. 11. he could tinker with America's <u>tools of the trade</u>. 12. when someone with a <u>pair of huge brass balls</u> posted a list of their names on the web 13. racist cowards are not so courageous when <u>the spotlight is suddenly upon them</u>. 14. <u>Lap it up</u> BNP, <u>lap it up</u>. Enjoy your just desserts. 15. 'Home is where the <u>dungeon is</u>' 16. Prince William is <u>setting sail</u> on a long Caribbean Navy cruise this summer 17. the future king to <u>ease the brakes</u> on Kate's constant endeavours to close the young prince off and <u>force his hand</u> into marriage. 18. prince William who is being <u>put into a corner</u> 19. because of the war that is <u>in progress</u> at the moment. 20. a privilege to have Prince William with the Royal Navy; <u>not a pain in the arse</u> 21. the ailing fantasist <u>living in cloud-cuckoo-land</u>, JK Rowling. 22. a ministry which <u>has lined its own pockets</u> with sleaze, scandal and pure lies. 23. they're all <u>crying their eyes out</u> 24. in a bid to "<u>keep the lid on</u> all the crimes coming from the front office". 25. there are Americans, of every race, every hue, and every culinary disposition who want Obama to succeed because <u>he eats the food and farts the fart</u> 26. they just <u>lost it</u>

27. he just wouldn't work that blue
28. airport security captured Iowa Senator Larry Craig tapping boots with Iranian president
29. this incident may shed new light on the U.S. visit
30. Supreme Court Justice Anthony Kennedy was found to have a Magic 8 Ball within the sleeves of his judicial vestments
31. It really sets my mind at ease
32. Donny has always been a sad boy, but he seems to be turning a corner
33. have been on strike for nearly three months.
34. New Year revellers will be putting themselves at risk
35. Migrant Watch report was to be read with a pinch of salt.
36. all this success has unfortunately gone to Mr Prescott's head
37. he doesn't have a sodding clue whether he's coming or going.
38. crime-fighting agencies crawling out of the woodwork
39. a careerist timeserver who swings with the political wind
40. some news that sneaked out largely unnoticed amid the smoke billowing from David Cameron's alleged erstwhile misdemeanours
41. Tony finally takes the hint and leaves Number 10
42. the top dogs at the Labour party have tried to limit the damage
43. Prescott urged everyone to give themselves a pat on the back
44. unaware that he was serving only to make a tit of himself
45. that could be construed as making a public scene.
46. the public hanging of John Prescott to get the ball, or rather the head, rolling
47. All this and entertaining to boot: what more could you want?
48. The Laptop Herring. All this focus on laptops as the problem is a red herring
49. Green Herring: Green herring is an environmental rhetoric and discourse that legitimates the capitalist world-system.
50. No turn unstoned for pens that are mightier than the sword
51. Jarvis wants to turn back the clock, stuff the cat back in the bag, repack Pandora's box - choose your metaphor - by the public act of returning his iPad to the Fifth Avenue Apple Store.
52. We pay the piper; but we don't call the tune! The fortunes we shell out on banking fees, internet connection charges, cell phone calls and municipal rates buy little respect from the companies providing the so-called service.
53. A recent 'bucket list' adventure holiday has provided a terminally ill Auckland woman with memories she hopes will sustain her family through the final stages of her disease
54. he would get this restriction on use lifted, but only after the transfer of funds, since he would need to utilize some of them to grease the bureaucratic wheels. The wheels were never greased and the counterparty is now nowhere to be found.

Short stories (13,796 words)

1. Fate had played an evil trick
2. she could hear her other grandchildren running amok through the big, century old home
3. but I don't have a deep sleep due to having to keep an ear open for on coming foot and motor traffic
- 4 + 5. One day I will stand tall and hold my head up
6. I stopped killing of my own accord
7. My heart skips a beat every time I think about seeing you again.
8. I lacked the nerve to explain my charitable spirit to my wife

Entertainment/ Celebrity news (21,177 words)

1. when it comes time for the guild to cut a new deal
2. The SAG Awards itself is decidedly lower key than the Golden Globes
3. and we're having a ball doing it
4. According to the Golden Globes givers and the Golden Globes host, all is well
5. Certainly, in this case, he pushed the envelope and occasionally went too far,
6. there are no hard feelings between him and Hollywood
7. he has a good head on his shoulders
8. they missed the big picture

9. Bruno Mars would like to clean off his slate at least a little bit before the Grammys
10. his record will be wiped clean
11. the possibility of prison remains if he somehow runs afoul of the law in the next year
12. Natalie Portman sure seemed to have fallen prey to dark forces at the box office
13. their journeys are more about exploring possibilities, their paths are less set in stone
14. It's another African team who bite the African dust
15. The door remained locked on Cameron's closet
16. I wanted to kill two birds with one purse. The bag should not be black. It should be messenger- or courier-style
17. It was raining cats and dogs and then I stepped in a poodle
18. Whatever, nonbelievers, I say these control pants are magic, they leave no stern untuned
19. Our 1,000 subscribers will be happy to let these priceless pearls be shared with your millions of readers
20. Holly Willoughby looks a proper darlin' in her clementine-coloured frock. Let's hope Holly had fun and painted the town, er, orange...
21. jump on board this bargain bandwagon
22. With just one win in the bag and just two weeks to go, she's pulling out all the fashion stops
23. Americans have been suffering with the foot-in-mouth disease
24. Leaving no scone unspurned
25. She took it like a squirrel to water
26. Here's a bud that needs nipping
27. So, waffles for your thought?
28. It's a great feeling when you suddenly find that the cloud is silver-lined.
29. Vanessa Minnillo & Nick Lachey: Baby In the Works?
30. Sheen's rep is hoping to keep details under wraps
31. Since Charlie's latest incident, filming for his CBS comedy has been brought to a halt
32. Brittany Murphy's home auction put on hold
33. The former "Hills" babe has just released a new single called "Heartbeat," though it's not likely to raise many pulses.
34. I must say the Super Bowl is a dream come true
35. Charlie's dad Martin Sheen is beside himself with worry
36. Charlie is already treading on this ice [alcohol abuse and prostitutes] with CBS execs due to previous scandal-ridden tales coming from "bender" trips to NYC and Las Vegas
37. January Jones and Jason Sudeikis Call It Quits
38. So many people would kill to be in my shoes
39. It seems that he's popped the lock on the closet and come out
40. Once finished up, Paris was ready to hit the sack
41. Jess created a bit of a stir among film directors and writers
42. two billion people watching her tie the knot with Prince William this spring
43. The bride-to-be is getting into top shape
44. Jennfier Aniston: No Adoption in the Works
45. the bundle of joy comes after the "Vicky Cristina Barcelona" co-star couple tied the knot during a secret ceremony in July 2010.
46. She's been overjoyed to have a bun in the oven, and Jewel has announced that she's having a boy later this year.

Academic essays (23,069 words)

1. in terms of age
2. lend itself to intensification
3. in terms of cohesion
4. It is also important to bear in mind that
5. in terms of music
6. in terms of architecture
7. avoid the the pitfalls of the bird's eye view

APPENDIX 2

Idiomatic uses in spoken texts
<p>Scripted comedy show 'The Office' (9,280 words)</p> <ol style="list-style-type: none"> 1. I can make <u>that dream</u> come true 2. you talk the talk but <u>do not</u> walk the walk 3. every bloke here has woken up at the <u>crack</u> of <u>Dawn!</u> 4. c'est la vie 5. Not to her face. 6. <u>you're pulling</u> together to do something 7. present company excepted 8. can we keep a lid on <u>this</u> 9. I couldn't give a shit. 10. <u>Into</u> the fray 11. <u>it</u> (paper) doesn't grow on trees 12. Can't put a price on comedy 13. Gareth's my right hand 14. what source your little bird is from 15. to put the record straight- 16. it will be out of your hands 17. it won't be out of my hands 18. it could be this branch that <u>gets</u> the chop 19. I could <u>teach</u> her the ropes 20. I don't give a monkey's 21. out of bounds 22. worse for morale in the long run 23. <u>Makes</u> sense 24. he's gonna be hard by the book 25. it's like shitting on your <u>own</u> doorstep 26. <u>opening</u> a can of worms there 27. you back me into a corner 28. a word with you now 29. that was out of order 30. you're on <u>very</u> thin ice as it is 31. <u>you're taking</u> the piss
<p>Improvised comedy 'For your consideration' (13,757 words)</p> <ol style="list-style-type: none"> 1. make <u>such</u> a big deal out of it 2. complete waterworks 3. you shouldn't just keep it to yourself 4. I've <u>done all my</u> homework on this guy 5. I don't give <u>two hoots</u> what Dr Silverstein says 6. that doesn't make sense 7. ring a bell? 8. I'm gonna put you <u>across</u> my knee 9. it's a breath of fresh air out here 10. the long and the short of it is some bloke got on the set 11. an Academy Award nomination in the offing 12. it's a big deal 13. it's no big deal 14. you straightened me out 15. you're a sight for sore eyes 16. nice of you to keep <u>my sister</u> company 17. that just don't make <u>no</u> sense 18. I don't want to throw the baby out with the bath water 19. you've lost your minds 20. you need to light a fire under <u>an actor</u> 21. Can I have an order of humble pie 22. she <u>lets</u> it all go

23. I'm a peeled nerve
24. not a big deal
25. plant the seeds of new life
26. dust off the mantle
27. you're off to the races
28. this ship's turning round
29. this is on the tip of my tongue
30. I should be so lucky
31. Marilyn made quite a splash
32. they are not only alive and well
33. this hot little movie that seems to be creating a stir
32. you're so full of yourself
35. I see right through you
36. babes in the woods
37. there'll be babes in the woods and in the trailers
38. don't shove it down people's thro-
39. I don't shove it down your throat
40. the dogs bark, but the caravans move on.
41. with the Oscars around the corner
42. it turned this town on its ear
43. it's a dream come true
44. I've got morning wood this morning
45. that's the reason for my wood
46. the director calls the shots
47. we're gonna pump this beat now
48. we've watched this man's ball drop ignominiously
49. I was rooting for you
50. Let's sneak inside like a fly on the wall
51. You're striking while the iron is cold
52. to give us inside track on this year's favourites

TV Interviews (30,000 words)

1. I've put my foot down
2. it gets in the way of being funny
3. you've lost your minds
4. he's in his head
6. my ears pricked up
7. people didn't bat an eyelid
8. I'd take my chances here
9. you get out of your head
10. it's been a stumbling block*
11. ego gets in the way
12. I'll never have to pull another set of balls like that again*
13. Jeffrey Tambor really working hard under the water
14. he was in the show, out of the gate
15. just keep it in mind
16. you wanna throw your weight around
17. that's the deal that we struck
18. this limo is getting wolf whistle
19. spick and span
20. to fish for compliments
21. put the light on me sergeant
22. here's another nice kettle of fish you pickled me in
23. he was gonna keep it in the dark
24. I didn't keep my mouth shut
25. nobody'd be any the wiser
26. keep your trap shut
27. samie's shell has been on my mind
28. you get something off your chest
29. lying about keeping things to myself

