

NANYANG TECHNOLOGICAL UNIVERSITY

LINGUISTICS AND MULTILINGUAL STUDIES

FINAL YEAR PROJECT

**Losing One's Mind Over Meaning:
Analysing the Behaviour of English
Possessive Idioms**

Author:

Jia Qian Ho

Supervisor:

Associate Professor Francis BOND

August 10, 2015

Contents

Acknowledgements	vi
Abstract	vii
1 Introduction	1
2 Literature Review	2
3 Scope and Aims	4
4 Motivations	5
5 Methodology	7
6 Results and Discussion	16
6.1 Syntactic analysis	16
6.1.1 Alternation in Idioms	16
6.1.2 Prepositional phrasal idioms	19
6.1.3 Transformation in idioms	20
6.2 Decomposability	20
6.3 Plausibility	23
6.4 Further semantic analysis	24
6.4.1 Integration into WordNet	24
6.4.2 Semantic variation in the BNC	26
7 Future Work	28
8 Conclusion	28
9 Bibliography	30
10 Appendix A	34
11 Appendix B	39

Acknowledgements

This thesis paper emerged from my Undergraduate Research Experience on CAmpus (URECA) project and provides a much deeper analysis of possessive idioms, which would not have been possible if not for the help of and support from several individuals.

Foremost, I would like to thank my supervisor, Francis Bond whose idea formed the basis of this project. I was fortunate to work with him on my URECA project, which was in turn extended into my FYP-URECA. The URECA project piqued my interest in possessive idioms and inspired me to investigate further and provide a more quantitative and qualitative analysis. Although I was unfamiliar with computational linguistics, Francis guided me patiently and helped me in very technical areas. Thank you, Francis, for your guidance and help.

I also wish to extend my gratitude to Dan Flickinger, who added templates into the ERG; for which without this, we would not have been able to automatically construct and parse the idioms and I would not have been able to carry out my analysis proper. Thank you for your help and for tolerating our multiple Skype sessions.

My thanks go out to my family – for giving me the much-needed moral support for this semester-long project. Despite not knowing much about the university education system, much less about Linguistics, your presence and confidence in me do mean a lot.

Lastly, I would like to thank you readers, for taking an interest in my thesis and the time to read through it. I hope that by the end of this thesis, you would have gotten a deeper insight into the syntactic and semantic complexities of possessive idioms. So let's put on our thinking caps and begin with the paper!

Ho Jia Qian

Abstract

Idioms are commonly found in everyday language and reflect the conventionalisations in speech communities. Regarding English idioms, past research have examined the syntactic and semantic analysis of idioms (Villavicencio & Copestake, 2002), along with their decomposability (Nunberg, 1978; Gibbs, 1989a,b) and comprehension (Titone & Connine, 1994; Cacciari & Tabossi, 2014). However, there has been little research on English possessive idioms despite their uniqueness and interesting properties. This thesis thus seeks to analyze the syntax and semantics of possessive idioms and describe their behaviour in terms of their decomposability and plausibility. A total of 514 idioms were categorized into co-indexed and separate possessive idioms and then grouped syntactically in order to be incorporated into new templates in the English Resource Grammar (Flickinger, 2011). Subsequently, the meaning of either each idiom component or paraphrase component was linked to WordNet (Fellbaum, 1998a) by choosing the most appropriate sense. The resulting comprehensive syntactic and semantic idiom descriptions allowed for analyses of their syntax, semantics, decomposability and plausibility. Results demonstrated the interplay between syntax and semantics and revealed novel aspects of possessive idioms, such as alternation and transformation in idioms. Furthermore, results confirmed that a degree of decomposability exists and suggested that possessive idioms could be categorized into four groups according to their projectability. The comprehensive idiom database will be released under an open license where it can be used as a dictionary and to further improve natural language processing applications.

1 Introduction

Despite being commonly found in everyday language, idioms cannot be integrated into any theory of grammar in a straightforward manner (Fellbaum, 2014: 777) and are thus still poorly represented in lexicons such as WordNet (Fellbaum, 1998a) and in grammars like the English Resource Grammar (Flickinger, 2011).

Idiomatic expressions, for instance *by and large*, *devil's advocate*, *play with fire*, *a taste of one's own medicine*, clearly exemplify the wide spectrum of idioms, syntactically, semantically and discursively (Langlotz, 2006: 1). Consequently, there is no agreed-upon definition and classification of idioms (Cigconi et al., 1999: 280; Langlotz, 2006: 2). Traditionally, idioms have been viewed as conventional multiword expressions (MWEs) that are structurally fixed and semantically opaque (Langlotz, 2006: 2). The opposing camp views idioms as expressions that allow for lexical and syntactic variation, albeit each to a different degree (Gibbs, 1989a; Villavicencio & Copestake, 2002).

Adopting the view that idioms are not frozen semantic units, this thesis seeks to further analyze this group of MWEs. Specifically, amongst the thousands of English idioms, this thesis focuses on **possessive idioms**, which are termed as “verb-based prototypical idioms” (O’Grady, 1998; Nenonen, 2007; both cited in Sameha, 2012: 5). However, we will see in Section 6.1.2 that possessive idioms are not always verb-based, and may instead be preposition-based.

Nonetheless, at this point, possessive idioms are expressions whereby the verb or verb phrase (VP) precedes a prepositional phrase (PP) or noun phrase (NP), within which the noun is possessed by an entity, which is typically the subject (Sameha, 2012: 5). A **possessive adjective** (*my*, *her*, *our*...) marks this possessor-possessed relationship. (1a) illustrates a **co-indexed possessive idiom** whereby the subject *Mary* co-indexes with the pronoun *her* to demonstrate that the noun *mind* is possessed by *Mary* and conveys “Mary becomes crazy”. Contrastingly, (1b) is ungrammatical in the intended reading since the gender and number of the possessive adjective (neutral and plural) disagrees with that of *Mary* (feminine and singular).

- (1) a. Mary loses her mind. “Mary becomes crazy”
 b. #Mary loses their mind.

The subject need not always co-index with the possessive adjective, as in a **separate possessive idiom**. For instance in *break someone's heart* (2a), the noun *heart* is possessed by the pronoun *my* which is not co-indexed with the subject *Mary*, to express “causes someone great sorrow”. In contrast, (2b) is ungrammatical in the intended reading since the possessive adjective *her* co-indexes with the subject *Mary*.

- (2) a. Mary breaks my heart. “Mary causes me great sorrow”
 b. #Mary_i breaks her_i (own) heart.

By integrating the co-indexed and separate possessive idioms into the **English Resource Grammar (ERG)** and **WordNet**, this paper seeks to analyze the syntactic and semantic behavior of possessive idioms. In addition, the newly created syntactic templates would be run through the British

National Corpus (The British National Corpus, 2007) to determine variants and understand the behaviour of possessive idioms. Lastly, this paper also would examine the relationship between **decomposability** and **plausibility**, which are two dimensions said to affect idiom comprehensibility (Titone & Connine, 1994, 1999).

The next section examines past linguistics research; following which, the methodology is explained. Section 6 presents the results and a syntactic and semantic discussion of the idioms while Section 7 suggests areas for future work.

2 Literature Review

The study of idioms has been long-standing due to their heterogeneous nature which has posed problems to linguists aiming to describe their internal structure. Thus far, there have been many studies on English idioms covering areas such as their syntactic categorisation (Villavicencio & Copestake, 2002), their encoding for computational linguistics (Bond et al., 1996; Sag et al., 2002), their decomposability (Gibbs et al., 1989a,b; Nunberg et al., 1994), and their comprehension and processing (Titone & Connine, 1999; Cacciari & Tabossi, 2014).

The **noncompositional** and **compositional approach** are two approaches to the study of idioms. The former, traditional approach treats idioms as lexical entries whose figurative meanings are arbitrarily associated with their constituent words (Chomsky, 1980; Fraser, 1970; van der Linden, 1992; all cited in Titone & Connine, 1999: 1656). A prototypical example is the idiom *kick the bucket*, whose literal meaning “to hit a bucket with one’s foot” has no semantic relation with the figurative meaning “to die suddenly” (Titone & Connine, 1999: 1656). Thus, the noncompositional approach states that the meaning of an idiom cannot be derived from that of its components.

In comparison, the compositional approach emphasizes on the non-arbitrary internal semantic and syntactic structures of idioms in that the literal meanings of component words do contribute to the figurative meaning (Titone & Connine, 1999: 1661). Nunberg (1978) first posited the idea that idioms may differ in their semantic compositionality (Titone & Connine, 1999: 1661). Correspondingly, Nunberg categorized the idioms into three groups. **Normally decomposable** idioms are those in which a component is used literally in relation to the figurative meaning (such as *question* in *pop the question* which refers to a marriage proposal). **Abnormally decomposable** idioms have components whose meanings are metaphorically linked to the figurative meaning (such as *maker* in *meet one’s maker* which metaphorically relates to gods). Lastly, for **nondecomposable** idioms, their figurative meanings are not compositionally obtained from their constituents (such as *shoot the breeze* “to chat idly and without purpose”).

Adopting this classification, Gibbs et al. (1989a) proposed the “idiom decomposition hypothesis” which posited that idioms vary in the degree to which their individual components constitute the figurative meaning, and that this semantic decomposability affects the syntactic flexibility of idioms. For instance, the decomposable idiom *John laid down the law* can be passivised into *The law was laid down by John*, without changing the figurative meaning “to enforce the rules”. However, the non-decomposable idiom *John kicked the bucket* cannot be passivised into *the bucket was kicked by John*

without affecting its idiomatic meaning “to die suddenly”. The relationship between semantic decomposability and syntactic variability is not always direct. For instance, the abnormally decomposable idiom *carry a torch* “to be in love with” tended not to produce the figurative meaning in its passivised form *A torch for Sally was carried by Jim*. Gibbs et al. (1989b) also suggested that decomposability is related to lexical flexibility. For example, the decomposable idiom *button one’s lips* can be changed to *fasten one’s lips* without altering its figurative meaning, “to not say anything”. However, if the nondecomposable idiom *kick the bucket* is lexically changed to *kick the pail*, its figurative meaning, “to die suddenly” is lost.

In a more recent study, Nunberg et al. (1994) refined their argument and distinguished between idiomatically combining expressions (ICEs) and idiomatic phrases (IPs) – a distinction which could be determined using five tests, namely modification, quantification, topicalization, ellipsis, and anaphora. ICEs are idioms whose individual components contribute to the figurative meaning. For instance, in *pull strings*, *pull* is associated with “exploitation or exertion” while *strings* is metaphorically extended to “personal relations”. Contrarily, IPs are idioms whose figurative meaning cannot be derived from their constituents, such as *saw logs*, whose meaning “sleep” cannot be obtained from the idiom’s individual words.

However, there have been opponents to Nunberg et al.’s (1994) absolute, dichotomic separation of ICEs from IPs. As mentioned earlier, Gibbs et al. (1989a,b) view decomposability as a matter of degree. Additionally, Espinal and Mateu (2010) used the idiom *V one’s head off* to illustrate that the dichotomy view is empirically inadequate, since such idioms exhibit some compositionality but neither pass all the five syntactic tests nor retain the aspectuality of the literal meaning. Titone and Connine (1999) also proposed a hybrid model, which attributes both noncompositional and compositional characteristics to idioms.

