Feeling our way to an analysis of English possessed idioms

Francis Bond
Nanyang Technological University

Jia Qian Ho
NanyanUniversity

Dan Flickinger
Stanford University

Proceedings of the 22nd International Conference on
Head-Driven Phrase Structure Grammar

Nanyang Technological University (NTU), Singapore

Stefan Müller (Editor)

2015

CSLI Publications

pages 61–74

http://csli-publications.stanford.edu/HPSG/2015

Bond, Francis, Ho, Jia Qian, & Flickinger, Dan. (2015). Feeling our way to an analysis of English possessed idioms. In Stefan Müller (Ed.): *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore* (pp. 61–74). Stanford, CA: CSLI Publications.

Abstract

This paper describes an analysis for possessive idioms in English (e.g. *I twiddle my thumbs'* "I am idle"). The analysis relies on matching at the semantic level, to allow for syntactic variation. It has been implemented in the English Resource Grammar, and tested by parsing a subset of the British National Corpus. In addition to the syntactic analysis, we have linked the idioms to entries in the Princeton Wordnet, to allow for further lexical semantic analysis.

1 Introduction

Idiomatic constructions are very common in language, both at a type and token level. Despite considerable effort in categorizing and analyzing them (Nunberg et al., 1994; Moon, 1998; Sag et al., 2002) they are still not adequately represented in lexical resources, neither in lexicons such as wordnet (Fellbaum, 1998) or grammars such as the English Resource Grammar (Flickinger, 2000).

In this paper we focus on possessive idiomatic constructions: prototypically those in which one constituent is modified by a possessive pronoun co-indexed with a different constituent (typically the subject). A typical example is ***wrack one's brains*** "think hard", where the possessor of the brains must be the subject: *I wrack my brains*; *You wrack your brains*; *Kim wracks their brains*. These are interesting theoretically because of the interaction between syntax and semantics and are also of practical interest in translation (Bond, 2005). Most languages, even with similar idioms, do not include this possessive expression. For example, the equivalent phrase in Japanese is ***chie-wo shiboru*** "think hard: lit., squeeze knowledge". In this case it is a verb phrase with a fixed object, but there is no possessive.

The immediate motivation for this research was for machine translation: when translating out of English, typically the idiomatic possessive pronoun should be omitted. Going the other way, the possessive pronoun must be generated, and it must agree with the subject. Shallow statistical systems often get this wrong. A complete list of these idioms may also be useful for computer-assisted language learning. For example, an English learner can engage with the materials developed on corpora 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.

For example *Kim racks her brains* "Kim thinks hard" is given the unlikely literal translation by the statistical machine translations systems used by Google and Bing translate (1: translated on 2015-10-16).

(1) *Kim racks her brains*

a. キムは、 彼女の 脳を ラック
Kimu wa, kanojo no nou o rakku
Kim-TOP her-'s brain-ACC rack

Kim [dish] rack her brain (Google Translate)

a. キムは、 彼女の 脳を ラックします
Kimu wa, kanojo no nou o rakku shimasu
Kim-TOP her-'s brain-ACC rack do

Kim racks her brain (puts her brain in a [dish] rack) (Bing Translate)

In the following sections we present our idiom database, our analysis, and some corpus results,

2 The Idiom Database

In order to study their behavior we collected idioms from that included possession from a variety of sources, including WordNet (Fellbaum, 1998) and on-line lexicons such as Dictionary.com (2012). We ended up with 514 idioms:[1] very similar idioms have been merged into one entry (*to rack/wrack one's brains*) and idioms with two interpretations are treated as separate entries. These were categorized into different classes based on their syntactic and semantic structure. In addition, we attempted to give more literal paraphrases: ***wrack one's brains*** $\sim$ ***think hard***. Because of the variance in the possessive pronoun, it is hard to extract these automatically even using sophisticated methods (Zhang et al., 2006). For this reason, we are trying to cover as many as possible manually.

