Systematic Flexibility in Verb–Object Idioms[*]

Megan Schildmier Stone ☼ U of Arizona ☼ stonem@email.arizona.edu

BCGL 8: The Grammar of Idioms

Brussels, Belgium ☼ 5 June 2015

Preview

► I conducted two corpus experiments (one using Google Books and one using COCA) investigating the syntactic behavior of idioms in English.

► The results revealed a previously unnoticed empirical pattern: If an idiom was found in nominal gerundization, it was also found in object incorporation. And if an idiom was found in object incorporation, it was also found in passivization.

   o **gerundization ⇒ incorporation ⇒ passivization**

► This pattern supports the structural requirement approach to idiomaticity (Folli and Harley 2007, Stone 2008, Harley and Stone 2013, Punske and Stone 2014), which holds that differences in idiomatic flexibility are tied to differences in the functional structure required by different idioms in order to retain their idiomatic interpretations.

   o For example, if an idiom requires an active Voice head to license its idiomatic interpretation, passivization would be prohibited due to the mismatch in functional structure.

► Because the structural requirement approach relies on independently motivated functional structure—such as the articulated verb phrase—to account for idiomatic behavior, its plausibility is increased.

Outline

► Background on idiom flexibility

► Corpus experiments

► Interpretation of results

► Theoretical implications

► Conclusions and future work

Background

► **Idiomatic flexibility** refers to an idiom's interpretability in configurations other than its canonical form.

► **Manipulation** is intended to be a theory–neutral term referring to one of various noncanonical forms that idioms—specifically here, verb–object idioms—might take (e.g. passivization, topicalization).

► Of particular interest is the observation that some idioms are relatively flexible, while others are not (Chafe 1968, Fraser 1970, Katz 1973, Fiengo 1974, Newmeyer 1974, etc.).

► This can be seen in the contrast between idiomatic *spill the beans* ('divulge a secret') and *kick the bucket* ('die') with respect to passivization in (1) and (2) below.

   1) a. Eli spilled the beans.
       b. The beans were spilled (by Eli).
   2) a. Lauren kicked the bucket.
       b. # The bucket was kicked (by Lauren).

► The corpus experiments outlined below contribute to the growing body of corpus work on idioms (Moon 1998, Fellbaum 2007, etc.). Specifically, I'm interested in investigating the range of idiomatic flexibility.

Google Books Corpus Study

Research Questions

► Do idioms fall into two classes—flexible and inflexible—or is there a greater range of variation such that there are subclasses of idioms?

► Are there any patterns within the variation?

► 38 verb-object idioms were selected based on a familiarity survey of University of Arizona undergraduates.

► The Google Books corpus, which has over 360 billion words of English, was used to search for each idiom in each of the following 6 manipulations (examples from the COCA corpus unless otherwise noted):

  o Object modification

    3) You're gonna help murder this guy, you **draw the moral line** at extramarital sex?

  o Object quantification

    4) We **draw another line**. And they blow up an entire village…

  o Passivization

    5) **The line was drawn**, and I think it was a turning point for the president…

  o Nominal gerundization

    6) I -- I'm seeking some -- a **drawing of the line** by the companies themselves…

  o Object incorporation

    7) Cases that fall between the poles may present hard **line-drawing** questions…

  o Topicalization

    8) **The line** I was hesitant to **draw**. (constructed example)

► I used a binary 1/0 coding for manipulations. An idiom was coded as '1' for a given manipulation if I found at least one example in the corpus; otherwise, it was coded as '0'.

► If a search returned more than 500 hits, I looked at only the first 500 results.

► I had difficulty finding an appropriate search string to use to search for topicalization, and thus to find relevant results, so it has been excluded from further discussion.

► The distribution of 1s and 0s makes it clear that there are not simply two classes of idioms (see Appendix 1).

  o 22 of the 38 idioms participated in some but not all of the manipulations.

► There is an implicational relationship between gerundization, incorporation, and passivization: If an idiom was found in gerundization, it was also found in incorporation; if it was found in incorporation, it was also found in passivization.

  o Two exceptions to be discussed below.

► In contrast to this implicational relationship, the quantification and modification manipulations seem to be independent of each other, as well as the other manipulations.

► Fisher's Exact Test was used to evaluate dependence between each pair of manipulations. Only the three implicationally related manipulations meet the minimum significance threshold, with Bonferroni correction, of $p < 0.005$ (Table 1).

Table 1: Fisher's Exact Test results for paired manipulations of idioms in Google Books

	Modification	Passivization	Incorporation	Gerundization
Quantification	$p=0.0062$	$p=0.0099$	$p=0.15$	$p=0.037$
Modification		$p=0.19$	$p=0.049$	$p=0.025$
Passivization			$p=4.74e-05$•	$p=2.20e-07$•
Incorporation				$p=1.89e-08$•

► Multiple Correspondence Analysis (MCA; Joliffe 1986) was also used to investigate the relationships among these manipulations.