The aforementioned approaches clearly illustrate that **compositionality** and **decomposability** are often used synonymously, hence necessitating their delineation. According to Abel (2003: 332), the former is a theoretical concept important for generative grammar and which focuses on the combination of syntactic components and their phrasal and sentential meanings. In contrast, decomposability is an attribute of idioms that involves a psycholinguistic aspect since it concerns speakers’ perceptions of how individual word meanings contribute to idiomatic meanings. Unlike a nondecomposable idiom, a **decomposable** idiom is one whose component words contribute to its idiomatic meaning. To differentiate between compositionality and decomposability, Abel (2003: 333) used the idioms *miss the boat* and *kick the bucket*. In the former decomposable idiom, the verb *miss* contributes compositionally to the idiomatic meaning, in that something is missed, or more specifically, “an opportunity is missed”. Whereas in a truly compositional (or literal) reading, “a boat is missed”. In the latter nondecomposable idiom *kick the bucket*, the literal meaning is compositional while the idiomatic meaning *to die suddenly* is noncompositional. There are also nondecomposable idioms that do not have any compositional reading, such as *shoot the breeze* “to chat idly”. The awareness of the difference between compositionality and decomposability would in turn influence the scope and methodology that would be further elaborated on in subsequent sections.

Besides compositionality, idioms could be analyzed according to **transparency** and **convention-**

ality. Transparency refers to the ease of recovering the motivation for the figurative meaning (Nunberg et al., 1994: 496), while conventionality is the degree to which figurative meanings cannot be predicted purely from knowledge of the constituent words and of the rules of a language environment (Nunberg et al., 1994: 492). Titone and Connine (1994) posited a fourth dimension – **literality** – which refers to the possibility for a literal reading. Literality in this sense deals with truth-conditional literality (Cacciari, 2014: 29), otherwise termed as **plausibility** (Fellbaum, 2014: 787).

Studies have examined the relationship between compositionality and literality in influencing idiom processing. Titone and Connine (1999) demonstrated that idiom comprehension was faster when literal word meanings contributed to both figurative and literal interpretations of the expression (i.e. decomposable idioms), than when they contributed only to the literal interpretation (i.e. nondecomposable idioms). Libben and Titone (2008) showed that idiom comprehension was more difficult when an idiom is literally plausible and nondecomposable, than when the idiom is literally implausible. However, other studies demonstrated that literality plays a lesser role in idiom activation. A study by Caillies and Butcher (2007) demonstrated that ambiguous decomposable idioms were activated faster than nondecomposable idioms, thus suggesting that the figurative meaning of decomposable idioms are stored in people’s memory and can be automatically recovered from the semantics of the expression, regardless of the literality degree – results that were also supported by another study (Caillies & Declerq, 2011: 19). However, Caillies and Butcher (2007: 100) mentioned that further research should explore the influence of compositionality on the activation of implausible idioms.

Regarding possessive idioms in particular, a previous study (Sameha, 2012) examined the syntactic and semantic structures of 307 possessive idioms. Idioms were categorized into six VP idiom types which were further analyzed into nineteen syntactic templates. Sameha concluded that conceptual metaphors and image schemas could be used to explain how literal readings of idioms were linked to their figurative senses. There was an intention to add the idioms into the ERG, which was not executed in the end. Moreover, the idioms were not run through a large enough corpus to determine the accuracy of the analysis.

There was a follow-up study (Ho & Bond, 2014) which explored the decomposability of possessive idioms and linked them to WordNet. Results demonstrated that most idioms were nondecomposable and the most common patterns for co-indexed and separate possession idioms were $X_{NP} V_I X's N_I$ and $X_{NP} V_I Y_{NP}'s N_I$ respectively. However, the authors acknowledged that more research had to be done, since the frequency results of the decomposable and nondecomposable idioms conflicted with that obtained in other studies (Gibb et al., 1989a; Villalencio & Copestake, 2002; both cited in Ho & Bond, 2014: 8). Furthermore, the research lacked a precise quantitative analysis, and similar to Sameha’s study, analyses were not implemented in a grammar such as the ERG.

3 Scope and Aims

This thesis thus draws upon previous studies (Sameha, 2012; Ho & Bond, 2014) with the main aim of describing the syntax and semantics of possessive idioms in good detail, so as to better analyze their behaviour and work towards a broader goal of incorporating them into the ERG and WordNet.

A point to note is the large syntactic variation in idioms. In a study which looked at the most frequent English idioms (Villavicencio & Copestake, 2002), the idioms were first grouped syntactically which led to too much variation. Subsequently, a semantic classification which separated idioms into decomposable and non-decomposable ones was deemed as a better approach. Nevertheless, this thesis would still be examining the syntactic variation of possessive idioms due to the broader aim of incorporating them into the grammar and lexicon. This involves describing the syntax and semantics of idioms in great detail, such as their syntactic templates, figurative meanings and WordNet senses of their components and paraphrase. This would permit a better quantitative and qualitative analysis of the behaviour of idioms. Using the syntactic templates to integrate the idioms into the ERG would also allow the idioms to be constructed for analysis and parsed in a corpus. Additionally, by representing the idioms using WordNet, they could be run across a larger corpus to obtain variants for analysis.

In addition, this paper adopts the compositional approach. Within which, instead of taking the strict dichotomy as espoused by Nunberg et al. (1994), this paper views decomposability as a matter of degree. Moreover, as mentioned in Section 2, this paper distinguishes between compositionality and decomposability, whereby the former is a theoretical concept while the latter refers to how constituents contribute to an idiom's figurative meaning. Henceforth, this paper would examine the decomposability of English possessive idioms, which could in turn help to explain the relatively high percentage of nondecomposable idioms in Ho and Bond's (2014) study.

Another area of focus is the plausibility of idioms, which was one of the four dimensions in which idioms can be analyzed (Titone & Connine, 1994), and which has a possible relation with compositionality (refer to Section 2). This paper hypothesizes that a more decomposable idiom would be more plausible. This is because a less decomposable (or nondecomposable) idiom would have to have its meaning interpreted as a whole and thus it would be difficult to obtain a literal reading. For instance, while the decomposable idiom *earn one's keep* means "to earn one's livelihood", it can also have a literal reading. In contrast, the words in the nondecomposable idiom *make one's bones* do not combine to constitute the figurative meaning, "to take actions to establish achievement, status or respect". Besides, it is impossible in reality to create one's bones. Another reason for focusing on plausibility as opposed to transparency and conventionality is due to the relative ease in judging plausibility.

Therefore, this paper has three main purposes centering around possessive idioms: first, to describe and analyze their syntax and semantics so as to integrate them into the ERG and WordNet; second, to describe their decomposability; third, to study the relation between decomposability and plausibility.

4 Motivations

There are several reasons for focusing on possessive idioms. Firstly, they are one of the more common types of English idioms, as demonstrated in Villavicencio et al.'s study (2002), whereby 16.9% of the most frequent English idioms were possessive idioms. Their relative high frequency thus necessitates that they should be analysed more thoroughly.

Secondly, there is discontinuity in the possessive idiom, whereby the genitive slot can be replaced with any possessive adjective or noun (Fellbaum, 1998b), thereby changing the figurative meaning.

For example, *lose one's mind* can appear as *I lose my mind* “I become crazy”, or *Kim loses her mind* “Kim becomes crazy”. Evidently, there is interaction between syntax and semantics which makes possessive idioms hard to represent in the grammar and lexicon. Moreover, there might be long-distance dependencies, such as in *Kim never intended to lose her mind*, where the subject *Kim* does not appear right next to the idiom *lose her mind*. A simple pattern matcher is unable to recognise the discontinuity, which makes it difficult for possessive idioms to be represented in the grammar and lexicon (Wehrli, 1998: 1389). This thesis thus hopes to resolve this by shedding light on the syntactic and semantic behaviour of possessive idioms.

Possessive idioms are also problematic for Natural Language Processing (NLP) applications such as parsing (Sag et al., 2002) and machine translation (Shutova et al., 2012; cited in Muzny & Zettlemoyer, 2013: 1417). For instance, while the possessive adjective usually have to be omitted when translating from English to another language, it must be produced and agree with the subject when the translation goes the other way (Bond et al., 2013). The following examples which compare English with French [(3)-(4)], Japanese [(5)-(6)], and German (7) illustrate that while the possessive adjective is used in the English idioms, there is no equivalent in the target languages.

- (3) *El presidenti treia el fetgei per la boca*
 the president got+out the liver through the mouth
 “The president worked his guts out.” [French: Mateu and Espinal, 2007: 42]
- (4) *A la gerentai lii sortien els ullsi de la cara*
 to the manager CL went+out the eyes of the face
 “The manager cried her eyes out.” [French: Mateu and Espinal, 2007: 42]
- (5) *kanojo-wa chie-o shibotta*
 she-TOP knowledge-OBJ wrung
 “She racked her brains.” [Japanese: Bond et al., 1996: 2]
- (6) *kare-wa kono-shigoto-kara te-o hiita*
 he-TOP this-work-from hands-OBJ pulled
 “He washed his hands of this work.” [Japanese: Bond et al., 1996: 7]
- (7) *jemanden eine lange Nase machen*
 to-somebody a long nose make
 “thumb one's nose at somebody” [German: Fellbaum, 2014: 785]

Clearly, possessive idioms are a relatively unique characteristic of English. Consequently, a machine translation system has to determine when and which pronouns to use. A clearer syntactic and semantic description of possessive idioms would allow for a better understanding of them and in turn enhance NLP applications, such as machine translation and computer-assisted language learning. It

would also aid in incorporating idioms into dictionaries, which would be beneficial especially for non-native speakers, due to the difficulty in understanding idioms.

5 Methodology

The initial list of possessive idioms was obtained from Sameha's (2012) paper on English possessive idioms. These were in turn sourced from WordNet, augmented by the Collins Cobuild Idioms Dictionary and checked against Dictionary.com to give 370 idioms (Sameha, 2012: 13). I then enhanced the list of idioms by verifying against two online dictionary resources, Dictionary.com and Oxforddictionaries.com. The former is a popular online dictionary and contains data from multiple dictionaries (Sameha, 2012: 13), while the latter has been documenting the English language for many centuries. This thus ensured the comprehensiveness and reliability of the data. Similar to Sameha's selection criteria, the additional idioms were chosen based on the main criteria of the NP containing the possessive adjective *one's*. This resulted in about hundred more idioms. However, the final count was at a grand total of 514 idioms. This was because some of these idioms were essentially the same idioms, except that they either appear in both possession types, for instance *change one's mind* versus *change someone's mind*, or were semantic variants, such as *collect one's wits* and *gather one's wits*, or conveyed more than one idiomatic meaning, such as *make one's way*.

These 514 idioms were broadly categorized into co-indexed and separate possessive idioms and further grouped syntactically, and arranged according to their frequency as shown in Tables 1 and 2 respectively. A range of symbols was used to represent the various syntactic templates. X_{NP} denotes variable noun phrases since it can refer to either *Kim* (8a) or *John* (8b). The same applies to Y_{NP} , as evident in (9). While N denotes invariable nouns that predominantly belong to the possessive idiomatic construction, V denotes verb, P for preposition or particle, A for adjective, R for adverb, D for determiner, aux for auxiliary and neg for negation. These abbreviations also have numbers appended to them as there may be more than one of each syntactic category in a single idiomatic expression, as shown in example (9), whereby N_1 refers to *piece* whereas N_2 refers to *mind*. Additionally, referring to Tables 1 and 2, square brackets [] were used in the syntactic templates to denote prepositional phrases (PP) within the idioms. Within these brackets, P denotes a preposition since it heads a PP, whereas in other cases, P represents a particle.

- (8) a. *Kim loses her mind.*
 X_{NP} V_1 X 's N_1
 "Kim becomes crazy."
- b. *John loses his mind.*
 X_{NP} V_1 X 's N_1
 "John becomes crazy."

(9) a. *Kim gives John a piece of her mind.*

$X_{NP} V_1 Y_{NP} D_1 N_1 P_1 X's N_2$

“Kim scolds John harshly.”

b. *John gives Kim a piece of his mind.*

$X_{NP} V_1 Y_{NP} D_1 N_1 P_1 X's N_2$

“John scolds Kim harshly.”

