Idioms, Collocations, and Structure
Syntactic Constraints on Conventionalized Expressions

Benjamin Bruening (University of Delaware)

rough draft, July 7, 2018; comments welcome

Abstract

Phrasal idioms have been used as evidence in syntactic theorizing for decades. A common assumption, occasionally made explicit (e.g., [Larson 2017]), is that non-literal phrasal idioms differ significantly from completely literal collocations in the kinds of syntactic structures they can be built from. I show with a detailed empirical study that this is false. In fact, the syntactic constraints on idioms and collocations are identical. In particular, patterns that are missing from one are missing from the other, most strikingly ditransitives with a fixed first object and open second object (*throw the wolves X). Idioms and collocations should therefore be treated the same, as a broad class of conventionalized expressions. I propose a new analysis of the syntactic forms that conventionalized expressions can take. Unlike all previous analyses, I take the open slots to be part of the expression, and the task then becomes to explain the limited distribution of the open slots. A structural constraint on the open slots accounts for the missing ditransitive pattern. It also explains why expressions with a fixed subject but open object are rare, but also why certain examples of this pattern do exist in English and in other languages.

1 Introduction

There is a long tradition in generative syntax of examining the syntactic patterns of non-literal, conventionalized expressions, and using those patterns to inform our theories of syntax, or even to decide between competing analyses of some phenomenon. As an example, patterns of non-literal phrasal idioms have been used as arguments for the proper analysis of ditransitive constructions (e.g., [Green 1974, Larson 1988, Richards 2001, Harley 2002, Bruening 2010]. Larson (2017) dismisses all of these arguments, claiming that expressions that were taken to be phrasal idioms are actually not. Instead they are mere collocations, combinations of words that frequently co-occur but do not involve any special, non-literal meaning. Larson argues that collocations and idioms obey very different structural constraints. According to Larson, there are no true ditransitive idioms, and so idioms are irrelevant for understanding the structure of ditransitives. As for collocations, Larson implies that they are subject to no structural constraints whatsoever, and so are also irrelevant to syntactic theory.

Claims like this ought to be very surprising, given that lexicographers and even many generative syntacticians have long recognized that there is no clear dividing line between non-literal idioms and literal collocations (e.g., [Everaert 2010 and references there). As I will discuss in more detail in section 2, expressions can be partly literal and partly non-literal. In addition, no other property correlates with literal versus non-literal interpretations: even purely literal expressions can be completely fixed and invariant, while expressions that are entirely non-literal can allow permutations and substitutions.

Nevertheless, in this paper I will attempt to investigate whether there is any empirical basis for the claim that non-literal idioms obey different structural constraints from literal collocations. That is, for the sake of argument I will adopt as a starting point what seems to be the view advocated by Larson (2017), according to which we should expect literal expressions and non-literal expressions to occur in very different syntactic
patterns. I will ask whether this is true, by dividing conventionalized expressions into literal and non-literal ones, and seeing what patterns they occur in. The empirical investigation reported here shows that they do not differ. In fact they are identical. This is exactly what we would expect, since there really is no clear dividing line between literal collocations and non-literal idioms.

Furthermore, having done this investigation and laid out what the patterns are that conventionalized expressions fall into, I will explore the question of whether there are any structural constraints on conventionalized expressions, and if there are, what those structural constraints are. It turns out that there are indeed constraints, but they are completely different from what has been proposed in the syntactic literature heretofore. I will propose a new way of looking at conventionalized expressions that explains and predicts the forms they can take. Unlike all previous analyses, I propose that the open slots are a crucial part of the conventionalized expression (e.g., the object of the preposition in light a fire under X, ‘stimulate X to action’). Once we make this move, then the task becomes to explain the limited distribution of the open slots. I propose that the patterns that are missing are ruled out by a constraint that bans open slots from having a hierarchically higher case competitor (in the sense of case competition models like that of [Marantz 1991]). This explains why possessors and objects of prepositions can freely be open slots (they do not have case competitors), while objects are much more constrained. A second object cannot be an open slot if the first object is fixed (*throw the wolves X), and an object cannot be an open slot if the subject is fixed (*the farmer bought X). There are some exceptions to the latter restriction, in English and other languages, and this way of formulating the constraint allows for them in certain specific conditions.

I begin in section 2 with a discussion of how we might attempt to distinguish idioms from collocations. This section also discusses whether they differ in needing to be underlying constituents. Section 3 presents an empirical examination of the syntactic patterns idioms and collocations each appear in. The detailed empirical study reported in this section shows that they are almost identical in the forms that they take. Section 4 argues that there must be syntactic constraints on conventionalized expressions by examining the patterns found with ditransitives in the empirical study. Sections 5–6 then explore what those structural constraints might be. Section 5 shows that previous proposals fall short, while section 6 proposes a novel account that makes the right distinctions. Section 7 concludes with some speculations about why open slots in conventionalized expressions would be constrained in precisely the way that they are.

2 Idioms Versus Collocations

In this section I first discuss criteria by which we might try to distinguish idioms from collocations, and then dismiss the idea that either of them needs to be an underlying constituent.

2.1 Distinguishing Criteria

Phrasal idioms are typically identified as fixed expressions consisting of multiple morphological (or prosodic) words that, just when they occur together, have a non-literal interpretation. For instance, kick the bucket has the non-literal meaning ‘die,’ and when used in this sense, does not involve any literal kicking or any actual buckets. In contrast, phrasal collocations like cost a fortune are interpreted literally (even if somewhat hyperbolically), and their component parts merely co-occur with a high degree of frequency and are recognized as conventionalized expressions by speakers of the language.

Note that the special meaning of idioms is sometimes referred to as a non-compositional interpretation, meaning that the phrase as a whole has an interpretation that does not follow from the composition of its parts. It is not clear that this is accurate, as many have noted. For instance, [Nunberg et al. 1994] point out that many idioms in fact do seem to be compositional, just not literal. For instance, spill the beans seems to involve composition of a non-literal meaning for spill, ‘divulge’ (which also occurs in other expressions),
and a non-literal meaning for *the beans* as ‘the secret’. Additionally, [Gazdar et al. 1985] and [Kratzer 1996] suggest ways of compositionally interpreting special meanings of words and phrases in the context of other words and phrases. This means that even the most semantically opaque expressions can be viewed as fully compositional (see also [Fellbaum 1993] [Everaert 2010] [Harley & Stone 2013] [Lichte & Kallmeyer 2016] [Bargmann & Sailer 2018] among many others). Since it is not clear at all that idioms are non-compositional, I will avoid use of this term completely.

How, then, can we distinguish between idioms and collocations? The only distinctions that seem to carry any water are literal versus non-literal meaning and substitutability. Take the phrasal idiom *kick the bucket* versus the collocation *cost a fortune*. The words in *kick the bucket* do not have their literal meaning, as noted above: when someone kicks the bucket (i.e., dies) there is no kicking motion and there is no actual bucket. In contrast, if something costs a fortune, it does literally cost an amount of money, and this amount is a large amount, like a fortune (hyperbolic but still literal). Additionally, the special meaning of *kick the bucket* does not survive when near synonyms are substituted for the parts: *kick the pail* and *boot the bucket* do not mean ‘die’. In contrast, one can also *charge a fortune* and something can *cost a bundle* or a *king’s ransom* (or any large amount).

However, even the criteria of literal meaning and substitutability fail. There are many expressions where some of the words are interpreted literally, but others are not. Consider the following three examples:

(1) a. miss the boat (‘miss an opportunity’)
   b. foot the bill (‘end up paying for something’)
   c. Y steal X’s thunder (‘use someone else’s ideas or words and take away their recognition’)

In *miss the boat*, the object is not a literal boat, but the verb has its literal meaning. Nevertheless, the meaning of missing an opportunity only arises with this particular combinations of words (Riehemann 2001; see also Nicolas 1995, Schenk 1995). One cannot, for example, say that someone *caught the boat* to mean that they seized an opportunity. In *foot the bill*, the verb is not literal and only seems to be used as a verb in this one phrase. But *the bill* is completely literal. Nevertheless one cannot *foot the charges* or *foot the check*. Similarly for *Y steal X’s thunder*, where the verb is used in its literal sense. One cannot instead say *Y rob X of X’s thunder* or *Y swipe X’s thunder*. Since these three expressions are partly literal and partly non-literal, the criterion of literal versus non-literal meaning cannot sort them into either the idiom or collocation categories.

The criterion of substitutability also fails, since it does not correlate with literal/non-literal meanings. It is possible to find some substitution with phrases that are not literal at all:

(2) a. grasp/clutch/seize at straws (‘desperately seize at even the most unlikely means of salvation’)
   b. give/get the boot/sack/cold shoulder/… (‘fire someone/ intentionally ignore someone’)
   c. beat/kick/knock the crap/shit/bejeezus/tar/(living) daylights out of X
   d. the roof/ceiling cave in on X (‘disaster strike X’)

Some non-literal expressions allow multiple choices of certain verbs within a certain semantic range, as in (2). Many expressions also permit both *give and get* (2b), as has been discussed heavily in the literature (e.g., Richards 2001, Harley 2002, Bruening 2010, Larson 2017). Others allow substitution of verbs and/or nouns, as in (2c–d). Conversely, many literal collocations do not admit substitution at all. For instance, it is very hard to see how *X blow X’s nose* and *X would do well to Y* could have other words substituted in them, even though they are completely literal (see also the discussion of expressions like *can-do* and *devil-may-care* in Bruening 2017).

1 Some idioms with open slots for possessors require that the possessor be identical to the subject, while others require disjointness. I indicate this throughout this paper with the variables X and Y.
The criteria of literal/non-literal interpretation and substitutability fail, then: there are expressions that are partly literal and partly non-literal, and the two criteria do not pick out the same expressions. I take this to show again that there really is no dividing line between non-literal idioms and literal collocations.

Nevertheless, to even ask whether there is a difference between two categories of expressions (which in the extremes might be called ‘idioms’ versus ‘collocations’) in the syntactic structures they can occur in, we need some way to sort expressions into two categories. For the purposes of the study I undertake here, I will take literal interpretation of all words in the expression as the dividing line. If I judge all words in the expression to be literal, I will classify it as a collocation. If any of the words are not interpreted literally, then I will classify it as an idiom. This makes expressions like miss the boat an idiom rather than a collocation, which is in keeping with treatment of it in the literature (e.g., Riehemann 2001). Even judging whether something is literal is tricky, however; many verbs and nouns have a large range of meanings, some of them more or less metaphorical. In such cases, I will classify the meaning in the expression as literal if it occurs outside of that expression as well. That is, I will classify an expression as an idiom if one or more words within it has a non-literal meaning that is limited to that particular expression. Otherwise, it is a collocation. To illustrate with some tricky examples, I classify come to grips with as a collocation because the verb is used within its normal semantic range, while grips has the meaning of understanding or acceptance outside of this expression as well, for instance in I can’t seem to get a grip on this concept. In contrast, I classify keep X posted as an idiom because posted does not have the meaning of being informed outside of this one expression. I will attempt to exemplify each syntactic pattern with the clearest extremes that I can find and with multiple examples, with the hope that there will be enough examples for people to agree on the sorting of at least some of them.

I should stress here that the point of my study is not to quibble about what counts as an idiom and what a collocation, but only to get some rough means of classifying them so that we can see what syntactic patterns occur. If we find significant non-overlap, then Larson’s contention will be vindicated: literal collocations and non-literal idioms do differ, and they will have to be carefully distinguished in future work. However, if we find overwhelming overlap in distribution (which is what we do find), then his contention will have been shown to be false, and there will be no reason to distinguish conventionalized expressions into two different categories.

Before getting to the empirical study, in the next subsection I will simultaneously illustrate the criterion of literal interpretation and address the one way that Larson (2017) claims with arguments that idioms and collocations differ.

2.2 Constituency

Larson (2017) claims that true idioms must be underlying phrasal constituents that exclude all non-idiomatic material, but this is not true of collocations. He does not say what structural constraints collocations are subject to, but implies that there are none.

Larson (2017) illustrates his claim that collocations do not have to be underlying constituents with the example of rancid butter. According to Larson, this is a collocation. However, it can be freely disrupted by various modifiers:

(3) rancid yellow creamery butter (Larson 2017: 402 (25))

This is a particularly poor example of a collocation, as few contemporary speakers of English would recognize rancid butter as a conventionalized expression (most, having refrigerators, have never experienced rancid butter). Nevertheless, Larson is correct in that true collocations like cost a fortune can be freely disrupted (cost a small fortune, cost a royal fortune).

2Note that this seems to be the definition of idiom in Weinreich (1969).
However, this was shown long ago to be possible with non-literal phrasal idioms, as well (e.g., Ernst 1981, Nunberg et al. 1994, Nicolas 1995, O’Grady 1998). They also admit adjectival modification:

(4) a. The Feminist Cat is out of the Marxist Bag (https://www.youtube.com/watch?v=374dTEMGii0; idiom: cat out of the bag, ‘facts are revealed’)
   b. First let me cast the usual pall on proceedings! (https://math.mit.edu/~rbm/papers/anulec/anulec.pdf; idiom: cast a pall on X, ‘give X an unhappy mood’)
   c. The Bulletin herewith announces its intention to turn over many new leaves. (Bulletin of the Garden Club of America; idiom: turn over a new leaf, ‘start to act in a better way’)

Note that these examples also serve to illustrate the criterion of literal interpretation: there is no literal cat or bag, no literal pall of either the cloth or smoke variety, and no literal leaf or turning over.

In addition, a very common idiom pattern is for the idiom to consist of a verb and the head of its object, with an open slot for the possessor of the object:

(5) (O’Grady 1998 (4d–e))
   a. Y get X’s goat (‘drive X to anger/annoyance’)
   b. Y fill X’s shoes (‘satisfactorily take over X’s duties’)

No theory of phrase structure that I am aware of posits a constituent that includes a verb and the head of its object, while excluding the possessor of the object. (Again note that there is no literal goat and no literal shoes in the situations where these idioms are used.)

Another common idiom pattern is one where the idiom consists of the verb, an object, and a preposition, but excludes the object of the preposition (O’Grady 1998):

(6) a. Y beat the bushes for X (‘search thoroughly for X’)
   b. Y cast a pall on X (‘give X an unhappy mood’)
   c. Y NEG hold a candle to X (‘not measure up to X’)
   d. Y light a fire under X (‘stimulate X to act’)

Again, no theory that I am aware of would posit a constituent that consists of a verb, an NP object, and preposition, but excludes the object of that preposition. (Regarding the criterion of literal interpretation: there are no literal bushes and no literal beating; no literal pall; no literal candle; and no literal fire.)

To begin the comparison with collocations, they occur in exactly these same patterns. They allow adjectival modification, as noted above (these words are all interpreted completely literally):

(7) a. declare war: Anonymous collective declares total war on Donald Trump.
   b. answer the door: Go answer the front door!

They also occur with an open slot for a possessor (again, these words are all completely literal, catch within its normal semantic range):

(8) a. X blow X’s nose
   b. Y break X’s heart
   c. X catch X’s breath

And they can consist of a verb, an object, and a preposition, excluding the object of that preposition:

(9) a. Y break the news to X
b. Y call/draw attention to X

c. Y stand trial for X

(Again, all of these words are literal, break, draw, and stand being used within their regular semantic ranges, as in break the story, draw one’s gaze, stand as an example.)

It is clear from the above that neither collocations nor idioms need to be underlying constituents. Not only that, so far collocations and idioms look identical. As the next section will show, they are identical in almost all respects.

3 Empirical Study of Collocations and Idioms

The empirical study of collocations and idioms undertaken here involved the use of the following print collections of collocations:


The introduction to the Oxford dictionary identifies 13 patterns of collocations. The introduction to the Kjellmer dictionary lists 19 patterns of collocations. The BBI dictionary lists fifteen major types, plus nineteen sub-types involving verbs. All of these are represented in the patterns identified below. Some of the patterns these works view as collocations I treat as simple selection, for instance verbs or adjectives selecting particular prepositions (e.g., depend on), or adjectives or other categories selecting infinitives (e.g., ready to) or that clauses. I do not include these here. I also exclude compounds and verb-particle combinations, and include only expressions that are clearly phrasal combinations of multiple words (I do include both compounds and verb-particle combinations if they are part of a larger phrase).

