Like Finding a Needle in a Haystack: Annotating the American National Corpus for Idiomatic Expressions

Laura Street, Nathan Michalov, Rachel Silverstein, Michael Reynolds, Lurdes Ruela, Felicia Flowers, Angela Talucci, Priscilla Pereira, Gabriella Morgon, Samantha Siegel, Marci Barousse, Antequa Anderson, Tashom Carroll, Anna Feldman

Montclair State University
1 Normal Avenue, Montclair, New Jersey, 07043, USA
{streetl1, michalovn1, silversteir2, feldmana}@mail.montclair.edu

Abstract

This paper presents the details of a pilot study in which we tagged portions of the American National Corpus (ANC) for idioms composed of verb-noun constructions, prepositional phrases, and subordinate clauses. The three data sets we analyzed included 1,500-sentence samples from the spoken, the non-fiction, and the fiction portions of the ANC. This paper provides the details of the tagset we developed, the motivation behind our choices, and the inter-annotator agreement measures we deemed appropriate for this task. In tagging the ANC for idiomatic expressions, our annotators achieved a high level of agreement (> .80) on the tags but a low level of agreement (< .00) on what constituted an idiom. These findings support the claim that identifying idiomatic and metaphorical expressions is a highly difficult and subjective task. In total, 135 idiom types and 154 idiom tokens were identified. Based on the total tokens found for each idiom class, we suggest that future research on idiom detection and idiom annotation include prepositional phrases as this class of idioms occurred frequently in the nonfiction and spoken samples of our corpus

1. Introduction

Researchers have been investigating idioms and their properties for many years. According to traditional approaches, an idiom is "in its simplest form…a string of two or more words for which meaning is not derived from the meanings of the individual words comprising that string" (Swinney and Cutler, 1979: 523). As such, the meaning of *to kick the bucket* (which is 'to die') cannot be obtained by breaking down the idiom and analyzing the meanings of its constituent parts, *to kick* and *the bucket*. In addition to being influenced by the principle of compositionality, the traditional approaches are also influenced by theories of generative grammar (Flores d'Arcais, 1993: 80-82; Langlotz, 2006: 15-16). The properties that traditional approaches attribute to idiomatic expressions are also the properties that make them difficult for generative grammars to describe. For instance, idioms can be syntactically ill-formed (e.g., *by and large*), resistant to grammatical transformations (e.g., *the bucket was kicked by him* ≠ 'he died'), impervious to lexical substitutions (e.g., *to kick the pail* ≠ 'to die'), and semantically ambiguous without context (e.g., *Charles kicked the bucket and was buried last night* vs. *Charles kicked the bucket through the window).*

In recent years, post-generative approaches to idioms have argued that idioms, like all linguistic expressions, are a type of construction. As Goldberg (2003: 219) states: "Constructions are pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns." To be more precise, a construction is any linguistic pattern where an aspect of its form or function is unpredictable based on its component parts or other known constructions. When conducting research on idiomatic expressions, a constructionist approach is advantageous because it seeks to describe language in general. That is to say, it does not disregard unusual linguistic phenomena (such as idioms) because they are 'peripheral' rather than part of the 'core' of language. Constructionist approaches are also beneficial because they claim that constructions are acquired based on their frequency in the input and this works well with a corpus-based approach where data and frequency counts are used to capture properties of linguistic phenomena such as collocations, multi-word expressions, etc. For a more detailed discussion of the approach we take to idioms in this paper, please refer to (Goldberg, 2003).

Constructions	**Examples**
Morpheme	*pre-*, *-ing*
Word	*avocado*, *anaconda*, *and*
Complex Word	*dare-devil*, *shoo-in*
Idiom	*going great guns*
Idiom (Partially Filled)	*jog* <someone's> *memory*
Covariational Conditional	*the more you think about it, the less you understand* (Form: The Xer, the Yer)
Ditransitive	*he gave her a Coke*, *he baked her a muffin* (Form: Subj [V Obj1 Obj2])
Passive	*the armadillo was hit by a car* (Form: Subj aux VPpp (PPby))

Table 1. Examples of constructions (adapted from Goldberg 2003: 220)

Previous work on automatic idiom classification has typically been of two types: those which make use of type-based classification methods (Lin, 1999; Baldwin et al. 2003; Fazly and Stevenson, 2006; Bannard, 2007; Fazly et al., 2009) and those which make use of token-based classification methods (Birke and Sarkar,