Table 1
Types of Co-indexed Possessive Idioms

Structure	Example	Frequency
$X_{NP} V_1 X's N_1$	lose one's mind	137
$X_{NP} V_1 [P_1 X's N_1]$	fly off one's handle	40
$X_{NP} V_1 X's N_1 [P_1 Y_{NP}]$	cast one's lot [with someone/thing]	39
$X_{NP} V_1 X's N_1 [P_1 D_1 N_2]$	have one's head [in the clouds]	27
$X_{NP} V_1 X's N_1 P_1$	cry one's eyes out	22
$X_{NP} V_1 X's own N_1$	blow one's own horn	18
$X_{NP} V_1 + P_1 X's N_1$	pull up one's socks	17
$X_{NP} be [P_1 X's N_1]$	off one's rocker	13
$X_{NP} V_1 X's N_1 [P_1 X's N_2]$	scratch one's ear [with one's elbow]	13
$X_{NP} V_1 D_1 N_1 [P_1 X's N_2]$	a dose [of one's medicine]	10
$X_{NP} V_1 X's N_1 A_1$	get one's hands dirty	10
$X_{NP} V_1 Y_{NP} [P_1 X's N_1]$	wind someone [around one's finger]	10
$X_{NP} V_1 X's N_1 (est)$	do one's best	8
$X_{NP} V_1 [P_1 X's N_1 [P_2 Y_{NP}]]$	pour out one's heart [to someone]	7
$X_{NP} aux+neg V_1 X's N_1$	not mince one's words	5
$X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 X's N_2]$	give someone a piece [of one's mind]	4
$X_{NP} V_1 R_1 A_1 [P_1 X's N_1]$	too big [for one's boots]	3
$X_{NP} V_1 [P_1 D_1 N_1 P_2 X's N_2]$	by the skin of one's teeth	2
$X_{NP} V_1 N_1 [P_1 X's N_2]$	have egg [on one's face]	2
$X_{NP} V_1 X's N_1 [P_1 X]$	have one's wits [about one]	2
$X_{NP} V_1 X's N_1 and V_2 N_2$	have one's cake and eat it	2
Remainder	let grass grow under one's feet	30
Total		421

This table lists the co-indexed possessive idioms, arranged in order of frequency with the exception of the last group, *remainder*.

The idioms elicited twenty-one syntactic structures for the co-indexed possessive idioms and thirteen for the separate possessive idioms, with both groups comprising an additional category titled *remainder*, which consists of idioms that belong to none of the other categories. Referring to Table 1, most of the categories follow the basic structure $X_{NP} V_1 X's N_1$ but with modifications such as PPs,

adjectives and negation. Specifically, some templates would be elaborated upon.

The first is $X_{NP} V_1 + P_1 X's N_1$, for instance *pull up one's socks* “to make an effort to improve one's behaviour or performance”, in which + denotes that the particle can also appear at the end of the idiomatic phrase without affecting the figurative meaning, as in *pull one's socks up*. In $X_{NP} be [P_1 X's N_1]$, such as *off one's rocker* “to be crazy”, the verb *be* is used instead of V_1 as in the basic syntactic category $X_{NP} V_1 X's N_1$. The difference here is that idioms in the former group usually appear with the *be*-verb, for instance *Kim is off her rocker* “Kim is crazy” and *Kim is beyond her depth in physics class* “Kim is outside of her knowledge in physics”. Thus, they are placed under a separate template. The idioms in the group $X_{NP} V_1 X's N_1(est)$, for example *do one's best* “to do one's utmost possible”, also warrant their own category instead of being placed in the basic group. This is because the nouns are the superlative forms of their respective adjectives and are decomposed as such in the ERG. For instance, “best” is analyzed as “most good thing” while “worst” as “most bad thing”. Grouping these idioms with those that belong to the group $X_{NP} V_1 X's N_1$ would make their incorporation into the ERG difficult. Instead, having their own template would help in the subsequent ERG incorporation.

As for Table 2, one syntactic template to note is $X_{NP} V_1 Z_{NP} [P_1 Y_{NP}'s N_1]$, which has an additional denotation Z_{NP} that refers to a variable NP in the idiomatic expression. For instance, the idiom *ram something down someone's throat* can appear as *Kim rams ideas down Jim's throat* or *they rammed nonsense down my throat*.

Table 2
Types of Separate Possessive Idioms

Structure	Example	Frequency
$X_{NP} V_1 [P_1 Y_{NP}'s N_1]$	breathe [down someone's neck]	21
$X_{NP} V_1 Y_{NP}'s N_1$	catch someone's eye	18
$X_{NP} V_1 D_1 N_1 [P_1 Y_{NP}'s N_2]$	a thorn [in someone's flesh]	10
$X_{NP} V_1 N_1 [P_1 Y_{NP}'s N_2]$	send shivers [down someone's spine]	6
$X_{NP} V_1 Y_{NP}'s N_1 [P_1 N_2]$	put someone's mind [at rest]	5
$X_{NP} V_1 Y_{NP}'s N_1 P_1$	bite someone's head off	5
$X_{NP} V_1 neg N_1 [P_1 Y_{NP}'s N_2]$	no skin [off someone's back]	3
$X_{NP} V_1 X [P_1 Y_{NP}'s N_1]$	put oneself [in someone's shoes]	3
$X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 Y's N_1]$	give someone a run [for someone's money]	3
$X_{NP} V_1 Y_{NP} [P_1 Y's N_1]$	bring someone [to someone's knees]	2
$X_{NP} V_1 Y_{NP} R_1 [P_1 Y's N_1]$	set someone back [on someone's heels]	2
$X_{NP} V_1 Y_{NP} Y's N_1$	give someone someone's dues	2
$X_{NP} V_1 Z_{NP} [P_1 Y_{NP}'s N_1]$	ram something [down someone's throat]	2
Remainder	leave a bad taste in someone's mouth	11
Total		93

This table lists the separate possessive idioms, arranged in order of frequency with the exception of the last group, *remainder*.

With reference to (10) and (11) – henceforth termed as **idiom entries** for reference purpose – the next step involved providing the definition of each idiom, by referring to the Collins Cobuild Idioms Dictionary, Dictionary.com and Oxforddictionaries.com. Thereafter, I assessed the literal senses of the individual components of each idiom and should the idiom be regarded as decomposable, a corresponding synonym or metaphorical extension for each component word was decided upon and marked with an asterisk as illustrated in (10). In all instances, the literal and figurative senses were determined and carefully chosen from WordNet. The decomposability of an idiom was listed using the feature *@type*.

To reiterate, idiom decomposability refers to how individual components contribute to the figurative meaning. In this thesis, this was determined by **semantic substitution**, which is the possibility of substituting each lexical component with an appropriate word without altering figurative meanings. For instance, in (10), *eat* can be metaphorically extended to mean “swallow/withdraw” (**V_I*) while *words* with “statement” (**N_I*), to give “withdraw one’s statement”. The figurative meaning of *eat one’s words* generally remains unchanged, thus rendering this idiom decomposable. In contrast, in (11), *twiddle* and *thumb* cannot be replaced with suitable synonyms nor metaphorical extensions, without altering the syntactic structure and figurative meaning “to do nothing”. Consequently, *twiddle one’s thumb* is a nondecomposable idiom. To further exemplify, semantic substitution is the replacement of each lexical component with an equivalent synonym. Thus, in *nail one’s colour to the mast*, its main lexical parts – *nail*, *colour* and *mast* – cannot be semantically substituted without affecting the figurative meaning “to explicitly voice one’s opinion regarding an issue”, hence rendering this idiom nondecomposable. Thus, semantic substitution is important for determining decomposability in this thesis.

Lastly, all the idioms were given paraphrases which were restricted, linked to WordNet and marked with @ in the idiom entries. In cases whereby the paraphrase was headed by a verb (10), the verb paraphrase (@*V*) would be the hypernym¹ of the idiom; whereas when the paraphrase involved the be-form (11), the adjective paraphrase (@*A*) would be the hypernym of the idiom. Thus, by restricting the paraphrase and linking its senses to WordNet in the idiom entries, this allowed the basic essence of each idiom to be captured and illustrated its hyponymy relation to lexical entries already listed in WordNet.

¹Fellbaum (1998b: 56) suggested using the concept of “troponymy” to accommodate verb phrase idioms into WordNet. However, I would replace this term with “hyponymy/hyponymy” which would be further explained in Section 6.4.1.

(10)	<i>Idiom entry</i>
Index form	eat one's words
Template	$X_{NP} V_1 X's N_1$
Example	Kim eats her words
Example	Kim is going to have to eat her words
Definition	to retract one's statement, especially with humility
V_1	S: (v) eat (take in solid food)
N_1	S: (n) words (the words that are spoken)
$*V_1$	S: (v) swallow, take back, unsay, withdraw (take back what one has said)
$*N_1$	S: (n) statement (a message that is stated or declared; a communication (oral or written) setting forth particulars or facts etc)
@type	decomposable
Paraphrase	X retracts statement
@template	X V N
@V	S: (v) abjure, recant, forswear, retract, resile (formally reject or disavow a formerly held belief, usually under pressure)
@N	S: (n) statement (a message that is stated or declared; a communication (oral or written) setting forth particulars or facts etc)

(11)	<i>Idiom entry</i>
Index form	twiddle one's thumbs
Template	$X_{NP} V_1 X's N_1$
Example	Kim twiddles her thumbs
Definition	to do nothing
V_1	S: (v) twiddle, fiddle with (manipulate, as in a nervous or unconscious manner)
N_1	S: (n) thumb, pollex (the thick short innermost digit of the forelimb)
@type	Nondecomposable
Paraphrase	X is idle
@template	X BE A
@A	S: (adj) idle (not in action or at work)

Referring to (10) and (11), it is evident that the idiom entries clearly describes the syntax and semantics of the idioms by encompassing their syntactic templates, sentence examples, definitions, literal and figurative meanings, decomposability and paraphrases. With a total of 514 idioms, I thus created the most substantial, detailed description of possessive idioms in existence – a major contribution to the study of idioms.

The various syntactic templates were then incorporated into the English Resource Grammar (ERG: Flickinger, 2011), which is a broad-coverage, declarative English grammar built upon the Head-driven Phrase Structure Grammar schema, with the help of Flickinger himself. To enable the recognition of the idioms in parsing, the verbs and nouns in the idiomatic expression had to be linked together. This was done with the **id.rel** marker on the verb to link it with its argument. Another marker was **'s** **poss.rel** which marks the possessive relationship between the possessor and possessed noun, either lexicalized as *'s* or incorporated into the possessive determiner (*my, her, our...*).

Using these newly created ERG syntactic templates, we were able to automatically construct, compile and parse the idioms. This was done by combining the ERG templates with the lexicon in the Type Description Language (Krieger & Schäfer, 1994) and idioms in the Machine Translation Rules (Oepen, 2008), before running the idioms through the Answer Constraint Engine (Packard, 2013). Constituent trees, feature structures and Minimal Recursion Semantics (MRSs: Copestake et al., 2005) of the idioms were obtained from the Linguistic User Interface (LkbLui, n.d.).

Example (12a) illustrates a co-indexed possessive idiom with the structure $X_{NP} V_1 X's N_1$, with its MRS in (12b) and its simplified version in (12c) for easier understanding. There are several important elements present in the grammar of *get one's break*. The first is the idiomatic rule (12a) that is obtained from incorporating the syntactic template listed in its idiom entry into the ERG, and which ensures that the predicates and the possessive relation are present. Referring to (12b), the second element would be the lexical entries *_get_v_i* and *_break_n_1_rel*. In (12c), the entry *_get_v_i* associates the proper noun *Kim* with the verb *break*, while *poss.rel* indicates that the noun *break* is possessed by the subject, which co-indexes with the possessive adjective. The last element is *id_rel* which links the verb *get* with the proper noun *Kim* (the subject) and the pronoun *her* (the possessor of the object). This association thus captures minor variations, such as number and spelling variations (Bond et al., 2013: 4). For instance, in *(w)rack one's brains*, both the singular and plural forms of *brain* are accepted in dictionaries. During the construction of such idioms, the number would be underspecified so as to permit both the singular and plural entries. Similarly, *rack* can be spelled as *wrack*. Assuming that both words carry the same meaning, *id_rel* thus allows two orthographically different words to have the same predicate and be recognised as the same idiom during parsing.