I also made use of two on-line lists of collocations (both accessed December 2017):


The point of this survey is to identify syntactic patterns. For each pattern I identify, I provide representative examples of each. The lists are not meant to be exhaustive, but I do try to give a sense for how large each set is. For each pattern that I identify, I also provide a list of phrasal idioms that occur in that pattern (or note that none occur in that pattern). Phrasal idioms were culled from various publications (especially [Nunberg et al. 1994], [O’Grady 1998] and [Bruening et al. 2018]), as well as the following idiom dictionaries:


[Larson 2017] note 15 also claims that book titles can be non-constituents, citing data from a 2008 Language Log post by Geoffrey Pullum [http://itre.cis.upenn.edu/myl/languagelog/archives/005333.html]. This is supposed to show that formulaic language more broadly is subject to no syntactic constraints whatsoever. However, the point of that Language Log post is how rare non-constituent titles are and how strained they often feel (many also appear to me to be elliptical). I conclude, with Pullum, that these titles show nothing about the syntactic constraints that conventionalized expressions are subject to.
I also used the following syntactically annotated corpus:


This dataset actually included far more literal collocations than idioms. Collocations and idioms were not distinguished in any of the above sources; as noted, lexicographers do not distinguish idioms from collocations (so, all of the above sources were sources for both).

I also added collocations and idioms that I noted in speech, radio, television, and print over the course of several months (so there may be expressions in the lists below that do not appear in any of the sources listed above).

3.1 Pattern 1: Modification

The first pattern I identify involves modification, broken down into various subtypes. Pattern 1a has an adjective modifying a noun:

(10) Collocation Pattern 1a: Modification Adjective-Noun
active ingredient, adverse effects, big deal, breaking news, bright idea, broken home, clean energy, comfort food, concerted effort, conventional wisdom, crushing defeat, cursory glance, direct quote, distinguishing feature, early days, early riser, easy money, empty promises, false impression, first impression, frayed nerves, free speech, front page, guilty party, happy ending, heavy drinker, heavy losses, heavy smoker, hidden extras, honest mistake, ill effects, informed consent, innocent victim/bystander, large number, little bit, long time, long way, mixed feelings, mysterious circumstances, nasty shock, natural causes, nervous wreck, next time, odd socks, old age, old friend, open wound, painful reminder, plain yogurt, poor health, popular belief, private life, public opinion, quick fix, real life, rhetorical question, ringing endorsement, rough draft, rough estimate, rough idea, round number, runny nose, safe distance, second thoughts, second opinion, sexually transmitted disease, short memory, social life, spare time, stiff competition, straight answer, sure sign, sure thing, tight grip, top priority, top speed, ulterior motive, ultimate goal, unconditional love, unfair advantage, unrequited love, upper echelons, upper limit, used car, vague idea, valid point/reason, valuable information, valuable lesson, vast majority, violent crime, vital organs, vocal critic, warm welcome, warmest regards, weak point, welcome change, wide range, wild animal, wrong number, wrong way, youthful enthusiasm

For comparison, there are also many idiomatic phrases of this sort:

(11) Idioms: Adjective-Noun
bad apple, big guns, bitter pill to swallow, busy (little) bee, cold feet, cold fish, cold shoulder, cold turkey, dead duck, eager beaver, good egg, hot button, hot potato, hot rod, hot water, little bird, the little woman, loose cannon, the old ball and chain, the last straw, X’s old man, pretty penny, red herring, shrinking violet, smooth sailing, straight arrow, thin ice, tough row to hoe, white elephant

Pattern 1b has an adverb modifying an adjective:

(12) Collocation Pattern 1b: Modification Adverb-Adjective
absolutely necessary, acutely aware, all alone, also known as X, best described as X, completely different, critically acclaimed, crystal clear, dead tired, deeply concerned, deeply held, diametrically
opposed, every single, far removed from X, fast asleep, generally accepted, heavily armed, heavily guarded, highly regarded, highly successful, ideally suited (to X), little known, long overdue, painfully shy, painfully slow, perfectly normal, perfectly safe, quite enough, quite good, quite right, quite sure, readily available, reasonably priced, sound asleep, still alive, strongly opposed, utterly ridiculous, well aware of X, well worth X, wide awake, wide open, widely used

I have only been able to find two phrasal idioms of this sort:

(13) *Idioms: Modification Adverb-Adjective*

tickled pink, writ large

An alternative analysis of these idioms is that they are adjectival passives formed from a verb plus secondary predicate, which would put them in pattern 2j below. As for the scarcity of these idioms, it should be noted that idioms are metaphors, and adjectives, being purely descriptive, do not typically take on metaphorical meanings themselves, but only in combination with nouns (as in pattern 1a).

One other thing to note before moving on is that the patterns can combine. Pattern 1a can combine with 1b to produce a collocation *adverb-adjective-noun*, as in *sexually transmitted disease*. I listed this in pattern 1a above; in what follows I will include such complex examples under just one or sometimes multiple headings, depending on how in need of exemplification a pattern is.

Collocation pattern 1c involves modification of a verb or verb phrase by an adverb:

(14) *Collocation Pattern 1c: Modification Adverb-Verb*

barely able to X, communicate effectively, do better, eat properly, fail miserably, figure prominently in X, get nowhere fast, get off lightly, go ahead, go smoothly, greatly appreciate, know best/better, NEG last long, laugh out loud, love dearly, need badly, never knew (that X), quite agree, recommend X highly, run amok, run rampant, strictly speaking, strongly support X, NEG take long, take X seriously, talk freely, travel light, try hard, turn around, vary widely, word hard, work well

I am only aware of two phrasal idioms with a simple adverb, those in (15a), but there are plenty with adverbial phrases (15b):

(15) *Idioms: Adverb-Verb*

a. Y only have eyes for X, (NEG) was born yesterday
b. be that as it may, close the stable door after the horse has bolted, X count X’s chickens before they hatch, X cross that bridge when X comes to it, Y nip X in the bud, rob Peter to pay Paul, strike while the iron is hot

Collocation pattern 1d involves modification of an adverb or a preposition by another adverb (I treat *just a minute* as an adverb, since it seems to be used as such as a temporal phrase):

(16) *Collocation Pattern 1d: Modification Adverb-Adverb or P*

hardly any, hardly ever, just a minute, nowhere near, only just, reasonably well, right away, right now, straight after, straight ahead, straight away, way ahead, well after/ahead/before/behind, yet again

I am not aware of any phrasal idioms of this type, but again adverbial phrases are unlikely to take on metaphorical uses by themselves.

Collocation pattern 1e has a noun plus PP. The noun may also be modified by an adjective:
(17) Collocation Pattern 1e: (Adjective)-Noun-PP
article of clothing, bit/piece of advice, blessing in disguise, call of duty, call of nature, cold light of day, enormous amount of X, great deal of X, honor among thieves, kick in the pants, lapse of/in judgment, momentary lapse of reason, pat on the back, (X’s) point of view, prisoner of war, public display of affection, random acts of violence/kindness, ray of light, recipe for disaster, room for improvement, sense of direction, sense of humor, speed of light, thing of beauty, vision of loveliness, waste of time

There are phrasal idioms of this type:

(18) Idioms: (Adjective)-Noun-PP
apple of X’s eye, big fish in a small pond, bolt from the blue, boots on the ground, butterflies in the stomach, clean bill of health, dog in the manger, elephant in the room, hair of the dog, land of nod, line in the sand, manna from heaven, pie in the sky, place at the table, school of hard knocks, shot in the arm, shot in the dark, skin in the game, slap on the wrist, snake in the grass, a snowball’s chance in hell, stick in the mud, tall glass of water, tit for tat, wolf in sheep’s clothing

Collocation pattern 1f has an adjective followed by a PP (or an NP, in the case of worth):

(19) Collocation Pattern 1f: Adjective-PP/NP
ahead of time, easy on the eyes, hard of hearing, missing in action, open to suggestions, worth a fortune, worth a try, worth X’s while

There are also at least a few idioms of this form:

(20) Idioms: Adjective-PP
blue around the gills, caught in the crossfire, hot under the collar, loaded for bear, long in the tooth, running on empty, running on fumes, X too big for X’s britches

Collocation pattern 1g involves comparison. These typically involve a comparative adjective plus a standard of comparison, but a few have verbs rather than adjectives:

(21) Collocation Pattern 1g: Comparison
as Adj as the next man, as best X can, as blind as a bat, better than a kick in the pants, colder than a witch’s teat, drink like a fish, dumb as a stump/rock, free as the wind/a bird, high as a kite, more dead than alive, old as the hills, pure as the driven snow, quick as a flash/wink, scarcer than hen’s teeth, sick as a dog, sleep like a baby/log, slippery as an eel, swear like a trooper, weak as a kitten, white as a sheet, work like a beaver

There are a few idioms of this type:

(22) Idioms: Comparison
X’s bark is worse than X’s bite, may as well be hung for a sheep as a lamb, thick as thieves

Comparison serves the same purpose as an idiom: analogizing one thing to something else. It is therefore not surprising that there are few idioms involving comparison (and one could argue about whether the collocations are truly literal; I take them to be in their lexical semantics, even if their application in the world is not).

Collocation pattern 1h has a possessor modifying a noun. Some have an open slot for the possessor, while in others the possessor is fixed (I thank a reviewer for some of these):

(23) Collocation Pattern 1h: Possessor N
a. X’s better nature, X’s finest hour, X’s good lady, X’s good deed for the day, X’s guardian angel, X’s knowledge and consent, X’s last gasp, X’s old man
b. anybody’s guess, beginner’s luck, child’s play, a fool’s errand, a fool’s paradise, God’s gift (to X), a hair’s breadth, (in/out of) harm’s way, nobody’s fool

There are also non-literal idioms of this form:

(24) **Idioms: Possessor N**

a. X’s better half, X’s birthday suit

b. Davy Jones’s locker, the emperor’s new clothes, the bee’s knees, the cat’s meow, the cat’s pajamas, a snowball’s chance in hell, the widow’s mite

Finally, collocation pattern 1i has a relative clause modifying a noun. Relative clauses can be finite or non-finite:

(25) **Collocation Pattern 1i: N Relative Clause**

the first thing that comes to X’s mind, for all X care/know, the last thing X want, the least X can do, the next thing X knew, X practice what X preaches, room to spare, something/nothing to lose, things that go bump in the night

There are non-literal idioms of this form, as well (including as objects of verbs, pattern 2a, next):

(26) **Idioms: N Relative Clause**

a bitter pill to swallow, a tough row to hoe, X bite the hand that feeds X, have bigger fish to fry, kill the goose that lays the golden eggs, the straw that broke the camel’s back

3.2 Pattern 2: Verbs and Arguments/Adjuncts

The second pattern I identify involves a verb and one or more of its arguments, and possibly adjuncts as well (typically PPs). Pattern 2a has a verb and its direct object:

(27) **Collocation Pattern 2a: Verb-Object**

abuse drugs, accept a challenge, accept defeat, answer the door, balance a budget, X be more trouble than X is worth, break a promise, break a record, break the law, catch fire, catch a bus, change the subject, commit a crime, cost a fortune, cover the costs (of X), cut costs, declare war, deliver a baby, do a good/bad job (of X), do the dishes, earn a living, eke out a living, enter a plea, face a challenge, face facts, find time, follow directions/instructions, get a joke, get the message, give up hope, have a baby, have a big heart, have a chat, have fun, have a good time, have an idea, have no idea (that X), have room, have sex, hold a referendum, hold an election, hold hands, join a club, keep a diary/journal, keep a promise, keep a secret, keep an appointment, keep score, keep the change, keep a low profile, kill time, lead the field, lead the way, leave a message, lose hope, make the bed, make a decision, make a difference, make a fuss, make a living, make a mess, make a mistake, make progress, make a scene, meet a need, meet the requirements, meet a target, miss an opportunity, obey an order, open an account, open fire, pass the time, pass a law, pass a test, pay a bill, pay attention, pull a muscle, put on weight, quit a job, raise money, reach a verdict, reach an agreement, return fire, run a business, satisfy a need, satisfy a requirement, save money/space/time, see reason, serve a purpose, set a date, set a record, set a table, solve a crime, solve a problem, take a call, take a risk, take a seat, take a step, take a test, take ages, take care, take drugs, take liberties, take notes, take notice, take place, take time, take turns, take up space, tell a lie/the truth, tell time, undergo surgery/treatment, waste an opportunity, win an award, yield results
It is well-known that this is also a very common pattern with idioms. Here are a few examples:

(28) **Idioms: Verb-Object**

back the wrong horse, bark up the wrong tree, beat/flog a dead horse, bite the big one, bite the bullet, bite the dust, X bite the hand that feeds X, blow/let off steam, break the ice, bring home the bacon, bury the hatchet, buy the farm, call off the dogs, chew the fat, close up shop, come a cropper, cry wolf, cut corners, dodge a bullet, draw a blank, drive a hard bargain, drop the ball, eat crow, eat humble pie, face the music, fly the coop, foot the bill, gild the lily, NEG give a crap/hoot, go the whole hog, haul ass, have a ball, have an itchy palm, NEG have a leg to stand on, hit home, hit the ceiling/roof, hit the sack/hay, hit the spot, jump the gun, jump ship, kick the bucket, look the other way, lose steam, make tracks, mean business, miss the boat, pack a punch, pass muster, pay the piper, put out fires, rock the boat, ruffle feathers, saw logs, scrape the bottom of the barrel, screw the pooch, shake a leg, shoot the breeze/the bull, sing a different tune, smell a rat, speak volumes, spill the beans, stack the deck, talk turkey, take a powder, throw in the towel, toe the line, turn over a new leaf, turn tail, turn tricks

One thing to note is that the object is almost always inanimate, with both collocations and idioms. The only animates that occur are *baby* and animal names. This is yet another way in which idioms and collocations behave alike. On this tendency with idioms, see [Nunberg et al. (1994)](http://example.com), [Harley & Stone (2013)](http://example.com) and section 4 below.