These idioms were categorized into co-indexed and separate possessive idioms and further grouped syntactically. We list the most common types of co-indexed idioms in Table 1. X_{NP}, Y_{NP} and Z_{NP} denote variable noun phrases, N for invariable noun, V for verb, A for adjective, R for adverb, D for determiner, aux for auxiliary and neg for negation. Square brackets [ ] denote prepositional phrases (PP). Within these brackets, *P* denotes a preposition; elsewhere, *P* represents a particle.

We give two examples of individual **idiom entries** in (2) and (3).

Definitions were written based on online dictionaries. Individual open-class words were linked to senses wordnet (by intuition, no deep etymological search was made).

[1] Available from `http://compling.hss.ntu.edu.sg/idioms/possessed`.

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 type frequency, with the exception of the last group, *remainder*

If the idiom is decomposable, a synonym or metaphorical extension for each component was chosen (marked with *) as in (2) and also linked to synsets in WordNet. Idiom decomposability is shown in @type.

Idiom decomposability was determined by **semantic substitution**: whether a lexical component can be replaced by appropriate word without altering its syntactic structure. In (2), *eat* is metaphorically extended to mean "withdraw" (*V_1) while *words* with "statement" (*N_1), to give "withdraw one's statement". This is the idiomatic meaning of the expressions, it is thus deccomposable. In contrast in (3), *twiddle* and *thumb* cannot be replaced with suitable synonyms nor metaphorical extensions, without altering the syntactic structure. The figurative meaning is "to be idle". Consequently, *twiddle one's thumb* is nondecomposable.

(2)

Idiom entry — fully projected	
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	(v) eat (take in solid food)
N_1	(n) words (the words that are spoken)
$^*V_1$	(v) swallow, take back, unsay, withdraw (take back what one has said)
$^*N_1$	(n) statement (a message that is stated or declared; a communication (oral or written) setting forth particulars or facts etc)
@type	decomposable

All non-decomposable idioms were given paraphrases, also linked to WordNet, marked with @ in their idiom entries. Decomposable idioms are paraphrasable with the extensions, so there is no need to list a separate paraphrase. In this case, the idiomatic meaning of the head (*V) will be the hypernym of the idiom. For non-decomposable examples, the head will also be the hypernym. However, where the paraphrase involves a copula and adjective, as in (3), the adjective paraphrase (@A) will be the hypernym of the idiom. This paraphrase captures the basic essence of each idiom and illustrates its hyponymy relation to lexical entries already listed in WordNet.

(3)

Idiom entry — non-projected	
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	(v) twiddle, fiddle with (manipulate, as in a nervous or unconscious manner)
N_1	(n) thumb, pollex (the thick short innermost digit of the forelimb)
@type	Nondecomposable
Paraphrase	X is idle
@template	X BE A
@A	(adj) idle (not in action or at work))

3 Analysis

The syntactic analysis uses idiom machinery inspired by Copestake (1994) and extended in Riehemann (2001); Copestake et al. (2002); Sag et al. (2002). It is implemented in the latest version of the English Resource Grammar (ERG: Flickinger, 2000, 2011). The relationship between the words in the idiom is captured using a fundamentally semantic mechanism, in our case encoded using Minimal Recursion Semantics (MRS: Copestake et al., 2005). Special lexical items introduce idiomatic predicates (marked as such in the lexicon). Idioms are treated as bags of predicates, with relations between them partially specified. If the semantics of a sentence can match this, then it has the idiomatic reading. This allows for considerable syntactic flexibility. During parsing, if a word has an idiom in it, a final check is made by the grammar when it enforces the root condition. Each idiomatic predicate must be licensed by at least one rule, otherwise the idiomatic interpretation is rejected.