2006; Katz and Giesbrecht, 2006; Fazly et al., 2009; Sporleder and Li, 2009). Type-based classification methods generally rely on the notion that many idioms share unique properties with one another. For instance, several idioms are composed of verb-noun constructions (e.g., *break a leg*, *get a grip*, *kick the bucket*) that cannot be altered syntactically or lexically (e.g., **break a skinny leg*, **a grip was got*, **kick the pail*). These unique properties are used to distinguish idiomatic expressions from other types of expressions in a text. Token-based classification methods are different in that they make use of the surrounding context and/or search for lexical cohesion. When given a cohesive text, sentences in one part of the text are assumed to be related to sentences in another part of the text. Idiomatic expressions, therefore, stand out because they are semantically unrelated to the surrounding context. To take a simple example, several words in the following text are highlighted as sharing the semantic property [+cold] with the underlined expression. The underlined expression is taken to be literal rather than non-literal in this context: *He had **cold** feet while standing in the bus stop, but he wouldn't be deterred by the **frigid** air, the **chilling** wind, or the unending **snow**.* If the underlined expression did not share semantic properties with the words that surround it, the chances of it being non-literal would be high.

While the work on automatic idiom detection has been groundbreaking, these methods typically have two major limitations: 1.) they require many resources such as dictionaries and/or annotated corpora to be implemented, and 2.) they are limited to detecting only one type of idiom (i.e., verb-noun constructions). If the goal is to identify and extract idioms automatically, more work needs to be done on other types of idiom constructions. We have initiated this type of research by annotating the ANC for various types of idioms. Thus far, Hashimoto & Kawahara (2008) have noted that there are few corpora annotated for idiomatic expressions even though researchers have taken an interest in this topic recently. The SAID corpus (Kuiper et al., 2003) and the TroFi Example Base (Birke and Sarkar, 2006), however, are two of the few examples available in English. Like the SAID corpus and the TroFi Example Base, the corpus we developed is intended to be the foundation for a more in-depth study of idioms. As such, the corpus is a useful tool for researchers working in the fields of theoretical linguistics, applied linguistics, and language technology.

2. Methodology

Our sample consisted of approximately 4,500 sentences (68,915 tokens) extracted from the American National Corpus (ANC), a developing corpus of American English currently containing approximately 22 million words. We chose to work with the ANC because English varies notably from one brand to another in lexicon, phrasing, semantics, and—most importantly for our study—idiomatic expressions (Fillmore et al., 1998), and as speakers of American English we felt it important to work with the corpus that represents the brand of English with which we are most familiar. Furthermore, while the ANC Second Release has already been annotated for word and sentence boundaries, parts-of-speech, and noun and verb chunks, it does not yet include any annotations for figurative language. [1]

The ANC Second Release includes data from a variety of domains ranging from government, academic, and technical writing to journal articles, fiction, blogs, magazines, and spoken interviews. Although the corpus as a whole will not be considered balanced until it is complete,[2] we selected a sample that we believe represents a wide range of idiomatic language. Approximately one third of our sample data comes from the written nonfiction domain, one third comes from transcribed spoken narratives, and one third comes from written fiction. In a trial run of the study, the annotators identified and tagged idioms in a 111-sentence sample (1644 tokens) according to the tagset elaborated below. After the trial run, the annotators discussed the numbers and types of idiomatic expressions they found and re-evaluated the tagset, ultimately deciding to use the same tagset for the pilot study as they had used for the trial run. For the pilot study, the annotators tagged a 4,500-sentence sample in the same manner as the trial run. The annotators worked individually, but three or four were assigned to each sample. The task was completed in roughly six to ten hours.

To expand upon the work of previous studies and to facilitate future work in the area of automatic idiom detection, we developed a tagset comprised of three tags. Our tagset included tags for idioms composed of verb-noun constructions (VNCs), prepositional phrases (PPs), and subordinate clauses (SCs). We included less prototypical structures such as PPs and SCs because one of our goals was to illustrate the many cases of figurative language that are missed by approaches that focus exclusively on idioms containing select verbs.

