Losing one's mind over meaning: Analysing the decomposability of possessive idioms

Ho Jia Qian, Francis Bond, Dan Flickinger, Christiane Fellbaum

Abstract

Although found in everyday language, idioms are not well handled in natural language processing applications. By syntactically and semantically analysing **possessive idioms** (such as *X loses X's mind*), this paper seeks to determine their decomposability and semantic class distribution, so as to enhance machine translation. Three hundred and eighty idioms were grouped into various syntactic structures and co-referenced with templates in a precise grammar (the English Resource Grammar --- ERG: Flickinger, 2011). Thereafter, the meaning of either the paraphrase or each component of the paraphrase was linked to the appropriate WordNet sense (PWN: Fellbaum, 1998). Results show that both groups of idioms (co-indexed and separate possession) share similar syntactic structures and vary in their degree of decomposability, with most being non-decomposable.

Introduction

Multiword idiomatic expressions are ubiquitous in everyday language and already well-studied (Nunberg et al., 1994, Moon, 1998, Sag et al., 2002). Nevertheless, idioms still pose a problem to computational linguistics due to a great variation in syntax (Villavicencio et al., 2004, Fellbaum, 2011). While the constituents of many decomposable idioms have been linked to their appropriate WordNet entry (Osherson and Fellbaum, 2010), idioms are still inadequately represented in lexicons and grammars like the English Resource Grammar (ERG: Flickinger, 2008).

This paper focuses on English possessive idioms which are defined as expressions in which the verb or verb phrase (VP) precedes a prepositional phrase (PP) or noun phrase (NP) (O'Grady, 1998; Nenonen, 2007), within which the noun is possessed by an entity, which is typically the subject, and the relationship is marked by a pronoun which must be coindexed with this entity, as demonstrated in examples (1) and (2). In (1), the subject *Mary* co-indexes with the pronoun *her* to demonstrate that the noun *mind* is possessed by *Mary*. In contrast, (2) is considered ungrammatical since the gender and number of the possessive pronoun (neutral and plural) does not agree with that of *Mary* (feminine and singular).

(1) Mary loses her mind. (Mary becomes crazy: one argument – Mary)
(2) *Mary loses their mind.\\

However, the subject does not always necessarily co-index with the reflexive pronoun. For instance in (3), the noun *heart* is possessed by the reflexive pronoun *my* which is not co-indexed with the subject *Mary*. In such instances where the noun is co-referenced with the object instead of the subject, it is a case of separate possession rather than co-indexed possession.

(3) Mary breaks my heart. (Mary causes me great sorrow: two arguments – Mary and me)

It is also important to note that the nouns in English possessive idioms are not always only **inalienable possession** nouns such as body parts, but also include semantically opaque constituents, as in *be off one's rocker* or *have a bat in one's belfry.* This paper thus looks at a whole range of English possessive idioms, broadly categorized into co-indexed and separate possession, to determine their decomposability and semantic class distribution by analyzing their syntax and semantics.

Motivation

There are several reasons for focusing on possessive idioms. Firstly, they require interaction between syntax and semantics, and are thus hard to represent in the lexicon and grammar. They also present a problem to machine translation, as the possessive pronoun is typically not translated, even in decomposable idioms (Bond, 2005). Adopting a multilingual perspective, for instance by comparing English to French, it is not always the case that

French uses the possessives in idioms when its English equivalents do. For instance, the idiom in (4) illustrates the use of the possessive in both French and English; however in example (5), French uses determiners instead of possessives and also does not have the preposition 'with'. Moreover, in certain cases, while the syntactic behavior of idioms is more fixed in English, the French equivalents exhibit more flexibility as in example (6), which can undergo syntactic operations such as passive, raising, dislocation and adverbial modification (Wehrli, 1998, pg. 1389). Similarly, the idioms in (7) and (8) illustrate the difference between languages, in this instance, between English, German and Japanese - where the possessive pronoun is used in English, there is no equivalent in the other two target languages. In a reciprocal relationship as illustrated in (9), German uses the possessive pronoun together with a reflexive while English uses neither.

(4) French: *parler à travers son chapeau* (Vinay & Darbelnet, 1995)
Gloss: speak to through his hat
English: 'to talk through's one hat'

(5) French: *les mains dans les poches* (Vinay & Darbelnet, 1995)
Gloss: the hands in the pockets
English: 'with his hands in his pocket'

(6) French: *Jean a forcé la main à Luc* (Wehrli, 1998)
Gloss: Jean has forced the hand to Luc
English: 'Jean twisted Luc's hand'

(7) Japanese: *kanojo-wa chie-o shibotta* (Bond et al., 1995)
Gloss: she-TOP knowledge-OBJ wrung
English: 'she racked her brains'

(8) German: *jemanden eine lange Nase machen* (Fellbaum, 2014)
Gloss: to somebody (Dative) a long nose make
English: 'thumb one's nose at somebody'

(9) German: *X ist sich X's Sache sicher*
Gloss: X is X-self X's matter sure
English: 'X be sure/certain about what one says or does"

Examples (1)-(6) demonstrate the syntactic variability between idioms across different languages despite them having similar semantics. This lexical transfer is of importance to machine translation, since it is difficult to choose the correct equivalent in the target language (Santos, 1990). For instance in Japanese, *X blows X's stack [at Y]* can be learned as X-が Y-に 怒る *okoru* "Y angers X/X gets angry at Y" (equivalent to the paraphrase) or as an idiom X-が Y-の 頭 に くる *X-ga Y-no atama-ni kuru* "Y angers X (lit: X come to Y's head)". We can exploit this to translate as a normalized form (the paraphrase) or an equivalent linked idiom. By linking the idiom and its paraphrase to the full multilingual WordNet, we also link to idioms in other languages. Thus, by analyzing the various English possessive idioms, with various comparisons to other languages, this project seeks to create a database which researchers can use for further research and application into more sophisticated machine translation systems. Moreover, having a clearer picture of possessive idioms will also aid in computer-assisted language learning. For example, an English learner can use the material to understand figurative language, which is a more difficult aspect of language to learn; and to understand how pronouns operate in both literal and figurative English.