30. that was completely made up on the spot
31. the weight of the world on your shoulders
32. that makes my blood boils
31. many people couldn't put a face to the name
32. it's as good as it gets
33. that's the top of the tree
34. (the) top of the line
35. right across the board
36. not really much to call my own
37. that used to drive my mother crazy
38. it made sense
39. we need that extra push over the cliff
40. take my word for it
41. that could have gone around in circles for fifteen minutes
42. hoping the little man would snap out of it in the end
43. that's obviously through words of mouth
44. I'm a dark horse
45. *I don't kiss and tell*
46. you don't normally kiss you've got nothing to tell
47. I don't shove it down your throat
48. you jumped to a conclusion
49. this is a monthly at ease weekend
50. a chance to let down our hair
51. you all have a head start
52. I got a real rain wagon
53. this made all the sense in the world
54. the ball's in the enemy's court
55. it's that thing that's dangerous around the corner
56. you're not thinking things through
57. don't have a cow mate
58. just try and think things through
59. i don't want to throw the baby out with the bath water

Tourist commentaries (16,000 words)

1. doesn't the word careless springs to mind
2. this sort of set the tone for the next 15 years
3. he never actually broke and entered
4. the politicians who repealed that tax called it daylight robbery
5. I'd almost (0.2) almost put money on it that you'll be going straight to that
6. he was righ- Hitler's right hand man
7. he passed the point of no return

Supervision meetings (30,000 words)

1. I lost the plot towards the end
2. I was going to make an absolute fool of myself
3. pull my finger out
4. I'm a one-man band, aren't i?
5. you can reel it off
6. it gets you off the hook
7. you make a virtue of the necessity
8. completely out of the window
9. that's on the right lines
10. out of place
11. first time I've even met him in the flesh
12. I lost the thread of the argument
13. what brought it home is that
14. after our talk finished everyone upped sticks
15. it's nice to have a few under the belt
16. We like to sort of fl= fly the flag

17. you've set the scene
18. we touched upon er the pros and cons of of of corpus work
19. your ears had pricked up
20. firing paper bullets
21. ring alarm bells

APPENDIX 3: IDIOMS DATA SHEET

Idioms	Phrase type	Motivation	Type of Motivation	Analysability	Literal well-formedness	Transparency type	Number of content words	Number of phrase	Invariable usage	Tokens	Types	Types including web
<i>alive and kicking</i>	AdjP	Full	Conceptual metonymy + etymology	Analysable	Well-formed		1 2	2	26	5	3	8
<i>be spoilt for choice</i>	AdjP	Partial constituent	Literal meaning	Analysable	Ill-formed		2 2	3	60	1	1	6
<i>full of one's self</i>	AdjP	Partial global	Conceptual metaphor	Unanalysable	Ill-formed		6 1	2	17	0	0	2
<i>spick and span</i>	AdjP	Unmotivated	None	Unanalysable	Ill-formed		8 2	2	25	1	1	2
<i>a little bird told me</i>	Cl	Partial constituent	Conceptual metaphor + general knowledge	Partial analysable	Well-formed		1 3	4	3	5	4	6
<i>that ship has sailed</i>	Cl	Full	Conceptual metaphor + metonymy	Analysable	Ill-formed		2 2	2	0	0	0	3
<i>the top of the tree/line</i>	NP	Full	Conceptual metaphor	Analysable	Well-formed		1 2	2	13	10	3	3
<i>a night owl</i>	NP	Global	Conceptual metaphor + emblem	Analysable	Well-formed		1 2	2	19	9	1	2
<i>a dark horse</i>	NP	Constituential	Conceptual metaphor + etymology	Analysable	Well-formed		1 2	2	20	13	1	2
<i>inside track</i>	NP	Full	Conceptual metaphor + general knowledge	Analysable	Well-formed		1 2	2	24	0	0	0
<i>a babe in the woods</i>	NP	Full	Conceptual metaphor + emblem	Analysable	Well-formed		1 2	3	1	9	5	5
<i>a far cry from st</i>	NP	Constituential	Conceptual metaphor + etymology	Partial analysable	Well-formed		1 2	3	178	0	0	2
<i>a fly on the wall</i>	NP	Global	Conceptual metaphor + general knowledge	Partial analysable	Well-formed		1 2	3	8	3	3	6
<i>a breath of fresh air</i>	NP	Full	Conceptual metaphor + lexicalised figurative meaning	Analysable	Well-formed		1 3	2	88	2	3	6

<i>someone's right hand man/tuoman</i>	NP	Full	Conceptual metaphor + metonymy + etymology + general knowledge	Analysable	Well-formed		1	3	2	25	0	0	0
<i>a sight for sore eyes</i>			Conceptual metaphor + metonymy + lexicalised figurative meaning										
<i>a wolf whistle</i>	NP	Full		Analysable	Well-formed		1	3	3	8	1	1	4
<i>a stumbling block</i>	NP	Constituent	Etymology + emblem	Isomorphic	Ill-formed		2	2	2	7	9	1	2
	NP	Partial constituent	Conceptual metaphor + etymology	Partial analysable	Ill-formed		2	2	2	94	45	1	5
<i>tools of the trade</i>	NP	Constituent	Conceptual metaphor	Analysable	Ill-formed		2	2	3	19	15	4	5
<i>the cream of the crop</i>	NP	Constituent	Conceptual metaphor + lexicalised figurative meaning + etymology	Analysable	Ill-formed			2	3	9	1	1	5
<i>a fine/pretty kettle of fish</i>	NP	Unmotivated	None	Partial analysable	Ill-formed		4	3	2	2	9	1	5
<i>the long and the short of it</i>	NP	Unmotivated	None	Unanalysable	Ill-formed		8	2	3	6	0	0	2
<i>(up) to the hilt</i>	PrepP	Global	Conceptual metaphor	Analysable	Well-formed		1	1	2	48	2	2	3
<i>in the spotlight</i>													
<i>out of bounds</i>	PrepP	Full	Conceptual metaphor + general knowledge	Analysable	Well-formed		1	1	2	37	55	6	7
			Conceptual metaphor + lexicalised figurative meaning										
<i>out of one's hands</i>	PrepP	Full	Conceptual metaphor + metonymy	Analysable	Well-formed		1	1	2	0	4	2	5
<i>in the long/short run</i>	PrepP	Full	Conceptual metaphor	Analysable	Well-formed		1	1	2	0	4	1	1
<i>in cold blood</i>	PrepP	Full	Conceptual metaphor + metonymy	Analysable	Well-formed		1	2	3	637	23	4	6
<i>in a pickle</i>	PrepP	Full		Partial analysable	Ill-formed		2	2	3	46	3	2	7
<i>by the book</i>	PrepP	Unmotivated	None	Analysable	Well-formed		3	1	2	5	4	4	8
<i>on one's doorstep</i>	PrepP	Global	Conceptual metaphor	Unanalysable	Well-formed		4	1	2	60	6	4	4
							5	1	2	154	8	3	3

<i>around the corner</i>	PrepP	Global	Conceptual metaphor	Unanalysable	Well-formed	5	1	2	207	37	1	1
<i>to one's face</i>	PrepP	Full	Conceptual metaphor + metonymy	Unanalysable	Well-formed	5	1	2	50	3	2	3
<i>out of the (starting) gate</i>	PrepP	Global	Conceptual metaphor + etymology	Unanalysable	Well-formed	5	1	3	0	0	0	4
<i>off to the races</i>	PrepP	Unmotivated	None	Unanalysable	Ill-formed	8	1	3	2	0	0	2
<i>in the offing</i>	PrepP	Unmotivated	None	Unanalysable	Ill-formed	8	1	2	68	0	0	0
<i>none the wiser</i>	N/A	Partial constituent	Literal meaning	Unanalysable	Ill-formed	6	1	1	31	10	3	4
<i>plant/sow the seeds of st</i>	VP	Full	Conceptual metaphor	Analysable	Well-formed	1	2	2	42	22	7	8
<i>make a splash</i>	VP	Global	Conceptual metaphor	Analysable	Well-formed	1	2	2	26	66	5	7
<i>get the chop</i>	VP	Full	Conceptual metaphor + lexicalised figurative meaning	Analysable	Well-formed	1	2	2	14	1	1	3
<i>do one's homework</i>	VP	Full	Conceptual metaphor + lexicalised figurative meaning	Analysable	Well-formed	1	2	2	159	21	7	7
<i>turn on the waterworks</i>	VP	Full	Conceptual metaphor + lexicalised figurative meaning + general knowledge	Analysable	Well-formed	1	2	2	1	1	1	2
<i>turn back the clock</i>	VP	Full	Conceptual metaphor + metonymy	Analysable	Well-formed	1	2	2	23	86	5	6
<i>join/enter the lists</i>	VP	Full	Conceptual metaphors + etymology	Analysable	Well-formed	1	2	2	29	24	2	6
<i>ring a bell</i>	VP	Global	Etymology	Analysable	Well-formed	1	2	2	35	15	3	7
<i>kiss and tell</i>	VP	Constituent	Etymology + emblem	Analysable	Well-formed	1	2	2	11	5	1	3
<i>line one's (own) pockets</i>	VP	Partial constituent	Conceptual metaphor + metonymy + etymology	Partial analysable	Well-formed	1	2	2	15	12	5	7
<i>put the brakes on st</i>	VP	Full	Conceptual metaphor	Analysable	Well-formed	1	2	3	19	57	5	5
<i>light a fire under someone</i>	VP	Global	Conceptual metaphor + emblem	Analysable	Well-formed	1	2	3	0	0	0	4