Tags	Examples
<PP> … </PP>	*at death's door, to a fault*
<VN> … </VN>	*pass the buck, learn the ropes*
<SC> … </SC>	*when hell freezes over, if the price is right*

Table 2. The tagset with examples

In an effort to make the tagging process more efficient and reliable, we limited our tagset to a very small number of tags and gave our annotators the following guidelines: 1.) Always tag for the narrowest scope possible, 2.) Use the same syntax throughout the annotation process, and 3.) Do not include punctuation such as periods or quotation marks.

3. Measuring Inter-Annotator Agreement & Reliability

To ensure reliability of the results two alpha scores were calculated for each of the three sections using the following variation of Krippendorff's Alpha[1]:

$$\alpha = 1 - \frac{D_o}{D_e} = \frac{(n-1)\sum_c O_{cc} - \sum_c n_c(n_c-1)}{n(n-1) - \sum_c n_c(n_c-1)}$$

Where Do represents observed disagreement among the annotators, De is the disagreement which would be expected to occur if tags had been selected by chance. ∑ indicates summation of all data of the given type, n represents the number of items associated with the given subscript (where no subscript is present n indicates the total number of tags), and subscripts relate to the coincidence matrices which are described in more detail below.

Krippendorff's Alpha was used because it was appropriate for our data set. It is specifically designed to analyze nominal data sets generated by multiple annotators, including instances where one or more annotators did not provide a rating for some of the units in the data set. While Fleiss' Kappa and Cohen's Kappa are commonly used to analyze inter-annotator agreement, they were inappropriate in the current study as Cohen's Kappa can only be used with two annotators and Fleiss' Kappa cannot be used when data sets contain units that are not rated by all of the annotators.

The first alpha evaluated annotator reliability in the detection of idioms within the corpus and accounted for annotator reliability regarding the status of a phrase as idiomatic or not. The second alpha determined reliability in the categorization of detected idioms. In cases where annotators disagreed on the scope of idioms and overlapping tags were noticed, these tags were grouped together for alpha scoring and noted for annotators. Given the size of the corpus and the fact that tag lists were obtained by machine search, it is possible that not all of the overlapping tags were noticed, which could have affected the preliminary alpha score. Given that subsequent modifications were made manually by the annotators, however, such artifacts are not present in the final corpus and all tags agree in all relevant domains.

The specific methodology of the alpha scoring was to form two tables of annotator responses. The first table (Table 3) was used to generate the first alpha score. It contains all the phrases (units) that were marked as idioms by at least one annotator. Columns beneath each unit record whether the corresponding annotator marked the phrase as an idiom or not.

[1] Information regarding Krippendorff's Alpha can be found at Klaus Krippendorff's website: http://www.asc.upenn.edu/usr/krippendorff/dogs.html

	Unit 1	Unit 2	...	Unit U
Annotator A	Idiom	Idiom	.	Non-Idiom
Annotator B	Non-Idiom	Idiom	.	Non-Idiom
Annotator C	Idiom	Idiom	.	Idiom

Table 3. Table of phrases marked as idioms

The second table (Table 4) was used to generate the second alpha score. It contains all the phrases (units) that were tagged as idioms along with their corresponding tags. (Note: '.' indicates that the annotator did not mark the phrase as being an idiom.) The final row records the total number of tags for a given phrase, and the final column records the number of each type of tag a given idiom received. Mu is the total number of tags and is present only in the second table as the reliability of idiom categorization can only be determined in cases where annotators have marked a statement as a type of idiom. In the first case, all annotators will have marked any given phrase as idiomatic or not, so Mu will always equal the number of annotators and its inclusion in the table would be redundant.

	Unit 1	Unit 2	Unit 3	.	Unit U	
Annotator A	VN	SC	PP	.	SC	
Annotator B	VN	SC	PP	.	.	
Annotator C	VN	PP	.	.	.	
Number of values in unit (Mu)	3	3	2	.	1	∑ Mu

Table 4. Presents tags assigned to each given phrase.

Data within the tables were then used to generate coincidence matrices of the form:

Values:	**1**	.	**k**	.	.	
1	o_{11}	.	o_{1k}	.	.	n_1
.	.	.				.
.			.	.		.
c	o_{c1}	.	o_{ck}	.	.	$n_c = \Sigma_k o_{ck}$
.	.	.				.
	n_1	.	n_k	.	.	$n = \Sigma_c \Sigma_k n_{ck}$

Figure 1. Sample matrix used to calculate the number of tag pairs that occur.