Collocation pattern 2b was illustrated above (29). It has a verb plus an object, but with an open slot for a possessor of the object:

(29) **Collocation Pattern 2b: Verb-X’s Object**

X bare X’s teeth, X blow X’s nose, Y break X’s heart, X catch X’s breath, Y catch X’s eye, Y catch X’s attention, Y come to X’s rescue, X cross X’s mind, X do X’s best, X do X’s duty, X get X’s kicks from Y, X give (X’s) consent, Y hurt X’s feelings, Y kick up X’s heels, Y lose X’s place, X lose X’s temper, X lose X’s way, X make up X’s mind, X pay X’s respects (to), X pick X’s nose, X put up X’s hand, X raise X’s voice, Y save X’s life, Y save X’s strength, Y take X’s advice, Y take X’s temperature, Y take X’s place, X take X’s (sweet) time, X thank X’s lucky stars, Y upset X’s stomach, X wait X’s turn, X watch X’s weight

Phrasal idioms of this type were also illustrated above (30). Here is a longer list of such idioms:

(30) **Idioms: Verb-X’s Object**

Y bend X’s ear, X bet X’s bottom dollar, X blow X’s stack, Y breathe down X’s neck, X burn X’s bridges, Y clean X’s clock, Y clip X’s wings, Y cook X’s goose, Y cramp X’s style, Y fill X’s shoes, Y get X’s goat, X gird X’s loins, X hold X’s tongue, X let X’s hair down, X lose X’s cool, X lose X’s head, X keep X’s shirt on, Y pull X’s leg, X put X’s foot down, X spill X’s guts, Y steal X’s thunder, X tighten X’s belt

Collocation pattern 2c was also illustrated above (31). It involves a verb, a direct object, and a preposition, with an open slot for the object of the preposition:

(31) **Collocation Pattern 2c: Verb Object P X**

bear/hold a grudge against X, bear a resemblance to X, break the news to X, call/draw attention to X, cast doubt on X, catch a glimpse of X, catch sight of X, come to terms with X, do damage to X, draw the line at X, express interest in X, gain/have access to X, get a bang out of X, get a kick out of X, get rid of X, give birth to X, give chase to X, give consideration to X, give credence to X, give form to X, give priority to X, give some thought to X, give way to X, have a word with X, have
an effect on X, impose conditions/restrictions on X, join forces with X, keep an eye on X, lay the blame on X, lay the groundwork for X, look forward to X, lose control of X, lose faith in X, make a note of X, make arrangements for X, make friends with X, make short work out of X, open fire on X, part company with X, pay attention to X, pay homage to X, pay respect(s) to X, pay tribute to X, place an order for X, pose a risk to X, pose a threat to X, put pressure on X, put a stop to X, raise doubts about X, raise questions about X, regain control of X, run the risk of X, say goodbye/hello to X, say sorry to X, set fire to X, Y set Y’s sights on X, stand trial for X, strike a balance between X, take a break from X, take a look at X, take action against X, take an interest in X, take care of X, take charge of X, take issue with X, take part in X, take pride in X, take the liberty of X-ing, tell the difference between X, wage war against X, wave goodbye to X

Idioms of this pattern were also illustrated above (6). A longer list of such idioms appears below:

(32) **Idioms: Verb-Object P X**

- beat the bushes for X
- carry a torch for X
- cast a pall on X
- do a number on X
- get the hang of X
- get a kick out of X
- get a load of X
- give it to X
- give rein to X
- give rise to X
- give vent to X
- give voice to X
- hang X’s hat on Y
- have a bone to pick with X
- have an axe to grind with X
- hold a candle to X
- keep a lid on X
- keep tabs on X
- light a fire under X
- lock horns with X
- make head or tail of X
- pay court to X
- pay lip service to X
- pour cold water on X
- put a stop to X
- raise doubts about X
- raise questions about X
- regain control of X
- say goodbye/hello to X
- say sorry to X
- set fire to X
- strike a balance between X
- take a break from X
- take a look at X
- take action against X
- take an interest in X
- take care of X
- take charge of X
- take issue with X
- take pride in X
- think the world of X
- throw the book at X
- turn a blind eye to X

Collocation pattern 2d involves ditransitives. Most of them are listed in my sources as the double object variant. These include a verb and second object, but with an open slot for the first object. I list them in two groups below, depending on whether or not they alternate, in my judgment. Those that alternate alternate with a prepositional variant, for instance, *do a favor for X* or *give advice to X*. (Individual speakers may vary in which they allow to alternate.)

(33) **Collocation Pattern 2d: Verb X NP, Non-Alternating**

- call X names
- deny X access to
- do X a disservice
- do X good/harm
- give X permission
- give X a bad name
- give X a break
- give X a call
- give X a chance
- give X a go
- give X a lift
- give X a minute
- give X a round of applause
- Y give X Y’s word
- give X the benefit of the doubt
- give X the impression that Y
- save X the trouble
- tell X the time

(34) **Collocation Pattern 2d: Verb X NP, Alternating**

- do X a favor
- give X advice
- give X credit
- give X meaning
- give X some guidance
- give X some thought
- give X short shrift
- lend/loan X money
- make X an offer
- offer X a job
- offer X an explanation
- offer X (Y’s) condolences
- offer X Y’s services
- owe X an apology
- owe X an explanation
- pay X a compliment
- pay X a visit
- save X a seat
- teach X a lesson
- tell X a thing or two

Idioms also occur in this pattern. Larson (2017) denies that these exist, but he only discusses those in (35) (I would not classify all of these as idioms, in particular *give X a kick* and *show X the ropes*). There are numerous others, some of which are listed in (36):

(35) **Idioms: V X NP**

- give X the boot
- give X the creeps
- give X flak
- give X a kick
- Y give X a piece of Y’s mind
- promise X the moon
- read X the riot act
- show X the door
- show X the ropes

(36) **Idioms: V X NP**

- Y bet X Y’s bottom dollar, bet X dollars to donuts, cut X some slack, cry X a river, drop X a line, give X a free hand, give X a hand (‘applaud’ or ‘help’), give X pause
- Y give X the rough edge of
Y’s tongue, give X a turn, give X a whirl, give X no quarter, give X the once-over, give X the slip, give X a wide berth, give X what-for, give X some skin, lend X color, Y lend X Y’s ear, lend X a hand, NEG pay X any mind, throw X a bone, throw X a curve

I have not indicated which of these idioms alternate and which do not (in my judgment); see [Bruening (2010)] and section 6.5 below.

Collocation pattern 2e consists of a small number of double object collocations where both NPs are part of the expression:

(37) *Collocation Pattern 2e: V NP NP*
    give it a rest, give it your best shot, give it the old college try, don’t give me that (Imperative)

There are four idioms of this form listed in my sources, as well (the fourth forms a relative clause on the second object, see pattern 1h above):

(38) *Idioms: V NP NP*
    give the devil his due, X give X’s self airs, give a dog a bad name and hang him, that’s the stuff to give the troops

The dividing line between literal collocations and non-literal idioms is not so clear here; some of the expressions I listed as collocations could just as easily be considered non-literal idioms. I will return to conventionalized expressions involving ditransitives in sections 4 and 6.5.

Collocation pattern 2f has a verb, a direct object, and a PP, all fixed (but possibly with an open possessor):

(39) *Collocation Pattern 2f: V NP PP*
    Y bring a tear to X’s eye, X get a lump in X’s throat, X give a good account of X’s self, take the rough with the smooth

There are many more idioms of this sort:

(40) *Idioms: V NP PP*
    add fuel to the fire, build castles in the air, burn the candle at both ends, cast pearls before swine, change horses in midstream, fight fire with fire, get a foot in the door, get the show on the road, give hostages to fortune, go to hell in a handbasket, have a bun in the oven, X have X’s ear(s) to the ground, hit the nail on the head, Y hold X’s feet to the fire, X keep X’s eye on the prize, keep the wolf from the door, kill two birds with one stone, X lay X’s cards on the table, X paint X’s self into a corner, X place X’s head in the lion’s mouth, Y pull the wool over X’s eyes, X put all X’s ducks in a row, put the cart before the horse, take/seize/grab the bull by the horns, Y take the wind out of X’s sails, Y take the word’s out of X’s mouth, throw out the baby with the bath water

Collocation pattern 2g has just a verb and a PP:

(41) *Collocation Pattern 2g: Verb PP*
    come into view, come to a stop, come to a standstill, come to an agreement, come to an end, fall in love, fall into the category of X, get into trouble, go out of fashion, fly by night, jump at shadows, jump to conclusions, look for trouble, meet with approval, resort to violence

There are also many idioms of this form:

(42) *Idioms: Verb PP*
    bark up the wrong tree, beat around the bush, brush up on X, cut to the chase, dance with the devil,
fall in(to) line, fly off the handle, fly under the radar, get in on the ground floor, get into the swing of things, get to first base, Y get under X’s skin, go overboard, fall in line, is in the wind, jump on the bandwagon, knock on wood, pay through the nose, play with the big boys, Y rain on X’s parade, read between the lines, X sit on X’s hands, Y slip through X’s fingers, start from scratch, X stew in X’s own juices, X stick to X’s guns, swim against the current, swim with the sharks, X take to X’s heels

Collocation pattern 2h has a verb and a PP, but an open slot for an object in between (in all of these collocations, the PP appears to be an argument of the verb):

(43) Collocation Pattern 2h: Verb X PP
bring X to an end, bring X to justice, drive X to distraction, keep X in mind, keep X under wraps, put X out of X’s misery, take X to court, Y tell X to X’s face

There are also many idioms of this form (with both argument and adjunct PPs):

(44) Idioms: Verb X PP
carry X to extremes, can cut X with a knife, drag X over the coals, feed X to the lions, hoist X with X’s own petard, Y keep X under Y’s hat, kick X to the curb, lay X on the line, Z lay X at Y’s door, leave X in the lurch, push X to the edge, put X out to pasture, rake X over the coals, sell X down the river, send X to the devil, send X to the showers, stab X in the back, sweep X under the rug, take X in stride, take X to the cleaners, take X to task, Y take X under Y’s wing, take X with a grain of salt, throw X to the wolves, throw X under the bus

The subject is always disjoint from the object, in both collocations and idioms.

Collocation pattern 2i has two PPs, with an open slot for the object of the second:

(45) Collocation Pattern 2i: Verb PP P X
come to grips with X, go on a date with X, keep in touch with X

I have not found any idioms of this form.

The next two patterns involve verbs and adjectives. Collocation Pattern 2j has a verb plus adjective, where the adjective is either selected by the verb or is a secondary predicate modifying the subject of the verb:

(46) Collocation Pattern 2j: Verb Adjective/Secondary Predicate
a. be dragged kicking and screaming, become apparent (that X), call in sick, come alive/ to life, come close, come in first/last, come true, drop dead, fall asleep, get ready, get upset, get worried, go bald, go bankrupt, go crazy, go unchallenged, go unnoticed, go wrong, keep busy, keep/stay calm, keep fit, keep mum, keep quiet, keep still, keep warm, look nice, make sense, make sure, run late, run low, sit crosslegged, stay awake, stay put, stay tuned
b. get something straight, get/set the record straight, X keep X’s ears/eyes open, X keep X’s mouth shut, make a long story short, take it easy

I also include here as a sub-case some collocations that also include a direct object (46b). There are some idioms consisting of a verb and adjective or secondary predicate:
Some also include an object (some with an open slot for a possessor). I include *X keep X’s shirt on* here although it has a particle rather than an adjective, because the order particle-NP is not possible.

Collocation pattern 2k has a verb and adjective or secondary predicate, but also has an open slot for a direct object (the subject of the secondary predicate, necessarily disjoint from the subject of the verb):

(48) **Collocation Pattern 2k: Verb X Adjective/Secondary Predicate**

beat X to death, bring X (a)round, drive X crazy, get X ready, have X removed, hold X hostage, hold X prisoner, keep X waiting, keep X quiet, leave X alone, make X better, make X easy (for), make X public, put X together, read X aloud/out loud, serve X right, set X on fire, set X free, take X apart, take X hostage/prisoner, tell X apart

There are few idioms of this form; I have found the following:

(49) **Idioms: Verb X Adjective/Secondary Predicate**

catch X napping, give it to X straight, keep X posted, knock ’em/X dead, pull it/X together

The second one has a fixed object but open object of P.

Collocation pattern 2l has a verb plus an infinitive, sometimes with an open slot for an object or with a fixed object:

(50) **Collocation Pattern 2l: Verb (NP/X) Infinitive**

give X to understand Y, X have it in X to Y, have the nerve to X, lead X to believe Y, let (X) go, let X know (that Y), quit drinking/smoking, talk X into/out of Y

There are a few idioms of this form that I have found:

(51) **Idioms: Verb (NP/X) Infinitive**

let the dog see the rabbit, let sleeping dogs lie, let them/X eat cake, make X’s blood boil, make ends meet, set the ball rolling

Finally, there is a small number of collocations that include a verb and a CP argument. One of these has an NP argument in addition to a CP argument:

(52) **Collocation Pattern 2m: V (NP) CP**

X is as X does, X looks like X has seen a ghost, NEG know what to do with X, say/tell me it isn’t/ain’t so, tell/know which is which/what is what

In *X is as X does*, the clause with *as* seems to be the complement of the verb *be*. It is a full finite clause. The subject of the lower clause is obligatorily covalued with the matrix subject (and typically repeats it, e.g., *famous is as famous does*). There are a few idioms that have CP arguments, possibly in addition to an NP argument:

(53) **Idioms: V (NP) CP**

know what the score is (or know the score), NEG X know where X’s head is (at), X know which side X’s bread is buttered on, know which way the wind blows, look what the cat dragged in, X teach X’s grandmother to suck eggs, tell X where to get off, wait for the other shoe to drop
Know what the score is also occurs as know the score, with just an NP object (but that NP object is semantically a concealed question).

Finally, collocation pattern 2n has a verb, an object, and an open clausal constituent. In (54a), this is probably a complement of the verb, while in (54b), it is probably the complement of the noun that is the object:

(54) Collocation Pattern 2n: V (X’s) NP Y
    a. X spend X’s life Y-ing, have a hard time X-ing
    b. get/give the impression that X

I have found one idiom of this form:

(55) Idioms: V (X’s) NP Y
    X cut X’s teeth Y-ing

This concludes the patterns of collocations and idioms that I have found involving a verb and its arguments and adjuncts.

3.3 Patterns 3–9

The rest of the patterns I have identified are of various different forms. Collocation pattern 3 involves coordinations (subdivided by syntactic category):

(56) Collocation Pattern 3: Coordinations
    a. aches and pains, an X or two, bits and pieces, (this) day and age, everybody and his uncle, for all intents and purposes, fun and games, health and wellbeing, law and order, nooks and crannies, nothing but net, peace and quiet, time and time again, this and that, thunder and lightning, wear and tear
    b. bait and switch, cut and paste, give and take, grin and bear it, kiss and tell, stop and go (traffic), suck it up and deal, take it or leave it (Imperative), tax and spend, toss and turn
    c. alive and kicking, black and blue, black and white, bound hand and foot, hot and bothered, more or less, neat and tidy, null and void, safe and sound
    d. above and beyond, down and out, out and about, over and above, through and through
    e. and so forth, and yet, but again, or else, or so it seems

There are also idioms that are coordinations:

(57) Idioms: Coordinations
    a. the birds and the bees, fire and brimstone, kit and caboodle, night and day, the old ball and chain, pins and needles, (on) the up and up
    b. bow and scrape, cross my heart and hope to die, dot the I’s and cross the T’s, (NEG) make head or tail of X, shit or get off the pot, sink or swim
    c. cut and dried, one foot in and one foot out, spic and span
    d. betwixt and between, by and large, by hook or by crook, through thick and thin

Collocation pattern 4 has a preposition followed by its NP object:
Collocation Pattern 4: Preposition NP

above all, above/under ground, against all odds, against the law, after all, along those/the same lines, at first sight, at some point, at X’s own expense, beside the point, between jobs, beyond recognition, beyond reproach, by accident, for all practical purposes, for the fun of it, for argument’s sake, from now on, in addition to X, in advance, in agony, in agreement with X, in any case, in black and white, in short, NEG in the least, of a personal nature, on (the) alert, on loan (from X), out of order (‘not functioning’ or ‘out of sequence’), over the edge, under glass, under the microscope, under the circumstances

There are idioms of this form, as well:

Idioms: Preposition NP

after a fashion, at sixes and sevens, behind the eight ball, below the belt, beside X’s self, between the devil and the deep blue sea, between a rock and a hard place, beyond the pale, down in the dumps, down that road, for crying out loud, from pillar to post, in the black/red, in the can, in the doghouse, in the same boat, off the grid, on the hop, on the dole, on the same page, out of hand, out of line, out on a limb, over a barrel, over X’s head, over the hill, over the hump, to a T, under the weather, under wraps, with flying colors

Collocation pattern 5 has an adjective plus infinitive:

Collocation Pattern 5: Adjective Infinitive

enough to go on, hard to believe, hard to say, hard to take, hard to tell, needless to say, too good to be true, too hot to handle

There is at least one idiom of this form:

Idioms: Adjective Infinitive

fit to be tied

Collocation pattern 6 is predicative. It has an NP that is the subject of a small clause:

Collocation Pattern 6: Predicative (NP + Small Clause)

back to front, hand in hand, hands off, head over heels, no money down, no strings attached, time off, upside down

There are some idioms of this form as well:

Idioms: Predicative (NP + Small Clause)

back (is) against the wall, blind leading the blind, cat (is) out of the bag, fat (is) in the fire, noses (are) to the grindstone, tail (is) wagging the dog, writing (is) on the wall

Collocation pattern 7 involves one or more functional elements plus an open slot, possibly with other specified elements as well (like a main verb or main verb plus object):

Collocation Pattern 7: F V (NP P) X

a. NEG bet on X, has had enough of X, had better X, might want to X, would do well to X, would rather X
b. can’t afford X, can’t do any harm, can’t help X-ing, can’t stand X, couldn’t/don’t give X (where X is a minimizer)
c. might as well X_{VP}
I list as a separate subtype collocations that have specifically negation, a modal, and a verb (64b), as there are several of these. There is also at least one case with no main verb, but a modal and adverb with an open slot for a VP (64c). There are some idioms that include functional elements, some have open slots while others do not:

(65) **Idioms: F V (NP P X)**
   a. be cruisin’ for a bruisin’ (progressive), break a leg! (imperative), can cut X with a knife, has been around the block a few times, has had it, has gone to the dogs, no can do
   b. NEG breathe a word about X, NEG have a leg to stand on, NEG have a pot to piss in, (can/could) NEG help it/Xself (‘unable to stop oneself’), NEG see the forest for the trees, NEG stop at anything

I list examples with negation specifically as a separate subcategory; see section 6.9.

