Seeking the Nature of Idioms:
A Study in Idiomatic Structure

Tim Ifill
Haverford College
December 10, 2002

1.0 Introduction

Idioms are odd ducks[1]. They seem to act very much like normal language, but they are

quite different in many ways. It's been said that "If natural language had been designed

by a logician, idioms would not exist" (P. N. Johnson-Laird in a foreword to a collection

of works on idioms). But exist they do, and not only that, they represent a rich vein in

language behavior, so they cry out for explanation. Though they have been dismissed by

many theorists to various extents, some have seen the need for theories of language to

adequately explain the behavior of idioms, and, moreover, the potential for what analysis

of idioms can tell us about language itself.

   Put as simply as possible, an idiom is a fixed expression whose meaning can not

be taken as a combination of the meanings of its component parts. Thus, the common

phrase *kick the bucket* has nothing to do with either kicking or buckets, but means simply,

"to die." In other words, idioms are not literal expressions. They are also, as mentioned,

fixed expressions, to the extent that the elements which make up the idioms are limited in

the kinds of variability they are able to demonstrate. Idioms can also be found

comprising nearly any kind of syntactic phrase, right up to a full sentence (DiSciullo

1987), and it has been widely noted that they tend to exhibit similar syntax to non-

idiomatic phrases (Van Gestel 1995, Fellbaum 1993, Abeillé 1995)[2].

   But where do idioms come from, and what kind of structure do they have, if any?

This study will attempt to provide insight into those questions. I intend to argue, first of

---

[1] I would like to thank my advisor Ted Fernald, as well as my second reader Kari Swingle, both of whom
provided valuable input. Sarah Lindholm and Rebecca Weinberger were likewise extremely helpful with
their suggestions. I would also like to thank Maddy Yovanoff for lending logistical support and Katie
"Kunger" Unger for her encouragement and for unwittingly providing inspiration for this study of "hokey
expressions." All errors remain my own.
[2] For a thorough explanation and categorization of the kinds of syntactic structure idioms can have, see
Makkai 1972.

all, that idioms are a very important aspect of language, and that they should be included in the lexicon. I will then argue that idioms are not frozen structureless atoms –they do have significant internal structure, which is responsible for much of their behavior. This structure consists of a relationship between the syntax of the idiomatic phrase and the idiom's literal meaning (which will then give us a division into two types of idioms) as well as a continuity chain of head to dependent relations within idioms.

2.0 The Importance of Idioms and the Case for the Lexicon

Ray Jackendoff (1997) proposed an interesting argument for the importance of fixed expressions in natural language in the form of his *Wheel of Fortune* Corpus, a compendium of over six hundred solutions to the word puzzles on the popular game show. Roughly categorized, the solutions fall into the following groups: compounds, such as

    (1) black and white film
          frequent flyer program

idioms,

    (2) a breath of fresh air
          they've had their ups and downs

proper names,

    (3) a. Clint Eastwood
          b. Boston, Massachusetts
          c. John Deere tractor
          d. Boston Pops Orchestra

clichés,

    (4) any friend of yours is a friend of mine
          gimme a break

titles of songs, books, etc.,

    (5) City Lights
          The Price Is Right

brief famous quotations,

    (6) beam me up, Scotty
          a day that will live in infamy

and foreign phrases.

    (7) au contraire
          persona non grata

There are six hundred such fixed expressions in the Corpus. In order for the game show to work, all of these must be readily available to nearly any American speaker of English, and there must be a vast number of such expressions out there in order to have enough fodder for the shows to avoid repetition. It is the sheer size of this group which Jackendoff finds most intriguing. Jackendoff (1997) cites one estimate of 25,000 for the total number of such fixed expressions in English (there is a similar estimate for French), which is the same order of magnitude as the number of individual words in the lexicon. Given these facts – not to mention the very fundamental fact that these forms are made up of linguistic components and common components themselves in larger pieces of linguistic discourse – it would be shortsighted not to include serious analysis of the nature of idioms, being an important category of fixed expressions, in any theory of our knowledge of language. In other words, we can't simply right them off as rule-breaking anomalies of little significance.

So we know that idioms are worth studying, but where do they come from? It turns out that there is ample reason to believe that idioms are stored in the lexicon. To begin with, let us look at the ideas of the lexicon put forward by DiSciullo and Williams

in *On the Definition of Word* (1987). For them, the criteria for being part of the lexicon is listedness, and objects that are listed have been dubbed listemes by the authors. In order for an object to be listed in the lexicon, it must have a meaning which can not be computed compositionally, thus its meaning must be "memorized" (their quotation marks). They make this claim starting from a fairly simple premise:

> Knowledge of language involves in some way a knowledge of particular linguistic objects – for example, the word *transmission* and the knowledge that it (1) has a certain morphological form and (2) refers to a part of a car; that *take to task* has a certain syntactic form and means "rebuke". To the extent that an object does not have the form or interpretation specified by the recursive definitions of the objects of the language, that object and its properties must be "memorized." . . . . Our overall point is that listedness is no more intrinsically characteristic of words than it is of phrases. Some words and some phrases are listed, but infinitely many of each are not.