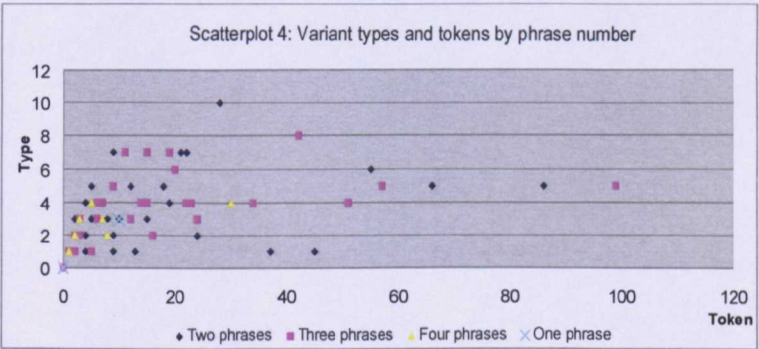
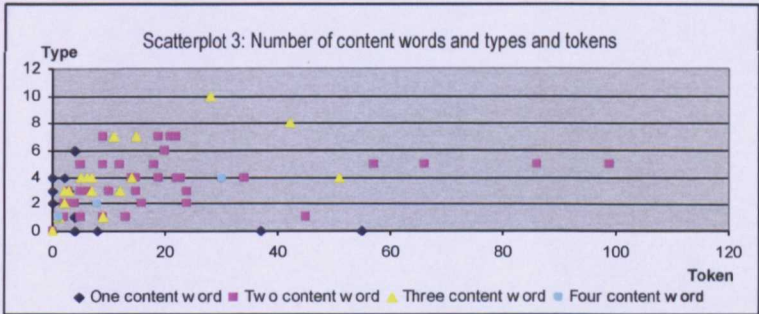
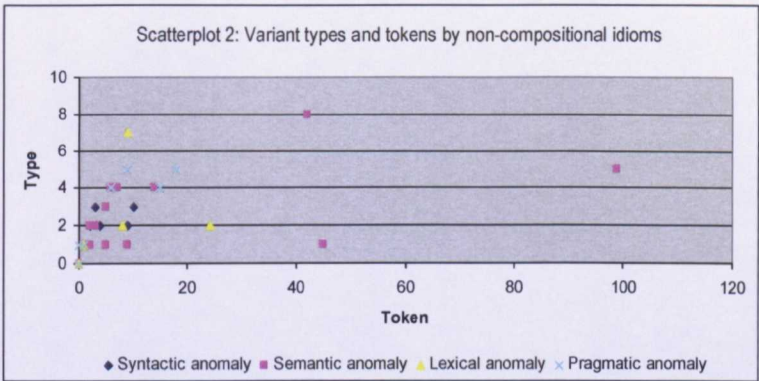
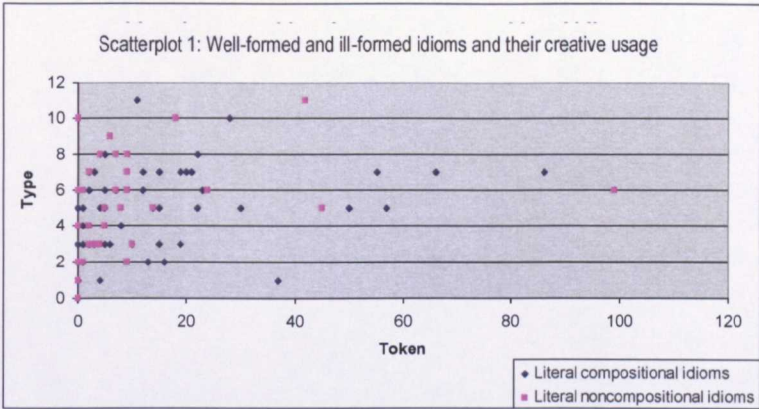
<i>back someone into a corner</i>	VP	Global	Conceptual metaphor + general knowledge	Analysable	Well-formed		1	2	3	5	22	4	5
<i>keep a lid on st</i>	VP	Full	Conceptual metaphor + metonymy	Analysable	Well-formed		1	2	3	7	19	7	7
<i>come/ crawl out of the woodwork</i>	VP	Constituential	Conceptual metonymy + etymology	Analysable	Well-formed		1	2	3	0	2	1	6
<i>have st up one's sleeves</i>	VP	Full	Conceptual metonymy + etymology + general knowledge	Partial analysable	Well-formed		1	2	3	45	34	4	6
<i>open (up) a can of worms</i>	VP	Full	Conceptual metaphor + emblem	Analysable	Well-formed		1	3	2	5	28	10	10
<i>roll out the red carpet</i>	VP	Full	Emblem	Partial analysable	Well-formed		1	3	2	8	2	3	6
<i>skating/ walking on thin ice</i>	VP	Global	Conceptual metaphor	Analysable	Well-formed		1	3	3	15	11	7	11
<i>start/set the ball rolling</i>			Conceptual metaphor + emblem + lexicalised figurative meaning										
<i>strike while the iron's hot</i>	VP	Global	Conceptual metaphor + general knowledge	Analysable	Well-formed		1	3	3	41	15	7	7
<i>pad (someone) on the back</i>	VP	Global	Conceptual metaphor + general knowledge	Analysable	Well-formed		1	3	3	1	0	0	5
<i>throw the baby out with the bathwater</i>	VP	Global	Conceptual metonymy	Partial analysable	Well-formed		1	3	3	37	51	4	5
<i>see which way the wind blows</i>	VP	Full	Conceptual metaphor + metonymy	Analysable	Well-formed		1	3	4	2	7	3	6
<i>sell one's soul (to the devil)</i>	VP	Constituential	Conceptual metaphor + metonymy	Analysable	Well-formed		1	4	4	1	30	4	5
<i>turn the/a corner</i>	VP	Full	Conceptual metaphor + etymology	Analysable	Ill-formed		2	2	2	44	5	3	4
	VP	Global	Conceptual metaphor + etymology + general knowledge	Analysable	Ill-formed		2	2	2	27	18	5	9

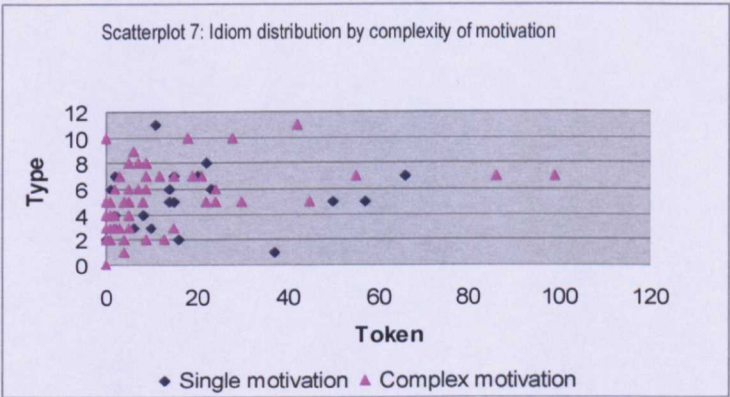
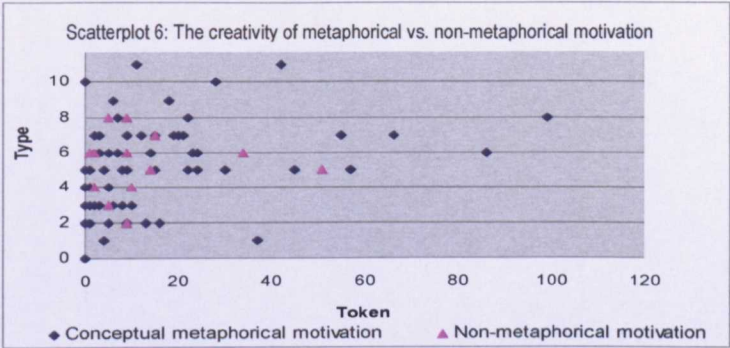
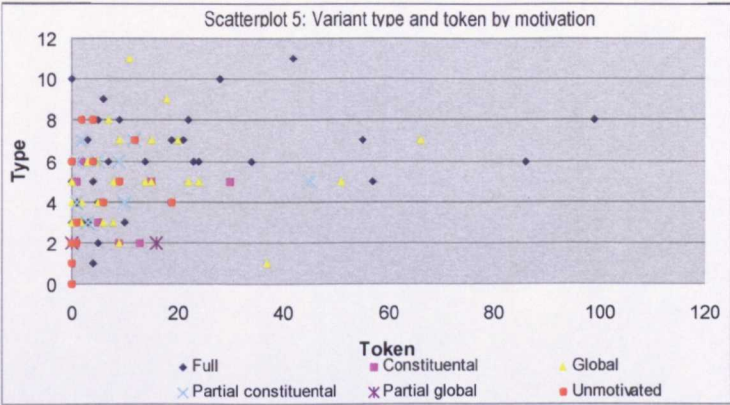
<i>lose one's nerve</i>	VP	Full	Conceptual metaphor + lexicalised figurative meaning	Analysable	Ill-formed		2	2	2	2	58	5	1	2
<i>lose one's mind</i>	VP	Full	Conceptual metaphor + metonymy	Analysable	Ill-formed		2	2	2	38	0	0	0	2
<i>show/teach someone the ropes</i>	VP	Full	Lexicalised figurative meaning + etymology	Analysable	Ill-formed		2	2	2	33	9	7	8	
<i>learn/know the ropes</i>	VP	Partial constituent	Lexicalised figurative meaning + etymology	Analysable	Ill-formed		2	2	2	45	9	5	6	
<i>call the shots</i>	VP	Full	Conceptual metaphor + emblem	Analysable	Ill-formed		2	2	2	189	99	5	8	
<i>shed/throw light on something</i>	VP	Full	Conceptual metaphor + metonymy + etymology	Analysable	Ill-formed		2	2	2	0	0	0	10	
<i>think outside the box</i>	VP	Full	Conceptual metaphor + metonymy + etymology	Analysable	Ill-formed		2	2	2	16	2	2	7	
<i>throw one's weight around</i>	VP	Partial constituent	Conceptual metaphor	Partial analysable	Ill-formed		2	2	2	39	5	1	4	
<i>cry one's eyes out</i>	VP	Global	Conceptual metaphor + metonymy	Partial analysable	Ill-formed		2	2	2	94	14	4	6	
<i>put/set the record straight</i>	VP	Full	Conceptual metaphor	Analysable	Ill-formed		2	3	3	5	0	0	5	
<i>have the brass balls/ neck to do st</i>	VP	Full	Conceptual metaphor + metonymy	Analysable	Ill-formed		2	3	3	7	7	4	8	
<i>make one's blood boil</i>	VP	Global	Conceptual metaphor + metonymy	Analysable	Ill-formed		2	3	3	9	42	8	11	
<i>touch/hit a raw nerve</i>	VP	Full	Conceptual metaphor + metonymy + general knowledge	Partial analysable	Ill-formed		2	3	3	6	3	3	3	
<i>put one's mind at ease</i>	VP	Partial constituent	Conceptual metaphor + metonymy	Analysable	Ill-formed		2	3	4	9	8	2	5	
<i>live in cloud-cuckoo land</i>	VP	Global	Conceptual metaphor + etymology	Analysable	Ill-formed		2	4	4	9	8	2	5	
<i>talk the talk and walk the walk</i>	VP	Partial constituent	Conceptual metaphor + lexicalised figurative meaning	Partial analysable	Ill-formed		2	4	4	0	1	1	4	

<i>make a scene</i>	VP	Unmotivated	None		Analysable	Well-formed		3	2	2	27	19	4	4
<i>lose it</i>	VP	Unmotivated	None		Analysable	Ill-formed		4	1	2	6	0	0	1
<i>rain cats and dogs</i>	VP	Unmotivated	None		Partial analysable	Ill-formed		4	3	3	3	0	0	6
<i>have a cow</i>	VP	Global	Conceptual metaphor + general knowledge		Unanalysable	Well-formed		5	2	2	2	0	0	3
<i>not bat an eyelid/ eyelash</i>	VP	Global	Conceptual metaphor + metonymy		Unanalysable	Well-formed		5	2	2	46	5	5	6
<i>force one's hand</i>	VP		Conceptual metaphor + metonymy + etymology											
<i>get st off your chest</i>	VP	Global			Unanalysable	Well-formed		5	2	2	20	15	4	5
<i>ram st down one's throat</i>	VP	Global	Conceptual metaphor		Unanalysable	Well-formed		5	2	3	36	6	3	3
<i>stop someone/st in their tracks</i>	VP	Global	Conceptual metaphor		Unanalysable	Well-formed		5	2	3	13	20	6	7
<i>turn st on its head</i>	VP	Partial global	Conceptual metaphor		Unanalysable	Well-formed		5	2	3	10	16	2	2
<i>put one's foot down</i>	VP	Global	Conceptual metaphor + metonymy + general knowledge		Unanalysable	Well-formed		5	2	3	87	24	3	5
<i>keep in mind</i>	VP	Global	Conceptual metonymy		Unanalysable	Well-formed		5	2	3	52	14	4	5
<i>eat humble pie</i>	VP	Full	Conceptual metaphor + metonymy		Unanalysable	Ill-formed		6	2	3	156	2	2	3
<i>put/set pen to paper</i>	VP	Full	Conceptual metaphor + etymology		Unanalysable	Ill-formed		6	3	3	13	6	4	9
<i>fall for st/sh hook line and sinker</i>	VP	Global	Conceptual metonymy		Unanalysable	Ill-formed		6	3	4	46	2	2	4
<i>not give a hoot</i>	VP	Global	Conceptual metaphor + etymology		Unanalysable	Ill-formed		6	4	2	6	9	2	7
<i>take st with a pinch/ grain of salt</i>	VP	Unmotivated	None		Unanalysable	Well-formed		7	2	2	20	2	2	8
<i>to boot</i>	VP	Unmotivated	None		Unanalysable	Well-formed		7	3	3	32	12	3	7
<i>take the piss</i>	VP	Unmotivated	None		Unanalysable	Ill-formed		8	1	1	5	0	0	0
	VP	Unmotivated	None		Unanalysable	Ill-formed		8	2	2	122	1	1	3