► MCA is a form of dimensionality reduction that tries to account for a large amount of variance using the smallest number of uncorrelated dimensions, called Principal Components (PCs).

► Figure 1 shows the results of MCA on the Google Books data. These results are summarized in Table 2.

Figure 1: MCA of idiom flexibility in Google Books
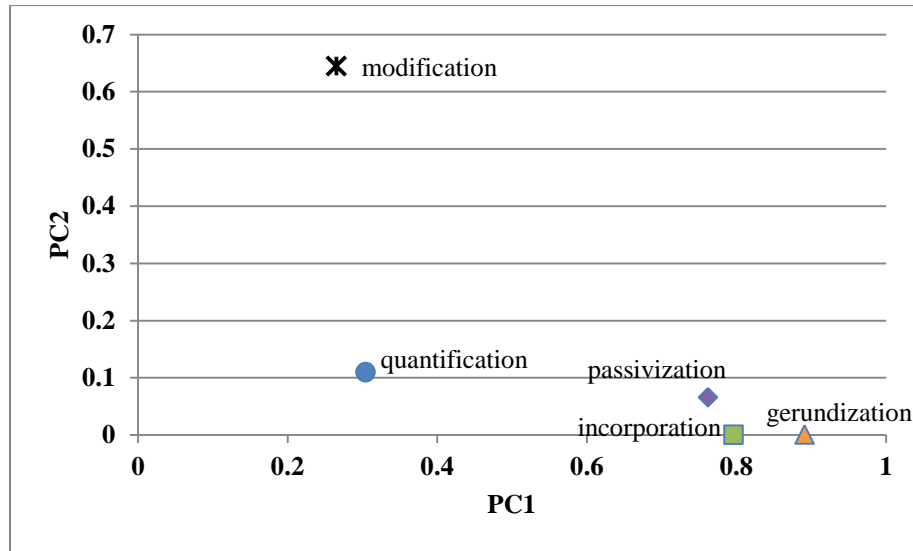
Table 2: MCA summary for Google Books

	PC1	PC2	PC3	PC4	PC5
Percentage of Variance	60.36	16.46	16.17	5.50	1.51
Cumulative Percentage	60.36	76.82	92.99	98.49	100.00

► In Figure 1, we see the first two PCs of the MCA. What's striking here is the fact that passivization, incorporation, and gerundization are closely grouped, while modification and quantification appear to be independent. The interpretation of these results will be discussed below.

COCA Corpus Study

Motivation

► The large size of Google Books (350 billion words) made it a good choice for investigating a relatively rare phenomenon like idiom flexibility.
► However, it also had some drawbacks:
  o Ambiguity of "hits"
  o Binary data rather than actual counts due to inability to sort through every search result
  o Terrible search interface
► The Corpus of Contemporary American (Davies 2008–) overcomes all three:
  o Finite corpus with no ambiguous results
  o Smaller size (450 million words) allows for gathering of actual counts of idiomatic uses
  o Very user-friendly and flexible search interface
► At three orders of magnitude smaller, however, there's a real concern that COCA won't have enough data.

Research Questions

► Is the implicational hierarchy from Google Books present in COCA?
► Are there any additional patterns?

Methodology

► 40 verb-object idioms were investigated; 32 were shared between the two studies.
► 6 idioms from the original study were excluded.
  o *go the distance*, *hit the spot*, and *see the light* experienced "interference" during the search process from related metaphorical uses (e.g. is "see the light of God" idiomatic?)
  o *cross paths*, *give a hand*, and *lend a hand* were excluded due to difficulties with argument structure and the search interface (e.g.

searching for "give a hand" omits all ditransitive uses and captures only the prepositional dative uses)

► 8 additional idioms were added to bring the total to an even 40.
  o Most were selected based on the results of the original idiom familiarity survey.

► COCA, with 450 million words, was used to search for each idiom in the same manipulations as above, excluding topicalization.

► In most cases, the exact number of idiomatic search results for each string was recorded.

► Five of the searches conducted for this experiment yielded too many results to be sorted manually. In those cases, I examined 100 random strings and extrapolated to the total number of results to get my final count.

Results

► Manipulated idioms are, overall, quite rare in the COCA data; zeros fill over 50% of the cells in the chart, and only 13.5% contain numbers greater than single digits (see Appendix 2).

► The implicational hierarchy from Google Books is less obvious here, although the data are suggestive.
  o Four exceptions are discussed below.