Thus according to this view, there's really nothing special about idioms in terms of the lexicon. Some older views of idioms may have accepted them as being part of the lexicon, but they were odd exceptions – anomalies that were not words, but were reluctantly included with them in the lexicon. To DiSciullo and Williams, "there is nothing more to say about them [idioms] than that (1) they are syntactic objects and (2) they are listed because of their failure to have a predictable property (usually their meaning)."

This assessment of idioms as listemes makes sense in light of Jackendoff's (1997) assertions of the importance of idioms and other fixed expressions in terms of our knowledge of language. (DiSciullo and Williams (1987) also acknowledge this fact, pointing to "the great wealth of such expressions" as evidence that they're "nothing special.") Jackendoff (1997) as well as Van Gestel (1995) agrees with the placement of idioms in the lexicon, and adopts their term *listeme*.   Their commonality, abundance, and

availability, as evidenced by the *Wheel of Fortune* corpus, combined with the fact that they are composed of linguistic levels of representation and seamlessly integrate into normal sentences, makes it illogical that they are part of some "general-purpose" memory – they must be lexical.

3.0 Fixedness of Idioms and Internal Structure

There may be nothing special about idioms in terms of their listedness, but that does not mean there is nothing special about them at all, and our analysis of them does not simply end here by accepting them as listemes. One consequence of admitting idioms to the lexicon is that it permits them to have internal structure of some kind. If an idiom is a phrasal listeme, it represents a linguistic unit – it will have internal linguistic structure: syntax, semantics, morphology, phonology (Jackendoff 1997). This makes sense – a phrasal listeme is by definition a phrase, so it will have the same kind of linguistic makeup that a phrase has. The task now is to show that they must have internal structure, the existence of which can be used to explain idiomatic behavior. We will see that idioms are fixed phrasal expressions, but they are not completely frozen forms. Examining just how fixed they are, and in what ways, is the most direct method of refuting the notion that idioms are atomic units which lack internal structure. And ideally, this kind of examination will shed light on just what any internal structure to idioms might look like.

3.1 Minimal Idiomatic Variability

Let's begin with the most minimal way in which an idiom can be altered from its base

form (i.e. the actual listeme itself, which is stored in the lexicon): morphology.

(8) a. We will take them to task for their irresponsibility.
     b. We are taking them to task for their irresponsibility.
     c. We took them to task for their irresponsibility.
     d. We have taken them to task for their irresponsibility.
(9) a. John and Sue have their ups and downs.
     b. John and Sue are having their ups and downs.
     c. John and Sue had their ups and downs.

In these example sets, we will take *take* NP *to task* and *have one's ups and downs* to be

the listed forms of the idioms in (8) and (9), respectively[3]. We can already see from

looking at the verb forms that these phrasal listemes are not completely frozen in form.

These differences in the conjugation of the verbs may seem minor at first, but they are the

first key piece of evidence which indicates that these idioms must have an internal

structure of some kind. In other words, it can not be the case that the idiom *take to task* is

a completely frozen atomic unit of language, free of internal structure, somewhere in the

memory (either within the lexicon or elsewhere) which then gets inserted wholesale into

a sentence. If this were so, there would be no way for the individual verb *take* to accept

the proper tense. Indeed, if this were so, *take* would not even exist as an individual verb,

let alone be able to undergo such modification. This represents the minimal evidence

indicating internal structure to idioms (Jackendoff 1997). Makkai (1972) worked with a

model of idioms which described levels of frozenness, but even those idioms which were

classified as "completely frozen" exhibit this kind of behavior (*trip the light fantastic* vs.

tripping the light fantastic vs. *tripped the light fantastic*).

3.2 Synonymy and Idioms

---

[3] The open NP in *take to task* and the possessive in *have ups and downs* will be dealt with in a later section.

It has been widely noted that the individual words in an idiom can not be replaced by synonyms and still retain the idiomatic reading of the phrase. This is what qualifies them as fixed forms. In most non-idiomatic discourse, a speaker can use synonymy to create a new sentence with the same semantic meaning. This is not the case for idioms. Thus the sentences below indicate how the individual words of an idiom, not just their normal semantic denotation, are part of the fixed form of such idioms:

(10)   a. John kicked the bucket.
         b. #John kicked the pail[4].
(11)   a. They have had their ups and downs.
         b. They have had their differentials in elevation.