Methodology

This project studies a total of 300 possessive idioms which vary in syntactic structure, ranging from the simple *X loses X's cool* (X V X's N structure) to those with a more deeply embedded MWE such as X escaped by the skin of X's teeth (X V P1 D N1 P2 X's N2). A total of 33 syntactic templates were identified. Through these syntactic templates, this project aims to produce a representation so that each idiom can be identified by the ERG parser, with the resulting idiom linked to appropriate WordNet concepts. For decomposable idioms, such as *X faces X's demons*

"X confronts X's fears", each component is linked to the appropriate WordNet synsets. For non-decomposable idioms, like *X twiddles X's thumbs* "X is idle", the words in the paraphrase are linked to the corresponding WordNet concepts. For each idiom, we provide the index form, ERG idiom type (X V X's NP), links to synsets for the idioms (if decomposable) or the paraphrase (otherwise), as well as examples and a sentiment value between -1 and +1.

By linking to synsets, we can model variability, such as *X blows X's own trumpet* "X brags", with variants *X toots X's own horn; X sounds X's own trumpet; X blows X's own bugle*. Not all combinations are found: one current research goal is to try to automatically search and identify possible variants.

The decomposability of idioms lies on a continuum (*idiom decomposition hypothesis:* Gibbs et al., 1989). For example *X holds X's temper* is decomposable whereby *holds* means restrain and *temper* means outburst. However in some decomposable idioms, the relationship between components and idioms is less clear, as in *X recharges X's batteries*, whereby the idiomatic meaning would be understood only with the knowledge that batteries provide energy. In this case, we link both the components and the paraphrase to WordNet, thus capturing the dual nature. By taking this perspective, we seek to investigate the decomposability of possessive idioms and determine if there are any patterns in their decomposability based on their syntactic and semantic structure.

The idioms will also be analyzed according to the distribution of their components, for instance, whether they literally or figuratively refer to sources such as body parts or to targets such as emotions. This distribution of semantic classes would be evaluated based on idiom decomposability to determine if these two areas are related, whereby we posit that idioms with body part components would be more decomposable than idioms with a different lexical or syntactic structure. Finally, we will mark these idioms in the NTU-multilingual Corpus (Liling and Bond 2012) and investigate how they are translated there.

References

Bauer, Laurie. (1983). English Word-formation, Cambridge: Cambridge University Press.

Bond, F., Ogura, K., & Ikehara, S. (1996). Possessive pronouns as determiners in Japanese-to-English machine translation. *arXiv preprint cmp-lg/9601006*.

Fellbaum, Christiane. (ed.) (1998). WordNet: An electronic lexical database. MIT Press.

Fellbaum, Christiane. (2011). Idioms and Collocations. In Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.): Handbook of Semantics. Berlin: deGruyter, 441-456.

Fellbaum, Christiane. (2014). The syntax and grammar of idioms and collocations. In: Kiss, Tibor and Alexiadou, Artemis (eds.) Handbook of Syntax. Berlin: de Gruyter, 776-802.

Flickinger, Dan. (2011). Accuracy vs. Robustness in Grammar Engineering. In Emily M. Bender and Jennifer E. Arnold (eds.) Language from a Cognitive Perspective: Grammar, Usage, and Processing. CSLI Publications: Stanford, 31-50.

Francis Bond, Sheefa Samara Sameha, and Dan Flickinger. (2013). Making English possessed idioms our own. Paper presented at *The 20th International Conference on Head-Driven Phrase Structure Grammar (HPSG 2013.)* Berlin.

Gibbs Jr, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of memory and language*, *28*(5), 576-593.

Liling Tan & Francis Bond (2012). Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). In *International Journal of Asian Language Processing,* 22(4), pp 161–174

Moon, Rosamund. (1998). Fixed expressions and idioms in English: A corpus-based approach. Oxford University Press.Nunberg, Geoffrey, Ivan A. Sag & Tom Wasow. (1994). Idioms. Language 70, 491–538.

Osherson, Anne & Fellbaum, Christiane. (2010). The Representation of Idioms in WordNet. Proceedings of the Fifth Global WordNet Conference, Mumbai.

Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander Gelbuk (ed.), Computational linguistics and intelligent text processing: Third international conference: Cicling-2002, 1–15. Heidelberg/Berlin: Springer-Verlag.

Santos, D. (1990, August). Lexical gaps and idioms in Machine Translation. In *Proceedings of the 13th conference on Computational linguistics-Volume 2* (pp. 330-335). Association for Computational Linguistics.

Villavicencio, A., Baldwin, T., & Waldron, B. (2004). A Multilingual Database of Idioms. In *LREC*.

Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: a methodology for translation* (Vol. 11). John Benjamins Publishing.

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.