(12) a. Kim_i gets her_i break.

[$X_{NP} V_1 X's N_1$]

b.

<i>mrs</i>				
TOP	[0] <i>h</i>			
INDEX	[2] <i>e</i>			
RELS	$\left[\begin{array}{l} \textit{proper_q_rel} \\ \text{LBL} \quad [4] \textit{h} \\ \text{ARG0} \quad [3] \textit{x} \\ \text{RSTR} \quad [5] \textit{h} \\ \text{BODY} \quad [6] \textit{h} \end{array} \right],$	$\left[\begin{array}{l} \textit{named_rel} \\ \text{LBL} \quad [7] \textit{h} \\ \text{CARG} \quad \textit{"kim"} \\ \text{ARG0} \quad [3] \textit{x} \end{array} \right],$	$\left[\begin{array}{l} \textit{_get_v_i_rel} \\ \text{LBL} \quad [1] \textit{h} \\ \text{ARG0} \quad [2] \textit{e} \\ \text{ARG1} \quad [3] \textit{x} \\ \text{ARG2} \quad [9] \textit{x} \end{array} \right],$	$\left[\begin{array}{l} \textit{id_rel} \\ \text{LBL} \quad [1] \textit{h} \\ \text{ARG0} \quad [10] \textit{i} \\ \text{ARG1} \quad [3] \textit{x} \\ \text{ARG2} \quad [11] \textit{x} \end{array} \right],$
RELS	$\left\langle \begin{array}{l} \textit{poss_rel} \\ \text{LBL} \quad [15] \textit{h} \\ \text{ARG0} \quad [16] \textit{e} \\ \text{ARG1} \quad [9] \textit{x} \\ \text{ARG2} \quad [11] \textit{x} \end{array} \right\rangle,$	$\left\langle \begin{array}{l} \textit{pronoun_q_rel} \\ \text{LBL} \quad [17] \textit{h} \\ \text{ARG0} \quad [11] \textit{x} \\ \text{RSTR} \quad [18] \textit{h} \\ \text{BODY} \quad [19] \textit{h} \end{array} \right\rangle,$	$\left\langle \begin{array}{l} \textit{pron_rel} \\ \text{LBL} \quad [20] \textit{h} \\ \text{ARG0} \quad [11] \textit{x} \end{array} \right\rangle,$	$\left. \right\rangle$
RELS	$\left[\begin{array}{l} \textit{def_explicit_q_rel} \\ \text{LBL} \quad [12] \textit{h} \\ \text{ARG0} \quad [9] \textit{x} \\ \text{RSTR} \quad [13] \textit{h} \\ \text{BODY} \quad [14] \textit{h} \end{array} \right],$	$\left[\begin{array}{l} \textit{_break_n_l_rel} \\ \text{LBL} \quad [15] \textit{h} \\ \text{ARG0} \quad [9] \textit{x} \end{array} \right]$		
HCONS	$\left\langle \begin{array}{l} \textit{qeq} \\ \text{HARG} \quad [0] \textit{h} \\ \text{LARG} \quad [1] \textit{h} \end{array} \right\rangle,$	$\left\langle \begin{array}{l} \textit{qeq} \\ \text{HARG} \quad [5] \textit{h} \\ \text{LARG} \quad [7] \textit{h} \end{array} \right\rangle,$	$\left\langle \begin{array}{l} \textit{qeq} \\ \text{HARG} \quad [13] \textit{h} \\ \text{LARG} \quad [15] \textit{h} \end{array} \right\rangle,$	$\left\langle \begin{array}{l} \textit{qeq} \\ \text{HARG} \quad [18] \textit{h} \\ \text{LARG} \quad [20] \textit{h} \end{array} \right\rangle$
ICONS	$\langle \rangle$			

c.

get	(Kim _i , break)
poss_rel	(her _i , break)
id_rel	(Kim _i , her _i , get)

The other idiom types were constructed in a similar manner. However, some were done manually either because their frequency was too low to render the creation of their ERG templates or because of the way constituents are written in the ERG. For instance, in (13), the idiom *do one's best* cannot properly fit into the ERG template $X_{NP} V_I X's N_I$ because of two reasons; the first being that *best* is the superlative of the word *good* while the second being that there exists an additional relation, whereby *good* refers to a “thing”, which is absent from the idiom itself. This posed a problem to its automatic construction and compilation and hence had to be manually added into the database.

(13) a. Kim_i does her_i best.

b.
$$\left[\begin{array}{l} \text{do} \quad (\text{Kim}_i, \text{thing}) \\ \text{best} \quad \left[\begin{array}{l} \text{most} \quad (\text{good}) \\ \text{good} \quad (\text{thing}) \end{array} \right] \\ \text{poss} \quad (\text{her}_i, \text{thing}) \\ \text{id_rel} \quad (\text{Kim}_i, \text{her}_i), \text{do} \end{array} \right]$$

In this thesis, ERG templates and the automatic construction and compilation were done only for the co-indexed possession idioms, since they are more problematic for translation. However the same framework can be applied to the separate possessive idioms as future work. Moreover, due to the complex syntactic nature of many co-indexed possessive idioms, analyses were only provided for those with frequencies higher than ten (Table 1).

After automatically constructing and compiling the idiomatic entries, the entries in the ERG were combined with WordNet to obtain variants. This was possible since the senses of the component words were linked to WordNet right from the beginning during the creation of the idiom entries (refer to (10) and (11)). Consequently, a depth command was introduced and restricted in order to generate synonyms and hyponyms which were one level down from the verbs and nouns in each idiom. The variants were then run through the British National Corpus (BNC: The British National Corpus, 2007), a large corpus of 100 million word samples of written and spoken British English, in order to determine the syntactic and semantic flexibility of the possessive idioms. Due to time constraints, only the 137 idioms belonging to the category $X_{NP} V_I X's N_I$ were run through the BNC. This entailed two preliminary experiments. The first involved manually determining how many listings of two particular idioms were lost during parsing while the second was a fully automatic parsing of the 137 idioms and their WordNet variants to examine how many were interpreted as idioms. It also entailed manually analyzing the sentences which were parsed as containing idioms, and judging if the parses were accurate.

A decomposability scale was also generated from the idiom entries so as to better examine the decomposability of idioms. The scale was tabulated based on two aspects – first being whether the individual component had a corresponding synonym or metaphorical extension (represented by *); second being the similarity between the resulting senses and the original senses of the individual words. To illustrate, referring back to (11), *twiddle one's thumb* is considered nondecomposable since its component parts cannot be paraphrased, as evident from the absence of $*V$ and $*N$.

In comparison, (14) to (16), which are the simplified versions of their respective idiom entries, illustrate decomposable idioms, albeit to different degrees. In (14), the main lexical items *button* and *lip* are metaphorically extended to “control” and “speech” respectively. Whereas in (15), *change* retains its meaning while *tune* extends metaphorically to “opinion”. Lastly, in (16), the idiom *keep one's temper* is decomposable but its component words have the exact same senses in both the literal and figurative senses.

As the methodology involved semantic substitution, the concept of **projection** was used to group the idioms. Idioms such as that in (16) were labelled *non-projected* since the senses of their con-

stituents remained unchanged during semantic substitution. Moving along the scale, idioms similar to (15) were termed as *partially projected* since some of its components retained their original meaning while others mapped onto other words. The next group was similar to (14), which was labelled *fully projectable* since all their components had a one-to-one connection with the words that they were mapped to. On the other end, nondecomposable idioms such as (11) were termed as *non-projectable*. Therefore, the decomposability scale would reflect idiom decomposability by looking at the concept of *projectability*.

In order to examine how decomposability and plausibility correlate, five idioms from each syntactic template were randomly obtained for analysis. Since each idiom had an identification number attached to it, this allowed an integer set generator to generate five numbers at random (Random Integer Set Generator, 1998). For templates which had less than five idioms, analysis involved all the idioms belonging to that particular template. Subsequently, the idiom samples were examined for their decomposability based on the decomposability scale. Finally, they were analysed for their plausibility and a matrix was outlined to determine if the two dimensions interact.

(14)	<i>Idiom entry – fully projected</i>
Index form	button one's lip
Template	$X_{NP} V_1 X's N_1$
Definition	to keep quiet, especially to refrain from revealing information
V_1	S: (v) button (fasten with buttons)
N_1	S: (n) lip (either of two fleshy folds of tissue that surround the mouth and play a role in speaking)
$*V_1$	S: (v) control, hold in, hold, contain, check, curb, moderate (lessen the intensity of; temper; hold in restraint; hold or keep within limits)
$*N_1$	S: (n) speech (something spoken)
@type	decomposable

(15)	<i>Idiom entry – partially projected</i>
Index form	change one's tune
Template	$X_{NP} V_1 X's N_1$
Definition	to change one's opinion, especially since it benefits one
V_1	S (v) switch, shift, change (lay aside, abandon, or leave for another)
N_1	S: (n) tune, melody, air, strain, melodic line, line, melodic phrase (a succession of notes forming a distinctive sequence)
$*V_1 = V_1$	
$*N_1$	S: (n) opinion, sentiment, persuasion, view, thought (a personal belief or judgment that is not founded on proof or certainty)
@type	decomposable

(16)	<i>Idiom entry – non-projected</i>
Index form	keep one's temper
Template	$X_{NP} V_1 X's N_1$
Definition	to refrain from showing anger and remain poised
V_1	S: (v) restrain, keep, keep back, hold back (prevent the action or expression of)
N_1	S: (n) pique, temper, irritation (a sudden outburst of anger)
$*V_1 = V_1$	
$*N_1 = N_1$	
@type	decomposable

6 Results and Discussion

6.1 Syntactic analysis

The comprehensive idiom entries led to the discovery of several novel syntactic aspects of possessive idioms, which would be further discussed in the following subsections.

6.1.1 Alternation in Idioms

An analysis revealed that some of the possessive idioms appeared in **alternation** pairs, which typically have verbs in the *give-get* relationship, as exemplified by six (S/N 1-6) out of the eleven idioms in Table 3.

Though each alternation pair could be found in both their co-indexed and separate possessive structure, it is apparent that possession types and idiom alternation have no relation. This is exemplified in Table 3, whereby both co-indexed (e.g. *get one's dues*) and separate possessive (e.g. *get a piece of someone's mind*) idioms are grouped under *cause-type idiom*; and vice versa for the *result-type idiom* category. Thus, instead of possession type, I argue that it is the verb type that results in the alternation pairs. Furthermore, rather than a reciprocal relationship between the alternation pairs, I contend that the association is one-way, whereby the idioms typically with *give* verbs – henceforth referred as **cause-type idioms** – result in those with *get* – hereafter termed as **result-type idioms**.

An analysis of the 514 idioms revealed that many other idioms have *get* or *have* verbs, but do not have their cause-type alternations. Examples include *get one's hands dirty* “to do something shameful or illegal” and *have one's eye on something* “to have as one's objective”. The same applies to separate possessive idioms, like *get under someone's skin* “to annoy or to impact someone deeply” and *get the monkey off someone's back* “to end a problem”. In comparison, the result-type alternation could be found for all the idioms with the *give* verb. Consequently, we can conclude that it is the cause-type idioms that result in their result-type alternations.