It's not just a matter of the concept of dying being associated with knocking over small open-topped water vessels with one's foot, but the words *kick* and *bucket* are also required to convey that particular sense.

I would argue that this fixed quality of idioms is analogous to individual morphemes in words being non-interchangeable. For instance, in the word *hardness,* the morpheme *–ness*, meaning roughly "the quality of being", which is roughly the same meaning as the morpheme *–ity* has, but the two are certainly not interchangeable, as demonstrated by the non-existence of the word *hardity*. DiSciullo and Williams (1987) refer to this process as "blocking," where the existence of one word with a certain form keeps another from being used in its place (as in, *hardness* blocks the generation of *hardity*). If phrasal idioms are truly lexical, then perhaps it is this same mechanism that keeps them fixed (Zeevat 1995). This aspect of idioms may seem to support the notion that idioms are frozen atomic linguistic units, but it does not necessarily do so. The fact that the words of the idioms are fixed is what makes them idioms, first of all. And if

---

[4] I have adopted the convention used by several writers including Gazdar at al. (1985) of using # to indicate a sentence which, while essentially grammatical, does not have a possible idiomatic reading.

idioms are really listemes, then it makes sense that they will have an analogous structure, in terms of fixedness, to their companions in the lexicon, individual words. And just as words can have internal structure, so can idioms.

Before we close the book on synonymy and idioms, however, it is necessary to take a look at a few exceptions to this kind of fixed behavior. Some idioms, it seems, can accommodate interchangeable synonyms (or closely related words with different senses), for instance these data offered by Gazdar et al. (1985:239):

(12)   hit the sack/hay
(13)   pack a punch/wallop
(14)   get off one's ass/butt/rear/etc.
(15)   stretch/strain a point
(16)   stop/turn on a dime[5]
(17)   pick/punch/poke/shoot holes in an argument
(18)   lay/throw/place/put one's cards on the table

These idioms, however, do not tolerate unlimited interchangeability:

(13)   b. #*pack a blow
(18)   b. #toss one's cards on the table

Thus these forms are still largely fixed, yet they demonstrate a capacity for some idioms to show individual lexical variability. In a few idioms, therefore, there can be some lexical variability, but that variability is limited to only a few options. All of the options for variability are semantically similar, but then not all semantically similar words can fit. To me, this also points to some kind of internal structure, if the individual parts of idioms are allowed to behave as individuals. Moreover, these examples also indicate that multiple levels of linguistic representation are at work within the idioms. There is the

---

[5] It has been suggested that the verbs are not a part of this idiom, but merely the "on a dime" portion is idiomatic – after all, there is an actual turning or stopping event – though Gazdar does include the verbs in his list.

phonological level of the actual words used, but then the exceptional idioms in (12) through (18) indicate that the semantic level also plays a role.

Thus the fixedness of idioms is a quality which not only characterizes them, but can fit in with a view of idioms as internally structured lexical items.

3.3 Passivization and a Division in Idiomatic Structure

One thing that is readily noticeable about idioms is that many seem to resist undergoing transformations that similar non-idiomatic constructions can readily undergo while retaining the same sense. For instance:

(19)   a. John kicked the bucket.
         b. # The bucket was kicked by John.

(20)   a. They have had their ups and downs.
         b. # Ups and downs were had by them.

Were sentence (19a) to have a non-idiomatic reading – that is, if it described an actual bucket-kicking event performed by John – then passivization of the sentence would lead to a new sentence with the same sense (that is, with a non-idiomatic reading (19a) and (19b) both mean the same thing). But this does not happen with the idiomatic reading – sentence (19b) does not mean that John died. *Have one's ups and downs* works the same way. A passive version of the idiomatic verb phrase just doesn't make sense idiomatically.

This might at first seem to be an argument for the frozenness of idioms, but the story doesn't end here, for there are many idioms which are not frozen in this way, and are able to undergo passivization. For instance:

(19)   a. Roger kept tabs on them.
         b. Tabs were kept on them (by Roger).

(20)   a. I spilled the beans.
         b. The beans were spilled (by me).
(21)   a. He laid his cards on the table.
         b. His cards were laid on the table.

In these cases, the passive versions of the verb phrases do retain their idiomatic status,

and the (a) and (b) pairs of passive and active sentences can be said to have the same

sense.

It seems strange that only some idioms are capable of this syntactic mobility.

Does that mean that there is something different about the structure of an idiom like *kick*

the bucket and one like *spill the beans*? I believe that there is. The difference seems to

have to do with issues of transparency, an idea suggested by Jackendoff (1997) and

others[6].