► There is a new implicational relationship in the COCA data that was not present in the Google Books data: If an idiom was found in any other manipulation, it was also found in modification.

► The same statistical tests were used to examine these data.

► Using Fisher's Exact Test, only the quantification-passivization relationship meets the $p < 0.005$ threshold of significance, again with Bonferroni correction (Table 3).

Table 3: Fisher's Exact Test results for paired manipulations of idioms in COCA

	Modification	Passivization	Incorporation	Gerundization
Quantification	$p=0.014$	$p=0.0011$●	$p=0.052$	$p=0.37$
Modification		$p=0.029$	$p=0.075$	$p=0.57$
Passivization			$p=0.0059$	$p=0.029$
Incorporation				$p=0.014$

► Principal Component Analysis (PCA) was used instead of MCA on these data, because MCA is a statistical measure for categorical data, while PCA is the corresponding measure for continuous data.

► Figure 2 shows the PCA results for the COCA data. These results are summarized in Table 4.

Figure 2: PCA of idiom flexibility in COCA
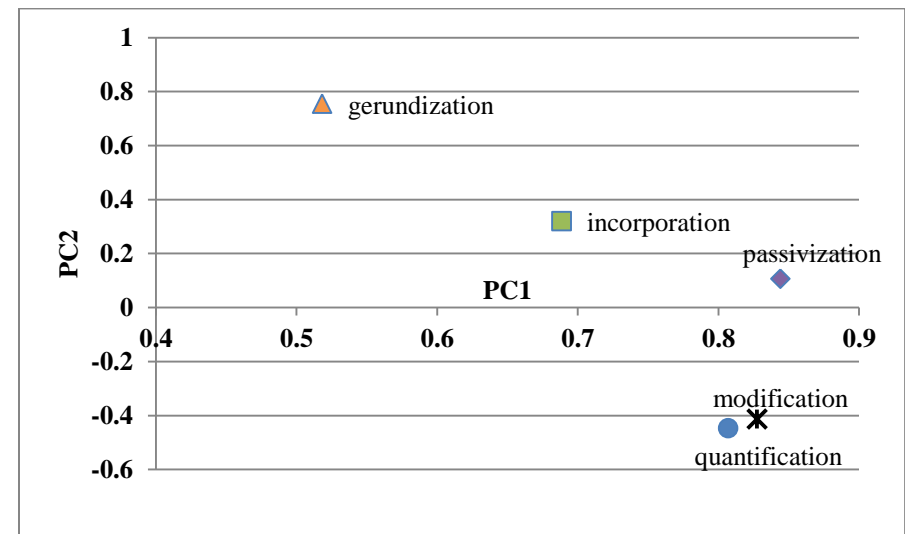
Table 4: PCA summary for COCA

	PC1	PC2	PC3	PC4	PC5
Percentage of Variance	55.81	21.10	11.68	6.96	4.46
Cumulative Percentage	55.81	76.90	88.58	95.54	100.00

► In Figure 2, modification and quantification pattern together with respect to both PC1 and PC2, while passivization, incorporation, and gerundization seem to be set apart from the others by PC2 only. The interpretation of these results follows.

Interpretation

► Overall, these results tentatively support the conclusion that three manipulations are implicationally related: gerundization implies incorporation, which in turn implies passivization.

► In other words, if an idiom allows gerundization, it should also allow incorporation; and if it allows incorporation, it should also allow passivization.

► The Google Books data support this conclusion most robustly; Fisher's Exact Test highlights the significant dependences among these three manipulations, and MCA groups them into a tight cluster on both the PC1 and PC2 dimensions.

► I hypothesize that PC1 reflects argument structure, given that the division here is between manipulations that affect argument structure (passivization, incorporation, and nominalization) and those that don't (modification and quantification).

► The nature of PC2 is less clear.

► In Google Books, PC1 accounts for 60% of the variance; in other words, grouping passivization, incorporation, and gerundization separately from quantification and modification explains over half of the variance in these data.

► There were two exceptions in the Google Books data, which could be accidental gaps or might require more explanation:

- o *hit the spot* was found in object incorporation but not passivization; there was quite a bit of skewing from the high frequency of the idiomatic NP *sweet spot*, which could have affected the results here
- o *bust a move* was also found in object incorporation but not passivization; I'm not sure what could be motivating this inconsistency

► The COCA data were less straightforward.

► Fisher's Exact Test didn't tell us much about anything, which is likely due to the nature of the test itself. Fisher's Exact Test identifies dependence only with lopsided distributions across both variables (that is, an X shape in the contingency table). For example, in this case it would require both the number of idioms appearing in neither manipulation and the number of idioms appearing in both manipulations to be larger than the number of idioms appearing in each manipulation to the exclusion of the other. But this distribution does not fit all implicational relationships, including those found in the COCA data. There are also a lot of zeros potentially skewing the data.