<i>not give a shit</i>	VP	Unmotivated	None		Unanalysable	Ill-formed	8	2	2	65	4	2	6
<i>turn/set st on its ear</i>	VP	Unmotivated	None		Unanalysable	Ill-formed	8	2	3	1	0	0	6
<i>go to one's head</i>	VP	Unmotivated	None		Unanalysable	Ill-formed	8	2	3	37	2	1	3
Total										3962	1214		

APPENDIX 4: SCATTERPLOTS





APPENDIX 5: MANN-WHITNEY TESTS
FOR RANKED GROUPS OF TRANSPARENCY AND
VARIANT TYPE AND TOKEN

Test Statistics ^b		
	Variant type	Variant token
Mann-Whitney U	24.000	24.000
Wilcoxon W	69.000	69.000
Z	-2.217	-2.202
Asymp. Sig. (2-tailed)	.027	.028
Exact Sig. [2*(1-tailed Sig.)]	.034 ^a	.034 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of three factors 1 2		

Test Statistics ^b		
	Variant type	Variant token
Mann-Whitney U	61.500	53.000
Wilcoxon W	106.500	98.000
Z	-3.010	-3.215
Asymp. Sig. (2-tailed)	.003	.001
Exact Sig. [2*(1-tailed Sig.)]	.002 ^a	.001 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of three factors 1 3		

Test Statistics ^b		
	Variant type	Variant token
Mann-Whitney U	54.000	49.500
Wilcoxon W	99.000	94.500
Z	-3.338	-3.427
Asymp. Sig. (2-tailed)	.001	.001
Exact Sig. [2*(1-tailed Sig.)]	.000 ^a	.000 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of three factors 1 4		

Test Statistics ^a		
	Variant type	Variant token
Mann-Whitney U	183.000	171.500
Wilcoxon W	261.000	249.500
Z	-1.036	-1.289
Asymp. Sig. (2-tailed)	.300	.197
a. Grouping Variable: Combination of three factors 2 3		

Test Statistics ^a		
	Variant type	Variant token
Mann-Whitney U	165.500	158.500
Wilcoxon W	243.500	236.500
Z	-1.728	-1.866
Asymp. Sig. (2-tailed)	.084	.062
a. Grouping Variable: Combination of three factors 2 4		

Test Statistics ^a		
	Variant type	Variant token
Mann-Whitney U	677.000	689.500
Wilcoxon W	1418.000	1430.500
Z	-1.012	-.880
Asymp. Sig. (2-tailed)	.312	.379
a. Grouping Variable: Combination of three factors 3 4		

APPENDIX 6: MANN-WHITNEY TESTS
FOR INDEPENDENT GROUPS OF TRANSPARENCY AND
VARIANT TYPE AND TOKEN

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	49.500	54.000
Wilcoxon W	94.500	99.000
Z	-3.427	-3.338
Asymp. Sig. (2-tailed)	.001	.001
Exact Sig. [2*(1-tailed Sig.)]	.000 ^a	.000 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 1 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	9.500	8.000
Wilcoxon W	54.500	53.000
Z	-2.140	-2.344
Asymp. Sig. (2-tailed)	.032	.019
Exact Sig. [2*(1-tailed Sig.)]	.036 ^a	.026 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 6 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	38.500	45.000
Wilcoxon W	83.500	90.000
Z	-2.845	-2.614
Asymp. Sig. (2-tailed)	.004	.009
Exact Sig. [2*(1-tailed Sig.)]	.004 ^a	.010 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	1.500	.500
Wilcoxon W	46.500	45.500
Z	-1.863	-2.128
Asymp. Sig. (2-tailed)	.062	.033
Exact Sig. [2*(1-tailed Sig.)]	.073 ^a	.036 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 7 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	.500	.000
Wilcoxon W	45.500	45.000
Z	-2.112	-2.253
Asymp. Sig. (2-tailed)	.035	.024
Exact Sig. [2*(1-tailed Sig.)]	.036 ^a	.036 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 3 and 8)		

Test Statistics ^a		
	Variant token	Variant type
Mann-Whitney U	402.500	405.500
Wilcoxon W	702.500	705.500
Z	-1.220	-1.190
Asymp. Sig. (2-tailed)	.223	.234
a. Grouping Variable: Combination of characteristics (group 1 and 2)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	14.000	16.500
Wilcoxon W	59.000	61.500
Z	-2.897	-2.732
Asymp. Sig. (2-tailed)	.004	.006
Exact Sig. [2*(1-tailed Sig.)]	.003 ^a	.006 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 5 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	40.500	30.000
Wilcoxon W	901.500	891.000
Z	-.029	-.641
Asymp. Sig. (2-tailed)	.977	.521
Exact Sig. [2*(1-tailed Sig.)]	.977 ^a	.567 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 1 and 3)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	43.000	40.000
Wilcoxon W	53.000	50.000
Z	-1.560	-1.694
Asymp. Sig. (2-tailed)	.119	.090
Exact Sig. [2*(1-tailed Sig.)]	.128 ^a	.099 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 1 and 4)		

Test Statistics ^a		
	Variant token	Variant type
Mann-Whitney U	245.500	219.500
Wilcoxon W	323.500	297.500
Z	-.011	-.569
Asymp. Sig. (2-tailed)	.992	.569
a. Grouping Variable: Combination of characteristics (group 1 and 5)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	82.000	91.500
Wilcoxon W	103.000	112.500
Z	-1.310	-1.014
Asymp. Sig. (2-tailed)	.190	.311
Exact Sig. [2*(1-tailed Sig.)]	.202 ^a	.324 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 1 and 6)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	33.500	34.000
Wilcoxon W	36.500	37.000
Z	-.433	-.408
Asymp. Sig. (2-tailed)	.665	.683
Exact Sig. [2*(1-tailed Sig.)]	.678 ^a	.718 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 1 and 7)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	17.000	13.000
Wilcoxon W	317.000	313.000
Z	-.676	-1.076
Asymp. Sig. (2-tailed)	.499	.282
Exact Sig. [2*(1-tailed Sig.)]	.554 ^a	.345 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 3)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	33.500	29.000
Wilcoxon W	43.500	39.000
Z	-.959	-1.272
Asymp. Sig. (2-tailed)	.338	.203
Exact Sig. [2*(1-tailed Sig.)]	.355 ^a	.235 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 4)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	115.000	129.000
Wilcoxon W	415.000	429.000
Z	-.977	-.510
Asymp. Sig. (2-tailed)	.329	.610
Exact Sig. [2*(1-tailed Sig.)]	.344 ^a	.631 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 5)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	64.500	72.000
Wilcoxon W	85.500	93.000
Z	-.391	.000
Asymp. Sig. (2-tailed)	.696	1.000
Exact Sig. [2*(1-tailed Sig.)]	.705 ^a	1.000 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 6)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	22.500	20.500
Wilcoxon W	322.500	320.500
Z	-.145	-.342
Asymp. Sig. (2-tailed)	.885	.732
Exact Sig. [2*(1-tailed Sig.)]	.886 ^a	.745 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 2 and 7)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	2.000	1.000
Wilcoxon W	12.000	11.000
Z	-.939	-1.500
Asymp. Sig. (2-tailed)	.348	.134
Exact Sig. [2*(1-tailed Sig.)]	.533 ^a	.267 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 3 and 4)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	12.000	6.000
Wilcoxon W	90.000	84.000
Z	.000	-1.115
Asymp. Sig. (2-tailed)	1.000	.265
Exact Sig. [2*(1-tailed Sig.)]	1.000 ^a	.352 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 3 and 5)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	9.000	7.500
Wilcoxon W	19.000	17.500
Z	-.654	-.987
Asymp. Sig. (2-tailed)	.513	.324
Exact Sig. [2*(1-tailed Sig.)]	.610 ^a	.352 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (gropu 4 and 6)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	3.000	1.000
Wilcoxon W	24.000	22.000
Z	-1.006	-1.752
Asymp. Sig. (2-tailed)	.314	.080
Exact Sig. [2*(1-tailed Sig.)]	.429 ^a	.143 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 3 and 6)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	2.000	2.000
Wilcoxon W	12.000	12.000
Z	-.939	-.939
Asymp. Sig. (2-tailed)	.348	.348
Exact Sig. [2*(1-tailed Sig.)]	.533 ^a	.533 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 4 and 7)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	1.000	.000
Wilcoxon W	4.000	3.000
Z	-.775	-1.633
Asymp. Sig. (2-tailed)	.439	.102
Exact Sig. [2*(1-tailed Sig.)]	.667 ^a	.333 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 3 and 7)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	13.000	15.500
Wilcoxon W	58.000	60.500
Z	-.840	-.426
Asymp. Sig. (2-tailed)	.401	.670
Exact Sig. [2*(1-tailed Sig.)]	.503 ^a	.710 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 4 and 8)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	12.500	13.500
Wilcoxon W	22.500	23.500
Z	-1.406	-1.292
Asymp. Sig. (2-tailed)	.160	.196
Exact Sig. [2*(1-tailed Sig.)]	.170 ^a	.212 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 4 and 5)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	21.500	28.500
Wilcoxon W	42.500	49.500
Z	-1.362	-.717
Asymp. Sig. (2-tailed)	.173	.474
Exact Sig. [2*(1-tailed Sig.)]	.180 ^a	.494 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 5 and 6)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	4.000	5.000
Wilcoxon W	25.000	26.000
Z	-.683	-.358
Asymp. Sig. (2-tailed)	.495	.721
Exact Sig. [2*(1-tailed Sig.)]	.643 ^a	.857 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 6 and 7)		

Test Statistics ^b		
	Variant token	Variant type
Mann-Whitney U	8.000	10.500
Wilcoxon W	11.000	13.500
Z	-.731	-.279
Asymp. Sig. (2-tailed)	.465	.780
Exact Sig. [2*(1-tailed Sig.)]	.549 ^a	.791 ^a
a. Not corrected for ties.		
b. Grouping Variable: Combination of characteristics (group 5 and 7)		