It has been noted that some idioms seem to have a more transparent structure than

others. In other words, certain idioms seem to closely resemble their non-idiomatic

counterparts syntactically. It is these transparent idioms which can undergo

passivization, and syntactically opaque idioms which can not. Consider the following

examples:

(22)   keep tabs on NP = maintain surveillance of NP
(23)   spill the beans = reveal a secret
(24)   lay one's cards on the table = make one's feelings known
(25)   kick the bucket = die

Examples (22) through (24) represent idioms which are syntactically transparent. The

syntax of the non-idiomatic version of the phrase directly maps to the syntax of the

idiomatic phrase. Each of these phrases is made up of the same components, and the

components have the same thematic roles within the phrases in each form. In (22) *keep*

tabs on x is a VP consisting of a transitive verb, the verb's direct object theme, and a

---

[6] This observation is also made by Gazdar et al. (1985) and others.

prepositional phrase which relates the idiom to the person(s) it is referring to. This is the same structure as its non-idiomatic correspondent *maintain surveillance of x* has. The phrases map onto each other syntactically, and thus the idiom can be considered transparent. The same kind of transparency relationships hold in (23) and (24) as well.

But *kick the bucket* in example (25) is different. Its structure is not transparent in this way, it is syntactically opaque and thus can not undergo passivization. In other words, the elements of the idiom do not have a one to one mapping with the elements of the phrase's actual non-idiomatic meaning. The idiomatic version consists of a transitive verb and its direct object, while the non-idiomatic version is simply an intransitive verb – *the bucket* corresponds to nothing because *die* does not, and indeed can not, take any complements. This is why it can not be passivized. It is important to note here that no synonyms for *die* can take a direct object either (e.g. perish, expire), so it is not a case that just the choice of the word *die* as a paraphrase is what holds this opacity up. Moreover, two other idioms which simply mean "die" spring to mind, and they both have the same syntactic structure as *kick the bucket*, so the hypothesis that syntactic transparency is necessary for an idiom to be able to undergo passivization predicts that they too would not have an idiomatic reading in the passive:

(26)   a. John bought the farm.
           b. #The farm was bought by John.
(27)   a. John bit the dust.
           b. #The dust was bitten by John.

So now we have seen another example where the degree of fixedness points to some kind of internal structure existing for idioms, only this time we can observe something of what that structure might be. We also observe that there is a split in the kinds of internal structure that we see. One class of idioms is syntactically transparent,

exhibiting a one on one mapping between the syntax of the idiomatic phrase and the

syntax of its non-idiomatic paraphrase. Such idioms retain their idiomatic readings when

in the passive. Syntactically opaque idioms lack this mapping, and thus they also lack the

ability to remain idiomatic when in the passive voice[7]. It is unclear whether transparency

is the only condition necessary for an idiom to be able to undergo passivization, but it

seems to be a necessary one, and an important aspect of the emerging internal structure of

idioms.

3.4 Modification Within Idioms – A Further Demonstration of Transparency

Another way that idioms can show variability from a frozen form is through

modification, which idioms can accept to various degrees. Let's start with two examples

showing fairly straightforward adverbial modification (in the examples in this section,

words in italics represent modification of idiomatic phrases):

    (28)   He *certainly* kicked the bucket.
    (29)   The shit *really* hit the fan.

Sentence (28) is quite straightforward. The idiom in question comprises a complete verb

phrase, and the adverb *certainly* simply modifies the whole thing. The adverb works the

same way on the VP whether the sentence has an idiomatic reading or not. It's pretty

simple, and doesn't tell us anything about internal structure. Sentence (29), on the other

hand, is a bit more complicated and does reaffirm some of what I have already argued for

about idioms' structures. *The shit hit the fan* is a full sentence-length idiom, so the

modification comes within the idiom itself. What this minimally shows is that the idiom

does have some internal structure – at the very least the structure of a regular phrase in

---

[7] Fellbaum's (1993) studies of the determiner in English idioms supports this – she found that determiners
are constrained by the same issues of transparency as passivization.

order for this modification to find its way to the right part of the sentence. But beyond

that, the idiom seems to work in much the same way as in *He certainly kicked the bucket*,

emphasizing the sense of the idiom as a whole. One can imagine that any idiom could

accept modification like this.

There are more ways to modify an idiom than just with an adverb. Individual

words within an idiom can also accept modification, but there are limits as to which ones

can. It seems like the explanation for this behavior can once again come from

transparency.

> (30)   We pulled *a lot of* strings to get these tickets.
> (31)   My brother kicked that *filthy* habit years ago.
> (32)   They shot *huge* holes in my argument.
> (33)   The prosecution left no *legal* stone unturned.
> (34)   #He kicked the *big* bucket.