► PCA was a bit more interesting. It does look as though PC2 is distinguishing quantification and modification from passivization, incorporation, and gerundization—the former are tightly clustered around –0.4, while the remaining three manipulations are in the positive range.

► PC1 accounts for 56% of the variance, while PC2 accounts for 21%.

► The proposed implicational hierarchy had four exceptions in COCA:

- o *jump the gun* is found in incorporation but not passivization; *gun-jumping* has an established meaning in business (trading securities on the basis of information that is not yet publicly disclosed) and is therefore expected to be more frequent than its passive counterpart
- o *rock the boat* is found in incorporation but not passivization; *boat-rocking* is also frequent in newsprint and the inconsistency might be attributable to frequency concerns as well
- o *take a chance* is found in incorporation but not passivization; this might be a light verb construction and might therefore be less likely to passivize for independent reasons (c.f. #a walk was taken; Folli and Harley 2013)

- o *spill the beans* is found in nominal gerundization but neither incorporation nor passivization; the missing passive form is almost certainly an accidental gap, given this idiom's canonical status in the literature as a passivizable idiom, and many native speakers and trained linguists corroborate my intuition that it can also undergo incorporation
- ► I conclude that the size of COCA was likely the primary factor in the discrepancy between corpora, being three orders of magnitude smaller than Google Books. Manipulated idioms are surprisingly rare even in the larger corpus.

Theoretical Implications

- ► What kind of theory could account for the implicational hierarchy established by the corpus data?
- ► It must have machinery capable of explaining this subset behavior.
- ► How well do proposed theories of idiomatization stack up?

Lexical Specification

- ► Some accounts of idiomatic flexibility posit that an idiom's ability to appear in a given manipulation is idiosyncratic and must be lexically specified (Weinreich 1969, Katz 1973, Chomsky 1981).
- ► According to this view, each idiom's flexibility is independently marked.
- ► As a result, any patterns of flexibility would be accidental.
- ► It's highly unlikely that the kind of systematic behavior observed in the data would be accidental in this sense.
- ► One exception is Fraser's (1970) Frozenness Hierarchy.
- ► Fraser claims that *all* idiomatic flexibility exhibits the subset behavior observed here for gerundization, incorporation, and passivization.
- ► According to Fraser, if an idiom is "frozen" at a given level, then it will allow all manipulations at the lower levels.

- ► However, while the Frozenness Hierarchy is able to capture the implicational hierarchy observed here, it doesn't seem to be right for the other manipulations observed (e.g. quantification).
- ► Lexical specification predicts random flexibility in idioms, not systematic behavior, as we have seen.
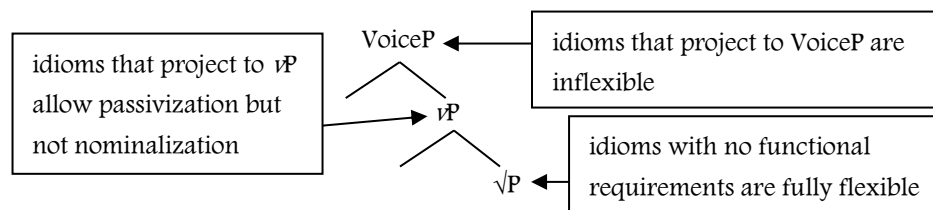
Semantic Mapping

- ► The most well-known approach to idiom flexibility is Nunberg, Sag, and Wasow's (1994) semantic mapping (see also Nunberg 1978; Wasow, Sag, and Nunberg 1984; Gazdar et al. 1985; Ruwet 1991; Pulman 1993).
- ► The basic proposal is that idioms come in two types: flexible and inflexible.
- ► Flexible idioms—their idiomatically combing expressions—can undergo manipulations because the elements of the literal expression can be mapped onto elements of the figurative meaning.
- ► In contrast, inflexible idioms—their idiomatic phrases—cannot undergo manipulations because no such mapping is possible.
- ► This mapping is represented schematically in (9) and (10) below.

9) **idiomatically combining expression**   10) **idiomatic phrase**

    spill   *the beans*   *kick* *the bucket*
    ↓   ↓   ✕
  divulge a secret   die
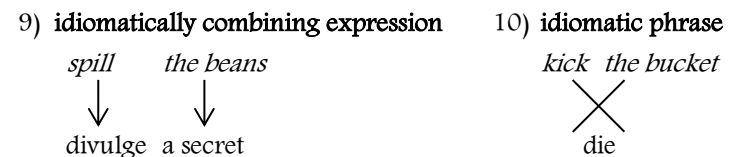
- ► Proponents of semantic mapping acknowledge that other factors may affect flexibility, e.g. a verb's ability to passivize independently of the idiom.
- ► However, this approach has no mechanism to account for subset behaviors like those reported here.
- ► Semantic mapping predicts two basic classes of idioms—flexible and inflexible—with exceptions based on other independent properties. It is unlikely that these exceptions would manifest themselves in a strict hierarchical relationship.