REFERENCES

- Abeille, A. (1995) 'The flexibility of French idioms: A representation with lexicalised tree adjoining grammar', in Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds) *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 15-42
- Ackerman, R. & Webelhuth, G. (1993) 'The composition of (dis)continuous predicates: lexical or syntactic?' Online. Available at <http://ling.ucsd.edu/~ackerman/lexblock.ps>
- Adolphs, S. (2008) *Corpus and Context. Investigating Pragmatic Functions in Spoken Discourse*, Amsterdam/ Philadelphia: Benjamins
- Altenberg, B. (1990) 'Spoken English and the dictionary', in Svartvik, J. (ed.) *The London-Lund Corpus of Spoken English. Description and Research*. Lund: Lund University Press: 177-191.
- Auer, P. (1995) 'From context to contextualization', *Links & Letters* 3: 11-28
- Auran, C., Bouzon, C. & Hirst, D. J. (2004) 'The Aix-MARSEC Project: An evolutive database of spoken British English,' in Bel, B. & Marlien, I. (eds) *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan: 561-564.
- Austin, J. L. (1962) *How to do Things with Words*, Oxford: Oxford University Press.
- Autor, D. H., Levy, F., and Murnane, R. J. (2001) 'The skill content of recent technological change: an empirical exploration,' *NBER Working Paper 8337*, Cambridge, MA: National Bureau of Economy Research
- Barcelona, A. (2003) 'Metonymy in cognitive linguistics. An analysis and a few modest proposals', in Cuyckens, H., Berg, T. & Dirven, R. (eds) *Motivation in Language: Studies in Honor of Günter Radden*, Amsterdam: John Benjamins Publishing Company
- Barkema, H. (1996) 'Idiomaticity and terminology: A multi-dimensional descriptive model', *Studia Linguistica*, 50 (2): 125-160.
- Becker, J. (1975) 'The phrasal lexicon', in Shank, R. & Nash-Webber, B.L. (eds) *Proceedings of Interdisciplinary Workshop on Theoretical Issues in Natural Language Processing*: 70-73
- Berger, A. A. (1998) *Anatomy of Humour*, New Jersey: Transactions Publishers
- Berlyne, D. E. (1969) 'Laughter, humour, and play', in Lindzey, G. & Aronson, E. *The Handbook of Social Science*, Vol 3, 2nd edn, Reading, MA: Addison Wesley
- Biber, D. (1988) *Variation across Speech and Writing*, Cambridge: Cambridge University Press
- Biber, D. (2003) 'Variation among university spoken and written registers: A new multi-dimensional analysis', in Meyer, C. & Leistyna, P (eds) *Corpus Analysis: Language Structure and Language Use*, Amsterdam: Rodopi: 47-70
- Biber, D., & Finegan, E. (eds) (1994) *Sociolinguistic perspectives on register*, New York: Oxford University Press
- Biber, D., and Finegan, E. (1997) 'Diachronic relations among speech-based and written registers in English', in Nevalainen, T. & Kahlas-Tarkka, L. (eds) *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, Helsinki: Societe Neophilologique: 253-275. (Reprinted in Conrad and Biber (eds.) (2001), 66-83.)
- Biber, D. & Conrad, S. (1999) 'Lexical bundles in conversation and academic prose', in Hasselgard, H. & Oksefjell, S. (eds) *Out of Corpora: Studies in Honor of Stig Johansson*, Amsterdam: Rodopi: 181-190
- Biber, D., Conrad, S., & Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use (Cambridge Approaches to Linguistics)*, Cambridge: Cambridge University Press
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999) *Longman Grammar of Spoken and Written English*, London: Longman
- Bloom, L. (1973) *One Word at a Time: The Use of Single-Word Utterances before Syntax*, The Hague: Mouton
- Bloom, D. E. (2000) 'Social capitalism and human diversity', *The Creative Society of the 21st Century*, Paris: OECD Publishing
- Bobrow, S. A. & Bell, S. M. (1973) 'On catching on to idiomatic expressions', *Memory and Cognition*, 1: 343- 346
- Bolinger, D. (1975) *Aspects of Language*, 2nd edn, NY: Hacourt Brace Jovanovich
- Bolinger, D. (1976) 'Meaning and memory,' *Forum Linguisticum* 1 (1): 1-14
- Bolinger, D. (1981) *Aspects of Language*, 3rd edn, NY: Harcourt Brace Jovanovich
- Bosch, A. V. D. & Daelemans, W. (1993) 'Data-oriented methods for grapheme-to-phoneme conversion,' *Proceedings of the 6th Conference of the EACL*: 45-53

- Botha, R. P. (1981) *The Conduct of Linguistic Inquiry*, The Hague: Mouton
- Buhler, K. (1918) (1930) *The Mental Development of the Child*, New York: Harcourt Brace.
- Burger, H. (1973) *Idiomatik des Deutschen*, Tübingen: Max Niemeyer
- Burger, H. (1998a) *Phraseologie: Eine Einführung am Beispiel des Deutschen*, Berlin: Erich Schmidt Verlag
- Burger, H. (1998b) 'Idiom and metaphor: their relation in theory and text', in Durco, P. (ed.) *Europhras '97: Phraseology and Paremiology*, Bratislava: Akadémia PZ: 30-36
- Burger, H., Buhofer, A. & Sialm, A. (1982) *Handbuch der Phraseologie*, Berlin: Walter de Gruyter
- Bybee, J. & Hopper, P. (eds.) (2001) *Frequency and the Emergence of Linguistic Structure*, Amsterdam: John Benjamins Publishing Company
- Cacciari, C. & Tabossi, P. (1988) 'The comprehension of idioms', *Journal of Language and Memory*, 27: 668-683
- Cacciari, C. & Glucksberg, S. (1991) 'Understanding idiomatic expressions. The contribution of word meanings', in Simpson, G. (ed.) *Understanding word and sentence*, Amsterdam: Elsevier Science Publishers: 217-240
- Cacciari, C. & Glucksberg, S. (1995) 'Understanding idioms: Do visual images reflect figurative meanings?' *European Journal of Cognitive Psychology*, 7(3): 283-305.
- Carruthers, P. (2002) 'Human creativity: Its cognitive basis, its evolution and its connections with childhood pretence,' *British Society for the Philosophy of Science* 53: 225-249
- Carter, R. (1998) *Vocabulary: Applied Linguistic Perspectives*, 2nd edn, London: Routledge
- Carter, R. (2004) *Language and Creativity – The Art of Common Talk*, London: Routledge
- Carter, R., & McCarthy, M. (1988) *Vocabulary and Language Teaching*, Essex: Addison Wesley Longman Ltd
- Carter, R. & McCarthy, M. (2004) 'Creating, interacting: Creative language, dialogue and social context,' *Applied Linguistics* 25: 162-88
- Carter, R. & McCarthy, M. (2006) *Cambridge Grammar of English: A Comprehensive Guide*, Cambridge: Cambridge University Press
- Carter, R. & Nash, W. (1990) *Seeing Through Language: A Guide to Styles of English Writing*, Oxford: Basil Blackwell
- Catford, J. C. (1965) *A Linguistic Theory of Translation*, London: Oxford University Press
- Cekaite, A. & Aronsson, K. (2005) 'Language play, a collaborative resource in children's L2 learning', *Applied Linguistics* 26 (2): 169-191
- Čermák, F. (2007) 'Idioms and morphology', in Burger, H., Dobrovolskij, D., Kuehn, P. & Norrick, N. R. (eds.) *Phraseology: An International Handbook of Contemporary Research*, Berlin: Walter de Gruyter: 20-26
- Chiaro, D. (1992) *The Language of Jokes: Analyzing Verbal Play*, London: Routledge.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press
- Cook, G. (1994) *Discourse and Literature: The Interplay of Form and Mind*, Oxford: Oxford University Press
- Cook, G. (2000) *Language Play, Language Learning*, Oxford: Oxford University Press.
- Coulmas, F. (ed.) (1981) *Conversational Routine*, The Hague: Mouton
- Cowie, A. P. (1988) 'Stable and creative aspects of vocabulary use', in Carter, R. & McCarthy, M. *Vocabulary and Language Teaching*, Essex: Addison Wesley Longman Ltd
- Cowie, A. P. (1993) 'Getting to grips with phrasal verbs', *English Today* 36, 9 (4): 38-41
- Cowie, A. P. (1994) 'Phraseology', in Asher, R. E. *The Encyclopedia of Language and Linguistics* vol. 6, Oxford: Pergamon Press: 3168-71
- Cowie, A.P. (ed.) (1998) *Phraseology: Theory, Analysis, and Applications*, Oxford: Clarendon Press.
- Cowie, A. P. & Mackin, R. (eds) (1975) *Oxford Dictionary of Current Idiomatic English*, Vol. 1, London: Oxford University Press
- Cowie, A. P. & Mackin, R. (eds) (1993) *Oxford Dictionary of Phrasal Verbs* (2nd edn.), Oxford: Oxford University Press
- Cowie, A. P., Mackin, R. & McCaig, R. (eds) (1983) *Oxford Dictionary of Current Idiomatic English*, Vol. 2, Oxford: Oxford University Press
- Cruse, D. A. (1986) *Lexical Semantics*, Cambridge: Cambridge University Press
- Crystal, D. (1998) *Language Play*, Harmondsworth: Penguin
- Cutting, J. C. & Bock, J. K. (1997) 'That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiomatic blends', *Memory & Cognition*, 25: 57-71
- Davidson, D. (1967) 'Truth and Meaning', *Synthese*, 17: 304-323