In all of these idioms just one noun within the idiomatic portion of the sentence is being

modified by an adjective[8]. Moreover, the modification is certainly not a standard part of

these well-known fixed expressions, yet it is easy to get a sense of what the modified

constructions mean. Transparency allows us to do this. Consider the following non-

idiomatic paraphrases of sentences (30)-(34):

> (30) We made use of *a lot of* contacts to get these tickets.
> (31) My brother discontinued that *filthy* habit years ago.[9]
> (32) They pointed out *huge* discrepancies in my argument.
> (33) The prosecution made use of all of their *legal* options.

Yet,

> (34) *He died *big*.

---

[8] Or in the case of pulling 'a lot of' strings, an idiomatic phrase that just means 'many'.
[9] In this example, *habit* is taken to be a part of the idiom even though it is the same word in both the paraphrase and the idiomatic version. The reason for this is that *habit* is a fixed part of the idiom, and seems irreplaceable by synonymy. Hence, you don't kick a custom or a practice, just a habit.

We can see that the idioms in which modification works are transparent, so their elements correspond to the elements of the non-idiomatic reading (Gazdar 1985). This makes some intuitive sense given what we know about transparency. In the idiom *kick the bucket*, the NP *the bucket* has no semantic sense as demonstrated by the fact that its only sensible non-idiomatic paraphrase, *die*, can take no NP argument. Therefore, were one to modify that NP, no sense could be made of it. This hypothesis predicts that *the bucket* can never accept adjectival modification (thus it is not the case that the only reason example (34) doesn't work is due to the choice of adjective)[10]. On the other hand, *holes* in (32) seems to correspond pretty closely with *discrepancies* or *flaws* in an argument. Thus, to modify *holes* is to give a finer sense of the nature of those discrepancies, as seen in the non-idiomatic paraphrase. The same applies for the other sentences in that group. If you modify an internal component of an idiom, you must be refining the sense of some internal structural element of the idiom (and thus it follows that the reason *the bucket* can not be modified is that there is no internal structural element that only *the bucket* corresponds to and whose sense can be refined because it is an opaque idiom – therefore if some speaker did not know the sense of *kick the bucket*, with this analysis it would be predictable that it corresponds to an intransitive verb).

The issue of compositionality comes back into play when looking at idiom modification. Notice that while positioned within non-compositional idioms, the meanings of the adjectives themselves are compositional and literal. They don't really modify the idiomatic non-compositional word, they modify the actual literal referent[11]. Example (33) is the best demonstration of that. The phrase *legal stone* doesn't really

---

[10] For an exception to this, see Section 3.4.1 below
[11] See Nicolas (1995) for an alternative argument that such adjectival modification semantically serves as adverbial modification of the whole idiom.

make a whole lot of sense. This is because the adjective *legal* actually is compositional and literal. In the sentence it modifies *stone*, but it is really describing the options or means that the prosecution is making use of . In other sentences, the adjectives can work either way (a hole can be huge, just like a flaw in an argument can, but you can shoot major holes in an argument too, and can a hole be major?), but it is pretty remarkable that the modifiers that work in the idioms make sense with the non-idiomatic paraphrases. Moreover, adjectives that would make sense with the words of the idioms don't work when used with idiomatic readings:

(35)   #The prosecution left no *hard* stone unturned.
(36)   #They shot *deep* holes in my argument.
(37)   #We pulled some *long* strings to get the tickets.

So now we see a continuing split between transparent and opaque idioms. And we know that the structure of the idiom is connected to the structure of its non-idiomatic counterpart, such that when you modify the idiom, you're talking about the literal version, and if you can't modify the literal version, then you can't modify the idiom either. (Nicolas (1995) notes that nearly all V-NP idioms can have adjectival modification of the NP, which might suggest that transparent idioms occur much more frequently than opaque ones.)

3.4.1   A Damn Exception

There is one notable exception to this account of the behavior of transparent and opaque idioms. Consider the following sentences:

(38)   The shit hit the *damn* fan.
(39)   John kicked the *fucking* bucket.

It appears that expletives can be inserted into any idiom, transparent or opaque, and don't

need to follow the constraints which normal adjectives must in order to be used with

idioms[12]. I would say that this tells us more about the nature of expletives than about the

structure of idioms, though. Expletives in general have quite a lot of freedom as to where

they can go, and that freedom allows them to find their way into opaque idioms where

such modification is normally not allowed. And once again we can see in idioms another

analogous process to what happens with their co-listemes, words. Expletives, after all,

can even find their way into some individual words as infixes, as in fan*fucking*tastic or

abso*friggin*lutely.

3.5   Conclusions Drawn From Idiomatic Variability

At the very least, what the variability of idioms discussed shows is that all idioms have

internal structure of some kind. They can not simply be dismissed as structureless, frozen

atomic units. Furthermore, the variability in passivization and modification shows that

the structure they do have is related to the structure of their non-idiomatic counterparts.