The initial analysis was that it was typically idioms which contain *give* that lead to their alternations which contain *get* or *have*. However, Table 3 shows that both groups involve other verbs, such as *put*

Table 3
Alternation in Idioms

S/N	Cause-type Idiom	Result-type Idiom
1	give someone someone's dues $X_{NP} V_1 Y_{NP} Y's N_1$	get one's dues $Y_{NP} V_1 Y's N_1$
2	give someone someone's walking papers $X_{NP} V_1 Y_{NP} Y's N_1$	get one's walking papers $Y_{NP} V_1 Y's N_1$
3	give someone a dose/taste of someone's own medicine $X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 Y's N_1]$	get/receive/have/need a dose/taste of one's own medicine $Y_{NP} V_1 D_1 N_1 [P_1 Y's N_2]$
4	give someone a run for someone's money $X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 Y's N_1]$	get/have a run for one's money $Y_{NP} V_1 D_1 N_1 [P_1 Y's N_2]$
5	give someone a piece of one's mind $X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 X's N_2]$	get a piece of someone's mind $Y_{NP} V_1 D_1 N_1 [P_1 X_{NP}'s N_2]$
6	give someone the rough edge of one's tongue $X_{NP} V_1 Y_{NP} D_1 A_1 N_1 [P_1 X's N_2]$	get the rough edge/side of someone's tongue $Y_{NP} V_1 D_1 A_1 N_1 [P_1 X_{NP}'s N_2]$
7	put someone's nose out of joint $X_{NP} V_1 Y_{NP}'s N_1 [P_1 N_2]$	get/have one's nose out of joint $Y_{NP} V_1 Y's N_1 [P_1 N_2]$
8	bring someone out of someone's shell $X_{NP} V_1 Y_{NP} [P_1 Y's N_1]$	come out of one's shell $Y_{NP} V_1 [P_1 Y's N_1]$
9	get/have/take someone under one's wing $X_{NP} V_1 Y_{NP} [P_1 X's N_1]$	be/come under someone's wing $Y_{NP} V_1 [P_1 X_{NP}'s N_1]$
10	have/keep someone under one's thumb $X_{NP} V_1 Y_{NP} [P_1 X's N_1]$	be under someone's thumb $Y_{NP} V_1 [P_1 X_{NP}'s N_1]$
11	let someone/thing slip through one's fingers $X_{NP} V_1 N_1 V_2 [P_1 X's N_2]$	slip through someone's fingers $Y_{NP} V_1 X_{NP}'s N_1$

The templates listed under *result-type idiom* differ slightly from the original templates listed in Table 1 and Table 2. This is because of the need to show the relation between the idioms in the two alternations and to show that the animate noun denoted by Y_{NP} in the *cause-type idiom* is essentially the same referent expressed in the *result-type idiom*.

and *bring* in the cause-type category, and *come* in the result-type category. The eleven idiom pairs should thus be further analyzed into two groups. Specifically, the cause-type verbs in idioms S/N 1-6 convey a change in possession while those in S/N 7-11 indicate a change in motion. The cause-type idioms thus express a causation of possession or motion, meanings which are also conveyed in double object or prepositional object dative constructions present in English (Levin & Hovav, 2007; Krifka,

2004: 13).

(17a) and (18a) illustrate that there exist three thematic roles – AGENT, OBJECT, THEME/GOAL (Krifka, 2004). In the cause-type idioms, verbs such as *give*, *put* and *bring* permit three thematic roles which in turn convey a change in possession or motion. Specifically, the AGENT does something to cause a change in possession (THEME) or motion (GOAL). Consequently, the “permissive agent” (Oehrle, 1976) acts as a causer which leads to the animate OBJECT receiving or experiencing a THEME or GOAL, as illustrated in (17b) and (18b). Correspondingly, such idioms are termed as *result-type idioms*, which are alternations of their cause-type idioms.

- (17) a. [Kim]_{AGENT} gives [Jim]_{OBJ} his [walking papers]_{OBJ_{THEME}}
 $X_{NP} V_1 Y_{NP} Y's N_1$
 “Kim fires Jim.” *cause-type idiom*
- b. [Jim]_{AGENT} gets his [walking papers]_{OBJ_{THEME}}
 $Y_{NP} V_1 Y's N_1$
 “Jim is fired.” *result-type idiom*
- (18) a. [Kim]_{AGENT} gets [Jim]_{OBJ} under her [wing]_{OBJ_{GOAL}}
 $X_{NP} V_1 Y_{NP} P_1 X's N_1$
 “Kim guides/protects Jim.” *cause-type idiom*
- b. [Jim]_{AGENT} comes under Kim’s [wing]_{OBJ_{GOAL}}
 $Y_{NP} V_1 P_1 X_{NP}'s N_1$
 “Jim is guided/protected by Kim.” *result-type idiom*

Distinguishing between idioms S/N 1-6 and S/N 7-11 revealed two senses of “get” amongst these idioms. The WordNet sense of *get* in the result-type idioms S/N 1-6 is “S: (v) get, acquire (come into the possession of something concrete or abstract)” while that for *have* is “S: (v) have, have got, hold (have or possess, either in a concrete or an abstract sense)”. Comparatively, that of *get* and *have* in idioms S/N 7-9 is “S: (v) get, let, have (cause to move; cause to be in a certain position or condition)”. This is aligned with the analysis of a change in possession or motion.

Additionally, at the beginning of this section, idioms with *give* were said to result in their *get* or *have* counterparts. The only exception found in this study is the idiom *give someone the shirt of one’s back* “to give anything and everything that one possesses”, which had no result-type alternation. This could be due to the focus of the figurative meaning, in that it describes the character of the AGENT as being generous, rather than an action per se. Consequently, it has no result-type alternation.

Regardless, it is interesting to note that the possessive idioms exhibit syntactic behaviour which is characteristic of typical English sentences. Particularly, the change in possession or movement expressed in cause-type idioms allow for the derivation of their result-type alternation. Additionally, two senses of *get* exist, depending on whether the verb suggests a change in possession or movement. This analysis could be applied to a wider group of idioms in future.

6.1.2 Prepositional phrasal idioms

Contrary to the initial definition of possessive idioms as “verb-phrase (VP) prototypical idioms” (Section ??), the results demonstrated the opposite, in that some idioms could have their verbs omitted and be headed by a preposition, and still convey the same figurative meaning.

Table 4 specifies the idioms with omissible verbs. Specifically, S/N 1-6 have their verbs replaced with the preposition *with*. According to Farrell (2009), *with* carries a locative, togetherness sense and has four related senses which are instantiations or metaphorical extensions of this central having-sense meaning. Farrell mentioned that the verb *have* in clauses which convey togetherness relationships, could be replaced by the preposition *with* and still retain the same meaning.

Table 4
List of PP-VP idioms

S/N	Prepositional phrase idiom	Verb phrase idiom
1	with one's back against/to the wall	have one's back against/to the wall
2	with one's head in the clouds	have one's head in the clouds
3	with one's nose to the grindstone	keep one's nose to the grindstone
4	with one's heart in one's mouth/throat	have one's heart in one's mouth/throat
5	with something under one's belt	have something under one's belt
6	with one/both hands tied behind one's back	do something with one/both hands tied behind one's back
7	by the skin of one's teeth	do something by the skin of one's teeth
8	in one's heart of hearts	know/believe in one's heart of hearts

Similarly, in the idioms in Table 4, the verb, which is typically *have*, conveys some locative sense and having relationship. Thus, it can be replaced by *with*, which expresses a togetherness sense, and hence changes from a VP to a PP idiom. This finding demonstrates that possessive idioms are not syntactically frozen, and instead exhibit aspects of the English grammar in terms of a transformation from a VP to PP, as illustrated in (19) and (20). One interesting aspect is the movement of the idiom from sentence-medial to sentence-initial position to retain its grammaticality and meaning. Besides, the preposition could even be omitted ((19c) and (20c)) without affecting the idiomatic meaning.

- (19) a. Kim had her back against the wall as she faced the strong competition.
 b. With her back against the wall, Kim faced strong competition.
 c. Kim's back was against the wall as she faced the strong competition.
- (20) a. Kim had her heart in her mouth when she watched the tennis match.
 b. With her heart in her mouth, Kim watched the tennis match.
 c. Kim's heart was in her mouth when she watched the tennis match.

6.1.3 Transformation in idioms

Table 5 illustrates transformation in terms of passivisation, which is another interesting aspect of possessive idioms.

One concern during analysis was whether these idioms were worth mentioning. As mentioned in Section 3, this thesis views some idiomatic expressions as decomposable. Consequently, they should have some syntactic flexibility and undergo syntactic processes such as modification, quantification, topicalization, ellipsis and anaphora (Nunberg et al., 1994). Thus, the question was whether the idioms listed in Table 5 were a result of syntactic flexibility due to decomposability or whether they were separate lexical entries.

A further examination revealed that only idioms S/N 3-5 were decomposable. Moreover, the five idioms only appear in restricted syntactic structures, namely those listed in Table 5. To illustrate, the idiom *burn one's fingers* never ever appears in the form "one's fingers are burnt" as opposed to *seal one's lips* which can appear as *keep one's lips sealed* or *one's lips are sealed*. Thus, these passivised forms should be recognized as individual idiomatic entries, and not as syntactic variations of each other.

Table 5
Transformation in Idioms

S/N	Basic	Passivised
1	burn one's fingers	get one's fingers burnt
2	cross one's fingers	keep/have/got/with one's fingers crossed fingers crossed (functions like an adverb)
3	dirty one's hands	get one's hands dirty
4	set one's heart on something	have one's heart set on something one's heart is set on something
5	seal one's lips	keep one's lips sealed one's lips are sealed

6.2 Decomposability

Table 6 illustrates the frequency of decomposable and nondecomposable idioms, whereby the former encompasses the three groups of idioms detailed in Section 5, namely the non-projected, partially projected and fully projected idioms, while the latter refers to the *non-projectable* idioms. The number of decomposable idioms remained stable at around one-third of the total number of idioms, regardless of whether they were co-indexed or separate possessive.

Similarly, in the randomized sample (Table 7), there were fewer decomposable idioms (41.1%) than nondecomposable (58.9%) ones. Furthermore, idioms were found in all four categories of projection (Table 7). This demonstrates that idiom decomposability is not a strict dichotomy, and is instead a matter of degree.

Table 6
Frequency of Decomposable and Nondecomposable Idioms

	Decomposable		Nondecomposable		Total
Co-indexed possessive idioms	159	(37.8%)	262	(62.2%)	421
Separate possessive idioms	32	(34.4%)	61	(65.6%)	93
Total number of idioms	191	(37.2%)	323	(62.8%)	514

Table 7
Decomposability Scale of Idiom Sample

	Decomposable			Nondecomposable
	Non-projected	Partially Projected	Fully Projected	Non-projectable
Frequency	8 (8.4%)	19 (20.1%)	12 (12.6%)	56 (58.9%)
Total frequency	39 (41.1%)			56 (58.9%)

Refer to Table 13 in Appendix A for a more precise listing of the 95 idioms along the decomposability scale.

Comparing the three categories – normally decomposable, abnormally decomposable and nondecomposable (Nunberg, 1978; Gibbs et al., 1989a) – with the four groups presented in this thesis – non-projected, partially projected, fully projected and non-projectable, I would like to argue for the latter categorization over the former.

In Section 2, it was mentioned that a component in a normally decomposable idiom is used literally, such as *save* in *save your skin*, while some constituents in an abnormally decomposable idiom are used metaphorically, such as *buck* in *pass the buck*. However, this categorization is inadequate since constituents in a normally decomposable idiom can also be used metaphorically. For instance, *skin* in *save your skin* does not literally refer to the natural body covering of an individual, but refers to the individual himself, to convey the meaning “protect oneself from danger or difficulty”. Similarly, some parts of an abnormally decomposable idiom could be used literally. For example, while *buck* in *pass the buck* is used metaphorically, *pass* is used literally. Therefore, categorizing decomposable idioms into normally and abnormally decomposable ones is inadequate in capturing the range of decomposable idioms.