Miyazaki et al. (1993) suggest that for some idioms we should allow nodes in a semantic hierarchy (so any noun with compatible semantics is allowed). We have linked the predicates in the idiom to their literal meanings (5) and the predicates in their paraphrases to the intended meaning using Wordnet synsets (6), but this is not used during parsing. Minor variations can easily be captured in the lexicon. For example, there are two alternative spellings of ***wrack***: *wrack* and *rack*. If we treat them as having no difference in meaning at all, then we represent them as two lexical items with different orthography, but the same predicate.

The interesting thing about the possessive idioms is that they also include an identity relation *id* to enforce the co-indexation. This is introduced by a special idiomatic verb-type, but could conceivable come from some kind of co-reference resolution. We give the bag of idioms that licenses ***wrack one's brains*** in (7).

(4) I_i rack my_i brains. [X Vs Y's Z; X=Y]

(5) Literal: $\boldsymbol{rack}_{v:9}$ "stretch on a rack"; $\boldsymbol{brains}_{n:1}$ "encephalon"

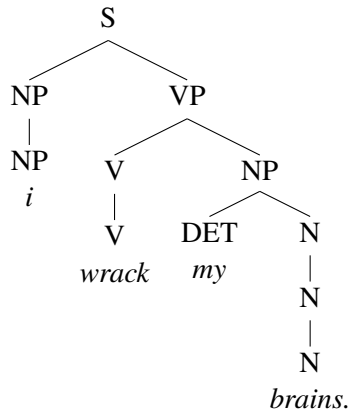
(6) Paraphrase: $\boldsymbol{think}_{v:1}$ "cogitate"; $\boldsymbol{hard}_{r:1}$ "with effort"

(7)
$$\begin{bmatrix} \textit{mrs} \\ \text{LTOP} \quad \boxed{h1}\ h \\ \text{INDEX} \quad \boxed{e3}\ e \\ \text{RELS} \quad \left\langle \begin{bmatrix} \textbf{\textit{\_rack\_v\_i}} \\ \text{LBL} \quad \boxed{hv} \\ \text{ARG0} \quad \boxed{v} \\ \text{ARG1} \quad \boxed{x} \\ \text{ARG2} \quad \boxed{z} \end{bmatrix}, \begin{bmatrix} \textit{id} \\ \text{LBL} \quad \boxed{hv} \\ \text{ARG0} \quad \boxed{id} \\ \text{ARG1} \quad \boxed{x} \\ \text{ARG2} \quad \boxed{y} \end{bmatrix}, \begin{bmatrix} \textit{poss} \\ \text{LBL} \quad \boxed{hz} \\ \text{ARG0} \quad \boxed{ps} \\ \text{ARG1} \quad \boxed{z} \\ \text{ARG2} \quad \boxed{y} \end{bmatrix}, \begin{bmatrix} \textbf{\textit{\_brain\_n\_i}} \\ \text{LBL} \quad \boxed{hz} \\ \text{ARG0} \quad \boxed{z} \end{bmatrix} \right\rangle \end{bmatrix}$$

The idiomatic ***wrack one's brains*** thus has three elements in the grammar: a lexical entry that introduces ***_brains_n_i***, a lexical entry that introduces ***_rack_v_i*** and *id* and links them appropriately; and an idiomatic rule that makes sure all the relevant elements are there: the above three predicates, and the possessive relation. The linking is crucial: the identity rel is linked to the external argument (XARG) of the verb (the subject) and to the external argument of the first element of the COMPS list (the determiner of the object). This links the subject to the possessor of the object. The idiom allows for variation in number: both *I rack my brain* and *I wrack my brains* are attested. In this case, we underspecify number in the construction, and allow both.

We give the parse tree and full MRS for (4) in (8) and (9), respectively.