Thus we see a division between transparent idioms whose syntactic structure is the same

as their counterparts and opaque ones whose syntactic structure is different from their

counterparts. This distinction dictates how the idioms are allowed to behave. This is not

to say that there is a distinction between fixed and flexible idioms, subject to different

rules as Gazdar et al. (1985) suggest. Rather, no idiom can be truly said to be fixed, and

all idioms have internal structure, but how that structure relates to their non-idiomatic

counterparts affects their behavior.

---

[12] There is another kind of insert that can also break the rule, as when one says, "He kicked the proverbial
bucket," but discussion of this phenomenon will be relegated to the section on idiom breaking.

4.0 Syntax and the Continuity Constraint

Since the variation that we have so far encountered within idioms can be shown to be indicative of some kind of internal structure, and even tell us a little bit about it, let us now examine that structure more closely. To do this, we can make use of a powerful constraint on the structure of idioms: the Continuity Constraint as put forward by O'Grady (1998). Working under the assumption that "a lexical item licenses its dependents via their heads," O'Grady proposed a relationship between heads and their dependents in idiomatic constructions, stating the constraint as "an idiom's component parts must form a chain."

Thus the analysis of an idiom would show a chain of heads to dependents. According to the constraint, this chain can not be broken within the idiom. The chain is not meant to replace all syntactic structure, but merely to serve as a bare minimum constraint on what the structure of an idiom can look like. One way to understand the continuity chain would be to say that if an idiom were placed in a tree structure, the continuity chain would go down the tree, from heads to their dependents. Let's start by looking at a few syntactically simple idioms and see how they fit in with the Continuity Constraint:
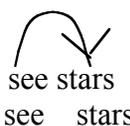
(40)   see stars
         see _ stars
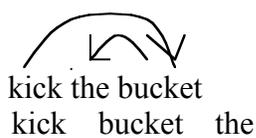
(41)   kick the bucket
         kick _ bucket _ the

(42)   beat around the bush
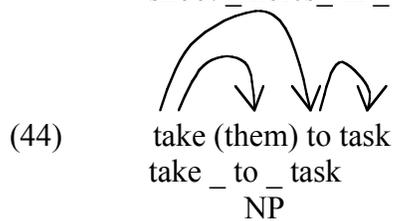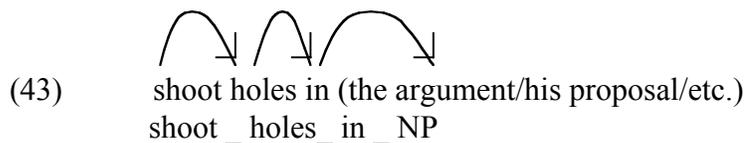   beat _ around _ bush _ the
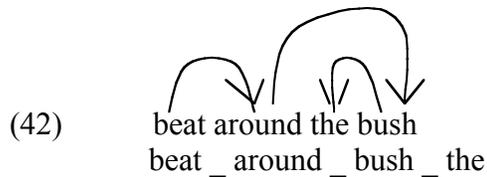
In example (40), the idiom consists of a VP, the head of which is *see* and it has the

NP dependent *stars*. Thus the chain of continuity goes from *see* to *stars*[13]. Examples

(41) and (42) are not very different, but they just have extra "links" in the continuity

chain. Note that the chains needn't follow the word order of the phrase

4.1 The Continuity Constraint and Incomplete Constituents

Notice that examples (40) through (42) all comprise complete constituents, in these

cases, all VPs. But as noted before, idioms are not always complete constituents.

Normal syntax doesn't explain that very easily, but the Continuity Constraint can. It

says that no idiom can exist which doesn't have a continuous chain between heads

and dependents, so let's see if it holds for idioms which aren't made up of complete

constituents:

(43)   shoot holes in (the argument/his proposal/etc.)
            shoot _ holes_ in _ NP

(44)   take (them) to task
            take _ to _ task
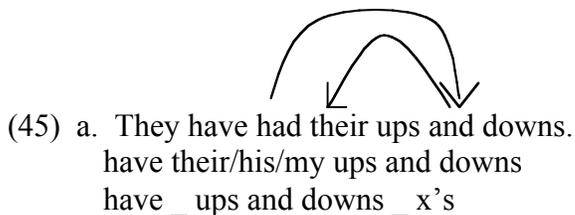   _ NP

In example (43) there are a number of NPs which can be an argument of the verb *shoot*

(e.g. argument, proposal, theory, etc.). Thus the idiom can only be taken to be *shoot*

---

[13] Again, it should be noted that this continuity does not stand in for all syntax, it only refers to the continuity chain. Thus, in the diagrams, the arrows mark only the chain from heads to dependents.