Collocation pattern 8 involves a full clause, including a subject:

(66) **Collocation Pattern 8: Full Clause**
   a. duty calls, easy does it, nature calls, the prodigal son returns, X’s heart is in the right place, there’s more to X than meets the eye, what you see is what you get, you know
   b. X’s gorge rises, X’s heart sinks, X’s jaw drops, X’s mouth waters, X’s spirits rise/sink/soar, X’s voice breaks
   c. Does a bear shit in the woods?, Is the pope catholic?, What/something is up, What can you do?, What gives?, What makes X tick?, What’s up with X?, Whatever happened to X?, How are you?, Where would we be without X?, Who does X think X is?, Who’s there?, Who would have thought that X?
   d. Don’t give me that, Don’t make me laugh, Don’t mention it
   e. as luck would have it, as things stand, when all is said and done, the fact remains that X, given time (adjunct clause with null controlled subject), if X knows what’s good for X, if you ask me, to be honest, to begin with, to say nothing of X

Some of these are main clauses: declaratives (66a–b), questions (66c), imperatives (66d). Others are adjunct clauses that attach to open main clauses (66e). A significant subset of the declaratives have a subject and verb, where the subject has an open possessor (66b). (There are far more examples of full clauses than I have represented here, as this category includes all kinds of sayings and proverbs, and also very common expressions like *How should I know?* and *as far as I know*; SAID includes long lists of these.) There are also idioms that are full clauses, on the same patterns:

(67) **Idioms: Full Clauses**
   a. all hell break loose, Bob’s your uncle, the bottom fell out of X, the die is cast, the fur fly, the jig is up, loose lips sink ships, the roof caved in on X, that ship has sailed, the shit hit the fan, something is on the wind, too many cooks spoil the broth
   b. X’s blood boils, X’s heart bleeds (for Y)
   c. What/something is eating X, What’s cooking?, What/something crawled up X’s ass and died
   d. put a sock in it, X put that in X’s pipe and smoke it
   e. come hell or high water, until the cows come home, when hell freezes over, when X’s ship comes in

Finally, collocation pattern 9 consists of two clauses in juxtaposition, with a conditional-like meaning. I will call these **covert conditionals**:  

18
Collocation Pattern 9: Covert Conditionals

easy come, easy go; no pain, no gain

There is at least one idiom of this type:

Idioms: Covert Conditionals
monkey see, monkey do

3.4 Summary of Empirical Study

I summarize the patterns identified here below, with an example of a collocation of that form and a matching idiom if there is one. The sub-types of Pattern 1 involve modification:

Pattern 1: Modification

a. Adj N: honest mistake (idiom: cold feet)
b. Adv Adj: acutely aware (idiom: writ large)
c. Adv V: go smoothly (idiom: only have eyes for X)
d. Adv Adv/P: hardly ever, way ahead (no idioms)
e. N PP: call of duty (idiom: elephant in the room)
f. Adv PP/NP: missing in action, worth a fortune (idiom: running on fumes)
g. Comparison: better than a kick in the pants (idiom: X’s bark is worse than X’s bite)
h. Possessor N: X’s finest hour, beginner’s luck (idiom: X’s birthday suit, Davy Jones’s locker)
i. N Relative Clause: the least X can do (idiom: a bitter pill to swallow)

The sub-types of Pattern 2 involve a verb and one or more of its arguments, or adjuncts to the VP:

Pattern 2: Verbs and Arguments/Adjuncts

a. V NP: answer the door (idiom: back the wrong horse)
b. V X’s NP: Y break X’s heart (idiom: Y clip X’s wings)
c. V NP P X: break the news to X (idiom: beat the bushes for X)
d. V X NP: give X permission, do X a favor (idiom: cut X some slack)
e. V NP NP: give it a rest (idiom: give the devil his due)
f. V NP PP: Y bring a tear to X’s eye (idiom: build castles in the air)
g. V PP: come into view (idiom: beat around the bush)
h. V X PP: bring X to justice (idiom: feed X to the lions)
i. V PP P X: keep in touch with X (no idioms)
j. V (NP) Adj: call in sick, make a long story short (idiom: getting busy, X get X’s feet wet)
k. V X Adj: drive X crazy (idiom: pull it/X together)
l. V (NP/X) Infinitive: lead X to believe (idiom: let sleeping dogs lie)
m. V (NP) CP: tell me it isn’t/ain’t so (idiom: X teach X’s grandmother to suck eggs)
n. V (X’s) NP Y: X spend X’s life Y-ing (idiom: X cut X’s teeth Y-ing)

Patterns 3–9 involve other structures:

(P3) Coordinations: law and order (idiom: the old ball and chain)

(P4) P NP: against the law (idiom: behind the eight ball)
(74) (P5) Adj Infinitive: hard to say (idiom: fit to be tied)
(75) (P6) Predicative: no money down (idiom: back against the wall)
(76) (P7) F V (NP P) X: can’t stand X (idiom: has had it)
(77) (P8) Full Clause: easy does it (idiom: the shit hit the fan)
(78) (P9) Covert Conditional: easy come, easy go (idiom: monkey see, monkey do)

As can be seen, collocations and idioms overlap significantly. The only collocation patterns where there
does not appear to be a corresponding idiom of that form are Patterns 1d and 2i. One of these patterns has
very few examples of collocations (2i), so it may not be surprising that there are no idioms of that form. My
guess is that there are examples of both that I have simply not found yet. One P NP idiom, out on a limb,
frequently occurs as go out on a limb for X, which fits this pattern, but it also occurs without the verb and
without the for PP. As for pattern 1d, it is headed by an adverb, and it is unlikely that adverbs would take on
metaphorical meanings on their own and so become idioms.

The overall correspondence between collocations and idioms is striking. There are few cases of non-
overlap, and those either have a good explanation (adverbs not being used metaphorically on their own) or
are probably not significant (low numbers for the one that does occur). Patterns with many collocations
typically also have many idioms, and vice versa, and the same holds for patterns with few collocations.
This indicates that any constraints on the structure of idioms are shared by collocations, and vice versa.
Importantly, there is no support for Larson’s [2017] contention that collocations and idioms are subject to
very different structural constraints. If Larson were correct, we would expect idioms to occur in a small
subset of the patterns that collocations occur in, but this is not what we find. It appears instead that idioms
and collocations are not distinct at all. That is to say, they belong to the same category, and do not need to be
distinguished. I will call this larger category conventionalized expressions. Note that this finding confirms
the remarks earlier, which showed that there really are no criteria that distinguish idioms from collocations.
As we saw, the criteria of literal/non-literal meaning and substitutability fail.5

In addition to this overall conclusion, it is also possible to make a few more observations concerning
the data presented in this section. First, thematic roles of all kinds occur in both collocations and idioms.
Consider PPs. There are PPs that encode goals (jump on the bandwagon), locations (build castles in the
air), recipients (feed X to the lions), sources (start from scratch), paths (pay through the nose), comitatives
(play with the big boys), instrumentals (can cut X with a knife), and temporal phrases (fly by night). Agent
thematic roles tend to be open slots in Pattern 2, but they occur with some frequency in full clauses, Pattern
8, and in multi-clause examples of Patterns 1 and 2, for instance Patterns 1h and 2m. For example, the
embedded null subject of have bigger fish [PRO to fry] is clearly agentive, as is the wh-phrase subject of
Who asked X?. I address agents in much more detail in section 5.1. Experiencer and benefactive thematic
roles are rarer, the only one I have found is an open slot in do a favor for X. Many of the expressions listed
here describe experiences, but they do so in terms of concrete metaphors (e.g., What’s eating X?, which asks
what’s bothering X but uses the metaphor of ingestion).

Similarly, it has often been noted that there is a tendency in idioms for fixed NPs to be inanimate, or
rather, non-human, since animal names feature prominently (e.g., Nunberg et al. 1994). We saw above that
this was true for collocations and idioms of the VP type, with an open subject and fixed internal arguments.

5Just as there is no legitimate distinction between idioms and collocations, distinctions that people have made within the class
of idioms are also spurious and of no grammatical relevance. Contra Nunberg et al. (1994) and numerous publications that follow
them, there is no distinction between “idiomatically combining expressions” and “idiomatic phrases”; see, for instance, Abeillé
see Felbaum (1993) and Bruening et al. (2018). Contra Stone (2016), there is no distinction in how much VP structure idioms have;
syntactic flexibility is larger a matter of pragmatics, see again Bargmann & Sailer (2018) (and note that Stone’s syntactic hierarchy
is paradoxical, since she has gerunds lower than passives but gerunds can be formed from passives). We only need to recognize
a single category of conventionalized expression, with no sub-categories.
However, NPs denoting humans do occur with some frequency in full-clause and multi-clause expressions, for instance X teach X’s grandmother to suck eggs, Who asked X?, and too many cooks spoil the broth. Any tendency toward non-human NPs therefore seems to be limited to internal arguments (objects of verbs and prepositions). It should be noted, however, that non-humans also do occur as external arguments (the shit hit the fan, that ship has sailed).

Second, we can also see that collocations and idioms occur in all kinds of syntactic patterns. From the above list of patterns, one might conclude that conventionalized expressions can take the form of any legitimate syntactic structure. This is almost true. Two particular patterns have been claimed to be missing. Marantz (1984) claimed that idioms with a fixed subject and an open object do not occur. This pattern is indeed rare in the data listed above, but there are a few examples of it, like What’s eating X? and Who asked X?. I will return to this point in section 6.7 where I propose an analysis that accounts both for the scarcity of this pattern and the few examples of it that do exist (which are systematic).

The other pattern that has been claimed to be missing is in fact categorically absent from the data above. In ditransitives, V X NP is well-attested (give X a hand), but V NP X is systematically missing, from both idioms and collocations (*give the little woman X; see, e.g., Rappaport Hovav & Levin 2008, Bruening 2010). There is not a single example of this pattern. This absence is especially striking, because the other sub-patterns of Pattern 2 indicate that basically every other combination of a verb with its arguments and adjuncts is possible. Verbs can combine with NPs, with PPs (both arguments and adjuncts), with CPs, with adverbial phrases, and in any combination of those. This missing ditransitive pattern is the only one that is absent. It is missing from both collocations and idioms, again indicating that they are the same and should be treated identically. In the next section, I examine the lack of this pattern in more detail, and argue that we need an analysis in which there are syntactic constraints on conventionalized expressions. In sections 5 and 6 I examine what those constraints might be and make a new proposal.

To sum up the results of the empirical study, literal collocations and non-literal idioms do not differ at all. There is no support for distinguishing them. They occur in exactly the same syntactic patterns and in roughly equal numbers, and the same patterns that are missing from one are also missing from the other (*V NP X). Any account of the syntactic form these expressions can take must treat them as a single class.

4 The Case for Syntactic Constraints: The Missing Ditransitive Pattern

As noted, conventionalized expressions seem to occur in almost every legitimate syntactic structure that there is. If they did occur in every possible structure, there would be very little for syntacticians to explain: we would just note that conventionalized expressions take the form that the syntax independently allows, and nothing more would need to be said. However, if there are syntactic patterns that are systematically missing, then this demands an account. I show here that the fact that one pattern of ditransitives is missing demands a structural account, and cannot be accounted for in terms of animacy or thematic roles.

Let me begin with Larson 2017 (some of the content of which also appeared in Larson 2014). According to Larson (2017), idioms tell us nothing about ditransitives, because there are no ditransitive idioms. Some phrases that have been treated as idioms, like give X the creeps, are actually collocations. According to Larson, the noun creeps has a meaning (‘a feeling of revulsion or fear’), which combines compositionally with the literal meaning of give. The reason that creeps does not occur with very many other verbs besides give and get, according to Larson, is that give X the creeps and X get the creeps are collocations. There is nothing more to explain, since collocations are not subject to any structural constraints, in Larson’s view.

Expressions like too many cooks spoil the broth are sometimes dismissed as proverbs rather than idioms. It is unclear why they should be, when the shit hit the fan is always discussed in the syntactic literature as a syntactically relevant idiom. Too many cooks spoil the broth participates in all the syntactic alternations that the shit hit the fan does, for instance raising: Too many cooks always seem to spoil the broth.
The study reported here shows that such a view cannot be sustained. There are numerous collocations of the form \( V \times NP \), with the verb and second object part of the collocation but the first object open (\( do \times good/harm, give \times permission, give \times a\ break, \ldots \)). If Larson were correct and nouns with particular (sometimes idiomatic) meanings can just combine with verbs as collocations, with no structural constraints, we should also expect there to be collocations of the form \( V \times NP \), with the first NP fixed and the second one open. There are plenty of NPs with special meanings analogous to the creeps but which are animate, and should therefore be able to combine with give or another verb as its first object. Here is a short list of some animate NPs that would fit the bill:

(79) a clockwatcher, a clothes horse, a Dutch uncle, everybody and his uncle, jailbait, the little woman, a living soul (NPI), a love child, a mother hen, the old ball and chain, X’s old man, poster child, sharkbait, shrinking violet, smartypants, a snake in the grass, snowbunnies, star-crossed lovers, a wallflower, a wannabe

On Larson’s reasoning, we should expect some of these to form collocations with verbs, perhaps like the following:

(80) a. * give everybody and his uncle X
    b. * give the smartypants X
    c. * give X’s old man Y
    d. * deny the wannabe X
    e. * give the little woman X
    f. * ask the/a shrinking violet X
    g. * ask the/a wallflower X

However, no collocations of this form exist, and it would be completely unnatural to conventionalize one of these.

Additionally, the first object of a double object construction can be inanimate in light verb constructions like the following:

(81) give the table a scrubbing, give the house a new coat of paint, give the car a tuneup, give it a kick

There then ought to be collocations consisting of \( V \times NP_{\text{man}} \times X \), where \( X \) is some sort of eventive NP that forms light verb constructions. No such collocations exist, they are all \( V \times X \times NP \) instead:

(82) \( V \times NP_{\text{man}} \times X \):
    give X the once-over, give X (e.g. rugs) a beating, give X a try, give X a whirl

(83) \( *V \times NP \times X \):
    * give the rugs X, * give the car X, * give the bed X

This shows both that Larson’s view is incorrect, and that animacy (or a constraint against human nouns) cannot be behind the lack of \( V \times NP \times X \) expressions. This latter point is further confirmed by the fact that there is a small but significant number of double object expressions with both NPs fixed (Pattern 2e repeated from above):

(84) Collocation Pattern 2e: \( V \times NP \times NP \)
    give it a rest, give it your best shot, give it the old college try, don’t give me that (Imperative)
What these show is that there is nothing wrong in principle with having the first object of a double object construction fixed in a conventionalized expression. The first object is banned from being fixed only when the second object is an open slot.

Let me say a little bit more about animacy, or a putative constraint against human nouns. The missing ditransitive pattern is sometimes claimed to be ruled out by a constraint against NPs denoting humans in idiomatic expressions (e.g., [Nunberg et al. 1994, Mishani-Uval & Siloni 2017]). Since the first object of a ditransitive is typically human, it is statistically unlikely, this reasoning goes, that idiomatic expressions will be created with a fixed NP in first object position. On this view, there is no structural constraint against expressions of the form *V NP X. There are many reasons this cannot be correct. First, as we just saw, there are expressions with both NPs fixed. There is nothing wrong with having a fixed NP as first object, this is only missing when the second object is an open slot. Second, appealing to a correlation between humanness and syntactic position yields little in the way of explanation. It is often claimed that there is a correlation between argument position and humanness: subjects and indirect objects tend to be human while direct objects tend to be non-human. I personally doubt that this is true in gross frequency. What humans talk about most is humans: what they do to each other, what they say to each other, and so on. My guess is that the most common NP in any syntactic position is a pronoun referring to a human. Direct objects are probably human just as often as subjects and indirect objects are in large corpora. Additionally, subjects and indirect objects can be non-human. In the conventionalized expressions that have a fixed subject, the subject is often inanimate, even though subjects are supposed to be associated overwhelmingly with humans (the shit hit the fan, that ship has sailed, the walls have ears, when X’s ship comes in, etc.). Since conventionalized expressions can easily be formed with non-human subjects, even though subjects are supposed to be highly correlated with humans, we should then expect that ditransitive expressions could also easily be formed with a fixed first object that is non-human, even though first objects are typically humans.