(8)



(9)
$$\begin{bmatrix} \textit{mrs} \\ \text{TOP} \quad \boxed{0}\,h \\ \text{INDEX} \quad \boxed{2}\,e \\ \text{RELS} \quad \left\langle \begin{bmatrix} \textit{pron\_rel} \\ \text{LBL} \quad \boxed{4}\,h \\ \text{ARG0} \quad \boxed{3}\,x \end{bmatrix}, \begin{bmatrix} \textit{pronoun\_q\_rel} \\ \text{LBL} \quad \boxed{5}\,h \\ \text{ARG0} \quad \boxed{3}\,x \\ \text{RSTR} \quad \boxed{6}\,h \\ \text{BODY} \quad \boxed{7}\,h \end{bmatrix}, \right. \\ \left. \begin{bmatrix} \textit{\_wrack\_v\_i\_rel} \\ \text{LBL} \quad \boxed{1}\,h \\ \text{ARG0} \quad \boxed{2}\,e \\ \text{ARG1} \quad \boxed{3}\,x \\ \text{ARG2} \quad \boxed{8}\,x \end{bmatrix}, \begin{bmatrix} \textit{id\_rel} \\ \text{LBL} \quad \boxed{1}\,h \\ \text{ARG0} \quad \boxed{9}\,i \\ \text{ARG1} \quad \boxed{3}\,x \\ \text{ARG2} \quad \boxed{10}\,x \end{bmatrix}, \right. \\ \left. \begin{bmatrix} \textit{def\_explicit\_q\_rel} \\ \text{LBL} \quad \boxed{11}\,h \\ \text{ARG0} \quad \boxed{8}\,x \\ \text{RSTR} \quad \boxed{12}\,h \\ \text{BODY} \quad \boxed{13}\,h \end{bmatrix}, \begin{bmatrix} \textit{poss\_rel} \\ \text{LBL} \quad \boxed{14}\,h \\ \text{ARG0} \quad \boxed{15}\,e \\ \text{ARG1} \quad \boxed{8}\,x \\ \text{ARG2} \quad \boxed{10}\,x \end{bmatrix}, \right. \\ \left. \begin{bmatrix} \textit{pronoun\_q\_rel} \\ \text{LBL} \quad \boxed{16}\,h \\ \text{ARG0} \quad \boxed{10}\,x \\ \text{RSTR} \quad \boxed{17}\,h \\ \text{BODY} \quad \boxed{18}\,h \end{bmatrix}, \begin{bmatrix} \textit{pron\_rel} \\ \text{LBL} \quad \boxed{19}\,h \\ \text{ARG0} \quad \boxed{10}\,x \end{bmatrix}, \begin{bmatrix} \textit{\_brain\_n\_1\_rel} \\ \text{LBL} \quad \boxed{14}\,h \\ \text{ARG0} \quad \boxed{8}\,x \end{bmatrix} \right\rangle \\ \text{HCONS} \quad \text{(omitted for simplicity)} \end{bmatrix}$$

The other idiom types are implemented in a similar way: the main predicate (verb or preposition) adds and links the identity relation. For some cases, such as ***keep one's cards close to one's chest*** (e.g. in *You$_i$ keep your$_i$ cards close to your$_i$ chest.*), it has to add two identity predicates. The idiomatic licensing rule for this is given in (11). A different kind of idiom observed was the double co-index idiom. The syntactic shape of such idioms is N_1 V N_1'S N_2 (PP) (CONJ) (V) N_1's N_3. These instances belong to the less frequently observed extended structure idioms, which were essentially basic in shape but modified through post-insertions or by embedding. In this case, a second possessive noun phrase was added to an idiom that would otherwise be of basic shape but lack the idiom's meaning. A double co-indexing idiom looks similar to a basic idiom with the exception of the embedded possessive noun phrase. One instance of such an idiom is shown below:

(10) You$_i$ keep your$_i$ cards close to your$_i$ chest. [X keeps Y'S cards close to Z's chest; X=Y=Z]