In contrast, referring to Table 13², the projection terms are better at reflecting the decomposability spectrum. Specifically, non-projected idioms have components which combine and map directly from the literal to the figurative meaning, such as *live beyond one’s means*. Contrarily, in partially projected idioms, some components retain their literal meaning while others are metaphorically extended or mapped onto a synonym. In *harden one’s heart*, *harden* retains its meaning “to make hard or harder”, while *heart* is metaphorically extended to “feeling”. In fully projected idioms, each of their main lexical components refers to a metaphorical relationship between itself and its referent. For instance in *button up one’s lip*, the act of *buttoning* can be metaphorically understood to mean “restraining

²Refer to Appendix A.

something”; while *lip* is symbolic of “speech” since it is involved in the act of speaking. Finally, non-projectable (nondecomposable) idioms do not have separate meaningful units which combine together to form the idiomatic meaning, for example *shake the dust off one’s feet* “to leave angrily or rudely”. Therefore, the categorization proposed in this thesis is more appropriate than that posited by Nunberg (1978) and Gibbs et al. (1989a) since the former clearly demonstrates idiom decomposability, in terms of how constituents contribute to the figurative meaning.

Table 7 shows that amongst the three categories of decomposable idioms, there are fewer non-projected idioms – only eight as opposed to nineteen partially projected idioms and twelve fully projected ones. This suggests that an understanding of idioms require some general knowledge or experience to be aware of the relation – whether concrete, abstract or metaphorical – between the individual parts and their figurative referents (Gibbs et al., 1989b: 65).

There is also a relative high percentage of nondecomposable (or non-projectable) idioms in both the initial frequency data (Table 6) and randomized sample (Table 7). This could be due to the particular group of idioms being analysed in this thesis, in other words, possessive idioms. To reiterate, idioms vary widely both syntactically and semantically. Thus, it is possible that most possessive idioms are more nondecomposable, unlike Gibbs et al.’s (1989a) results which revealed that approximately 60% were decomposable idioms (comprised roughly 30% normally decomposable and 30% abnormally decomposable idioms). A criticism to Gibbs et al.’s findings was that the idioms were chosen according to the three groups – normally decomposable, abnormally decomposable and nondecomposable. This resulted in a uniform distribution amongst the groups and reduced result validity (Abel, 2003). Moreover, a study of 171 idioms chosen randomly from the *Longman Dictionary of English Idioms* revealed that native English speakers assessed 41.9% as decomposable and 58.1% as nondecomposable (Titone & Connine, 1994). Therefore, it is possible that the higher number of nondecomposable idioms found in this thesis is representative of possessive idioms, and perhaps representative of idioms in general, as exemplified by Titone and Connine’s (1994) study.

Furthermore, the higher frequency of nondecomposable idioms could be due in part to their semantic variation (Table 8), whereby each variant is expressed as a separate lexicon entry, so as to incorporate it into the ERG and WordNet. This finding is contrary to the traditional approach, which views idioms as fully frozen semantic units. It is also not aligned with how the compositional approach treats nondecomposable idioms, in that they are syntactically inflexible (Gibbs et al., 1989a; Nunberg et al., 1994) and lexically inflexible (Gibbs et al., 1989b). Rather, Table 8 shows that there exist nondecomposable idioms which exhibit semantic flexibility in their verbs or nouns, while retaining their figurative meanings. This suggests that language use is a creative process and nondecomposable idioms would still be understood when they are used in an adequate pragmatic context (Gibbs et al., 1989b: 66).

Therefore, the decomposability of possessive idioms ranges over a scale, from non-projected to partially projected to fully projected and lastly, non-projectable, with the majority in this study belonging to the last group. One important finding was that nondecomposable idioms are not entirely frozen as claimed by proponents of both the noncompositional and compositional approach. Rather, they can exhibit semantic flexibility. This permits a new understanding of idioms in general, whereby

nondecomposable idioms notwithstanding, some of them are actually semantically flexible.

Table 8
Semantic Flexibility in Nondecomposable Idioms

Nondecomposable idiom	Figurative meaning
get one's knickers in a knot/twist	to become very upset about something, that is usually unimportant
blow/ sound/ toot one's own trumpet/ horn	to brag
off one's head/ nut/ rocker/ trolley	to be crazy
have one's heart in one's mouth/ throat	to be frightened or anxious
hold/ keep/ play one's cards close to one's chest/ vest	to be secretive and not reveal one's thoughts or plans
put/ set one's house in order	to improve one's behavior or correct one's faults
too big for one's boots/ breeches/ britches	to behave as if one is more important or clever than one really is
have/ keep one's wits about one	to remain alert and be prepared

6.3 Plausibility

In Section 3, it was hypothesized that a more decomposable idiom would also be more plausible. Results are shown in Table 9, whereby the label “decomposable idioms” includes the non-projected, partially projected and fully projected idioms. Results indicated that contrary to the hypothesis, there was little correlation between idiom decomposability and plausibility. Most possessive idioms were plausible (71.6%), regardless of their decomposability. Since the majority of both decomposability types can have a literally plausible reading, this suggests that decomposability and plausibility have little relationship.

Table 9
Decomposability and Plausibility Matrix

Plausible		Implausible	
Decomposable	Nondecomposable	Decomposable	Nondecomposable
23 (24.2%)	45 (47.4%)	16 (16.8%)	11 (11.6%)

Refer to Table 14 in Appendix B for a more detailed matrix of these 95 idioms.

Titone and Connine (1994) revealed that literality correlated negatively with abnormal decomposability and vice versa. However, the interrelation was very low and thus regarded as insignificant. In comparison, a study demonstrated that abnormally decomposable idioms were comprehended faster than normally decomposable idioms (Gibbs et al., 1989; cited in Burt, 1992: 601). This could be due

to literal agreement between verbs and their objects in most abnormally decomposable idioms which is absent in normally decomposable idioms (Burt, 1992: 601). This particular interaction between figurative and literal meanings was termed as *transparency*, and which was likewise supported by Abel (2003: 344).

A preliminary study was thus conducted on the sample to examine a possible association between decomposability and transparency. However, most decomposable and non-decomposable idioms were judged to be transparent, thus indicating little relationship between these two dimensions. Moreover, transparency was more difficult to reliably judge than plausibility. Henceforth, no further analysis was conducted for the separate possessive idioms and neither plausibility nor transparency was considered to correlate with the decomposability of possessive idioms.

6.4 Further semantic analysis

The following subsections examine the semantics of possessive idioms, which was made possible through the creation of idiom entries and the subsequent parsing through the BNC.

6.4.1 Integration into WordNet

To recap, this thesis aimed to provide an insight into the semantic behaviour of possessive idioms. To do so, not only was a definition provided for each idiom, the senses for each component, their synonyms (for decomposable idioms) and paraphrases were also selected from WordNet.

Particularly, the linking of paraphrases to WordNet is one meaningful contribution to improving NLP applications and machine translation. Idioms are commonly found and yet are not well documented in dictionaries. This includes WordNet, which has about 120,000 word forms lexicalising roughly 100,000 concepts, but still do not include many idiomatic VPs (Fellbaum, 1998b). Fellbaum (1998b) argued that VP idioms cannot be automatically integrated into WordNet due to several formal and semantic problems and offered suggestions to address these problems.

The first is a formal problem whereby the syntactic forms of idioms do not agree with those in WordNet (Fellbaum, 1998b: 54). Such idioms include VPs preceded by negation, such as *cannot believe one's eyes*, and idioms with two VPs joined together by a conjunction, like *pick up one's marbles and go home*. However, if such idioms are frozen, they would disallow any internal variation or modification, and could thus be incorporated into WordNet when they are synonymous with existing entries.

Another problem is the discontinuity in the fixed parts of many VP idioms (Fellbaum, 1998b: 54), particularly possessive idioms, whereby the genitive slot can be filled by any possessive adjective or noun, thus resulting in discontinuous constituents. This could be resolved by including a rule with a preprocessor that would allow the genitive slot to be filled with either a pronoun or a noun from WordNet.

The third problem is related to the conceptual-semantic quality of idioms. These idioms convey concepts which cannot be added into WordNet either as independent entries or members of current synsets, due to an absence of lexicalized WordNet concepts to which they can be associated. Examples

include *cook someone's goose* and *drown one's sorrows* whose particular meaning might be lost if they were linked to more general WordNet concepts. This could be resolved if idioms were decomposable which would allow their constituents to be assigned meanings that are already lexicalised in WordNet (Fellbaum, 1998b: 55).

Lastly, idioms which have a VP form but convey states are also problematic. Their figurative meanings are defined as *be/ become adjective*, such as *flip one's lid* “to be angry” and *hold one's tongue* “to be quiet”. WordNet separates verbs from adjectives since they convey different concepts. Moreover, copula-verbs can combine with many adjectives and is thus pointless for WordNet to list all combinations. Hence, such VP idioms are not integrated into WordNet. Fellbaum (1998b: 56) proposed using *troponymy*, which is the “manner relation that organizes the verb lexicon”, to incorporate figurative meanings as subordinates of copula-verbs without altering lexicon structure.

The methodology in this thesis targets the aforementioned four problems. Specifically, the syntactic templates and *id_rel* and *poss_rel* markers resolve the first two problems. By relating the necessary components of an idiom together, NLP systems would recognise constituents as a whole, such as for idioms with negation and conjunction. Moreover, they resolve the discontinuity problem since noun variables can be filled with any pronoun or noun and simultaneously agree with the genitive slot.

To represent idioms in WordNet without losing their specific meaning, Fellbaum suggested applying the decomposability concept. This thesis addresses this by creating idiom entries which examined the decomposability of idioms. Where idioms were deemed decomposable, the lexical and figurative meanings of their constituents were chosen carefully from WordNet. This extensive database would thus aid in the representation of decomposable idioms into WordNet.

To integrate VP idioms whose figurative meanings express states, Fellbaum proposed to use a troponymy relation. However, this thesis suggests adopting a hypernymy/hyponymy relation, such that idioms are linked as lexicalised entries to adjectives. To illustrate, Fellbaum suggested that *hold one's tongue* “be quiet” should be listed as a subordinate of the copula-verb *be* to mean “to be quiet is to be in some manner”. Instead, this thesis proposes to lexicalise *hold one's tongue* by adding it as an entry linked to the adjective *quiet* in a hypernymy relation, whereby the idiom is a hyponym of the adjective.

There are several advantages for doing so. Firstly, the hypernym functions as the baseline of the idiomatic expression. By expressing the idiom in a straightforward manner, the idiom becomes more accessible, thus benefitting dictionary users, especially non-native speakers. Moreover, a hypernymy relation would expose dictionary users to various hyponyms, which constitute different ways of expressing a lexical item, in this case, the adjective *quiet*.

Secondly, a hypernymy approach would reduce the amount of work, in that fewer entries are required for the incorporation of idioms. Instead of proceeding from the verb *be* to its troponym relation *be quiet* and finally to the idiom *hold one's tongue*, the proposed approach would go directly from the adjective *quiet* to the idiom itself.

Furthermore, the states conveyed in the figurative meanings can be expressed using various verb-adjective combinations, rather than a strict one-to-one pairing. For instance, *find one's feet* “to become confident in a new situation as one gains experience”, conveys “become confident”, which can also

appear as “get confident”, “grow confident”, “come to be confident”, among others. Should the troponymy solution be applied, subordinates for all possible copula-verb have to be listed, which makes the WordNet integration process tedious and unnecessary. Contrarily, relating idioms with adjectives through a relation of hypernymy would be more efficient.

6.4.2 Semantic variation in the BNC

To recapitulate, a depth command was applied to idioms in the $X_{NP} V_I X_{NP}$'s N_I category to generate variants which contained verbs and nouns one level down the WordNet hyponymy relation (Section 5). The original idioms and their variants were then run through the BNC in two preliminary studies to determine the kind and frequency of idioms in the current corpus, so as to better understand the semantic and syntactic flexibility of possessive idioms.

The first study involved two idioms *bite one's tongue* and *(w)rack one's brains*. Table 10 shows that not all the sentences with the two idioms were parsed from the BNC. This was due to parser limitation in that long sentences were simply ignored during parsing. Furthermore, for the parsed sentences, idiom identification was approximately 70% accurate because of how the ERG functions. For instance, the parser did not recognise *someone bit their tongue* as an example of *bite one's tongue* because the ERG considers the androgynous singular pronoun *someone* to be in disagreement with the plural pronoun *their*.