► The structural requirement approach to idiom flexibility (Folli and Harley 2007, Stone 2008, Harley and Stone 2013, Punske and Stone 2014) holds that there is a link between flexibility and independently motivated functional structure.

► The amount of functional structure required to license an idiom's special interpretation determines which manipulations that idiom allows.

► Punske and Stone (2014) provide the following schematic:

idioms that project to vP allow passivization but not nominalization

idioms that project to VoiceP are inflexible

idioms with no functional requirements are fully flexible

► Following Pylkkänen (2002) and Folli and Harley (2005), this schematic assumes that VoiceP is the control center for passivization.

► Following Harley (2008) and Punske (2012), nominal gerundization takes place below VoiceP, at the level of vP.

► We can now identify a "typology" of idioms (a la Punske and Stone 2014):
  o VoiceP idioms require both VoiceP and vP and are therefore inflexible with regard to both passivization and nominalization.
  o vP idioms require just the vP layer of functional structure and are therefore passivizable but not nominalizable.
  o √P idioms, true root + object idioms, place no requirements on functional structure and are therefore both passivizable and nominalizable.

► *Kick the bucket* is an example of a VoiceP idiom.

► Because *kick the bucket* requires an active Voice head, Voice$_{DO}$, it does not retain its idiomatic meaning in the passive (Figure 4).

► Additionally, because VoiceP is fixed, vP is also fixed, which means that the idiomatic meaning is not available in the nominal gerund either (Figure 5).

► (Trees simplified for the sake of space.)

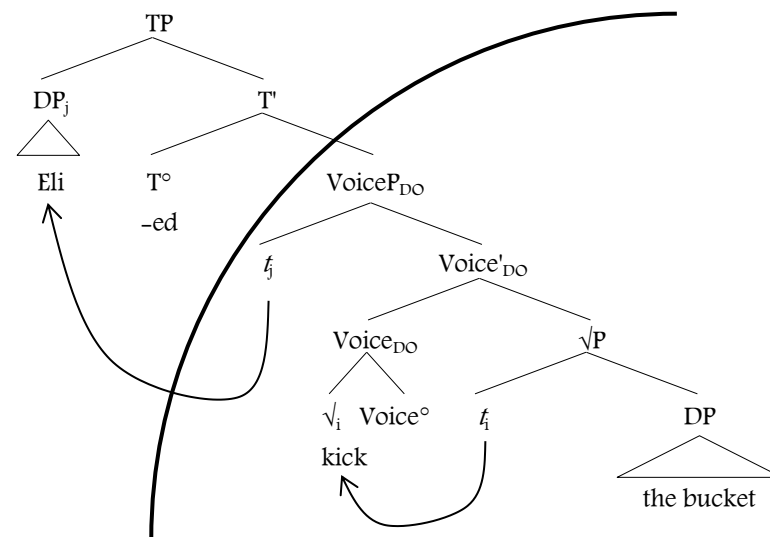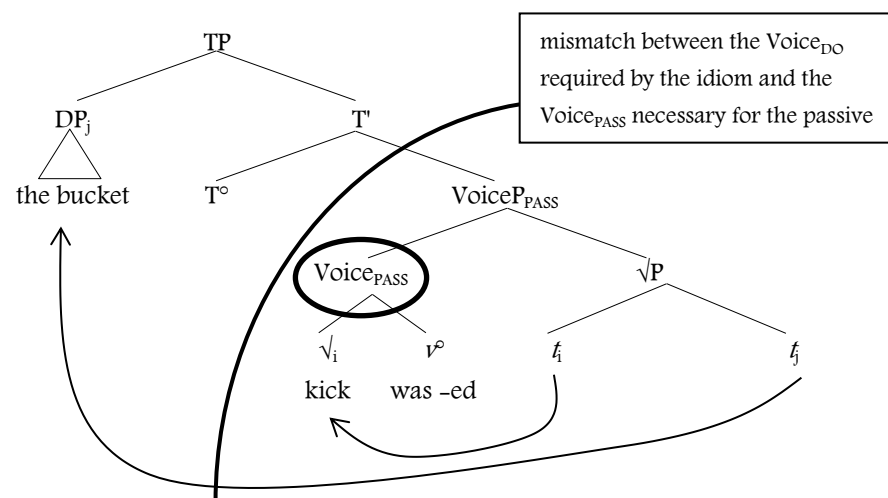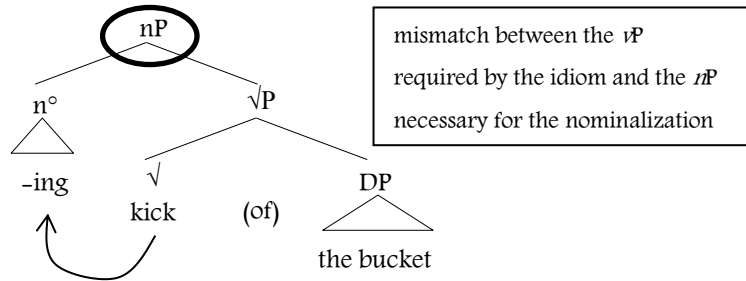**Figure 3:** VoiceP idiom – *Eli kicked the bucket*