Where Ock is determined using the following formula:

$$o_{ck} = \sum_u \frac{\text{Number of c-k pairs in unit u}}{m_u - 1}$$

For example, if a given phrase was tagged as a VN by two annotators and a PP by a third, there would be 6 possible pair combinations of the given tags: 2 VNPP, 2PPVN, and 2VNVN pairs. Note that in calculating ck pairs, the ordering matters and PPVN is not the same as VNPP. The number of pairs of each type is then normalized by dividing it by one less than the total number of tags for the unit. In the given example each pair would be divided by 2 (3 tags - 1) to create a value of 1VNPP, 1PPVN, and 1VNVN. The matrix then displays the total number of normalized tag pairs of each type. Matrices figures are then used in the alpha formula.

Values:

	PP	**VN**	**SC**	
PP	2	4	1	7
VN	4	10	3	17
SC	1	3	7	11
	7	17	11	35

Figure 2. An example of how the matrix appears for tag pairs

The alpha score is obtained by calculating the difference of observed disagreement over expected disagreement, or the amount of agreement that would be expected if tags were assigned at random, subtracted from one (1 – (Do/De)) so that perfect annotator agreement would result in $\alpha = 1$ (1 – (0/1) = 1) and if annotator agreement were equal to chance $\alpha=0$ (1 – (1/1) = 0). Specifically the form of the alpha for nominal data used here is obtained by subtracting the total possible number of agreeing tag pairs given the annotations made ($\sum$ nc(nc-1)) from the actual number of agreeing tag pairs made by the annotators ($\sum$ occ), and then multiplying by one less than the total number of tags in agreement (n-1). This number is then divided by the difference obtained by subtracting the number of possible tag pairs given the number of each tag type (again ((n-1)$\sum$occ)) from the maximum possible tag pair agreement for the overall number of tags (n(n-1)). Put more generally, the formula can be seen as calculating reliability in the following way:

$$\frac{\text{(\# of actually agreeing tag pairs) – (\# of possible pairs given tags)}}{\text{(Maximum \# agreeing pairs given tags) – (\# of possible pairs given tags)}}$$

An alpha of $\geq .80$ is considered reliable.

4. Results & Discussion

4.1 Inter-Annotator Agreement

Alpha scores for each section were as follows:

Spoken:
Tag agreement: 1.0 Idiom agreement: -0.34

Non-Fiction:
Tag agreement: 0.85 Idiom agreement: -0.199

Fiction:
Tag agreement: 0.83 Idiom agreement: -0.28

For each section of the annotated sample, alpha scores for tag agreement exceeded the .80 threshold of reliability, indicating that annotators agreed reliably and significantly above chance upon the types of idioms tagged. Tag agreement was most notable in the section of the sample drawn from spoken narratives, where a 1.0 alpha score indicates 100% agreement among annotators with regard to the type of tag. That is, when an idiomatic expression such as *drawing a blank* occurred, all annotators agreed that the idiom constituted a VN construction. Annotators also noted a few constructions for which they did not have tags, in particular suggesting that future research include a tag for verb + prepositional phrase constructions (VPPCs) such as *walking on air*.

Unlike the alpha scores for tag agreement, the alpha scores for idiom agreement were all very low, indicating that the level of agreement among annotators regarding which phrases were actually idioms was no better than chance. In fact, there were many instances in which only one annotator tagged a particular construction as an idiomatic expression. One possible explanation for the low inter-annotator agreement could be, as some studies suggest (Gibbs 1984), that there is no clear psychological distinction between literal and figurative expressions, so annotating for idioms is a rather difficult and subjective task, and the level of disagreement our annotators demonstrated on literal-idiom expressions is to be expected. The fact that we did not give our annotators a preset list of idioms to look for (as is usually done in corpus-based research on idioms) might have made the task more difficult as well.

Because the inter-annotator agreement for each idiom list was low, each group of annotators held a meeting following the analysis of the results to create a finalized list of idioms to be used as the official idioms for the corpus. The annotated corpus is currently available on Montclair State's server and will soon be available on the following website as well: http://netdrive.montclair.edu/~streetl1/index.htm. We hope that the tagset and the methodology we have developed for this pilot study can serve as the basis for developing a more complete tagset for annotating the entire ANC.