holes in x – this idiom is not a complete constituent. It is a VP, but it is missing an entire

NP which is the necessary argument of the PP. But this is not a problem, because the

variable portion, the missing piece of the chain, so to speak, is at the end of the chain. It

is not a complete constituent, but it is a complete idiom, with the variable portion at the

end of the continuity chain. Example (44) is similar, yet with some differences. There is

a variable NP, but here it is the direct object of the verb. This demonstrates a branching

in the continuity chain. This is still possible – the NP and PP are both arguments of the

verb *take*, but the head to dependency chain does not pass through either one to the other,

it just goes to both separately (the PP is not a dependent of the NP, nor vice versa, they

are simply both dependents of the verb). Thus the variable portion of this idiom does not

break the chain of continuity within the idiom, it remains at the end of the chain.

Not all empty slots in idioms are like these. A very common way in which

constituents of an idiom can be incomplete are through genitive gaps. These are also

"holes" in the idiom, but according to the Continuity Constraint, they can not be gaps in

the chain:[14]

(45) a. They have had their ups and downs.
          have their/his/my ups and downs
   have _ ups and downs _ x's

Again, there is an open spot in the idiom, but it falls at the end of the chain, and thus

adheres to the Continuity Constraint The idiom would not conform to the Continuity

Constraint if the open genitive spot were in the middle of the continuity chain and was

---

[14] Here *x* is taken to stand for a variable constituent, in this case an NP. The *'s* indicates a genitive
modifier, which could be a simple *s* with apostrophe or a possessive adjective.

followed (in the chain, not in normal word order) by other fixed components of the idiom.

4.2 The Continuity Constraint and Modification

This constraint also affects the modification of idioms. When an idiom does take some kind of modification, whether its syntax is transparent or not, it still is constrained by continuity. No modifier can break the chain of continuity, they can only come at the beginning or the end of it:

(46)   The shit *really* hit the fan.
          hit _ shit _ the
          _ fan _ the
          _ *really*
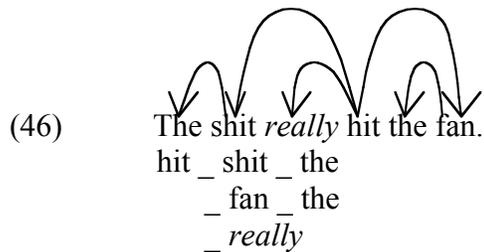
This is the simplest kind of modification, and as we've seen, it can work with transparent and opaque structures. But as we see, it does not break the chain of continuity. Its existence requires there to be an extra branch in the chain (from *hit* to *really*), but the chain that goes to the adverbial modifier does not connect further to other fixed elements of the idiom, so continuity is preserved.

Modification within transparent idioms also must adhere to the Continuity Constraint. Consider:

(47) They shot *huge* holes in my argument.
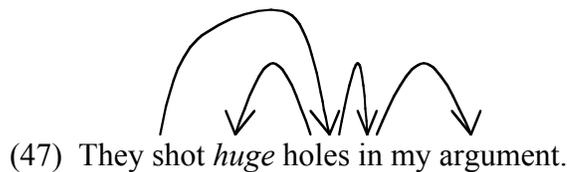
shoot _ holes _ in _ NP

_ huge[15]

This is an interesting example because it has two different variables. First, the NP which

is part of the incomplete PP is not part of the idiom, so this comes at the end of a

continuity chain. Then there is the adjectival modification, which also must come at the

end of the chain, and as the diagram shows, it does, from a new branch of the chain

coming from *holes*, which *huge* modifies and thus is a dependent of.

4.3 A Look at Discontinuity – What Doesn't Work

The Continuity Constraint not only describes the structure of existing idioms, but it

makes predictions about what kind of idioms can and can't exist. As we've seen, it has

predicted where variable positions in idioms can be – only at the end of chains, but it is

helpful to look at what a hypothetical idiom might look like which breaks the constraint.

Such an idiom could be imagined with a slight change to an existing idiom:

(48)   Play the devil's advocate.
(49)   #Play the devil's N.
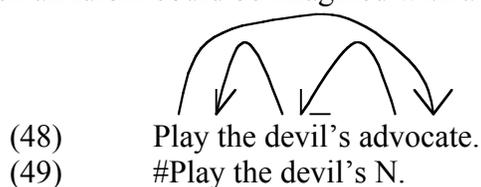
We can easily see that (48) adheres to the constraint:

(48)   Play _ advocate _ devil's _ the

---

[15] Cf. sentence (30), with the added modification, a *lot of*. This doesn't seem to fit, because here the noun
has been changed to the object of the preposition, which changes the relationship of heads to dependents.
However, *lots of* is an anomalous construction itself, which simply means 'many' and gets used
syntactically as though it were an adjective.