- Deignan, A. (2005) *Metaphor and Corpus Linguistics*, Amsterdam: John Benjamins
- Divay, M. & Vitale, A.J. (1997) 'Algorithms for grapheme-phoneme translation for English and French: Applications for database searches and speech synthesis,' *Computational Linguistics* 23(4)
- Dornyei, Z. (2007) *Research Methods in Applied Linguistics*, Oxford: Oxford University Press
- Dunbar, R. (1996) *Grooming, Gossip and the Evolution of Language*, London: Faber and Faber Ltd
- Engel, M. (1996) 'Loose cannons, the horses's mouth and the bottom line: idioms and culture change', paper presented at the XXVI International Congress of Psychology, Montreal, Canada
- Ernst, Th. (1981) 'Grist for the linguistic mill: idioms and 'extra' adjectives', *Journal of Linguistic research* 1 (3): 51-68
- Dobrovolskij, D. & Piirainen, E. (2003) *Figurative Language. Cross-Cultural and Cross-Linguistic Perspectives*, Oxford, UK: Elsevier Ltd.
- Drew, P. and Holt, E. (1998) 'Figures of speech: Figurative expressions and the management of topic transition in conversation', *Language in society* 27 (4): 495-522
- Evans, V. & Green, M. (2006) *Cognitive Linguistics: An Introduction*, Edinburgh: Edinburgh University Press
- Fauconnier, G. & Turner, M. (1998) 'Principles of conceptual integration', *Discourse and Cognition* (CSLI): 269-283
- Fellbaum, C. (ed.) (1998) *WordNet: An Electronic Lexical Database*, MIT Press
- Fellbaum, C. (1993) 'The determiner in English idioms, metaphors', in Cacciari, C. & Tabossi, P. (eds) *Idioms: Processing, Structure, and Interpretation*, Hillsdale, NJ: Lawrence Erlbaum Associates: 271- 295
- Fernando, C. (1996) *Idioms and Idiomaticity*, Oxford/NY: Oxford University Press
- Freeborn, D. (1995) *A Course Book in English Grammar. 2nd edn*, Palgrave
- Fernando, C. & Flavell, R. (1981) *On Idiom: Critical Views and Perspectives*, Exeter Linguistic Studies 5: University of Exeter
- Fillmore, C. (1975) 'An alternative to checklist theories of meaning', *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*: 123-131
- Fillmore, C. (1982) 'Frame semantics', *Linguistics in the Morning Calm*, Seoul, South Korea: Hanshin Publishing Co: 111-137
- Fillmore, C. & Atkins, B. T. (1992) 'Toward a frame-based lexicon: The semantics of RISK and its neighbors', in Lehrer, A. & Kittay, E. F. (eds.) *Frames, Fields and Contrasts*, Hillsdale, NJ: Lawrence Erlbaum: 75-102
- Fillmore, C., Kay, P. & O'Connor, M. C. (1988) 'Regularity and idiomaticity in grammatical constructions: The case of *let alone*,' *Language* 64(3): 501-538
- Finke, R. (1995) 'Creative realism', in Smith, S., Ward, T. & Finke, R. (eds.) *The Creative Cognition Approach*, Cambridge: Cambridge University Press: 301-326
- Finke, R., Ward, T., & Smith, S. (1992) *Creative Cognition*, Boston: MIT Press
- Firth, J. R. (1957) 'Modes of meaning', *Papers in Linguistics 1934-1951*. London: Oxford University Press: 190-215
- Fleming, L. and Marx, M. (2006) 'Managing creativity in small worlds', *California Management Review* 48(1): 8-9
- Fleischer, W. (1982) *Phraseologie der deutschen Gegenwartssprache*, Leipzig: Bibliographisches Institut, VEB Verlag Enzyklopädie
- Fletcher, W. H. (2004) 'Facilitating the compilation and dissemination of ad-hoc web corpora', in Aston, G., Bernardini, S. & Stewart, D. (eds) *Corpora and Language Learners*, Amsterdam: Benjamins: 275-302
- Fletcher, W. H. (2007) 'Concordancing the web: promise and problems, tools and techniques', in Hundt, M., Nesselhauf, N. & Biewer, C. (eds) *Corpus Linguistics and the Web*, Amsterdam: Rodopi: 25-45
- Fraser, B. (1970) 'Idioms within a transformational grammar', *Foundations of Language* 6: 22-42
- Frey, B. S. & Stutzer, A. (2002) 'What can economists learn from happiness research?' *Journal of Economic Literature* XL: 402-435
- Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice Hall
- Garside, R., Leech, G. & McEnergy, T. (eds) (1997) *Corpus Annotation: Linguistic Information from Computer Text Corpora* *Corpus Annotation*, London: Longman

- Geeraerts, D. (1995) 'Specialization and reinterpretation in idioms', in Everaert, M., van der Linden, E.-J., Schenk, A. & Schreuder, R. (eds) *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Erlbaum: 58-74
- Getzels, J. W. & Jackson, P. J. (1962) *Creativity and Intelligence: Explorations with Gifted Students*. New York: John Wiley and Sons, Inc
- Gibbs, R. W. (1980) 'Spilling the beans on understanding and memory for idioms in conversation', *Discourse Processes*, 9: 17-30
- Gibbs, R. W. (1985) 'On the process of understanding idioms', *Journal of Psycholinguistic Research*, 14 (5): 465-472
- Gibbs, R. W. (1986) 'Skating on thin ice: Literal meaning and understanding idioms in conversation', *Memory and Cognition*, 8: 149-156
- Gibbs, R. W. (1990) 'Psycholinguistic studies on the conceptual basis of idiomaticity', *Cognitive Linguistics*, 1: 417-451
- Gibbs, R. W. (1992) 'What idioms really mean', *Journal of Memory and Language* 31: 485-506
- Gibbs, R. W. (1993) 'Why idioms are not dead metaphors', in Cacciari, C. & Tabossi, P. (eds) *Idioms: Processing, Structure, and Interpretation*, Hillsdale, NJ: Lawrence Erlbaum Associates: 57-77
- Gibbs, R. W. (1994) *The Poetics of Mind*, Cambridge: Cambridge University Press
- Gibbs, R. W. (1995) 'Idiomaticity and human cognition', in Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds) *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 97-116
- Gibbs, R. W. (1998) 'The fight over metaphor in thought and language', in Katz, A. N., Cacciari, C., Gibbs, R. W. & Turner, M. (contribs) *Figurative Language and Thought*, New York: Oxford University Press: 88-108
- Gibbs, R. W. & Gonzales, G. P. (1985) 'Syntactic frozenness in processing and remembering idioms', *Cognition*, 20: 243-259
- Gibbs, R. W. & Nayak, N. (1989) 'Psycholinguistic studies on the syntactic behaviour of idioms', *Cognitive Psychology*, 21: 100-138
- Gibbs, R. W., Nayak, N. & Cutting, C. (1989) 'How to kick the bucket and not decompose: Analyzability and idiom processing', *Journal of Memory and Language*, 28: 576-593
- Gibbs, R. W. & O'Brien, J. (1990) 'Idioms and mental imagery', *Cognition*, 36(1): 35-68
- Gildea, P. and Glucksberg, S. (1983) 'On understanding metaphor: The role of context', *Journal of Verbal Learning and Verbal Behavior*, 22, 577-590
- Gläser, R. (1988) 'The grading of idiomaticity as a presupposition for a taxonomy of idioms', in Hüllen, W. & Schulze, R. (eds) *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*, Tübingen: Max Niemeyer
- Glewwe, P., Jacoby, H. and King, E. (1999) 'Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis,' *FCND Discussion Paper* 68, Washington, DC: International Food Policy Research Institute
- Glucksberg, S. (1993) 'Idiom meanings and allusional content', in Cacciari, C. & Tabossi, P. (eds) *Idioms: Processing, Structure, and Interpretation*, Hillsdale, NJ: Lawrence Erlbaum Associates: 3-26
- Glucksberg, S. (2001) *Understanding Figurative Language*, New York: Oxford University Press
- Glucksberg, S., Brown, M., & McGlone, M. S. (1993) 'Conceptual metaphors are not automatically accessed during idiom comprehension' *Memory & Cognition*, 21: 711-719.
- Glucksberg, S. & McGlone, M. (1999) When love is not a journey: what metaphors mean, *Journal of Pragmatics* 31: 1541-1558
- Goldman-Eisler, F. (1968) *Psycholinguistics: Experiments in Spontaneous Speech*, London/New York: Academic Press
- Grady, J. (1997) 'Foundations of meaning: primary metaphors and primary scenes,' University of California, Berkeley: Ph.D. Dissertation
- Granger S. & Paquot M. (2008) 'Disentangling the phraseological web', in Granger, S. & Meunier, F. (eds) *Phraseology: An Interdisciplinary Perspective*, Amsterdam & Philadelphia, Benjamins: 27-49
- Grant, L. and Bauer, L. (2004) 'Criteria for redefining idioms: Are we barking up the wrong tree?' *Applied Linguistics* 25/1: 38-61
- Greene, J. C., Kreider, H. and Mayer, E. (2005) 'Combining qualitative and quantitative methods in social inquiry', in Somekh, B. and Lewin, C. (eds) *Research Methods in the Social Sciences*, London: Sage
- Grice, H. P. (1975) 'Logic and Conversation,' in Cole, P. & Morgan, J. (eds) *Syntax and Semantics*, Volume 3, New York: Academic Press: 41-58

- Groom, N. (2006) 'Phraseology and epistemology in humanities writing: a corpus-driven study', English Department, University of Birmingham: PhD thesis
- Gumperz, J. J. (1982) *Discourse Strategies*, Cambridge: Cambridge University Press
- Gumperz, J. J. (1992) 'Contextualization and understanding', in Duranti, A. & Goodwin C. (eds) *Rethinking Context: Language as an Interactive Phenomenon*, Cambridge: Cambridge University Press: 229-52
- Hall, G. (2001) 'The poetics of everyday life', in McRae, J. (ed.) *Reading beyond Text: Processes and Skills*, Seville: University of Seville: 69-86
- Halliday, M. A. K. (1978) *Language as Social Semiotic*, London: Edward Arnold
- Halliday, M. A. K. (1985) *An Introduction to Functional Grammar*, London: Edward Arnold
- Hanushek, E. A. (2002) 'The Failure of Input-Based Schooling Policies,' *NBER Working Paper 9040*, Cambridge, MA: National Bureau of Economic Research
- Hanushek, E. A. (2004) 'Some Simple Analytics of School Quality,' *NBER Working Paper 10229*, Cambridge, MA: National Bureau of Economic Research
- He, Z. (1989) 'Creative idiomaticity', in Hall-Kira, H., Meacham, M. & Shapiro, R. (eds) *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society*, Berkeley: Berkeley Linguistics Society
- Hess-Biber, N. S. & Leavy, P. (2006) *The Practice of Qualitative Research*, Thousand Oaks, Calif.: Sage
- Hickey, T. (1993) 'Identifying formulas in first language acquisition', *Journal of Child Language* 20: 27-41
- Hockett, C. F. (1958) *A Course in Modern Linguistics*, NY: MacMillan
- Hoey, M. (1991) *Patterns of Lexis in Text*, Oxford: Oxford University Press
- Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*, London: Routledge.
- Hoey, M. (2007) 'Lexical priming and literary creativity,' in Hoey, M., Mahlberg, M., Stubbs, M. & Teubert, W. *Text, Discourse and Corpora – Theory and Analysis*, London: Continuum: 7-29.
- Holt, E. and Drew, P. (1995) 'Idiomatic expressions and their role in the organization of topic transition in conversation', in Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds), *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 117-131
- Holt, E. and Drew, P. (2005) 'Figurative pivots: the use of figurative expressions in pivotal topic transitions', *Research on Language and Social Interaction*, 38 (1): 35-61.
- Howarth, P. (1996) *Phraseology in English Academic Writing*, Tübingen: Max Niemeyer
- Howarth, P. (1998) 'Phraseology and second language proficiency,' *Applied Linguistics* 19/1, Oxford University Press: 24-44
- Hudson, J. (1998) *Perspectives on Fixedness: Applied and Theoretical*, Lund University Press
- Huizinga, J. (1939) (1949) *Homo Ludens*, London: Routledge
- Hummer, C. (2007) 'Synonymy, similarity and difference of multi-word expressions', in Fellbaum, C. (ed.) *Idioms and Collocations: Corpus-Based Linguistic and Lexicographic Studies*, Birmingham, UK: Continuum International Publishing Group Ltd.
- Hymes, D. (1962) 'The ethnography of speaking', in Gladwin, T. & Sturtevant, W. (eds) *Anthropology and Human Behaviour*, Washington, DC: Anthropological Society of Washington: 13-53
- Jakobson, R. (1960) 'Closing statement: Linguistics and poetics,' in Sebeok, T. (ed) *Style in Language*, Cambridge, MA: MIT Press: 350-77
- Jauss, H.R. (1970) 'Literary history as challenge to literary theory', *New Literary history* 2, reprinted in Rice, P. & Waugh, P. (eds) (1989) *Modern Literary Theory: A Reader*, New York: Routledge
- Johansson, S. (1991) 'Times change, and so do corpora', in Aijmer, K. & Altenberg, B. (eds) *English Corpus Linguistics*, London: Longman: 305-314
- Johnson, M. (1987) *The Body in the Mind. The Bodily Basis of Reason and Imagination*, Chicago, IL: University of Chicago Press
- Johnson, R. B. & Turner, L.A. (2003) 'Data collection strategies in mixed methods research', in Tashakkori, A. & Teddlie, C. (eds) *Handbook of Mixed Methods in Social and Behavioural Research*, Thousand Oaks, Calif.: Sage
- Jones, B. F. (2005) 'Age and great invention,' *NBER Working Paper 11359*, Cambridge, MA: National Bureau of Economic Research
- Katz, J. J. (1973) 'Compositionality, idiomaticity, and lexical substitution', in Anderton, S. & Kiparsky, P. (eds) *A Festschrift for Morris Halle*, New York: Holt, Rinehart, and Winston: 357-76