(11)
$$\begin{bmatrix}
\textit{mrs} & \\
\text{LTOP} & \boxed{h1}\ h \\
\text{INDEX} & \boxed{e3}\ e \\
\text{RELS} & \left\langle \begin{bmatrix} \textit{_keep_v_i_rel} \\ \text{LBL} & \boxed{h2}\ h \\ \text{ARG0} & \boxed{e3} \\ \text{ARG1} & \boxed{x} \\ \text{ARG2} & \boxed{card} \\ \text{ARG3} & \boxed{h9} \end{bmatrix}, \begin{bmatrix} \textit{id_rel} \\ \text{LBL} & \boxed{h2} \\ \text{ARG0} & \boxed{e3}\ i \\ \text{ARG1} & \boxed{x} \\ \text{ARG2} & \boxed{y} \end{bmatrix}, \begin{bmatrix} \textit{poss_rel} \\ \text{LBL} & \boxed{h13}\ h \\ \text{ARG0} & \boxed{e15}\ e \\ \text{ARG1} & \boxed{card} \\ \text{ARG2} & \boxed{y} \end{bmatrix}, \right. \\
 & \left. \begin{bmatrix} \textit{_card_n_i_rel} \\ \text{LBL} & \boxed{h14} \\ \text{ARG0} & \boxed{card} \end{bmatrix}, \begin{bmatrix} \textit{_close_a_to} \\ \text{LBL} & \boxed{h21}\ h \\ \text{ARG0} & \boxed{e22}\ e \\ \text{ARG1} & \boxed{card} \\ \text{ARG2} & \boxed{chest} \end{bmatrix}, \begin{bmatrix} \textit{id_rel} \\ \text{LBL} & \boxed{h2} \\ \text{ARG0} & \boxed{e4}\ i \\ \text{ARG1} & \boxed{x} \\ \text{ARG2} & \boxed{z} \end{bmatrix}, \right\rangle, \\
 & \begin{bmatrix} \textit{poss_rel} \\ \text{LBL} & \boxed{h27}\ h \\ \text{ARG0} & \boxed{e29}\ e \\ \text{ARG1} & \boxed{chest} \\ \text{ARG2} & \boxed{z} \end{bmatrix}, \begin{bmatrix} \textit{_chest_n} \\ \text{LBL} & \boxed{h27}\ h \\ \text{ARG0} & \boxed{chest} \end{bmatrix} \\
\text{HCONS} & \text{(omitted for simplicity)} \\
\text{ICONS} & \langle\ \rangle
\end{bmatrix}$$

There is a long tail of rare types: as Richter & Sailer (2009) point out, some of these idioms can even go across clause boundaries, for example: ***look as though butter wouldn't melt in one's mouth*** "appear innocent". Currently we have created idiom types for the most common classes of idiom (all those with a type frequency of greater than eight) and instantiated them with idiom rules for each of the entries in the database. In future work, we will keep working our way down the long tail.

While the two-place *id* predication appearing in the RELS lists of the above examples (7,11) was implemented and used for most of the empirical work reported here, we have also been developing an alternative representation of the identification of the possessor in our idioms with the external argument of the verb. Building on the notion of sets of constraints on pairs of individuals proposed for information structure by Song (2015), we can express the relevant identity in our idioms not

as a predication but as an ICONS ("individual constraint") pair. While binding constraints on intrasentential anaphors in general are still under development for the ERG, these ICONS pairs seem well-suited for expressing both coreference and non-coreference constraints imposed by the syntax, and that promise leads us to express these idiom-specific identities with the same formal mechanism. One advantage of removing the *id* predication from the RELS list is that we no longer have to engineer the assignment of the LBL for that predication; note that in our example above, that label value is identified with the label of *_rack_v_i*, but this is both awkward to ensure compositionally, and lacking in independent motivation. By using the ICONS representation instead, we clearly distingish coreference constraints between pairs of individuals from the contentful semantic predications that comprise the RELS list and are subject to scopal operators including quantifiers, modals, negation, and the like.

Sheinfux et al. (2015) also propose a method to handle idioms of this type in Hebrew. In their analysis, the verb selects for a special kind of argument, and the agreement properties are passed up using the XARG. This does not require our (independently motivated) idiom processing, but does require special lexical entries not just for the verb, but also the noun, the possessor and any prepositions involved in the idiom.