For all of these reasons, I conclude that a constraint against human NPs in conventionalized expressions cannot be the explanation for the missing ditransitive pattern. Contra Larson (2017), then, it is truly significant that there are no idioms or collocations of the form *V NP X. This pattern is strikingly missing from ditransitives, in contrast with V X NP, which is well represented with both collocations and idioms (as is V NP NP, with both NPs fixed). There must be a syntactic constraint against conventionalized expressions of this form, since there is no other reason they could not be formed.

Larson (2017) further argues that in all the collocations and idioms with give of the form V NP to X, like give way to X, give rise to X, give plus the noun after it form an inseparable compound and do not constitute a phrasal idiom. His arguments for this are that: (i) they cannot passivize; (ii) the noun cannot be modified; (iii) the noun cannot be questioned. None of these arguments is telling, and some are incorrect. At least one of them can passivize (86a), and some of them can be modified (86b):

(86) a. In build and coat and brush he was a huge timber-wolf; but the lie was given to his wolf-hood by his color and marking. (Jack London, Brown Wolf)
   b. . . . his courage gave immediate way to despair; (James Pettit Andrews, History of Great Britain from the death of Henry VIII to the accession of James VI of Scotland to the crown of England)

As for questioning, many if not most idioms do not easily allow questioning of their parts:

(87) a. * What did she kick? The bucket.
c. * What can’t he hold to his father? A candle.

This is because many sub-parts of idioms do not have independent meaning.

Larson’s arguments regarding V NP to X idioms like *give way to do not go through, then. I see no reason to treat these idioms as being in any way different from other idioms, many of which also do not passivize easily (e.g., *kick the bucket).

This gives us the following patterns of conventionalized expressions in ditransitives:

(88)  
double object construction  
a. V X NP: cry X a river, cut X some slack, give X a break, give X a go  
b. * V NP X: *throw the wolves X, *feed the lions X, *give X’s old man Y  
c. V NP NP: give the devil his due, give it a rest, give it the old college try

(89)  
prepositional dative construction  
a. V NP to X: give the lie to X, give rise to X, give priority to X, give some thought to X  
b. V X to NP: feed X to the lions, throw X to the wolves  
c. V NP to NP: give hostages to fortune

According to Larson (2017), it is important that we only look at prepositional datives that have a caused possession rather than a caused location meaning. This is because only caused possession prepositional datives have the potential to alternate with the double object construction, which uniformly encodes caused possession. I list in (89) only those collocations and idioms that do indeed seem to have a caused possession meaning. The tests that are supposed to distinguish caused possession from caused location PPs all show that these are caused possession. One such test involves substituting different directional prepositions for to. For instance, with throw, one can throw something to, at, toward, under, or over something. According to Hallman (2015), this test shows that *throw to the wolves is not caused possession, because throw can occur with other prepositions. However, I would contend that, because this idiom itself does not permit any other preposition, this test actually shows that this idiom does encode caused possession. Intuitively, the subject of throw in the idiom does have the intent of giving the object to the wolves (intended caused possession). Turning to other idioms, the preposition substitution test does show clearly that *give hostages to fortune and *feed X to the lions are caused possession: *give hostages to/*at/*toward; *feed X to/*at/*toward.

Another test for caused location versus caused possession is the ability to add halfway (Rappaport Hovav & Levin 2008): throw it halfway to the goal line is wellformed, showing that throw may involve location and not caused possession. But halfway is not possible with the idiom throw X to the wolves: *They threw him halfway to the wolves. This test also shows that give hostages to fortune and feed X to the lions are caused possession: *gave it halfway to Bill, *fed it halfway to the dog.

A third test is the ability to question the PP with where rather than to who(m). Where did you throw it? is wellformed, showing that throw is (or can be) a location verb (but again, this does not work with the idiom throw X to the wolves). This test also shows that give hostages to fortune and feed X to the lions are caused possession: *Where did you give it? *Where did you feed it? (where it is the food item).

In addition, according to Rappaport Hovav & Levin (2008) and Larson (2014), the verbs give, hand, and lend can only ever encode caused possession, so any conventionalized expression with those verbs is relevant (which makes lend X color, Y lend X Y’s ear, lend X a hand all examples of the V X NP pattern).

This means that, contra Hallman (2015) and Larson (2017), there are in fact many conventionalized expressions that are caused possession ditransitives. Among prepositional datives, all possible combinations of the verb and at least one of its objects are attested (89). In contrast, among double object constructions, one of the three possible patterns is systematically missing (see 88b). Given that the pattern with both objects exists (88c), it cannot be the case that something rules out the first object being part of a conventionalized
expression. It is only impossible for the first object to be part of the expression if the second is not. This calls for a structural account, as it is impossible to state this restriction in terms of thematic roles or a constraint against human NPs.

Furthermore, as just observed, all possible patterns of the prepositional dative exist. Both \( V \ NP \ P \ X \) (\textit{give rise to X}) and \( V \ X \ P \ NP \) (\textit{feed X to the lions}) exist. The former regularly alternates with the double object construction, as in \textit{give the lie to X} alternating with \textit{give X the lie}. Significantly, the second pattern never does: \textit{throw X to the wolves} and \textit{feed X to the lions} do not alternate with \textit{*throw the wolves X} or \textit{*feed the lions X} \cite{Rappaport Hovav & Levin 2008, Bruening 2010, Larson 2014}. This further reinforces the conclusion that some kind of syntactic constraint must rule out the missing pattern.

To sum up this section, the missing ditransitive pattern requires a syntactic account. It is not the case that conventionalized expressions are free to occur in any structure that the syntax can produce. In the next two sections, I explore what form syntactic constraints might take, starting with previous proposals.

5 Syntactic Constraints: Previous Proposals

Numerous proposals have been made on the topic of syntactic constraints on conventionalized expressions (typically for idioms). In this section I show that most of them are incorrect. In section 6 I propose a new account.

5.1 Constituency, Locality, etc.

As noted above, it has been common over the years for researchers to propose that idioms are underlying constituents, excluding all non-idiomatic material \cite{Larson 2017}. As also noted above, this view was shown long ago to be untenable \cite{Ernst 1981, Nunberg et al. 1994, Nicolas 1995, O’Grady 1998}. Idioms can be disrupted by modifiers 90; the determiner that usually appears can be exchanged for a different one 91, where the canonical form is \textit{beat around the bush}; numerous idioms have an open slot for a possessor 92; and many idioms consist of a verb, NP, and P, but exclude the object of that P 93:

\begin{itemize}
  \item \textit{Before the symbolic shit hit the colonial fan recently at the University of Cape Town,...} (http://www.thejournalist.org.za/spotlight/we-love-uct-says-student-who-covered-rhodes-in-shit)
  \item \textit{I’ve seen you beat around more bushes than a landscaper.} (http://www.kungfumagazine.com/forum/showthread.php?42550-OT-9-11-Truth-Symposium/page5)
  \item \textit{lose X’s cool, get X’s goat, fill X’s shoes, …} \cite{O’Grady 1998: (4c–e)}
  \item \textit{beat the bushes for X, cast a pall on X, NEG hold a candle to X}
\end{itemize}

In none of the cases above does the idiomatic material constitute a constituent that excludes all non-idiomatic material. We can now also add expressions like \textit{might as well X} and \textit{Who would have thought X?}, where the fixed material does not form a constituent excluding the open positions.

Researchers have made various other proposals, as well. \cite{Baltin 1989} proposes that idioms involve only the head of a phrase and the head of one of its complements, while \cite{van Gestel 1995} proposes that idioms consist of the selection of a lexical head by another lexical head. Neither of these proposals is correct, since specifiers can be involved too, as can modifiers \cite{play devil’s advocate, beat/flog a dead horse}. In addition, idioms can include functional material, like negation, aspect, and imperative mood \cite{pattern 7; see Weibelhuth & Ackerman 1994 on some of these points}. \cite{Koopman & Sportiche 1991: 224} propose a constraint that says that if X is the minimal constituent containing all the idiomatic material, the head of X is part of the idiom. This is correct, but it is too weak. For instance, it would permit any idiom that consisted of just the head C of a CP plus any subset of material within the CP, including discontinuous material. There
could be idioms like, *if X vs the wolves*, with open material except for the object. No such idioms exist. Finally, [Harley & Jung (2015: 724)](#) suggest that the head of an idiomatic constituent must combine with its sister as an idiom before idiomatizing with other constituents. This is also clearly false, as we have seen numerous examples where a preposition combines with other constituents as an idiom, leaving its own sister out (32, 93). There are also idioms that include a verb and an adjunct PP, but which exclude the object of the verb (*nip X in the bud, can cut X with a knife*). If arguments always combine with verbs before adjuncts do, as is commonly assumed, then these also falsify Harley and Jung’s proposed constraint.

There are two other approaches to idioms that I will rule out because they introduce notions that I believe syntactic theory can do without. These are the *catena* theory of Osborne et al. (2012), and construction grammar approaches (e.g., Riehemann 2001). I believe that syntactic theory can and therefore should do without the formal notions of a construction and a catena. As I will show below, it is possible to develop an adequate account without either of these formal constructs.

Other researchers have proposed that there are locality constraints on idioms. For instance, Marantz (1997) proposes that idiomatic interpretations are bounded by functional heads like v, the head that introduces the external argument of a verb. In a similar proposal, Harley & Stone (2013) propose that idioms may not include agents. Svenonius (2005) proposes that phase boundaries are absolute barriers to idiomatizations. That is to say, all parts of a phrasal idiom must be minimally dominated by the same phasal node. In the phase theory of Chomsky (2000), the two phasal nodes in the clause are vP and CP.

All of these proposals are incorrect, as Bruening (2017) showed. There are numerous idioms that include multiple clauses, in violation of any locality constraint. Additionally, many of these idioms include embedded clauses that have agentive subjects. Consider the following examples of embedded non-finite clauses:

(94) a. NEG have a pot [PRO to piss in] (‘lack any standing’)
   b. have an axe [PRO to grind] (‘have a (quarrelsome) reason’)
   c. NEG have a leg [PRO to stand on] (‘have no standing’)
   d. a bitter pill [PRO to swallow] (‘is unpleasant but must be accepted’)
   e. a tough row [PRO to hoe] (‘is very difficult’)
   f. play hard [PRO to get] (‘be coy’)
   g. rob Peter [PRO to pay Paul] (‘discharge one debt by incurring another’)
   h. X cut off X’s nose [PRO to spite X’s face] (‘do something self-destructive out of revenge’)
   i. Y teach X’s grandmother [PRO to suck eggs] (‘offer advice to someone with more experience’)

I indicate the null subject of a non-finite clause as PRO. All of the above non-finite clauses have agentive subjects; they therefore simultaneously counterexemplify both a locality condition and a constraint against having agentive subjects within idioms.

Idioms can also include or be full finite embedded clauses:

(95) a. kill the goose [that lays the golden eggs] (‘destroy a profitable thing out of greed’)
   b. X bite the hand [that feeds X] (‘turn against someone you depend on’)
   c. the straw [that broke the camel’s back] (‘a seemingly minor thing that causes a disaster’)
   d. strike [while the iron is hot] (‘act before it is too late’)
   e. close the stable door [after the horse has bolted] (‘act after it’s too late’)
   f. (NEG) count X’s chickens [before they hatch] (‘plan on something that might not happen’)
   g. X cross that bridge [when X comes to it] (‘wait to act until it is necessary’)

26
h. X bite off more [than X can chew] (‘take on more than one can handle’)
  i. kick X [when/while X is down] (‘treat someone badly who has already suffered a setback’)
  j. be that [as it may] (‘that is true, but it does not change things’)
  k. [when hell freezes over] (‘never’)
  l. [until the cows come home] (‘for a very long time’)

Such examples show that there are no locality constraints on idioms. Idioms can actually be quite large, and include multiple clauses. I will have more to say about such expressions in section 6.6. Note again that many of the embedded subjects are agentive. We can also find agentive subjects in expressions like Who asked X?

Note that the longer idioms shown in this subsection are just combinations of the patterns identified in section 3. For instance, NEG have a pot to piss in is the combination of a functional element plus a main verb (pattern 7), plus a verb and its object (pattern 2a), plus a non-finite relative clause modifying the N (pattern 1h); the relative clause is also a complete clause (pattern 8).

5.2 Previous Poposal: The Selection Theory

Two previous proposals do come close to capturing the facts adequately. These are the Dependency Theory account of O’Grady (1998) and the Selection Theory of Bruening (2010). I will briefly describe the selection theory, as the constraint it posits to rule out certain patterns as idioms captures the most data.

Bruening (2010) proposes the following as an account of the syntactic forms that idioms can take (Ev- eraert 2010 independently proposes something very similar):

(96) The Principle of Idiomatic Interpretation:
  X and Y may be interpreted idiomatically only if X selects Y. (Bruening 2010: 532, (24))

In this formulation, idiomatic interpretation is dependent on selection. Idioms consist of one element selecting another, and the two may then be interpreted as an idiom. Longer idioms are built up through chains of selectional relations.

As an example, one of the most common types of phrasal idiom consists of a verb and its direct object, for instance buy the farm (‘die’). This idiom consists of a verb, buy, which selects an NP object headed by farm. This follows the Principle of Idiomatic Interpretation. Another common idiom pattern consists of a verb and its object, but with an open slot for the possessor of the object. Possessors can be open slots because they are not selected. In Y get X’s goat (‘Y drive X to distraction’), for instance, the verb get selects an NP headed by goat, and so the two can be interpreted idiomatically. The possessor is not selected and does not disrupt this selection. Modifiers can also be added to idioms, as in kick the filthy habit, because they do not disrupt selection. In this example, kick selects habit, and a modifier can be added without disrupting this selection. Modifiers can also be part of an idiom, as in beat/flog a dead horse (‘waste energy on a lost cause’). This is possible because modifiers select what they modify, according to Pollard & Sag (1994) and Bruening (2010). In beat/flog a dead horse, the verb selects an NP headed by horse, and the adjective dead does so as well. All three elements are then connected through a chain of selection.

Importantly, the selection theory rules out certain types of phrases as idioms. In particular, it rules out discontinuities in idiomatic phrases, where X selects Y and Y selects Z but only X and Z are part of the idiom. Consider the example of X has gone to the dogs (‘X has worsened dramatically’). This idiom obligatorily occurs in the perfect aspect (Lebeaux 2009: 65). The idiom therefore consists of perfect aspect selecting a verb which selects a PP headed by to which selects an NP headed by dogs. An impossible idiom would have perfect aspect included plus something selected by the main verb, but where the main verb is not part of the idiom and is free to vary. This would be something like, *X has V-en an old shoe, perhaps
with a meaning like, ‘X has V-en every conceivable thing’. There is no such idiom, and the selection theory accounts for this.

Similarly, there could be no idiom that includes negation, imperative mood, and an object, but where the verb is free to vary. This would have a form like *NEG V! the golden goose, with a meaning like, ‘don’t V a profitable thing’. Again, no such idiom exists. The selection theory also rules out an idiom where the verb is fixed as is an adjective modifying the object, but the object itself can vary: *beat a dead X, *be in hot X. The verb does not select the adjective and the adjective does not select the verb, so such idioms violate the Principle of Idiomatic Interpretation. Also ruled out is an idiom consisting of a verb and the possessor of the object, but not the object itself: *play the devil’s X (O’Grady 1998: 287). Again, the verb does not select the possessor and the possessor does not select the verb. (In play devil’s advocate (‘argue for a contrary position’), the verb selects the object, and the possessor is selected by the head noun advocate, as it is a selected argument of that noun.)