Figure 4: VoiceP idiom – *#The bucket was kicked*

mismatch between the Voice$_{DO}$ required by the idiom and the Voice$_{PASS}$ necessary for the passive

Figure 5: VoiceP idiom – *#kicking of the bucket*

mismatch between the *v*P required by the idiom and the *n*P necessary for the nominalization

▶ *Make a killing* is a *v*P idiom—it can passivize but not nominalize (Punske 2010, Punske 2012, Bruening 2013).
    11) a.   The audience was killed (by Johnny Carson).
        b.   #The killing of the audience (by the witty comedian)…
▶ *Turn the tables* is a √P idiom—it can passivize and nominalize.
    12) a.   The tables were turned (by her opponent).
        b.   The turning of the tables (by his friend Jeannette)…

▶ Unlike the lexical specification and semantic mapping approaches, the structural requirement approach has the machinery to accommodate patterns of systematic flexibility in idioms.
▶ We have seen how three of the attested idiom types can be associated with established functional structure.
▶ The corpus results revealed a fourth type of idiom, one that allows passivization and object incorporation but not nominal gerundization.
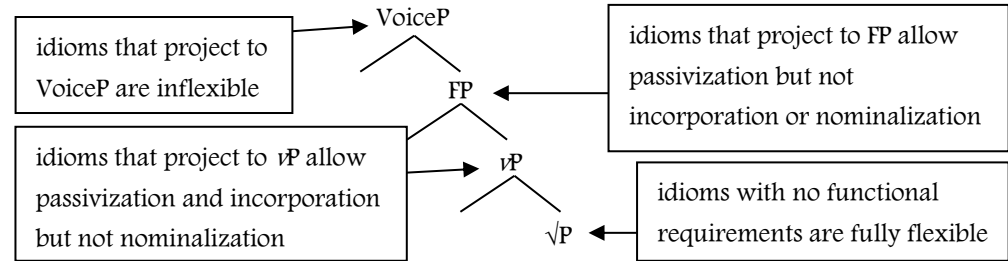▶ *Steal the show* is an example of this type of idiom.[1]
    13) a.   The show was stolen (by the flying dolphin).
        b.   The show-stealing flying dolphin…
        c.   #The stealing of the show (by the flying dolphin)…

---

[1] Under the new schematic *steal the show* would be a *v*P idiom. *Make a killing* would be an FP idiom, as it allows passivization but neither incorporation nor nominalization.
    (i). *The stockbroker's killing-making practices…

▶ This predicts the existence of another functional projection between the VoiceP and *v*P levels, as outlined in the revised schematic below:

idioms that project to VoiceP are inflexible

idioms that project to FP allow passivization but not incorporation or nominalization

idioms that project to *v*P allow passivization and incorporation but not nominalization

idioms with no functional requirements are fully flexible

▶ This FP projection should be the "control center" for object incorporation, if the predictions of the experiment are borne out.
▶ Note that all but one of the exceptions discussed above involve hierarchical reversals of passivization and incorporation, so it's possible that FP should ultimately be located above VoiceP (although this is problematic for phase-based accounts of idiomatic domains; c.f. Marantz 1984, Stone 2008, Svenonius 2005, Harley and Stone 2013, Harwood 2013).

▶ Of all the available theories of idiomatic flexibility, the structural requirement approach is best equipped to handle the subset relationship found among the passivization, incorporation, and gerundization manipulations.
▶ This approach utilizes the hierarchical syntactic structure of the verb phrase to account for this implicational hierarchy in idiom flexibility.
▶ The empirical data, in turn, make a prediction about the structure of the verb phrase and the presence of another as-yet-undiscovered functional projection intervening between VoiceP and *v*P.

Conclusions and Future Work

▶ The corpus experiments reported here tentatively support an implicational relationship among three idiomatic manipulations: gerundization implies incorporation implies passivization.