In future work, we will think further as to how to mark the idioms in the output semantic representation. Currently, the individual elements are marked as idiomatic. During processing we know which idiom was licensed (as we know which idiom rule applies), but this information is not part of the final MRS. Further, the possessive pronouns are not marked in any way, even though intuitively they are less meaningful than real referential pronouns. Both these issues are also relevant to the separate possessive idioms. One approach is to keep decomposed idioms as they are (but specify their predicates to have the idiomatic meanings) and paraphrase the non-decomposable ones, thus doing away with the non-referential pronouns altogether.

4 Testing on a corpus (the BNC)

We ran the extended ERG over the British National Corpus (Burnard, 2000) to identify actual examples of these idioms. We attempted to parse the first 3,494,381 sentences.[2] We were able to successfully parse 3,011,023 of the sentences (86%)

[2]This took 44 days on 20 CPUs, after which we had to stop to apply a security patch to the server. We are currently looking for a bigger server cluster.

and found 5,577 sentences with possible idioms (0.18%). We are the first to identify these idioms in the BNC. Up until now it has been hard to find these kinds of idioms, due to the complicated structure. With idioms implemented in a flexible grammar, they can be identified automatically.

A manual check of the first 319 idiom instances showed that 76.7% were being used idiomatically. The relatively high percentage shows that these complex idioms are typically used idiomatically. To distinguish between idiomatic and non-idiomatic uses we need to retrain the the parse ranking model with idiomatic examples and/or learn a special model to distinguish idiomatic from non-idiomatic uses (such as, Hashimoto & Kawahara, 2009).

The ten most common idiom types are shown in Table 2. The idiom ***shake one's head*** was the most common. In many cases, it was clearly referring both to the physical act of shaking one's head, and to the idiomatic meaning of "indicate disagreement". ***bite one's lip*** "forcibly prevent oneself from speaking" was similar: often both the literal and idiomatic meanings were applicable at the same time.

Table 2: Most common possessive idioms found in the British National Corpus

Idiom	**Frequency**	**Comment**
shake ones' head	2,055	
make one's way	359	often both idiomatic and literal
open one's eye	344	mainly non-idiomatic
find ones' way	205	
bite one's lip	145	
get one's way	131	
have one's way	139	
raise one's eyebrows	124	
shrug one's shoulders	118	
lose one's temper	113	

Current dictionaries rarely list idiom frequencies, this corpus-based study offers not just useful information for lexicographers, but also for improving translation systems by informing programmers which idioms to focus on. Future work can thus continue from this preliminary study and work on the other syntactic templates identified in section 2.

Finally, the BNC findings showed some interesting examples of syntactic flexibility, including modification, relativization and long distance dependencies, as shown (12). All of these were successfully identified by the ERG, although would be very hard to identify successfully using shallow chunk based systems. There

were many more examples of 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 we will continue to investigate by running a larger idiom sample through the corpus.

(12) a. *The butcher had lined his pockets too thickly in the past at their expense, and Faith's will had been a warning, a pointer to their future.*

b. *Now do thy speedy utmost, Meg,*

c. *Even if she is an overpaid brat in danger of losing her marbles, at least she provokes a reaction, and is 500 times more controversial than Madonna.*

d. *And if everybody starts getting very large discounts and the vendor loses control of the market, not only do the buyers lose all their advantage, but the vendor loses its corporate shirt.*

e. *Nor is it the case that the Federal Republic is using the issue of democratic accountability to drag its feet on EMU.*

f. *Mr Waddington, a former immigration minister and rightwinger, seems to have gritted his teeth at yesterday's meeting and stood by the compromise hammered out at Mrs Thatcher's insistence in a cabinet committee.*

g. *I'm starting to lose my bearings a bit—and my ball-bearings as well, come to that.*

With more data we can examine more reliably other aspects of syntactic flexibility, such as modification, quantification and topicalization, allowing us to test the claims of Nunberg et al. (1994). They distinguish idiomatically combining expressions (ICEs: our decompositional) and idiomatic phrases (IPs: our non-decompositional) with five tests: modification, quantification, topicalization, ellipsis, and anaphora.