- Katz, J. J. & Postal, P. M. (1963) 'Semantic interpretation of idioms and sentences containing them', *Quarterly Progress Report*, 70, MIT Research Laboratory of Electronics: 275-282
- Katz, J. J. & Postal, P. M. (1964) *An Integrated Theory of Linguistic Descriptions*, Cambridge, MA: MIT Press
- Keysar, B. & Bly, B. M. (1999) 'Swimming against the current: Do idioms reflect conceptual structure?' *Journal of Pragmatics*, 31, 1559-1576
- Koester, A. (2006) *Investigating Workplace Discourse*, London: Routledge
- Kövecses, Z. (2005) *Metaphor in Culture: Universality and Variation*, Cambridge: Cambridge University Press
- Kövecses, Z. & Szabó, P. (1996) 'Idioms: A view from cognitive semantics' *Applied Linguistics*, 17(3): 326-355
- Kuiper, K. (2004) 'Formulaic performance in conventionalised varieties of speech', in Schmitt, N. (ed.) *Formulaic Sequences: Acquisition, Processing and Use*, Amsterdam: John Benjamins Publishing Company
- Labov, W. (1972) *Language in the Inner City: Studies in Black English Vernacular*, Philadelphia: University of Pennsylvania Press
- Labov, W. (1975) 'The quantitative study of linguistic structure', in Dahlstedt, K. H. (ed.) *The Nordic Languages and Modern Linguistics*, Stockholm: Almqvist & Wiksell: 188-244
- Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Chicago, IL: University of Chicago Press
- Lakoff, G. (1990) 'The invariance hypothesis: Is abstract reason based on imageschemas?' *Cognitive Linguistics*, 1(1): 39-74
- Lakoff, G. (1993) 'The contemporary theory of metaphor', in Ortony, A. (ed.) *Metaphor and Thought* (2nd edn), Cambridge: Cambridge University Press: 202-51
- Lakoff, G. & Johnson, M. (1980) *Metaphors We Live By*, Chicago, IL: University of Chicago Press
- Lakoff, G. & Johnson, M. (1999) *Philosophy in The Flesh: the Embodied Mind and its Challenge to Western Thought*, NY: Basic Books
- Lamb, S. M. (1998) 'Pathways of the brain: The neurocognitive basis of language,' *Current Issues in Linguistic Theory* 170
- Landes, D. S. (2006) 'Why Europe and the West? Why not China?' *Journal of Economic Perspectives*, 20: 2 (Spring): 3-22
- Langacker, R. W. (1987) *Foundations of Cognitive Grammar, Volume I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press
- Langlotz, A. (2006) *Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom-Representation and Idiom-Variation in English*, Philadelphia, PA, USA: John Benjamins Publishing Company
- Lau, S., Hui, A.N.N., & Ng, G.Y.C. (2004) *Creativity: When East Meets West*, World Scientific Publishing Co Pte Ltd
- Lazaranton, A. (2005) 'Quantitative research methods', in Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning*, Mahwah, N.J.: Lawrence Erlbaum.
- Lecercle, J. J. (1990) *The Violence of Language*, London: Routledge
- Leech, G. N. (1983) *Principles of Pragmatics*, London: Longman
- Leech, G. (1990) *Semantics: An Introduction to the Study of Meaning*, Penguin Books.
- Leech, G. (2001a) 'The role of frequency in ELT: New corpus evidence brings a re-appraisal', in Wenzhong, H. (ed.) *ELT in China2001: Papers presented at the 3rd International Symposium on ELT in China*, Beijing: Foreign Language Teaching and Research Press: 1-23
- Leech, G. (2001b) 'The state of art in the corpus linguistics', in Aijmer, K. & Altenberg, B. (eds) *English Corpus Linguistics*, London: Longman: 305-314
- Levinson, S. C. (1983) *Pragmatics*, Cambridge: Cambridge University Press
- Lewis, D. (1969) *Convention: A Philosophical Study*, Cambridge: Harvard University Press
- Lewis, M. (1993) *The Lexical Approach: The State of ELT and a Way Forward*, LTP
- Lipka, L. (2000) 'English (and general) word-formation – The state of the art in 1999', in Reitz, B. & Rieuwerts, S. (eds) *Anglistentag 1999 Mainz*, Proceedings, Wissenschaftlicher Verlag: 5-20
- Lubart, T. I. (1999) 'Creativity across cultures', in Sternberg, R. J. (ed.) *Handbook of Creativity*, Cambridge: Cambridge University Press
- Lyons J. (1977) *Semantics, Volumes I and II*, Cambridge: Cambridge University Press
- Makkai, A. (1972) *Idiom Structure in English*, The Hague: Mouton

- Malinowski, B. (1923) 'The problem of meaning in primitive languages', in Ogden, C. K. & Richards, I. A. (eds) *The Meaning of Meaning*, London: Routledge & Kegan Paul: 146-152
- Maxwell, J. A. & Loomis, D.M. (2003) 'Mixed methods design: an alternative approach' in Tashakkori, A. & Teddlie, C. (eds) *Handbook of Mixed Methods in Social and Behavioural Research*, Thousand Oaks, Calif.: Sage
- McCarthy, M. & Carter, R. (1994) *Language as Discourse: Perspectives for Language Teaching*, London: Longman
- McCarthy, M. & Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching* (1st edn), Cambridge: Cambridge University Press
- McCracken, G. (1988) *The Long Interview*, Newbury Park, Calif.: Sage
- McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-based Language Studies: An Advanced Resource Book*, London: Routledge
- McEnery, T. & Gabrielatos, C. (2006) 'English corpus linguistics', in Aarts, B. & McMahon, A. (eds) *The Handbook of English Linguistics*, Oxford: Blackwell: 33-71
- McGlone, M. S. (1996) 'Conceptual metaphors and figurative language interpretation: food for thought?' *Journal of Memory and Language* 36: 50-67
- McGlone M. S. (2007) 'What is the explanatory value of a conceptual metaphor?' *Language & Communication* 27: 109-126
- McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994) 'Semantic productivity and idiom comprehension', *Discourse Processes* 17: 167-190
- Mel'cuk, I. (1998) 'Collocations and lexical functions', in Cowie, A. (ed.) *Phraseology. Theory, Analysis and Applications*, Clarendon Press
- Mey, J. L. (1993) *Pragmatics: An Introduction*, Oxford: Blackwell
- Meyer, R. E. (1999) 'Fifty years of creativity research', in Sternberg, R. (ed.) *Handbook of Creativity*, Cambridge: Cambridge University Press: 449-60
- Miles, M. B. & Huberman, A. M. (1994) *Qualitative Data Analysis, 2nd edition*, Thousand Oaks, Calif.: Sage
- Moon, R. (1996) 'Data, description, and idioms in corpus lexicography,' *Euralex '96 Proceedings*: 245-256
- Moon, R. (1997) 'Vocabulary connections: Multi-word items in English', in Schmitt, N & McCarthy, M. (eds) *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge: Cambridge University Press: 237-257
- Moon, R. (1998) *Fixed Expressions and Idioms in English – A Corpus-Based Approach*, New York: Oxford University Press
- Morgan, J. L. (1978) 'Two types of convention in indirect speech acts,' in Cole, P. (ed.) *Syntax and Semantics*, 9: 261-80
- Murnane, R. J. & Levy, F. (1996) *Teaching the New Basic Skills: Principles for Educating Children to Thrive in a Changing Economy*, Free Press
- Nash, W. (1985) *The Language of Humour*, New York: Longman
- Nattinger, J. (1980) 'A lexical phrase grammar for ESL', *TESOL Quarterly* 14: 337-44
- Nattinger, J. and DeCarrico, J. (1992) *Lexical Phrases and Language Teaching*, Oxford: Oxford University Press
- Nayak, N. P., & Gibbs, R. Jr. (1990) 'Conceptual knowledge in the interpretation of idioms', *Journal of Experimental Psychology: General*, 19(3): 315-330
- Newmeyer, F. J. (1974) 'The regularity of idiom behaviour', *Lingua* 34: 327-342
- Newmeyer, F. J. (2003) 'Grammar is grammar and usage is usage', *Language* 79: 682-707
- Nicholas, T. (1995) Semantics of idiom modification', in Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds), *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 233-252
- Nippold, M. A. & Martin, S. T. (1989) 'Idiom interpretation in isolation versus context a developmental study with adolescents', *Journal of Speech and Hearing Research* 32: 59-66
- Norrick, N. R. (1993) *Conversational Joking: Humour in Everyday Talk*, Bloomington: Indiana University Press
- Norrick, N. R. (2000) *Conversational Narrative: Storytelling in Everyday Talk*, Amsterdam: John Benjamins
- Nunberg, G. (1978) *The Pragmatics of Reference*, Bloomington, IN: Indiana University Linguistics
- Nunberg, G., Sag, I. A. & Wasow, Th. (1994) 'Idioms', *Language*, 70(3): 491- 538
- O'Grady, W., Dobrovolsky, M. & Aronoff, M. (1997) *Contemporary Linguistics: An introduction*, Boston: Bedford/St.Martin's