- ► Only the structural requirement approach outlined above can predict this type of subset behavior using independent theoretical machinery—the hierarchical structure of the verb phrase.
- ► It is clear that corpus size is an important factor in studying idiom manipulations; their relative rarity makes it difficult to draw conclusions about their behavior from a corpus of even 450 million words.
- ► Future work is needed to confirm the implicational relationship found here.
  - o Bigger corpora
  - o Better statistical test for implication
  - o Closer examination of exceptions
- ► Another future project involves investigating the full range of idiom flexibility, rather than limiting the scope to selected manipulations (c.f. Fellbaum 2007). The following are just a few examples found in COCA during my investigation.
  - 11) …she coulda been Melissa Mikawa's assistant, all the media **trash she was talkin'**.
  - 12) …you look great, but I got to tell you, **heads were turning** at the DNC.
  - 13) With Bob Hanssen, the church **dropped the ball**. They fumbled it.
  - 14) A kid who once **spilled every bean** may now say, "It's a secret" when you ask something as innocuous as "What did you do at school today?"

References

Bruening, Benjamin. 2013. *By* phrases in passives and nominals. *Syntax* 16: 1–41.

Chafe, Wallace. 1968. Idiomaticity as an Anomaly in the Chomskyan Paradigm. *Foundations of Language* 4: 109–127.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Davies, Mark. 2008–. *The Corpus of Contemporary American English: 450 million words, 1990–present*. Available online at http://corpus.byu.edu/coca/.

Fellbaum, Christiane, ed. 2007. *Idioms and collocations: Corpus-based linguistic and lexicographic studies*. London: Continuum.

Fiengo, Robert. 1974. *Semantic conditions on surface structure*. Doctoral dissertation, MIT.

Folli, Raffaella and Heidi Harley. 2005. Flavors of v: Consuming results in Italian and English. In *Aspectual Enquiries*, ed. by Paula Kempchinsky and Roumyana Slabakova, 95–120. Dordrecht: Springer.

Folli, Raffaella and Heidi Harley. 2007. Causation, obligation, and argument structure: On the nature of little v. *Linguistic Inquiry* 38: 197–238.

Folli, Raffaella and Heidi Harley. 2013. The syntax of argument structure: Evidence from Italian complex predicates. *Journal of Linguistics* 4: 993–125.

Fraser, Bruce. 1970. Idioms within a transformational grammar. *Foundations of Language* 6: 22–42.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. *Generalized phrase structure grammar*. Oxford: Basil Blackwell.

Harley, Heidi. 2008. The morphology of nominalizations and the syntax of *v*P. In *Quantification, definiteness, and nominalization*, ed. by Anastasia Giannakidou and Monika Rathert, 320–342. Oxford: Oxford University Press.

Harley, Heidi and Megan Schildmier Stone. 2013. The 'No Agent Idioms' hypothesis. In *Syntax and its limits*, ed. by Raffaella Folli, Christina Sevdali, and Robert Truswell, 251–275. Oxford: Oxford University Press.

Harwood, W. 2013. *Being progressive is just a phase: dividing the functional hierarchy*. Doctoral dissertation, Universiteit Gent.

Joliffe, I.T. 1986. *Principal component analysis*. New York: Springer.

Katz, Jerrold J. 1973. Compositionality, idiomaticity, and lexical substitution. In *A Festschrift for Morris Halle*, ed. by Stephen R. Anderson and Paul Kiparsky, 357–76. New York: Holt, Rinehart, and Winston.

Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.

Moon, Rosamund. 1998. *Fixed expressions and idioms in English: A corpus-based approach*. Oxford Studies in Lexicography and Lexicology. Oxford: Clarendon Press.

Newmeyer, Frederick. 1974. The regularity of idiom behavior. *Lingua* 34: 327–42.

Nunberg, Geoffrey D. 1978. *The pragmatics of reference*. Doctoral dissertation, City University of New York.

Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 1994. Idioms. *Language* 70: 491–538.

Pulman, Stephen G. 1993. The recognition and interpretation of idioms. In *Idioms: Processing, structure, and interpretation*, ed. by Cristina Cacciari and Patrizia Tabossi, 249–270. Hillsdale, NJ: Lawrence Erlbaum Associates.

Punske, Jeffrey. 2010. On nominal arguments. In *Proceedings of the Arizona Linguistics Circle 3* (Coyote Papers: Working Papers in Linguistics, Volume 17), ed. by Alan Hogue and Jessamyn Schertz. Tucson: University of Arizona.

Punske, Jeffrey. 2012. *Aspects of the internal structure of nominalization: Roots, morphology, and derivation*. Doctoral dissertation, University of Arizona.