5 Conclusions

We have implemented an analysis of co-indexed possessive idioms in HPSG, suitable for use in a computational grammar. We have tested an implementation of the major types of idiom in the English Resource Grammar and linked the predicates to wordnet. We are currently experimenting with expanding our variants and identifying corpus examples. As well as implementing in the ERG, the full idiom

lexicon, including definitions, examples and links to wordnets is freely available under an open licence (CC-BY) at: `http://compling.hss.ntu.edu.sg/idioms/possessed/`.

Acknowledgments

We would like to thank the reviewers and participants of HPSG 2015 for their helpful comments and discussion. This research was supported in part by the MOE Tier 2 grant *That's what you meant: a Rich Representation for Manipulation of Meaning* (MOE ARC41/13).

References

Bond, Francis. 2005. *Translating the untranslatable: A solution to the problem of generating English determiners* CSLI Studies in Computational Linguistics. CSLI Publications.

Burnard, Lou. 2000. *The British National Corpus users reference guide*. Oxford University Computing Services.

Copestake, Ann. 1994. Representing idioms. Presentation at the HPSG Conference, Copenhagen.

Copestake, Ann, Dan Flickinger, Ivan A. Sag & Carl Pollard. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation* 3(2). 281–332.

Copestake, Ann, Fabre Lambeau, Aline Villavicencio, Francis Bond, Timothy Baldwin, Ivan Sag & Dan Flickinger. 2002. Multiword expressions: Linguistic precision and reusability. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, 1941–7. Las Palmas, Canary Islands.

Dictionary.com. 2012. Free online English dictionary. `http://dictionary.reference.com/`.

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

Flickinger, Dan. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1). 15–28. (Special Issue on Efficient Processing with HPSG).

Flickinger, Dan. 2011. Accuracy vs. robustness in grammar engineering. In Emily M. Bender & Jennifer E. Arnold (eds.), *Language from a cognitive perspective: Grammar, usage, and processing*, 31–50. Stanford: CSLI.

Hashimoto, Chikara & Daisuke Kawahara. 2009. Compilation of an idiom example database for supervised idiom identification. *Language Resources and Evaluation* 43(4). 355–384.

Miyazaki, Masahiro, Satoru Ikehara & Akio Yokoo. 1993. Combined word retrieval for bilingual dictionary based on the analysis of compound word. *Transactions of the Information Processing Society of Japan* 34(4). 743–754. (in Japanese).

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.

Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th international conference on Head-Driven Phrase Structure Grammar, university of Göttingen, germany*, 297–317. Stanford: CSLI Publications. `http://cslipublications.stanford.edu/HPSG/2009/`.

Riehemann, Susanne Z. 2001. *A constructional approach to idioms and word formation*: Stanford dissertation.

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. Hiedelberg/Berlin: Springer-Verlag.

Sheinfux, Livnat Herzig, Tali Arad Greshler, Nurit Melnik & Shuly Wintner. 2015. Hebrew verbal multi-word expressions. In *Proceedings of the 22nd international conference on Head-Driven Phrase Structure Grammar (HPSG 2015)*, 123–136.

Song, Sanghoun. 2015. Representing honorifics via individual constraints. In *ACL 2015 workshop on grammar engineering across frameworks (GEAF 2015)*, .

Zhang, Yi, Valia Kordoni, Aline Villavicencio & Marco Idiart. 2006. Automated multiword expression prediction for grammar engineering. In *Proceedings of the workshop on multiword expressions: Identifying and exploiting underlying properties*, 36–44. Sydney, Australia: Association for Computational Linguistics. `http://www.aclweb.org/anthology/W/W06/W06-1206`.