- O'Keeffe, A., McCarthy, M. & Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press.
- Ohmann, R. (1971) 'Speech acts and the definition of literature,' *Philosophy and Rhetoric* 4: 1-19
- Ortony, A., Schallert, D. L., Reynolds, R. E. & Antos, S. J. (1978) 'Interpreting metaphors and idioms: Some effects of the context on comprehension', *Journal of Verbal Learning and Verbal Behaviour*, 17: 465- 477
- Palmer, H. E. (1933) *Second Interim Report on English Collocations*, Tokyo: Kaitakusha
- Palmer, F. R. (1981) *Semantics*, 2nd edn. Cambridge: Cambridge University Press
- Partington, A. (1998) *Patterns and Meanings*, Philadelphia, PA, USA: John Benjamins.
- Partington, A. (2006) *The Linguistics of Laughter: A Corpus-Assisted Study of Laughter-Talk*, London: Routledge
- Pawley, A. (1986) 'Lexicalization', in Tannen, D. & Alatis, J. E. (eds) *Language & Linguistics: The Interdependence of Theory, Data & Application*, Georgetown University Round Table on Languages & Linguistics 1985: 98-120
- Pawley, A. & Syder, F. H. (1983) 'Two puzzles for linguistic theory: native-like selection and native-like fluency', in Richards, J. C. & Schmidt, R. W. (eds) *Language and Communication*, London: Longman: 191- 225
- Pawley, A. & Syder, F. H. (2000) 'The one-clause-at-a-time hypothesis', in Riggenbach, H. (ed.) *Perspectives on Fluency*, Ann Arbor: University of Michigan Press: 163-199
- Peters, A. (1983) *The Units of Language Acquisition*, Cambridge: Cambridge University Press
- Philip, G. (2006) 'Reassessing the canon - 'Fixed' phrases in general reference corpora', in Granger, S. & Meunier, F. (eds) *Phraseology: An Interdisciplinary Perspective*, Amsterdam: John Benjamins
- Phillip, G. (2000) 'An idiomatic theme and variations,' in Heffer, C. & Sauntson, H. (eds) *Words in Context: A Tribute to John Sinclair on His Retirement* [ELR Monograph 18], Birmingham: The University of Birmingham: 221-233.
- Pope, R. (2005) *Creativity: Theory, History, Practice*, London: Routledge
- Pratt, M. L. (1977) *Toward a Speech Act Theory of Literary Discourse*, Bloomington: Indiana University Press.
- Prodromou, L. (2007) 'Bumping into creative idiomaticity', *English Today* 89, Vol. 3, No. 1, Cambridge: Cambridge University Press
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*, London: Longman
- Radden, G. (2000) 'How metonymic are metaphors?' in Barcelona, A. (ed.) *Metaphor and Metonymy at the Crossroads*, Berlin/New York: Mouton de Gruyter: 93-108
- Rayson, P., Archer, D., Piao, S. L., & McEnery, T. (2004) 'The UCREL Semantic Analysis System,' *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks in Association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal: 7-12.
- Raupach, M. (1984) 'Formulae in second language speech production', in Dechert, H. W., Möhle, D. & Raupach, M. (eds) *Second Language Productions*, Tübingen: Narr: 114-137
- Read, J. & Nation, P. (2004) 'Measurement of formulaic sequences', in Schmitt, N. (ed.) *Formulaic Sequences*, Amsterdam: John Benjamins: 23-35
- Redfern, W. (1984) *Puns*, Oxford: Blackwell Publication
- Renouf, A. (2007) 'Tracing lexical productivity and creativity in the British media: The chavs and the chav-nots', in Munat, J. (ed.) *Lexical Creativity, Texts and Contexts*, Amsterdam/Philadelphia: John Benjamins: 61-90.
- Rice, P. & Waugh, P. (eds) (1989) *Modern Literary Theory: A Reader*, New York: Routledge
- Richards, L. (2005) *Handling Qualitative Data: A Practical Guide*, London: Sage
- Roberts, M. H. (1944) 'The science of idiom: a method of inquiry into the cognitive design of language', *Publication of the Modern Language Association of America* 69: 291-306
- Ross, A. (1998) *The Language of Humour*, London: Routledge
- Rossmann, G. B. & Wilson, B. L. (1985) 'Numbers and words: combining qualitative and qualitative methods in a single large scale evaluation study', *Evaluation Review* 9/5: 624-43
- Rua, L. (2007) 'Keeping up with the times: lexical creativity in electronic communication,' in Munat, J. (ed.) *Lexical Creativity, Texts and Contexts*, Amsterdam/Philadelphia: John Benjamins: 137-162
- Sacks, H. (1992) *Lectures on Conversation: Volume I and II*, Oxford: Basil Blackwell

- Sadock, J. M. (1972) 'Speech act idioms', in Peranteau, P. M., Levi, J. N. & Phares, G. C. (eds) *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*: 329-39
- Sadock, J. M. (1974) *Towards a Linguistic Theory of Speech Acts*, New York: Academic Press
- Sampson, G. (1975) *The Form of Language*, London: Weidenfeld and Nicolson
- Sandelowski, M. (2003) 'Tables of tableaux? The challenges of writing and reading mixed methods studies', in Tashakkori, A. & Teddlie, C. (eds) *Handbook of Mixed Methods in Social and Behavioural Research*, Thousand Oaks, Calif.: Sage
- Sawyer, R. K. (2006) *Explaining Creativity: The Science of Human Innovation*, New York: Oxford University Press.
- Schank, R. C., & Abelson, R. P. (1977) *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, Hillsdale, NJ: Lawrence Erlbaum Associates
- Schenk, A. (1995) The syntactic behaviour of idioms. Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds) *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 253-271
- Schmitt, N. (ed.) (2004) *Formulaic Sequences*, Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Schmitt, N., & Carter, R. (2004) 'Formulaic sequences in action - an introduction', in Schmitt, N. (ed.) *Formulaic Sequences Acquisition, Processing, and Use*, Amsterdam; Philadelphia: John Benjamins: 1-22
- Searle, J. R. (1969) *Speech Acts. An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press
- Semino, E. and Short, M. (2004) *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*, London: Routledge
- Simpson, P. (2004) *Stylistics: A Resource Book for Students*, London: Routledge
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press
- Sinclair, J. (1997) 'Corpus evidence in language description', in Wichmann, A., Fligelstone, S., McEnery, T. & Knowles G. (eds) *Teaching and Language Corpora*, London and New York: Longman: 27-39
- Sinclair, J. (2004) 'Corpus and Text: Basic Principles - 'Developing Linguistic Corpora: A Guide to Good Practice'', *Text Corpora and Corpus Linguistics*, Tuscan Word Centre. Online. Available at: <<http://ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>>
- Shklovsky, V. (1917) 'Art as Technique,' *Russian Formalist Criticism; Four Essays*. Reprinted in Rice, P. & Waugh, P. (eds) (1989) *Modern Literary Theory: A Reader*, New York: Routledge
- Skandera, P. (2004) 'What are idioms?' in Allerton, D. J., Nesselhauf, N. & Skandera, P. (eds) *Fixed Expressions: Theory and Applications*, International Cooper Series in English Language and Linguistics 8. Basle: Schwabe
- Smith, L. P. (1925) *Words and Idioms*, London: Constable
- Smith, L. P. (1961) 'Four words: Romantic, originality, creative, genius', *SPE Tracts I-XX (I)*: 521-566
- Sperber, D. & Wilson, D. (1986) *Relevance: Communication and Cognition*, Cambridge, MA: Harvard U Press
- Sprenger, S. A., Levelt, W. J. M., & Kempen, G. (2006) 'Lexical access during the production of idiomatic phrases', *Journal of Memory and Language*, 54(2): 161-184
- Stankiewicz, E. (1960) 'Linguistics and the study of poetic language', in Sebeok, T. A. (ed.) *Style in Language*, Cambridge, Mass: MIT Press: 69-81
- Steen, G. J. (1999) 'From linguistic to conceptual metaphor in five steps', in Gibbs, R.W. & Steen, G. J. *Metaphor in Cognitive Linguistics* Amsterdam: 57-77
- Steward, D. (2005) 'Hidden culture: using the British National Corpus with language learners to investigate collocational behaviour, wordplay and culture-specific references,' in Barnbrook, G., Danielsson, P. & Mahlberg, M. (eds) *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*, London: Continuum: 83-95.
- Stockwell, P. (2006) 'Language and literature: stylistics', in Aarts, B. & McMahon, A. (eds) *The Handbook of English Linguistics*, Oxford: Basil Blackwell: 742-758
- Strassler, J. (1982) *Idioms in English: A Pragmatic Analysis*, Tübingen Verlag
- Stubbs, M. (1996) *Text and Corpus Analysis*, Oxford: Blackwell
- Stubbs, M. (2005) 'Conrad in the Computer: Examples of Quantitative Stylistic Methods,' *Language and Literature* 14(1)
- Sullivan, P. N. (2000) 'Playfulness as mediation in communicative language teaching in a Vietnamese classroom', in Lantolf, J. P. (ed.) *Sociocultural Theory and Second Language Learning*, Oxford: Oxford University Press: 115-32

- Swinney, D. A. & Cutler, A. (1979) 'The access and processing of idiomatic expressions', *Journal of Verbal Learning and Verbal Behaviour*, 18: 523- 534
- Sweet, H. (1899). *The Practical Study of Languages. A Guide for Teacher and. Learner*, London: Dent. [Reprinted London: OUP, 1964]
- Tabossi, P. & Zardon, F. (1993) 'The activation of idiomatic meaning in spoken language comprehension', in Cacciari, C. & Tabossi, P. (eds) *Idioms: Processing, Structure, and Interpretation*, Hillsdale, NJ: Lawrence Erlbaum Associates: 145- 162
- Tabossi, P., & Zardon, F. (1995) 'The activation of idiomatic meaning', in Everaert, M., van der Linden, E., Schenk, A. & Schreuder, R. (eds) *Idioms: Structural and Psychological Perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates: 273-282
- Tabossi P., Fanari, R. & Wolf, K. (2008) 'Processing idiomatic expressions: effects of semantic compositionality', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34 (2): 313-327
- Tannen, D. (1989) *Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse*, Cambridge: Cambridge University Press
- Tarone, E. (2000) 'Getting serious about language play: Language play, interlanguage variation and second language acquisition', in Swierzbis, B., Morris, F., Anderson, M. E., Klee, C., & Tarone, E. (eds) *Social and Cognitive Factors in Second Language Acquisition: Selected Proceedings of the 1999 Second Language Research Forum*, Somerville, MA: Cascadia Press: 31-54
- Tarone, E. (2002) 'Frequency effects, noticing, and creativity: Factors in a variationist interlanguage framework,' *Studies in Second Language Acquisition* 24: 287-296
- Tashakkori A. & Teddlie, C. (1998) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*, Thousand Oaks, Calif.: Sage
- Taylor, J. R. (2002) 'Category extension by metonymy and metaphor', in Dirven, R. & Parings, R. (eds) *Metaphor and Metonymy in Comparison and Contrast*, Mouton de Gruyter
- Turner, M. (1987) *Death is the Mother of Beauty: Mind, Metaphor, Criticism*, Chicago: University of Chicago Press
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*, Amsterdam/Philadelphia: Benjamins
- Titone, D. A. & Connine, C. M. (1999) 'On the compositional and noncompositional nature of idiomatic expressions', *Journal of Pragmatics*, 31: 1655- 1674
- Van Lancker, D. & Canter, G. J. (1981) 'Idiomatic versus literal interpretations of ditropically ambiguous sentences,' *Journal of Speech and Hearing Research*, 24/1: 64-69
- Veisberg, A. (2007) 'Occasional and systematic shifts in word-formation and idiom use in Latvian as a result of translation', in Munat, J. (ed.) *Lexical Creativity, Texts and Contexts*, Amsterdam and Philadelphia: John Benjamins: 239-263
- Verschueren, J. (1999) *Understanding Pragmatics*, London: Arnold
- Ward, T., Smith, S. and Finke, R. (1999) 'Creative cognition,' in Sternberg, R. (ed.) *Handbook of Creativity*, Cambridge: Cambridge University Press: 189-212
- Wasow, T., Sag, I. & Nunberg, G. (1983) 'Idioms: An interim report', in Hattori, S. & Inoue, K. (eds) *Proceedings of the XIIIth International Congress of Linguists*: 102-115
- Weinreich, U. (1963) *Languages in Contact: Findings and Problems*, The Hague: Mouton
- Weinreich, U. (1969) 'Problems in the analysis of idioms', in Puhvel, J. (ed.) *Substance and Structure of Language*, CA: University of California Press: 23-81
- Wilson, A & Thomas J A (1997) 'Semantic Annotation,' in Garside, R., Leech, G. & McEnery, T. (eds) *Corpus Annotation: Linguistic Information from Computer Text Corpora*, London: Longman: 53-65
- Wray, A. (2002) *Formulaic Language and the Lexicon*, Cambridge University Press
- Yusuf, S. (2007) 'From creativity to innovation', Washington, D.C.: World Bank, Development Research Group