4.2 Type and Token Counts

In total, 135 idiom types were found in the corpus with a total of 154 idiom tokens appearing. The following tables provide a breakdown of the idioms found within each section of the corpus. Since some types were found in more than one portion of the corpus, the types given sum to 140 rather than 135.

	Fiction		Nonfiction		Spoken	
	Types	Tokens	Types	Tokens	Types	Tokens
VNC	85	93	10	10	16	19
PP	3	3	14	14	10	12
SC	0	0	1	1	1	2
∑	88	96	25	25	27	33

Table 5. Type and token counts for each sample

	VNC	PP	SC	∑
∑ Tokens	122	29	3	154

Table 6. Total token counts for each idiom type

As one might expect, most of the annotated idioms occurred in the fiction sample. The total number of tokens in the fiction sample (n = 96) was almost three times the total number of tokens in the spoken sample (n = 33) and almost four times the total number of tokens in the nonfiction sample (n = 25). Variations between fiction, nonfiction, and transcribed speech are well documented, so it is not surprising that more idioms occurred in the fiction sample. Figurative language just tends to be more commonplace in fiction. Another major finding was that 79% of the idioms tagged were VNCs (n = 122). What was surprising, however, was that VNCs and PPs occurred in much closer numbers in both the nonfiction (n = 10, n = 14) and the spoken samples (n = 19, n = 12). PPs also occurred roughly 4.3 times more in the nonfiction and spoken samples than in the fiction sample. These findings indicate that many idioms are missed when researchers annotate texts solely for VNCs—particularly if those texts come from genres other than fiction. One possible explanation for why VNCs outnumber PPs in the fiction domain is that novel figurative language, which tends to occur more frequently in fiction, is likely to take the form of a VNC rather than a PP. More research needs to be done, however, to explain why VNCs occur more frequently than PPs in this domain. Finally, since there were so few instances of SCs, we recommend that future research on idiom detection prioritize other types of constructions over SCs. In particular, verb + prepositional phrase constructions (VPPCs) appear to be a good candidate as several of our annotators noted higher instances of them while tagging.

5. Conclusion

In a pilot study, portions of the American National Corpus (ANC) were tagged for idioms composed of verb-noun constructions, prepositional phrases, and subordinate clauses. The three data sets we analyzed included 1,500-sentence samples from the spoken, the non-fiction, and the fiction portions of the ANC. In tagging the ANC for idiomatic expressions, our annotators achieved a high level of agreement (> .80) on the tags but a low level of agreement (< .00) on what counted as an idiom. These findings support the claim that identifying idiomatic and metaphorical expressions in a text is a highly difficult and subjective task. In total, we identified 135 idiom types and 154 idiom tokens. Based on the total tokens found for each idiom class, we suggest that future research on idiom detection and idiom annotation include prepositional phrases. This class of idioms occurred frequently in the nonfiction and spoken portions of our corpus.

6. Appendix

The final list of idioms for all three groups of annotators:

at that point	PP
across history	PP
are spreading out	VN
as melted architecture	SC
at face value	PP
at the top of his lungs	PP
be a tall order	VN
be able to bring	VN
be better off	VN
be carried out for anyone	VN
be totally out of character	VN
behind-the-scenes	PP
being saddled with a wrong name	VN
being torn apart	VN
blurring the edges	VN
bugs me	VN
by the same token	PP
by the wayside	PP
call him on it	VN
calls to mind	VN
came across a passage	VN
came pat	VN
camp up	VN
can swing that	VN
can't figure out	VN
claim any deep roots	VN
come to terms with it	VN
come up with all this stuff	VN
conjures up an image	VN
conquered his fear	VN
cooling down	VN
count on your continued partnership	VN
cracked it	VN
cried out on that topic	VN
dawned on me	VN
deal with it	VN
depend on me	VN
didn't get it	VN
drawing a blank	VN