Punske, Jeffrey, and Megan Schildmier Stone. 2014. Idiomatic expressions, passivization, and gerundization. Handout, LSA 2014 Annual Meeting, Minneapolis, MN. 3 January 2014.

Pylkkänen, Liina. 2002. *Introducing arguments*. Doctoral dissertation, MIT.

Ruwet, Nicolas. 1991. *Syntax and human experience*. Chicago: University of Chicago Press.

Stone, Megan Schildmier. 2008. Passivizable and nonpassivizable idioms: A minimalist approach. Ms., University of Arizona.

Svenonius, Peter. 2005. Extending the extension condition to discontinuous idioms. *Linguistic Variation Yearbook* 5: 227–263.

Wasow, Thomas, Ivan A. Sag, and Geoffrey Nunberg. 1984. Idioms: An interim report. In *Proceedings of the XIIIth International Congress of Linguistics*, ed. by Shiro Hattori and Kazuko Inoue, 87–96. Tokyo: Nippon Toshi Center.

Weinreich, Uriel. 1969. Problems in the analysis of idioms. In *Substance and structure of language: Lectures delivered before the Linguistic Institute of the Linguistic Society of America, University of California, Los Angeles, June 17–August 12, 1966*, ed. by Jaan Puhvel, 23–81. Berkeley and Los Angeles: University of California Press.

Appendix 1: Results of Google Books corpus experiment

Idiom	Pass.	Inc.	Ger.	Mod.	Quant.
break the ice	1	1	1	1	1
break the news	1	1	1	1	1
bust a move	0	1	0	0	0
cross paths	1	1	1	1	0
draw a blank	0	0	0	1	0
draw the line	1	1	1	0	0
drop a bomb	1	1	1	1	0
give a hand	0	0	0	1	0
go the distance	0	0	0	1	1
hang a left	0	0	0	0	0
hang a right	0	0	0	0	0
hang a U-ey	0	0	0	0	0
have a go	0	0	0	1	0
have the last say	0	0	0	0	0
hit home	0	0	0	0	0
hit the hay	1	1	1	0	0
hit the road	0	0	0	1	0
hit the sack	0	0	0	0	0
hit the spot	0	1	0	1	0
jump the gun	1	1	1	0	0
lend a hand	1	1	1	1	0
make a killing	1	0	0	1	0
make amends	1	1	1	1	1
make way	1	0	0	0	0
pay attention	1	1	1	1	1
pop the question	1	1	1	1	0
pull strings	1	1	1	1	1
rock the boat	1	1	1	1	0
see the light	0	0	0	1	0
spill the beans	1	1	1	0	1
start a family	1	1	1	1	0
take a back seat	1	0	0	0	1
take a chance	1	1	1	1	1
take a chill pill	0	0	0	0	0
take a leak	0	0	0	0	0
take a raincheck	0	0	0	0	0
talk trash	1	1	1	1	1
turn heads	1	0	0	0	1

Pass. = Passivization, Inc. = Object Incorporation,

Ger. = Nominal Gerundization,

Mod. = Object Modification, Quant. = Object Quantification

Appendix 2: Results of COCA corpus experiment

Idiom	Pass.	Inc.	Ger.	Mod.	Quant.
break the ice	15	11	0	3	1
break the news	8	5	0	127*	11
bust a move	0	1	0	2	0
clear the air	5	1	3	3	0
draw a blank	1	0	0	9	2
draw the line	93	17	1	142*	11
drop a bomb	2	0	0	18	1
drop the ball	4	0	0	5	0
drop trou	0	0	0	0	0
hang a left	0	0	0	3	1
hang a right	0	0	0	5	0
hang ten	0	0	0	1	0
have a go	0	0	0	7	22
have the final say	0	0	0	1	0
have the last say	0	0	0	0	0
hit home	0	0	0	0	0
hit the hay	0	0	0	1	0
hit the road	0	0	0	6	0
hit the sack	0	0	0	0	0
jump the gun	0	1	0	2	0
kick the bucket	0	0	0	1	0
make a killing	1	0	0	26	1
make amends	6	2	1	15	8
make way	0	0	0	0	0
pay attention	713*	0	0	2242*	2402*
pop the question	2	0	0	5	0
pull a u-ey	0	0	0	0	0
pull strings	5	10	0	13	33
rock the boat	0	3	0	8	0
spill the beans	0	0	1	10	3
start a family	0	0	0	23	8
steal the show	1	7	0	6	1
take a back seat	0	0	0	6	0
take a chance	0	5	0	78	36
take a chill pill	0	0	0	1	0
take a leak	0	0	0	6	0
take a raincheck	0	0	0	0	0
talk trash	1	132	0	8	4
turn heads	2	52	2	14	29
turn the tables	89	4	6	3	0