In addition to the Principle of Idiomatic Interpretation, Bruening (2010: 532, (25–26)) proposes a constraint on possible idioms, as follows:

(97) Constraint on Idiomatic Interpretation:
If X selects a lexical category Y and X and Y are interpreted idiomatically, all of the selected arguments of Y must be interpreted as part of the idiom that includes X and Y.

(98) Lexical categories are V, N, A, Adv.

This constraint rules out certain patterns: having a fixed subject and verb but an open slot for an object (*the cat got X); and the missing ditransitive pattern, *V NP X (*throw the wolves X). In Bruening (2010), this is because external arguments and first objects of ditransitives are introduced not by the lexical verb, but by a functional head. The external argument is introduced by the head Voice (Kratzer 1996), and first objects of ditransitives are introduced by an Appl(icative) head (Marantz 1993). If a subject were to be included in an idiom, the idiom would also have to include Voice, which selects the subject; Voice then selects the V, which is allowed, but by the Constraint on Idiomatic Interpretation, all of V’s arguments would have to be included, too. There can therefore be idioms with both a subject and an object like the shit hit the fan, but not ones with a fixed subject but open object. Similarly, if the first object of a ditransitive were part of an idiom, Appl would have to be, too, since it is what selects the first object; Appl would then select the V, but then all of the arguments of V would also have to be included. Hence an idiom with fixed verb and first object but open second object is excluded.

At the same time as it rules out such idioms, the Constraint on Idiomatic Interpretation permits arguments of functional categories to be open slots. As O’Grady (1998: 300–301) notes, and as discussed above, there are numerous idioms that consist of a verb plus a PP, but where the object of the P is an open slot (beat the bushes for X). In the selection theory, the verb selects an NP and a PP headed by a particular preposition in these idioms. The complement of P does not have to be included, given the Constraint on Idiomatic Interpretation, because P is not a lexical category.

5.3 Problems Generalizing to All Conventionalized Expressions

The selection theory was proposed specifically for idioms. However, we could generalize the Principle of Idiomatic Interpretation to cover all conventionalized expressions, as follows:

(99) The Principle of Conventionalization:
X and Y may become a conventionalized expression only if X selects Y.

We could also generalize the Constraint on Idiomatic Interpretation to the following:
Constraint on Conventionalized Expressions:
If X selects a lexical category Y and X and Y are a conventionalized expression, all of the selected arguments of Y must be part of the conventionalized expression that includes X and Y.

Unfortunately, there are a few problems for the Constraint on Conventionalized Expressions in (100). The empirical study reported above reveals that conventionalized expressions exist that include a chain of selected elements, but then allow an open slot in a fairly low position. These include the following, where X appears to be selected by a lexical head (both idioms and collocations):

(101) a. give X to understand that Y, lead X to believe that Y, let X know that Y, get the impression that X, the fact remains that X, Who would have thought that X?
b. can’t afford X, can’t help X-ing, can’t stand X
c. Who asked X? What’s eating X? What makes X tick?
d. can cut X with a knife, nip X in the bud
e. recommend X highly, strongly support X, take X seriously, well worth X

The problem with these expressions is that one element of the expression selects a lexical category, but then not all of that lexical category’s selected arguments are included in the conventionalized expression. This violates the Constraint on Conventionalized Expressions in (100).

Consider for example the idiom *can cut X with a knife* (’X is a very strong/intense’). This always occurs with the modal *can*, and it has an adjunct PP. Suppose the modal selects VoiceP. The PP either selects VoiceP (Bruening 2013) or VP. Either way, VP is also selected, by Voice if by nothing else. Being a lexical category, all of its selected arguments have to be part of the conventionalized expression, but they are not.

The expressions listed in (101) that have an open clausal argument might not be problems. For instance, in *get/give the impression that X*, we could say that the N *impression* selects a CP headed by *that*. This is not a lexical category, so not all of the selected arguments of *that* need to be included. However, many of the open slots are objects of verbs, as in *can cut X with a knife, nip X in the bud* (’put a stop to X while it is small and manageable’), *Who asked X?*, etc. These definitely violate the CFV in (100).

Because of this issue with the selection theory, I will propose a completely different approach to the syntax of conventionalized expressions, with a very different constraint. The new proposal will account for all the patterns that exist, while ruling out the ones that do not.

6 Syntactic Constraints: New Proposal

All previous accounts of conventionalized expressions have tried to explain the syntactic form of the specified material, minus the unspecified material. I suggest instead that we view the unspecified material—the open slots—as part of the expression. So in *might as well VP*, the VP is part of the expression. In *can cut NP with a knife*, the open NP is part of the expression. Once we view open slots as part of the expression, then it is very simple to state what can form a conventionalized expression: any phrase XP can be a conventionalized expression, so long as all obligatory elements are present in XP. That is, all selectional and other syntactic requirements must be met. (This rules out an expression like *take seriously* with no open slot for an object, because *take* requires an object.)

(102) Any XP can form a conventionalized expression, so long as all selectional and other syntactic requirements imposed within XP are met within XP.

Syntactic requirements from outside XP do not have to be met within XP. For instance NPs can get case from outside XP, functional elements within XP can take their form based on elements outside XP, in some cases a wh-phrase within XP will have to undergo wh-movement once XP is embedded within a CP.
Additionally, elements within a conventionalized XP are one of four things:

(103) Elements of Conventionalized Expressions:
   a. Functional elements whose form is determined by the syntactic context;
   b. Elements specified for lexical content;
   c. Bound variables;
   d. Free variables.

Bound variables have a syntactic category but are bound by something else within XP. Their actual form is determined by the syntactic context (e.g., a pronoun). Free variables are also specified for a syntactic category, but they are filled in with items of the appropriate syntactic categories (items that are not part of the stored expression; these are the open slots). Additional, non-selected material that is not part of the expression can also be added to it when it is embedded in a larger syntactic structure (like adverbs, adjectives, etc.).

To illustrate with *might as well* VP, the conventionalized expression is the phrase depicted in (104). I am going to assume that all conventionalized expressions that do not include a subject are open predicates, \( \lambda x \ldots \). They all have Voice (Kratzer 1996), with a lambda operator and a bound variable as Spec-VoiceP. So in (104), the argument of Voice in Spec-VoiceP (the thematic position of the external argument) is a bound variable, while the VP is a free variable:

\[
\text{(104)} \quad \text{ModalP} \left( \lambda x \ldots \right)
\]

\[
\begin{align*}
\text{Modal} & \quad \text{VoiceP} \\
\text{might} & \quad \text{AdvP} \\
\text{as well} & \quad \text{VoiceP} \\
\text{NP}_x & \quad \text{Voice} \\
\text{Voice} & \quad \text{VP}_y
\end{align*}
\]

(A complete representation would indicate the morphological form of the verb, here the bare form since the VP is in the immediate scope of a modal. I will not include such details in the representations.)

All well-formed conventionalized expressions are now just XPs. \( \lambda x x \text{bring a tear to } y \text{NP’s eye} \) is a VoiceP with a free variable in a possessor position (of category NP). God’s gift to \( x \text{NP} \) is an NP with a free variable after a preposition. \( \lambda x x \text{spend x’s life } y \text{VP/PP} \) is a VoiceP with a bound variable in a possessor position and a free variable of category VP or PP. The shit hit the fan is a VoiceP with no bound or free variables (but functional elements like tense that will be determined by the syntactic context).

This is a maximally simple account. Any XP can be a conventionalized expression. This is something we might expect, and is probably what is behind the oldest intuition about idioms, namely, that they form underlying constituents. This is now true, if we view open slots as part of the underlying constituent.

---

30

An alternative is that many expressions with an open subject are just a bare VP. So buy the farm would just be a V and its object, which can (and probably must) be embedded under a Voice when used. Some expressions require variables bound by the external argument, though, so they must include VoiceP and a binder (e.g., \( X \text{ blow X’s stack} \), and might as well here, which includes a modal that is higher than VoiceP). I have therefore gone for a uniform analysis, where all expressions with open subjects have a bound variable as subject (except regarding alternating ditransitive idioms, in section 6.5 below).
6.1 Constraint: Full Phrases Only

Now, there are additional constraints. The first is that elements specified for lexical content, bound variables, and free variables are all necessarily full phrases. Only functional elements can be less than a full phrase. So discontinuous idioms of the type discussed in section 5.2 above are ruled out because only whole phrases can be free variables. These are repeated below:

(105) a. * X has V-en an old shoe
b. * NEG V! the golden goose
c. * beat a dead X, *be in hot X
d. * play the devil’s X
e. * if X Vs the wolves

In (105a–b), only the V is a free variable, but this is not allowed, since only whole phrases can be free variables. VPs can be free variables, but Vs cannot. In (105c–d), only part of an NP is a free variable, and again this is not permitted. In (105b), the subject is a phrase and so can be a free variable, but just the V of a VP cannot be.

As can be seen, this simple analysis, along with the reasonable assumption that all non-functional constituents of a conventionalized expression have to be full phrases, accounts for a large part of what can be a conventionalized expression and what cannot.

6.2 Constraint: Maximal Number of Free Variables

Once we allow any XP to be a conventionalized expression, and take open slots to be part of the expression, then the task becomes to explain the distribution of the open slots. We have shifted the focus from the fixed material to the open positions. One of the most obvious limitations on the open positions, or the free variables as I am treating them here, is that there is an upper limit on their number. The vast majority of the expressions collected here have zero or one (not counting the subject, since I am viewing predicate expressions as open predicates with a bound variable as subject, not a free variable). The most I have found is two, and I have found only five expressions that have two free variables:

(106) a. λx.x give yNP to understand zCP
b. λx.x lead yNP to believe zCP
c. λx.x bet (yNP) x’s bottom dollar (zCP) (‘X is certain that Z’)
d. λx.x talk yNP into/out of zNP
e. λx.x lay yNP at zNP’s door (‘X blames Z for Y’)

Other than the above five expressions, if there is more than one variable, the others are all bound variables, as in examples like the following:

(107) a. λx.x give yNP a piece of x’s mind
b. λx.x pay x’s respects (to yNP)
c. λx.x spend x’s life yVP/PP

---

8I am aware of two potentially problematic examples. These are the close synonyms X’s late, lamented Y and X’s dear departed Y. These seem to be NPs which have an open slot just for the head N. On closer inspection, however, there is reason to think that these do not have the structure of simple NP. The head N can be a proper name, for instance, with a definite determiner rather than a possessor: the late, lamented John Smith. The adjectives are not interpreted restrictively, but are rather non-restrictive. The dear departed can also occur by itself as an NP, with no other noun. I suggest that the pair of adjectives in both cases form a full phrase, and that phrase modifies another full NP appositively. The latter full NP is then what is a free variable.
I will state this constraint purely descriptively as the following:

(108)  Maximality of Conventionalized Expressions:
       A conventionalized expression may have a maximum of two free variables.

Conventionalized expressions can be quite long, as noted above, but when they are, they are entirely fixed or they have bound variables, not free variables. For instance, in $\lambda x.x \ cross \ that \ bridge, \ when \ x \ come \ to \ it$, the embedded subject is a bound variable, as are tense and the embedded object. There are no free variables. (Functional elements like tense can be bound variables without being full phrases because that is an option available to them in the syntax independently. That is, tense here is not really specified as a bound variable in the conventionalized expression, it is just a functional element; the syntactic context makes it a bound variable.)

6.3 A Structural Constraint on Free Variables

One of the goals of the present theory is to rule out the missing ditransitive pattern, *$\lambda x.x \ give \ the \ wolves \ y_{NP}$ (fixed first object, open second object). For the moment, I will also assume that the pattern *the farmer bought $x_{NP}$ (fixed subject, open object) is also to be ruled out, although I will examine whether it really is missing in great detail in section 6.7. There I will show that the constraint that rules out the missing ditransitive pattern also rules out this pattern in general, while also allowing the principled exceptions that exist, in English and in other languages.

Assuming that both of these patterns are to be ruled out, we might suggest that there is a constraint saying that a free variable cannot co-occur with a fixed NP if the fixed NP is hierarchically higher than the free variable. This would rule out both *$\lambda x.x \ give \ the \ wolves \ y_{NP}$ and *the farmer bought $x_{NP}$. This could not be right, however, because open possessors can freely co-occur with fixed NPs that are hierarchically higher than them, as in cat got $x_{NP}$’s tongue and $\lambda x.x \ bring \ a \ tear \ to \ y_{NP}$’s eye. Objects of prepositions can also be free variables in the presence of a hierarchically higher fixed NP, as in the bottom fell out of $x_{NP}$ and god’s gift to $x_{NP}$.

What seems to be banned is specifically a free variable co-occurring with a hierarchically higher fixed NP when both are core arguments, meaning subject, direct object, and indirect object. As just noted, possessors and objects of prepositions, which are not core arguments, are fine as free variables. It seems to be specifically the core arguments that are restricted.

There are numerous ways of stating such a constraint. I will state it in terms of case competition models like that of Marantz (1991), where NPs receiving structural case in a given domain compete with each other. In this kind of model, subject and object compete, with the lower argument being assigned accusative case in nominative-accusative languages only in the presence of another competitor (similarly, ergative is assigned to the higher of two competitors in ergative languages). Nominative is assigned otherwise. See Marantz (1991), Baker (2014) for details. I will state the constraint as follows:

(109) The Constraint on Free Variables in Conventionalized Expressions (CFV):
In a conventionalized expression XP, no phrase dominated by XP may be a free variable if it has a case competitor dominated by XP that precedes and phase-commands it and is not a bound variable.

Phase-command comes from Bruening (2014a) and is defined as follows:

(110) Phase-Command: X phase-commands Y iff there is no ZP, ZP a phasal node, such that ZP dominates X but does not dominate Y. (Bruening 2014a 343, (2))

(111) Phasal Nodes: CP, VoiceP, NP (modified from Bruening 2014a 343, (3))
Section 6.5 goes into more detail on phase-command, and why I adopt it and not c-command. For the moment, one can simply think of this relation as ‘hierarchically higher than’.

The CFV permits open subjects with fixed objects, because, as stated, such expressions are open predicates with a bound variable as subject ($\lambda x. buy the farm$). There is no free variable limited by the CFV. A fixed subject and open object, like *the farmer bought $x_{NP}$, is banned because the object is a free variable, but it has a case competitor in the subject that is not a bound variable. Similarly, the missing ditransitive pattern, *$\lambda x. give the wolves y_{NP}$, is banned because the free variable has a case competitor in the indirect object that is not a bound variable (the subject is a bound variable, which is allowed by the CFV). The pattern that is attested, $\lambda x. cut y_{NP} some slack$, is allowed because the free variable precedes and phase-commands the fixed NP, while the subject is a bound variable.

Objects of prepositions can be free variables quite freely, because they are not case competitors with any other NPs. In $\lambda x. light a fire under y_{NP}$, the object of the P is not a case competitor for a fire. Both ditransitive patterns with PPs are also allowed: In $\lambda x. throw y_{NP} to the wolves, the wolves is not a case competitor for y. In $\lambda x. give rise to y_{NP}$, again rise is not a case competitor for the free variable. It is also possible for both the direct object and the object of the preposition to be specified, as in $\lambda x. give hostages to fortune$, because there are no free variables at all. Similarly for $\lambda x. give the devil his due$.

Possessors can also be free variables rather freely, because they are not case competitors for any other NPs. In $\lambda x. bring a tear to y_{NP}$’s eye, the possessor is not a case competitor for either a tear or the NP it is part of (y’s eye). In cat got $x_{NP}$’s tongue, the possessor is not a case competitor for the subject. Inside an NP, the possessor can be specified while the object of a P is a free variable, as in God’s gift to $x_{NP}$, because objects of Ps are not case competitors for possessors.

As for the few expressions with two free variables, repeated below, they do not involve case competitors:

(112) a. $\lambda x. give y_{NP}$ to understand $z_{CP}$
   b. $\lambda x. lead y_{NP}$ to believe $z_{CP}$
   c. $\lambda x. bet (y_{NP}) x$’s bottom dollar ($z_{CP}$)
   d. $\lambda x. talk y_{NP}$ into/out of $z_{NP}$
   e. $\lambda x. lay y_{NP}$ at $z_{NP}$’s door

In (112a–c), the second free variable is a CP. I assume that CPs as objects do not enter into case competition with NPs (cf. the Case Resistance Principle of Stowell 1981). In (112d), the second free variable is the object of a preposition and so is not a case competitor for the first. In (112e), the second free variable is a possessor and so is also not a case competitor for the first.