Table 10
First Preliminary Study

	BNC sentences	Parsed sentences	Sentences with idioms
<i>bite one's tongue</i>	20	17	15 (75%)
<i>(w)rack one's brains</i>	62	49	38 (61.3%)

The percentages in *Sentences with idioms* were calculated as a percentage of *BNC sentences*.

The second study involved parsing about 100,000 sentences, of which 319 sentences (0.03%) contained idioms and variants. The low percentage could be attributed to several reasons. Besides the aforementioned parser and grammar limitation, it is possible that the current genres of spoken and written samples in the BNC have relatively few possessive idioms. Another reason was that many other possessive idioms were overlooked since only idioms belonging to one syntactic template were examined.

Table 11
Second Preliminary Study

	i-yes	i-no	Token count
Total	251 (76.7%)	68 (21.3%)	319

i-yes denotes parsed sentences that truly include idioms while *i-no* denotes sentences that were wrongly parsed as having idioms.

Regardless, the 319 sentences can be considered a lower bound, whereby if the possessive idioms in the other syntactic templates were included, the number is expected to perhaps double. Moreover, a manual check through the 319 idioms showed that 76.7% were correctly parsed as idioms (*i-yes*), as shown in Table 11. The relatively high percentage clearly validates our methodology, in terms of the semantic and syntactic descriptions of idioms. This allowed idioms to be incorporated into the ERG and subsequently constructed such that the BNC could identify idioms and their variants. An exception was the expression *earn one's way*, which was generated as a variant of the idiom *get one's way*. However, a manual check revealed that this phrase should be a variant of *earn one's keep*. Thus, the methodology of churning out variants still require some fine-tuning, in order to obtain the most accurate results.

Among the 319 parsed idioms, few were of the variants generated from the idioms with the $X_{NP} V_I X's N_I$ structure, which suggests little semantic flexibility. Conversely, Section 6.2 illustrated that some nondecomposable idioms exhibited semantic flexibility in their verbs or nouns (Table 8). Likewise, semantic flexibility was found in decomposable idioms as illustrated in Table 12. The findings thus seem to be in contradiction. However, it can still be concluded that possessive idioms are semantically flexible, except that the lexical substitution is relatively restricted, depending on aspects such as the context and verb characteristics, like telicity and verb-noun compatibility.

Table 12
Semantic Flexibility in Decomposable Idioms

Nondecomposable idiom	Figurative meaning
collect/ gather one's thoughts/ wits	to become mentally composed, especially after a state of distress; or to become mentally composed in preparation for a speech
drop/ lower one's guard	to reduce one's level of vigilance or caution to avoid danger or difficulty
bare/ open/ pour out one's heart to someone	to reveal one's deepest thoughts and feelings to someone
cast/ throw in one's lot with someone	to associate oneself with and share the fate of a person or group, no matter the outcome
be a thorn in someone's flesh/ side	to be a constant source of annoyance or problem to someone
put oneself in someone's place/ shoes	to improve one's behavior or correct one's faults
too big for one's boots/ breeches/ britches	to imagine being someone else and understand a situation from his/her point of view

The second study also revealed a high frequency of the idiom *shake one's head* (approximately 20%), which was often literally and figuratively used together. The relatively high percentage is partly because roughly half of the sentences containing this idiom were found in a particular literary text. This demonstrates how genres can affect the kind of idiom; in this case, *shake one's head* is very

common in literary genres. Since the BNC contains information about genres, such as medicine, sports and literature, future work could examine the relationship between genres and idiom frequency. This would in turn inform English-language learners of what idioms they have to understand, depending on their area of interest.

Moreover, considering that current dictionaries do not list idiom frequency, a corpus-based study of idiom frequency is important not just for enhancing dictionaries, but also for improving translation systems by informing NLP programmers what idioms to focus on. For this to happen, future work can continue from this preliminary study and work on the other syntactic templates identified in this thesis.

Lastly, though the focus of the BNC findings was on semantic flexibility, some syntactic flexibility was observed. This was manifested as modifications using adjectives such as *cannot believe my own bloody eyes*, *make one's unsteady way* and *have one's humorous moment*. This is an area worth investigating by running a larger idiom sample through the corpus to examine other aspects of their syntactic flexibility, such as quantification and topicalization as claimed by Nunberg et al. (1994).

7 Future Work

A comprehensive semantic and syntactic description of possessive idioms was created in this dissertation, within which, syntactic templates with higher frequencies were incorporated into the ERG. Future work should involve adding the other syntactic templates into the ERG to have a more inclusive idiom database.

The methodology also involved choosing the most appropriate WordNet sense and linking them to the literal, figurative and paraphrase senses in the idiom entries. Therefore, the entries can be used to integrate idioms into WordNet, via decomposability and hypernymy as discussed in Section 6.4.1. This could improve NLP applications that make use of WordNet as their lexical database. Idiom entries could also be used to develop automatic paraphrasing.

Evidently, the work accomplished in this thesis can be applied to a broader goal, in terms of representing idioms in the grammar and lexicon, so as to enhance NLP systems. Researchers could further investigate alternations in other idiom types, and also the relationship between genres and idiom types. All these would help to shed light on the behaviour of idioms, a group of MWE that has long posed a challenge to researchers.

8 Conclusion

This thesis emerged from my URECA project (2014) and Sameha's (2012) FYP in that the former focused on the decomposability of possessive idioms while the latter examined the implementation of such idioms into the grammar. However, both projects were neither thorough enough in the examination of the syntactic and semantic behaviour of possessive idioms, nor in the discussion of their inclusion into the grammar and lexicon.

Thus, this thesis set out with three aims (refer to Section 3): first, to provide a detailed semantic

and syntactic description of the possessive idioms with the broader aim of incorporating them into the ERG and WordNet; second, to evaluate their decomposability; third, to examine the relation between decomposability and plausibility.

Regarding the first objective, I have clearly produced an extensive description of the 514 possessive idioms through the creation of the idiom entries – a major contribution to the study of idioms.³ These were used in the subsequent methodology and analyses. Subsequently, with the help of Flickinger and Bond, the higher frequency co-indexed possessive idioms were added into the ERG then automatically constructed and compiled. In terms of incorporating them into WordNet, the linking of the literal senses, paraphrases and synonyms to WordNet helped in two aspects; one being the running of the idioms through the BNC to search for variants, and two being the relative ease of accommodating the idioms into WordNet in future.

I also examined the syntactic and semantic behaviour of possessive idioms, something which I deemed as incomplete in earlier research. The syntactic aspects include alternation and transformation in possessive idioms and the possibility of PP idioms. Semantic analysis revealed that in integrating the possessive idioms into WordNet, a hyponymy relation would be more appropriate than a troponymy relation. Some syntactic and semantic flexibility were also found from the idiom entries and BNC output.

This thesis proposed to use the projectability concept to better capture the degree of decomposability. Results also illustrated that there were more nondecomposable than decomposable idioms, in both the co-indexed and separate possession idioms. Furthermore, unlike past research, analysis showed that some nondecomposable idioms do undergo semantic variation.

Lastly, there appears to be little correlation between decomposability and plausibility, as opposed to the hypothesis (Section 3). Similarly, no relationship was found between decomposability and transparency. Thus, although these three dimensions affect idiom comprehension, they do not appear to interact meaningfully.

The findings exemplify that possessive idioms have a rich syntax and semantics. Evidently, this thesis is one step closer towards the ultimate aim of incorporating idioms into the English grammar and lexicon. Having laid down an extensive groundwork for future NLP research, hopefully less people will lose their mind when it comes to machine translation and second language learning.

³The idiom entries of the 514 idiom would be made available online through the Creative Commons BY 3.0 SG.

9 Bibliography

Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second language research*, 19(4): 329-358.

Bond, F., Ogura, K., & Ikehara, S. (1996). Possessive pronouns as determiners in Japanese-to-English machine translation. In *Proceedings of 2nd Pacific Association for Computational Linguistics Conference (PACLING '95)*. arXiv preprint cmp-lg/9601006.

Bond, F., Sameha, S. S., & Flickinger, D. (2013). *Making English Possessed Idioms Our Own*. HPSG Freie Universität, Berlin.

Burt, J. S. (1992). Against the lexical representation of idioms. *Canadian Journal of Psychology*, 46(4): 582-605.

Cacciari, C. (2014). The place of idioms in a literal and metaphorical world. In Cacciari, C & Tabossi P. (eds.) *Idioms: Processing, structure, and interpretation* (pp. 27-56). Psychology Press.

Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol*, 22(1): 79-108.

Caillies, S., & Declercq, C. (2011). Kill the song—steal the show: what does distinguish predicative metaphors from decomposable idioms?. *Journal of psycholinguistic research*, 40(3): 205-223.

Cignoni, L., Coffey, S., & Moon, R. (1999). Idiom variation in Italian and English: Two corpus-based studies. *Languages in contrast*, 2(2): 279-300.

Copestake, A., Flickinger, D., Sag, I., & Pollard C. (2005). Minimal Recursion Semantics. *An introduction Journal of Research on Language and Computation*, 3(2-3): 281-332.

LkbLui. (n.d.). In Deep Linguistic Processing with HPSG (DELPH-IN). Retrieved April 18, 2015, from <http://moin.delph-in.net/LkbLui>

Espinal, M. T., & Mateu, J. (2010). On classes of idioms and their interpretation. *Journal of Pragmatics*, 42(5): 1397-1411.

Farrell, P. (2009). The preposition with in role and reference grammar. *Studies in role and reference grammar*: 179-202.

Fellbaum, C. (ed.) (1998a). *WordNet: An electronic lexical database*. MIT Press.

Fellbaum, C. (1998b). Towards a representation of idioms in WordNet. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems [online]*, Montreal.

Fellbaum, C. (2014). The syntax and grammar of idioms and collocations. In: Kiss, T. & Alexiadou, A. (eds.) *Handbook of Syntax* (pp. 776-802). Berlin: de Gruyter.

Flickinger, D. (2011). Accuracy vs. robustness in grammar engineering. In Bender, E. M., & Arnold, J. E. (eds.). *Language from a Cognitive Perspective: Grammar, Usage, and Processing* (pp. 31-50). CSLI Publications: Stanford.

Geyken, A. (2007). The DWDS Corpus: A reference corpus for the German language of the 20th century. In Christiane F. (ed.) *Idioms and collocations: From corpus to electronic lexical resource* (pp. 23–39). Birmingham: Continuum.

Gibbs, R. W., & Nayak, N. P. (1989a). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive psychology*, 21(1): 100-138.

Gibbs, R. W., Nayak, N. P., Bolton, J. L., & Keppel, M. E. (1989b). Speakers' assumptions about the lexical flexibility of idioms. *Memory and Cognition*, 17(1): 58-68.

Ho, J. Q. & Bond, F. (2014). *Losing one's mind over meaning: Analysing the decomposability of possessive idioms*. Unpublished URECA paper, School of Humanities and Social Sciences, Nanyang Technological University, Singapore.

Krieger, H. U., & Schäfer, U. (1994). TDL: a type description language for constraint-based grammars. In *Proceedings of the 15th conference on Computational linguistics*, 2: 893-899. Association for Computational Linguistics.

Krifka, M. (2004). Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics*, 4: 1-32.

Langlotz, A. (2006). *Idiomatic Creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English (Vol. 17)*. John Benjamins Publishing, Amsterdam.

Levin, B., & Rappaport H. M. (2007). The Crosslinguistic study of dative alternations: A verb sensitive perspective. In handout, *Conference on ditransitive constructions*, Max Planck Institute for evolutionary anthropology, Leipzig.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory*

and Cognition, 36(6): 1103-1121.

Mateu, J., & Espinal, M. T. (2007). Argument structure and compositionality in idiomatic constructions. *The Linguistic Review*, 24(1): 33-59.