fell on the vermicelles	VN
fess up to how things had developed	VN
figure out	VN
find your way around	VN
fished around	VN
foot a bill	VN
for anything in the world	PP
freeze with horror	VN
from point to point	PP
from the bottom of my heart	PP
from time to time	PP
gave himself back	VN
get off his butt	VN
get off my back	VN
get over it	VN
get round the issue	VN
get through it	VN
get your goat	VN
give you a hand	VN
go on about that	VN
go to great lengths	VN
go up	VN
going my own way	VN
grabbed my attention	VN
grew out of a popular mass movement	VN
had a gift	VN
had my share	VN
have a clue	VN
have a foot to stand on	VN
have his memory in common	VN
have the same connection to him	VN
having such an impact	VN
hear about it	VN
held our attention	VN
higher in the feudal chain	PP
hit me	VN
hit me	VN
if I don't make it	SC
in a related vein	PP
in the right direction	PP
in trouble	PP
into boys	PP
is a world away	VN
keeled over	VN
keep them from selling it	VN
keep up to his example	VN
keep your head high	VN
land us in trouble	VN
learned first hand	VN
left for dead	VN
lies with you	VN
log off	VN
make it up	VN
of the gods	PP
off the hook	PP
on my dad's side	PP
on the downside	PP
on the other hand	PP
open your eyes	VN
pencil children	VN
pick out	VN
picked it apart	VN
picking up your contribution	VN
picture things	VN
playing a mean practical joke on them	VN
put on display	VN
putting your gift to work	VN
reach out to more girls	VN
round-the-clock	PP
run into	VN
sandwiching concentrated time	VN
seeing loved ones off	VN
set up camp	VN
shafted you	VN
skip school	VN
smack in the center	PP
sparks negative attitudes	VN
stir up a little mud in this matter	VN
sums it up	VN
sweeps away the vestiges	VN
take the first step	VN
take the sting out of them	VN
take your pledge	VN
taken aback	VN
throwing out onto him	VN
to be at odds	PP
took his eyes off the other	VN
torn in many directions	VN
turn a look of astonishment on him	VN
turned away	VN
turned back	VN
under god and under law	PP
up-to-the-minute	PP
walk in someone else's shoes	VN
was caught up in this idea	VN
was crowded with passengers	VN
was more or less	VN
work toward	VN

7. References

Baldwin, T., Bannard, C., Tanaka, T., Widdows, D. (2003). An empirical model of multiword expression decomposability. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pp. 89-96.

Bannard, C. (2007). A measure of syntactic flexibility for automatically identifying multiword expressions in corpora. In *Proceedings of the ACL-07 Workshop on a Broader Perspective on Multiword Expressions*, pp. 1-8.

Birke, J., Sarkar, A. (2006). A clustering approach for the nearly unsupervised recognition of nonliteral language. In *Proceedings of the EACL-06*, pp. 329-336.

Dekang, L. (1999). Automatic identification of noncompositional phrases. In *Proceedings of the ACL-99*, pp. 317–324.

Fazly, A., Stevenson, S. (2006). Automatically constructing a lexicon of verb phrase idiomatic combinations. In *Proceedings of the EACL-06*, pp. 337-344.

Fazly, A., Cook, P., Stevenson, S. (2009). Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics*, 35(1), pp. 61-103.

Fillmore, C., Ide, N., Jurafsky, D., Macleod, C. (1998). An American National Corpus: A Proposal. In *Proceedings of the First International Language Resources and Evaluation Conference*, pp. 965-70.

Flores d' Arcais, G. B. (1993). The comprehension and semantic interpretation of idioms. In C. Cacciari and P. Tabossi (Eds.) *Idioms: Processing, Structure, and Interpretation*, (pp. 79-98). New Jersey: Lawrence Erlbaum and Associates, Inc.

Gibbs, R. W. (1984). Literal meaning and psychological theory. *Cognitive Science* 8, pp. 275-304.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *TRENDS in Cognitive Sciences*, 7(5), pp. 219-224.

Hashimoto, C., Kawahara, D. (2008). Construction of an idiom corpus and its application to idiom identification based on wsd incorporating idiom-specific features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 992-1001.

Katz, G., Giesbrecht, E. (2006). Automatic identification of non-compositional multiword expressions using latent semantic analysis. In *Proceedings of the ACL/COLING-06 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pp. 12–19.

Kuiper, K., McCann, H., Quinn, H., Aitchison, T., van der Veer, K. (2003). A syntactically annotated idiom dataset (SAID). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Langlotz, A. (2006) *Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom- Representation*, (pp. 15-53). Philadelphia: John Benjamins North America.

Sporleder, C., Li, L. (2009). Unsupervised recognition of literal and non-literal use of idiomatic expressions. In *Proceedings of the EACL*, pp. 754–762.

Swinney, D. A., Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523-534.