6.4 Accounting for Patterns in the Study

I now go through the patterns that were identified in section 3 and show how the analysis accounts for them. I begin with modification in Pattern 1:

(113) Pattern 1: Modification
   a. Adj N: honest mistake (idiom: cold feet)—expression consists of NP with lexically specified AP and N
   b. Adv Adj: acutely aware (idiom: writ large)—consists of AP with specified AdvP and A, possibly with free variable PP/CP
   c. Adv V: go smoothly (idiom: only have eyes for X)—consists of VP with specified AdvP and V (and NP), possibly with a free variable within VP

---

9I assume that CPs as subjects are actually NPs, as in Davies & Dubinsky (2009) and others. As such, they are case competitors for objects. Note that this analysis makes predictions for languages where CP objects are nominalized and case marked.
d. Adv Adv/P: hardly ever, way ahead (no idioms)—consists of AdvP or PP with fixed AdvP
   modifying Adv/P

e. N PP: call of duty (idiom: elephant in the room)—consists of an NP with specified N and PP

f. Adj PP/NP: missing in action, worth a fortune (idiom: running on fumes)—consists of AP with
   specified A and PP (or NP)

g. Comparison: better than a kick in the pants (idiom: X’s bark is worse than X’s bite)—consists
   of AP with specified A, Deg(ree), and comparison (idiom: VoiceP with specified NP subject
   with free variable for possessor, specified V, AP with specified A, Deg, comparison, NP with
   bound variable possessor)

h. Possessor N: X’s finest hour, beginner’s luck (idiom: X’s birthday suit, Davy Jones’s locker)—
   consists of NP with free variable or specified possessor and specified N

i. N Relative Clause: the least X can do (idiom: a bitter pill to swallow)—consists of NP with CP
   modifier, possibly with a free variable within it

As can be seen, these all involve a phrase XP, and consist of elements that are lexically specified, bound
variables, or free variables. The same is true of Pattern 2, involving a verb and its arguments/adjuncts:

(114) Pattern 2: Verbs and Arguments/Adjuncts

a. V NP: answer the door (idiom: back the wrong horse)—consists of VoiceP with bound vari-
   able subject, specified V and NP

b. V X’s NP: break X’s heart (idiom: clip X’s wings)—consists of VoiceP with bound variable
   subject, specified V and NP, NP contains free or bound variable possessor

c. V NP P X: break the news to X (idiom: beat the bushes for X)—consists of VoiceP with
   bound variable subject, specified V and NP, PP with specified P and free variable NP object
   of P

d. V X NP: give X permission, do X a favor (idiom: cut X some slack)—consists of VoiceP with
   bound variable subject, ApplP with free variable NP argument, VP with specified V and NP

e. V NP NP: give it a rest (idiom: give the devil his due)—consists of VoiceP with bound
   variable subject, ApplP with specified NP argument, VP with specified V and NP

f. V NP PP: Y bring a tear to X’s eye (idiom: build castles in the air)—consists of VoiceP with
   bound variable subject, specified V and NP, PP with specified P and NP, NP may have free
   variable possessor

g. V PP: come into view (idiom: beat around the bush)—consists of VoiceP with bound variable
   subject, specified V and PP (unaccusative: assume non-thematic argument of VoiceP binds a
   bound variable NP object of V)

h. V X PP: bring X to justice (idiom: feed X to the lions)—consists of VoiceP with bound
   variable subject, specified V and PP with free variable NP object of V

i. V PP P X: keep in touch with X (no idioms)—consists of VoiceP with bound variable subject,
   specified V, PP, P with free variable NP object

j. V (NP) Adj: call in sick, make a long story short (idiom: getting busy, get X’s feet wet)—
   consists of VoiceP with bound variable subject, specified V and AP, may have specified NP
   in addition (which may have free variable possessor)

k. V X Adj: drive X crazy (idiom: pull it/X together)—consists of VoiceP with bound variable
   subject, specified V and AP, free variable NP
1. V (NP/X) Infinitive: lead X to believe Y (idiom: let sleeping dogs lie)—consists of VoiceP with bound variable subject, specified V and infinitive, V has NP object that may be specified or a free variable, infinitive may be entirely specified or may have free variable within it (CP)

m. V (NP) CP: tell me it isn’t/ain’t so (idiom: X teach X’s grandmother to suck eggs)—consists of VoiceP with bound variable subject, specified V and CP, may have NP object of V that is specified (may have bound/free variable possessor within it), specified CP may have bound variable within it

n. (P8) V (X’s) NP Y: X spend X’s life Y-ing (idiom: X cut X’s teeth Y-ing)—consists of VoiceP with bound variable subject, specified V, specified NP object with bound variable possessor, free variable gerund (or PP)

Patterns 3–9 also fall into line:

(115) (P3) Coordinations: law and order (idiom: the old ball and chain)—consists of XP with functional Conj(unction) and specified conjuncts

(116) (P4) P NP: against the law (idiom: behind the eight ball)—consists of PP with specified P and NP

(117) (P5) Adj Infinitive: hard to say (idiom: fit to be tied)—consists of AP with specified A and infinitive

(118) (P6) Predicative: no money down (idiom: back against the wall)—consists of small clause XP with specified NP subject and specified predicate

(119) (P7) F V (NP P) X: can’t stand X (idiom: has had it)—consists of a functional XP with bound variable subject, specified Modal/Aspect, VoiceP with bound variable subject, specified V with specified or free variable NP object

(120) (P8) Full Clause: easy does it (idiom: the shit hit the fan)—VoiceP with specified subject, specified V, specified NP object

(121) (P9) Covert Conditional: easy come, easy go (idiom: monkey see, monkey do)—consists of CP with CP adjunct clause, each CP has specified material within it

Some of patterns 3–9 involve functional elements in addition to lexically specified elements, bound variables, and free variables.

I will now go through some of the patterns above in more detail.10

6.5 More Detail: Ditransitives

Because ditransitives have been a substantial part of the motivation for the analysis here, I will go into more detail about how I analyze them. I will also explain the relation of phase-command that figured in the formulation of the CFV.

I will adopt the structure for double object constructions proposed in Bruening (2001, 2010). In this analysis, the first object is projected in the specifier of an Appl(icative) Phrase between Voice and VP (this analysis modifies a proposal in Marantz 1993):

---

10 A reviewer asks about the expression the X-er, the Y-er (Fillmore et al. 1988). This expression comes in two forms: it can have two variables which can be filled in extemporaneously (e.g., the more you eat, the hungrier you get), and it also occurs in several fixed expressions (e.g., the more, the merrier; the bigger they come, the harder they fall). These fit into the theory: in the fixed expressions, all elements of the constituent are specified for lexical content; in the variable version, there are two free variables, and neither is a case competitor for the other. I have not included this expression in the list because it is not clear to me how to classify it. Also unclear is long time, no see, which consists of two fully specified clauses in juxtaposition.
This structure underlies all double object constructions with a uniform caused-possession semantics, whether the corresponding prepositional construction would have to or for. See Bruening (2018).

The common conventionalized expression pattern $V X NP$, like $\lambda x. x \text{ cut } y_{NP} \text{ some slack}$ in this example, obeys all constraints on conventionalized expressions. It consists of an XP—VoiceP—and all elements within it are either specified for lexical content, are bound variables, free variables, or functional material. The free variable in the specifier of ApplP obeys the CFV, since the subject in Spec-VoiceP is a bound variable. The second object does not precede and phase-command Spec-ApplP. An expression with both NPs fixed, like $\lambda x. x \text{ give the devil his due}$, also obeys the CFV, since it has no free variables.

The missing pattern, *$V NP X$, in contrast, violates the CFV:

Such an expression would consist of a VoiceP, *$\lambda x. x \text{ throw the wolves } y_{NP}$. The free variable violates the CFV, because it has a case competitor that precedes and phase-commands it, the NP in Spec-ApplP. This is why there are no conventionalized expressions of this form, and why no conventionalized expression alternates with this form ($\text{throw } X \text{ to the wolves, } *\text{throw the wolves } X$).

I will adopt the structure in (124) for the prepositional dative construction. In this structure, the verb selects both the direct object and the PP. This structure is most consistent with constituency tests, since some speakers permit the verb and direct object to act as a constituent, excluding the PP (see Janke & Neeleman 2012). This structure has such a constituent, labeled $V$ here (but nothing hinges on this label, it could as well be VP).
The structure in (124) is also consistent with the binding facts discussed in Barss & Lasnik (1986), given that binding depends not on c-command but on precede-and-command (Bruening 2014a). Precede-and-command is the conjunction of two structural relations, precedence and phase-command. I repeat the definition of phase-command from above:

(125) a. Binding: A binds B iff A and B are coindexed and A precedes and phase-commands B. (Bruening 2014a: 344, (5))

b. Phase-Command: X phase-commands Y iff there is no ZP, ZP a phasal node, such that ZP dominates X but does not dominate Y. (Bruening 2014a: 343, (2))

c. Phasal Nodes: CP, VoiceP, NP (modified from Bruening 2014a: 343, (3))

In the structure in (124), the NP object precedes and phase-commands the object of the P, because there is no phasal node that dominates it that does not also dominate the object of the P. The first phasal node that dominates the NP object is the VoiceP node, but that also dominates the object of the P. The object of the P also phase-commands the NP object, but it does not precede it, so the NP object may bind the object of the P but not vice versa (give them to each other versus *give each other to them). See Bruening (2014a) for extensive discussion and arguments that precede-and-command is the correct structural notion for binding and c-command is not. Here, note that using precede-and-command rather than c-command makes the binding and constituency data consistent with each other, whereas with c-command binding and constituency conflict (see Pesetsky 1995 and Bruening 2014a).

Note also that the structure in (124) is the structure I assume for all V NP PP frames, regardless of the semantics of the PP. Recall that Hallman (2015) and Larson (2017) claimed that it was important to distinguish caused possession PPs from other types of PPs; this is not important here, and I see no reason to make any structural distinction between PPs with different semantics.

Hallman (2015) claims that there is a syntactic difference between caused possession PPs and locational PPs in their ability to control into non-subject-gap purpose clauses, but this is false. Locational PPs can control into purpose clauses (these examples are also from Larson Stromdahl for discussion and examples):

(i) a. I put it beside her [PRO to look over when she got the chance].

b. I set it in front of her [PRO to sign].

c. (What did you do with the microprocessor?) We inserted it in him [PRO to track himself with].

Hallman also claims that the purported difference in the ability to control is due to a difference in c-command. However, c-command is not required at all for non-subject-gap purpose clauses, as Whelpton (2002) showed. The possessor of the object of a locational P can also control into a purpose clause (these examples are also from Larson Stromdahl):

(ii) a. I put it in her mailbox [PRO to look over when she returned from the Bahamas].
In the structure in (124), it is possible to have conventionalized expressions of both the form V X PP (throw X to the wolves) and V NP P X (give rise to X), because the object of the P is not a case competitor for the object of the V. It is also possible to have idioms consisting of the verb and both of its arguments, as in give hostages to fortune. Not only that, is also possible for some ditransitive expressions with PPs to alternate with the double object frame. In my judgment, some of the ones that do are the following:

(126) a. collocations: do X a favor, give X advice, give X credit, give X meaning, give X some guidance, give X some thought
b. idioms: give X a free hand, lend X color, Y lend X Y’s ear, NEG pay X any mind, read X the riot act

I analyze alternating expressions as consisting solely of a VP with a specified V and specified direct object, and a free variable for a PP argument. In the semantics of the double object construction spelled out in [Bruening (2010)], the PP argument gets unified with the argument of the Appl head in the double object construction. Such a conventionalized VP can therefore appear either by itself (i.e., embedded under Voice), where the PP free variable gets filled in, or as the complement to Appl, where the PP argument is saturated in Spec-ApplP.

As this subsection has shown, the ApplP analysis of ditransitives in Bruening (2001, 2010), combined with the analysis proposed here for conventionalized expressions, succeeds in ruling out the missing *V NP X pattern and in accounting for the patterns of conventionalized expressions in ditransitives generally. In future work, I will show that the ApplP analysis of ditransitives is the only one that is successful.

### 6.6 More Detail: Multiple Clauses

In section 5.1, I presented examples of very long conventionalized expressions, consisting of multiple clauses. Here I go through some of them in more detail, including some examples from German from Richter & Sailer (2009).

It is striking that such long expressions have very few free variables. They instead have lexically specified elements or bound variables. In the case of non-finite embedded clauses, the embedded subject is often a bound variable, typically obligatorily controlled PRO, as in (127). Alternatively, PRO can be arbitrary, in which case it is like a free variable (128):

(127) a. λx.x have an axe [PROx to grind] (‘have a (quarrelsome) reason’)
   b. λx.x cut off x’s nose [PROx to spite x’s face] (‘do something self-destructive out of revenge’)
   c. λx.x teach x’s grandmother [PROx to suck eggs] (‘offer advice to someone with more experience’)

(128) a. a bitter pill [PROx to swallow] (‘is unpleasant but must be accepted’)
   b. a tough row [PROx to hoe] (‘is very difficult’)
   c. λx.x play hard [PROy to get] (‘be coy’)

In the cases with free variable PRO in (128), there is no other free variable, and the free variable has no case competitor. These multi-clause examples therefore follow the patterns we have seen for all conventionalized
expressions. (Note that free variable PRO can be expressed, as for instance *that would be a bitter pill for anyone to swallow.*) There is one expression with a free variable in an embedded non-finite clause in addition to a bound PRO:

\[(129) \quad \lambda x.x \text{ NEG know what } \text{PRO}_x \text{ to do with } y_{\text{NP}} \]

This expression still obeys the CFV, as the free variable has no case competitor that is not a bound variable. Expressions that include an embedded finite clause are similarly constrained. The embedded clauses are often entirely fixed, like those in (130), or the only variables they have are bound variables, as in (131)\[39\]

\[(130) \quad \begin{array}{l}
\text{a. } \lambda x.x \text{ kill the goose } [\text{that lays the golden eggs}] ('\text{destroy a profitable thing out of greed}') \\
\text{b. } \lambda x.x \text{ strike } [\text{while the iron is hot}] ('\text{act before it is too late}') \\
\text{c. } \lambda x.x \text{ close the stable door } [\text{after the horse has bolted}] ('\text{act after it’s too late}') \\
\end{array} \]

\[(131) \quad \begin{array}{l}
\text{a. } \lambda x.x \text{ bite the hand } [\text{that feeds x}] ('\text{turn against someone you depend on}') \\
\text{b. } \lambda x.x (\text{NEG}) \text{ count } x\text{'s chickens}_y [\text{before they}_y \text{ hatch}] ('\text{plan on something that might not happen}') \\
\text{c. } \lambda x.x \text{ cross that bridge}_y [\text{when x come to it}_y] ('\text{wait to act until it is necessary}') \\
\text{d. } \lambda x.x \text{ bite off more } [\text{than x can chew}] ('\text{take on more than one can handle}') \\
\text{e. } \lambda x.x \text{ kick y } [\text{when/while y is down}] ('\text{treat someone badly who has already suffered a setback}') \\
\end{array} \]

Again, this follows the pattern we have seen generally: there are no constraints on how large conventionalized expressions can be, but there are constraints on how many free variables they can contain. Richter & Sailer (2009) discuss some multi-clausal expressions in German. As in English, they are quite limited, and generally have bound variables rather than free variables. There are expressions where the whole embedded clause is specified, like (132a), and ones that include a bound variable, like (132b–c). Only one of the ones they discuss has a free variable, and there is only one (132d):

\[(132) \quad \begin{array}{l}
\text{(Richter & Sailer 2009: 299, (1a–d))} \\
\text{a. } \lambda x.x \text{ wissen, wo Barthel den Most holt} \\
\quad \quad \text{know where Barthel the young.wine gets} \\
\quad \quad \text{‘X knows every trick in the book’} \\
\text{b. } \lambda x.x (\text{NEG}) \text{ wissen, wo x(Dat) der Kopf steht} \\
\quad \quad \text{(NEG) know where x the.Nom head stands} \\
\quad \quad \text{‘X has a lot of stress’} \\
\text{c. } \lambda x.x \text{ glauben, x(Acc) tritt ein Pferd} \\
\quad \quad \text{believe x kicks a.Nom horse} \\
\quad \quad \text{‘X is very surprised’} \\
\text{d. } \lambda x.x \text{ wissen, wo y(Acc/Dat) der Schuh drückt} \\
\quad \quad \text{know where y the.Nom shoe presses} \\
\quad \quad \text{‘X knows what is worrying Y’} \\
\end{array} \]

\[\text{[An interesting case is as best X can, which is an adjunct clause. It attaches to a matrix clause, the subject of which binds the embedded subject X. It also has VP ellipsis in the embedded clause, where the missing VP is understood as having the matrix VP as its antecedent (e.g., *she will have to write the proposal as best she can*). Hence every variable is bound.} \]

39
It therefore appears that patterns of conventionalized expressions are very similar across languages, and obey the same constraints. They can be quite long, but there are strict limits on free variables within them.