Muzny, G., & Zettlemoyer, L. S. (2013). Automatic idiom identification in Wiktionary. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*. Seattle, Washington, USA, 1417-1421.

Nenonen, M. (2007). Unique, but not cranberries: idiomatic isolates in Finnish. In *Collocations and Idioms 1. Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, 19–20 May 2006 (Vol. 41)*. Studies in Languages.

Nunberg, G. (1978). *The pragmatics of reference*. Bloomington, IN: Indiana University Linguistics.

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*: 491-538.

Oehrle, R. T. (1976). *The grammatical status of the English dative alternation* (Doctoral dissertation). Massachusetts Institute of Technology.

Oepen, S. (2008). *The Transfer Formalism: General-Purpose MRS Rewriting*. (Draft of 2008-11-30). LOGON. URL: <http://www.emmtee.net/reports/11.pdf>

O'Grady, William (1998) The Syntax of Idioms. *Natural Language and Linguistic Theory*, 16(2): 279-312

Packard, W. (2013). *ACE: the Answer Constraint Engine*. Retrieved April 18, 2015, from <http://sweaglesw.org/linguistics/ace/>

Random Integer Set Generator. (1998). Random.org. Retrieved March 16, 2015, from <https://www.random.org/integer-sets/>.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In *Proceedings of Computational Linguistics and Intelligent Text Processing: Third International Conference: CICLing-2002*. Lecture Notes in Computer Science, 2276: 1-15.

Sameha, Sheefa. (2012). *Make up your mind: An analysis of idiomatic possessive verb phrase constructions in English*. Unpublished final year project. Department of Linguistics and Multilingual

Studies, Nanyang Technological University, Singapore.

Shutova, S. T., & Korhonen, A. (2012). Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>

Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5): 1126-1138.

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal Of Pragmatics*, 31: 1655-1674.

Villavicencio, A., & Copestake, A. (2002). On the nature of idioms. *LinGO Working Paper No. 2002-04*. Available online at: <http://lingo.stanford.edu/>.

Wehrli, E. (1998, August). Translating idioms. In *Proceedings of the 17th international conference on Computational linguistics - Volume 2* (pp. 1388-1392). Association for Computational Linguistics.

10 Appendix A

This table illustrates the decomposability scale of the randomized sample of 95 co-indexed possessive idioms, which is analyzed and referenced in section 6.2.

Table 13
Decomposability Scale of Randomized Sample⁴

Non-projected	Partially Projected	Fully Projected	Non-projectable
$X_{NP} V_1 X's N_1$			
	harden one's heart lose one's buttons		feather one's nest press one's luck gird one's loins
$X_{NP} V_1 X's N_1 [P_1 Y_{NP}]$			
	lose one's hold on/over someone/thing	bare one's soul to some- one keep one's eye on some- thing set one's mind on some- one/thing	get one's toes into something
$X_{NP} V_1 [P_1 X's N_1]$			
live beyond one's means		button up one's lip	laugh up one's sleeve rest on one's oars sit on one's hand
$X_{NP} V_1 X's N_1 [P_1 D_1 N_1]$			
		have one's ear to the ground	get one's ducks in a row have one's back against the wall get one knickers in a knot throw one's toys out of the pram

Continued on next page

⁴Some of the idioms are listed in their basic forms without a verb since they can appear with several verbs. An example is *a dose of one's own medicine* in $X_{NP} V_1 D_1 N_1 [P_1 X's N_2]$ which can appear as *get a dose of one's own medicine*, *receive a dose of one's own medicine* or *need a dose of one's own medicine*.

Table 13 – *Continued from previous page*

Non-projected	Partially Projected	Fully Projected	Unprojectable
$X_{NP} V_1 X's N_1 P_1$			
	bawl one's eyes out cry one's heart out		keep one's chin up keep one's head down stick one's neck out
$X_{NP} V_1 X's own N_1$			
mind one's own business			be one's own person cut one's own throat sound one's own trumpet toot one's own horn
$X_{NP} V_1 + P_1 X's N_1$			
		let down one's guard	dig in one's heels pull up one's socks put up one's feet stick in one's oars
$X_{NP} be [P_1 X's N_1]$			
out of one's head	out of one's depth		on one's last legs off one's rocker off one's trolley
$X_{NP} V_1 X's N_1 [P_1 X's N_2]$			
			have one's heart in one's boots have one's heart in one's mouth know one's arse from one's elbow put one's foot in one's mouth scratch one's ear with one's elbow
$X_{NP} V_1 X's N_1 A_1$			
	keep one's eyes open keep one's nose clean	get one's hand dirty	put one's house in order set one's house in order

Continued on next page

Table 13 – *Continued from previous page*

Non-projected	Partially Projected	Fully Projected	Unprojectable
$X_{NP} V_1 D_1 N_1 [P_1 X's N_2]$			
	a dose of one's own medicine a legend in one's own (life)time a legend in one's own mind	a run for one's money (sense 2)	shake the dust off one's feet the monkey off one's back
$X_{NP} V_1 Y_{NP} [P_1 X's N_1]$			
			have something up one's sleeve have something under one's belt keep something under one's hat under one's thumb wrap someone around one's finger
$X_{NP} V_1 X's N(est)$			
	do one's best do one's worst try one's best try one's damndest try one's worst		
$X_{NP} V_1 [P_1 X's N_1 [P_2 Y_{NP}]]$			
	throw in one's lot with someone		look down one's nose at someone/thing take off one's hat to someone turn up one's nose at someone/thing up to one's eyes in something

Continued on next page

Table 13 – *Continued from previous page*

Non-projected	Partially Projected	Fully Projected	Unprojectable
X_{NP} aux+neg V_1 X 's N_1	cannot get one's head around someone/thing cannot keep one's eyes off someone/thing not mince one's words	cannot believe one's ears cannot believe one's eyes	
X_{NP} V_1 Y_{NP} D_1 N_1 [PP P_1 X 's N_2]		the rough edge of one's tongue the rough side of one's tongue	give someone a piece of one's mind give someone the shirt off one's back
X_{NP} V_1 R_1 A_1 [P_1 X 's N_1]			too big for one's breeches too big for one's britches too big for one's boots
X_{NP} V_1 [P_1 D_1 N_1 P_2 X 's N_2]			by the skin of one's teeth fly by the seat of one's pants
X_{NP} V_1 N_1 [P_1 X 's N_2]			have bags under one's eye have eggs on one's face
X_{NP} V_1 X 's N_1 [P_1 X]			have one's wits about one keep one's wits about one
X_{NP} V_1 X 's N_1 and V_2 N_2			have one's cake and eat it too pick up one's marbles and go home

Continued on next page

Table 13 – *Continued from previous page*

Non-projected	Partially Projected	Fully Projected	Unprojectable
Remainder	at the top of one's game have one's heart set on something let one's emotions show let someone/thing slip through one's fingers		put all one's eggs in one basket
Total			
8 (8.4%)	19 (20.1%)	12 (12.6%)	56 (58.9%)

11 Appendix B

This table illustrates the decomposability and plausibility matrix of the randomized sample of 95 co-indexed possessive idioms. This table is explained and referenced in section 6.3.

Table 14
Decomposability and Plausibility Matrix of Randomized Sample⁵

Plausible		Implausible	
Decomposable	Nondecomposable	Decomposable	Nondecomposable
$X_{NP} V_1 X's N_1$			
lose one's buttons	feather one's nest gird one's loins	harden one's heart	press one's luck
$X_{NP} V_1 X's N_1 [P_1 Y_{NP}]$			
lose one's hold on/over something	get one's toes into something	bare one's soul to someone keep one's eye on something set one's mind on someone/thing	
$X_{NP} V_1 [P_1 X's N_1]$			
live beyond one's means	laugh up one's sleeve rest on one's oars sit on one's hand	button up one's lips	
$X_{NP} V_1 X's N_1 [P_1 D_1 N_1]$			
have one's ear to the ground	get one's ducks in a row get one knickers in a knot have one's back against the wall throw one's toys out of the pram		
$X_{NP} V_1 X's N_1 P_1$			
bawl one's eyes out cry one's heart out	keep one's chin up keep one's head down stick one's neck out		

Continued on next page

⁵Some of the idioms are listed in their basic forms without the verb since they can appear with several verbs. An example is *a dose of one's own medicine* in $X_{NP} V_1 D_1 N_1 [P_1 X's N_2]$ which can appear as *get a dose of one's own medicine*, *receive a dose of one's own medicine* or *need a dose of one's own medicine*.

Table 14 – *Continued from previous page*

Plausible		Implausible	
Decomposable	Nondecomposable	Decomposable	Nondecomposable
$X_{NP} V_1 X\text{'s own } N_1$			
mind one's own business	be one's own person cut one's own throat sound one's own trumpet toot one's own horn		
$X_{NP} V_1 + P_1 X\text{'s } N_1$			
let down one's guard	dig in one's heels pull up one's socks put up one's feet stick in one's oars		
$X_{NP} \text{ be } [P_1 X\text{'s } N_1]$			
	off one's rocker off one's trolley	out of one's depth out of one's head	on one's last legs
$X_{NP} V_1 X\text{'s } N_1 [P_1 X\text{'s } N_2]$			
	know one's arse from one's elbow put one's foot in one's mouth scratch one's ear with one's elbow		have one's heart in one's boots have one's heart in one's mouth
$X_{NP} V_1 D_1 N_1 [P_1 X\text{'s } N_2]$			
a dose of one's own medicine	shake the dust off one's feet		
a legend in one's own (life)time			
a legend in one's own mind			
a run for one's money (sense 2)			

Continued on next page

Table 14 – *Continued from previous page*

Plausible		Implausible	
Decomposable	Nondecomposable	Decomposable	Nondecomposable
$X_{NP} V_1 X's N_1 A_1$			
get one's hand dirty			put one's house in order
keep one's eyes open			set one's house in order
keep one's nose clean			
$X_{NP} V_1 Y_{NP} [P_1 X's N_1]$			
	have something up one's sleeve		wrap someone around one's finger
	have something under one's belt		
	keep something under one's hat		
	under one's thumb		
$X_{NP} V_1 X's N(est)$			
do one's best			
do one's worst			
try one's best			
try one's damndest			
try one's worst			
$X_{NP} V_1 [P_1 X's N_1 [P_2 Y_{NP}]]$			
	look down one's nose at someone/thing	throw in one's lot with someone	
	take off one's hat to someone	up to one's eyes in something	
	turn up one's nose at someone/thing		
$X_{NP} aux+neg V_1 X's N_1$			
		cannot believe one's ears	
		cannot believe one's eyes	
		cannot get one's head around someone/thing	
		cannot keep one's eyes off someone/thing	
		not mince one's words	

Continued on next page

Table 14 – *Continued from previous page*

Plausible		Implausible	
Decomposable	Nondecomposable	Decomposable	Nondecomposable
$X_{NP} V_1 Y_{NP} D_1 N_1 [P_1 X's N_2]$			
	give someone the shirt off one's back	the rough edge of one's tongue the rough side of one's tongue	give someone a piece of one's mind
$X_{NP} V_1 R_1 A_1 [P_1 X's N_1]$			
	too big for one's breeches too big for one's britches too big for one's boots		
$X_{NP} V_1 [P_1 D_1 N_1 [P_2 X's N_2]]$			
			by the skin of one's teeth fly by the seat of one's pants
$X_{NP} V_1 N_1 [P_1 X's N_2]$			
	have bags under one's eye have eggs on one's face		
$X_{NP} V_1 X's N_1 [P_1 X]$			
	have one's wits about one keep one's wits about one		
$X_{NP} V_1 X's N_1 \text{ and } V_2 N_2$			
	pick up one's marbles and go home		have one's cake and eat it too
Remainder			
at the top of one's game let one's emotions show let someone/thing slip through one's fingers	put all one's eggs in one basket	have one's heart set on something	
Total			
	23 (62.5%)	45 (80%)	16 (37.5%) 11 (20%)