Note that the fact that phrases embedded in such conventionalized expressions can be variables bound by phrases higher in the expression follows from the view of conventionalized expressions proposed here. Bound variables and free variables are as much a part of conventionalized expressions as lexically specified material. Such apparently open slots embedded deeply within an otherwise fixed expression are extremely problematic for all previous proposals, which try to account for sequences of fixed material minus the open slots. Example (132) with a completely open slot as the embedded object is particularly difficult for previous proposals; see immediately below for my analysis.

### 6.7 Expressions with Fixed Subjects

So far in this section, I have simply assumed that no conventionalized expressions have a fixed subject with an open slot for an object. I showed above that this follows from the CFV, since a free variable as an object is not allowed to have a case competitor that precedes and phase-commands it, which it would if the subject were anything besides a bound variable. Now it is time to examine whether this assumption was warranted.

The original observation of a subject-object asymmetry in idioms seems to come from Marantz (1984). The claimed asymmetry has been discussed in great detail in the literature, and has sometimes been denied to be true. This literature includes Bresnan (1982), Nunberg et al. (1994), Weibelhuth & Ackerman (1994), Everaert & Kuiper (1997), O’Grady (1998), Horvath & Siloni (2002), Bruening (2010), Harley & Stone (2013). A few of the counterexamples that have been cited in the literature do not seem to be real counterexamples, as Bruening (2010) and Harley & Stone (2013) discuss, among others. For instance, *a little bird told X Y* (Nunberg et al. 1994: 526) does not actually seem to include anything other than the subject *a little bird* in the expression. This expression is just a noun phrase with the meaning ‘an anonymous source of information’. This NP can occur with other verbs of communication, as in *a little bird is broadcasting that, a little bird whispered that to me, I heard it from a little bird*, or even with no verb of communication, as in, *Had Varys’s little birds failed him for once?* (George R.R. Martin, *A Clash of Kings*). Similarly, *the lovebug bit X* (Postal 2004: 255–256) is not a conventionalized expression with a fixed subject and verb, as the NP *the lovebug* can occur in many different expressions: *the lovebug has struck Bill, Bill has the lovebug* (and *lovebug* can also vary, with other things replacing *love*). *The devil drove X to it/Y also does not seem to include the verb as part of the expression, as it freely varies: the devil made him do it, the devil forced him to do it, the devil got into him.*

There are a few potentially real counterexamples, below (all from Bresnan 1982: 350):

(133) a. the bottom fell out of X (‘X collapsed’)
    b. the roof caved in on X (‘disaster struck X’)
    c. something/what is eating X (‘something/what is really bothering X’)

To these we can add the following collocations:

(134) a. Whatever happened to X?
    b. What would we do without X?
    c. What makes X tick?
    d. Who asked X? Nobody asked X!

The first thing to notice is that four of them have the open slot as the object of a preposition (133a–b, 134a–b). This is allowed by the CFV, since the free variable is not a case competitor for the subject. The only problematic examples are those in (133c) and (134d–d), which seem to have both a subject and a direct
object. The subject appears to be a case competitor for the object, because it makes the object accusative (What’s eating him?). I suggest that it is significant that the subject is a quantifier in all of these examples. The conventionalized expression is a larger XP with the quantifier displaced to an A-bar position, binding a variable in argument position. This makes the NP in subject position that acts as a case competitor for the direct object a bound variable (∃x.x is eating yNP). This satisfies the CFV. One could instead claim, as Harley & Stone (2013), that this particular expression is a psychological predicate and so not agentive, but this does not seem to be correct for the expressions in (134c–d). It appears that what is really significant about such expressions is that they always have a wh-phrase or a quantifier as a subject.

Some counterexamples have also been presented from other languages, for instance by Reis (1982), Sternefeld (1985), Webelhuth & Ackerman (1994) in German, Horvath & Siloni (2002) in Hebrew, Kiss (1987) in Hungarian, Ruwet (1991) in French, and Chtareva (2004) in Russian. Harley & Stone (2013) discuss many of these examples and show that they all involve a non-standard syntax. That is, the syntax of these expressions differs from a typical transitive clause. Harley and Stone interpret this as evidence that the expressions do not have agentive subjects (recall that they proposed, incorrectly, that idioms may not include agents). The current proposal provides a different perspective on this. It may be that such expressions have an unusual syntax for one of two reasons. The first reason is that they might involve non-standard ways of assigning case. If the subject and the object are not case competitors, then the CFV will permit a fixed subject and an open object. So, such expressions would be allowed if one of the subject or object could get non-structural case, making it not a case competitor for the other. This would result in a very different syntax from a typical transitive clause (for instance, it would not be passivizable, as Harley & Stone 2013 show to be true of many such expressions).

The second way the current proposal could allow a conventionalized expression with a fixed subject and an open object is if the language has an operation that can reverse the precede-and-command relation between the subject and the object. This is what I suggest is going on in German. In expressions like X(Acc) reitet der Teufel, ‘X rides the.Nom devil’ (‘X is going crazy’), the verb can be passivized (Harley & Stone 2013) and therefore appears to be agentive and to involve standard structural case. Discussions with native speakers of German indicate instead that the word order is significant: the open slot generally has to come first. I suggest that German, unlike English, has an operation (e.g., scrambling) that can move an object across the subject, creating an open predicate with a variable for the object: λx.the devil ride x. This is now allowed by the CFV, because the object is no longer a free variable that is preceded and phase-commanded by a case competitor. It instead precedes and phase-commands its case competitor, and it is also a bound variable, not a free variable. (I suggest the same analysis for the free variable in the embedded clause of example (132d) above; it is a free variable, not a bound variable, but it is not preceded and phase-commanded by its case competitor.)

Hence, if a language has ways of either assigning non-structural case or of inverting precede-and-command relations between subject and object, then it will have the means to build conventionalized expressions that have a fixed subject and an open object.

Note that, while English does not have an inversion operation like scrambling, it does have the passive. There are conventionalized expressions in English that only occur in the passive, like given time X has been had (‘X has been tricked’), and X is caught short (‘X is found to be lacking something’; see Bruening 2014b on these, and other putative examples). The passive operation results in an open predicate with a variable for the object. These expressions happen not to have logical external arguments at all, but they do illustrate the existence of the operation.

---

14 This expression must actually be a full main clause with an embedded clause, since the subject of the passive is a variable bound by the subject of the matrix clause: [λx.[PROx given time] xNP yZP]. It then has a free variable for some kind of clausal phrase, one that can be a VP up to a ModalP (given time, they would have caught the discrepancy).
6.8 Ditransitives Cross-Linguistically

Ditransitive idioms have been investigated quite systematically in several languages. Miyagawa & Tsujioka (2004) present data on ditransitive idioms in Japanese. Kim (2015) and Lee (2017) examine ditransitive idioms in Korean. Mishani-Uval & Siloni (2017) investigate ditransitive idioms in Hebrew. All of this literature is careful to distinguish a double object construction from a prepositional dative construction, even though the two often both involve one NP with dative case and one NP with accusative case. Once this distinction is made, the patterns are strikingly similar to English: the missing *V NP X pattern is missing in all three languages, but in the prepositional dative, either argument can be an open slot.

This finding has several implications. First, it confirms the CFV as a universal constraint on conventionalized expressions. Second, it requires that dative case be viewed as a structural case in a ditransitive, with the NP that receives dative case acting as a case competitor for the NP that receives accusative. Such a theory of dative case is presented in Baker & Vinokurova (2010), Baker (2014). Third, it appears that languages do not have ways of getting around the CFV with ditransitives the way they do with external arguments, since there are no exceptions. Once the structure is set as a double object construction, there is no way to assign case differently or to invert the lower object over the higher. The only option appears to be to base-generate a completely different structure, involving only one NP getting structural case while the other is a PP.

6.9 More Detail: Negation

We saw in section 3 that many conventionalized expressions include functional clausal elements within them, like aspect, modals, and imperative mood. A particularly common pattern is for negation to be included. A few examples of expressions that obligatorily include negation are repeated below:

(135) a. NEG last long, NEG take long, NEG bet on X  
   b. NEG breathe a word about X, NEG give a crap/hoop, NEG have a pot to pess in, NEG have a leg to stand on, (can/could) NEG help it/Xself (“unable to stop oneself”), NEG X know where X’s head is (at), NEG leave stone unturned, NEG see the forest for the trees, NEG stop at anything

Negation is particularly interesting because it can be spelled out in different ways. It is often spelled out as the adverb not adjoined to the highest VP of the clause. However, it can also be spelled out as the determiner no on an NP that is otherwise a fixed part of the expression. For instance, NEG leave stone unturned appears robustly as both leave no stone unturned and as NEG leave any stone unturned (NEG can even appear as a determiner in a non-fixed part of the expression, as for instance, No one is to leave any stone unturned!). The same is true for various other expressions with negation and an existential (e.g., stop at nothing ~ not stop at anything).

One analysis of this is to posit a clausal, abstract negation, NEG. Obligatorily negative conventionalized expressions would then consist of whatever XP includes this abstract functional NEG (perhaps a NegP, perhaps some other category with a NEG feature). This abstract negation is the actual semantic negation, but its presence needs to be signaled in some way. It can either be signaled by the adverb not, or by a negative determiner. Negative determiners on this view are not themselves negative, they are only existential but they are specified as having to occur in the domain of the abstract NEG (see, e.g., Penka 2012).
A second option is to view obligatorily negative idioms as negative polarity items. Then NEG would not necessarily be part of the idiom, but the idiom would be specified as having to occur in the scope of negation (or another downward-entailing operator). This is already necessary for certain NP expressions, like a living soul, which can occur with different verbs (don’t tell a living soul, I didn’t see a living soul) but must occur with negation. Negative verbs also license this expression: I doubt that there’s a living soul out there... as do conditionals (if there’s a living soul out there...), so sentential negation is not a crucial component of this idiom.

I will not try to decide between these two analyses here, or others, but merely point to some of the options that are available within the general approach to conventionalized expressions proposed here.  

6.10 Not Entirely Specified Material

The final remark I would like to make before leaving this section concerns substitutability. In section 2, I showed that many conventionalized expressions, including many non-literal ones, permit some substitution for material that is lexically specified. For instance, flag a dead horse is also used as beat a dead horse; clutch at straws is also attested as grasp at straws and seize at straws. There are also “families of idioms” like beat the crap out of. Here the verb can be one of several verbs like beat and kick, and the noun can also vary within a limited range, as crap, shit, tar, etc. Experimental work has also shown that speakers permit some substitution even with expressions that are not attested with that variation (McGlone et al. 1994). And of course, many completely literal expressions admit substitution quite readily.

This is an issue for a psycholinguistic theory of conventionalized expressions, not a theory like the present one that is only concerned with the syntax of such expressions. I have proposed that conventionalized expressions are XPs consisting of certain types of constituents, including constituents the lexical content of which is specified with the expression. I assume that the entire XP is stored as a unit, but other conceptions are possible as well. The issue that substitution raises is the issue of how such expressions are stored, or in what form, and what it means for lexical content to be specified. Apparently a set can be specified in addition to a single lexical item, or even just an abstract concept that will then be compatible with a small set of lexical items that match that concept (which is probably what is going on in the case of clutch/grasp/seize at straws). Even when an expression is stored with only a single lexical item, speakers will under some conditions permit a related one to take its place. I view this as something that is completely compatible with the current theory of the syntax of conventionalized expressions, and somewhat orthogonal to its concerns (although ultimately important).

6.11 Summary

In sections 3 and 4, we saw that collocations and idioms both seem to be subject to the same structural constraints. In this section, I have proposed a new theory of the syntax of all conventionalized expressions, both literal collocations and non-literal idioms. All conventionalized expressions consist of an XP constituent, the contents of which can be four different things: functional elements, lexically specified phrases, bound variables, and free variables. All but the first must be complete phrases, and this accounts for many missing patterns. There are in addition constraints on free variables: they are limited in number, and a free variable may not have a case competitor that precedes and phase-commands it that is not a bound variable (the CFV). This permits open slots as possessors and objects of prepositions (and certain other instances, like VPs and

---

16Andrew Murphy (p.c.) points out that obligatorily negative conventionalized expressions can occur with NEG raising, as in I don’t think he has a pot to piss in. If NEG is a specified part of the expression, then this would necessitate a movement derivation of NEG raising, as in Collins & Postal (2014). If negative conventionalized expressions are instead negative polarity items, then it would not.
CPs), but rules out the missing patterns, *NP V X and *V NP X. Other approaches to the syntax of conventionalized expressions fail because they do not view the open slots as part of the expressions. The current proposal also permits a new perspective on the counterexamples to *NP V X that have been noted in several different languages.

7 Conclusion and Speculation

The detailed empirical study of collocations and idioms undertaken here has shown that they appear in almost identical syntactic patterns. Contra Larson (2017), idioms and collocations behave the same. There are structural constraints on both idioms and collocations, and they are the same constraints. In particular, two patterns are systematically missing from both collocations and idioms: a pattern with a fixed subject and open object (with a few principled exceptions), and the missing *V NP X double object pattern (no exceptions). Also missing are open slots that are not full phrases. I have proposed a new way of looking at conventionalized expressions, where they include the open slots within them as free variables. Strict constraints on free variables explain why the missing patterns do not occur.

The next question is why conventionalized expressions are constrained by the CFV rather than in some other way. That there would be upper bounds on the number of variables makes perfect sense; after all, few (if any) verbs or other items take more than two open arguments (which can be viewed as free variables, too). Although I have not remarked upon it so far, it is clear that conventionalized expressions do not permit very many bound variables, either. Typically there is only one binder, and at most two (e.g., λx.x (NEG) count x’s chickens, before they, hatch). But why would free variables be constrained in exactly the way that the CFV constrains them, referring to case competitors?

At this point I can only speculate. Given that the number of free variables is constrained, we might suppose that such variables are computationally costly. This makes sense, since the grammar/processor has to do more work with a conventionalized expression that includes free variables. The grammar/processor has to combine a conventionalized expression XP with a larger structure in order to use it. If all of that XP is fixed and invariant, this will be easy. If there are open slots then the grammar/processor will have to construct further XPs of the appropriate type and insert them in the open slots. It would also make sense for case competition to be computationally costly: not only does the grammar have to put together a phrase with two (or more) NPs, it has to keep those NPs in working memory and evaluate them against each other in order to assign case. It would then make a lot of sense for two computationally costly things to cross a threshold when they combine. In the case of possessors and objects of prepositions, all the grammar/processor has to do is build and insert an appropriate XP. In the case of a subject or direct/indirect object, however, it not only has to construct and insert that XP, it has to keep it in working memory and evaluate it against other NPs in order to assign structural case.

Of course, spelling out such an analysis and explaining the directionality present within it (precede-and-command) will require much more detail than I can go into here. It is also possible that the motivation behind the constraint is something else entirely. I will leave explaining it to future research, and content myself here with having identified the patterns and modeled them within an analysis. The main point of this paper was to show that idioms and collocations do not differ and should not be distinguished, and to show that there are structural constraints on the class of conventionalized expressions that includes them.

Bibliography


Department of Linguistics and Cognitive Science
University of Delaware
Newark, DE 19716
(302) 831-4096
bruening@udel.edu