But were there an idiom of a similar structure, but with a variable in the noun position, it would violate the Continuity Constraint, because there would be a gap in the chain (O'Grady 1998). That's where example (49) comes in. It seems at first like it could be a conceivable idiom. Here, everything is fixed but the noun itself. It's always the devil's something, but what that something is could be a variable position. Depending on the situation, one could talk about playing the devil's mother, friend, confidant, advocate, or any number of other options. But as we see, this gap in the idiom is in such a place that it breaks the chain of continuity – the head to dependent chain necessary to go from *play* to *the devil's* (the two fixed parts of our hypothetical idiom) goes through the gap, and that can't be:

(49) Play the devil's N

         Play _*_ N _ devil's _ the
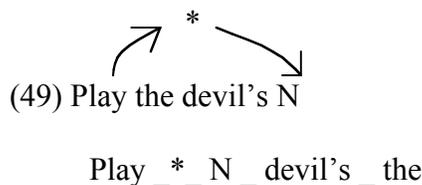
Of course, one could easily say that this example doesn't work simply because it's not an idiom, but the Continuity Constraint predicts that no idioms of this form, with a variable gap in such a position, can exist, while it so happens that there are many idioms of a form more like (45) above which do have gaps (often genitive gaps), but they always fall at the end of the continuity chain.

4.4 The Significance of Continuity

Section 3 showed us that idioms must have internal structure, and, for transparent idioms that structure is related to the structure of their non-idiomatic readings. The fact that idioms adhere to the Continuity Constraint is a further indication that structure exists, and

an explanation of what that structure minimally entails, namely, a continuous chain of heads to their dependents (continuity also makes it easy for idioms to be stored in the lexicon, because all the component parts are continuously connected).

One interesting aspect of continuity that O'Grady (1998) fails to point out is how it relates to the non-idiomatic portions of a sentence. Most idioms, after all, are not complete sentences in and of themselves, but occur in otherwise non-idiomatic sentences, and many idioms as we have seen have within them non-idiomatic elements like modification or open gaps taken up by variable literal elements. Thus the idiomatic part of the sentence must interact with the non-idiomatic part of the sentence, and as the Continuity Constraint demonstrates, the interface between an idiom and the rest of the sentence is at the end of these continuity chains.

5.0 Other Issues in the Structure of Idioms

This is by no means an exhaustive cataloging of the structural behavior of idioms, nor is it meant to be. What follows is a very brief look at other areas in which idioms are being looked at.

5.1 Compositionality Revisited

From the very beginning I have assumed as part of the definition of an idiom that it is a non-compositional form, that is, that its meaning can not be compositionally computed from its parts. This suggests that the way many idioms have found their way into language is as dead metaphors – once robust figurative tropes, their associations have now been severed rendering their meaning arbitrary to modern speakers.

Gibbs (1995) takes issue with this explanation, though. He argues that for many idioms, there still is a real figurative connection between idioms and the literal. It is inaccurate to call idioms non-compositional because people's metaphorical understanding plays a significant role in how they interpret idiomatic discourse. He cites several cognition experiments on idiom processing, as well as ideas of primary conceptual metaphors and visual schemas[16]. Moreover, he argues for a recognition that "many idioms are analyzable with their components *independently* contributing to what these phrases mean literally" (my italics). Compositional computation of an idiom's parts can lead to an interpretation of an idiom then. This is stronger than the explanation which I adopted of transparency – that many idioms have the same syntactic structure both in their idiomatic form and in a non-idiomatic form. Gibbs's view suggests that not only may there be an overlap in form, but a very real connection between the two forms. If components do have stronger independent meanings than has been before recognized, it can certainly have an effect on ideas of idioms' syntax, as well as the notion of fixedness. It presents a very different model of idiomatic behavior.

5.2 Syntactic Theories and Idioms

What I have presented has merely comprised some minimal conditions that constrain the syntax of idioms, without getting into thorough syntactic analysis of idioms within one theory of syntax.. Others have done the latter, though, many starting from different assumptions and winding up with different conclusions. Van Gestel (1995) asserted that "idioms are a matter of X-bar syntax," and analyzed them as such. He takes the idea of listemes from DiSciullo and Williams (1987) and finds that idioms comprise partial trees

---

[16] For discussions see Lakoff and Johnson 1980.

within X-bar theory. He also notes that levels of fixedness of idioms corresponds to X-bar levels of idiomatization. Abeillé (1995) studied French idioms using Lexicalized Tree Adjoining Grammar, and argued that an idioms syntactic structure is not fully predictable from its semantic representation. But when idioms are studied within these theories, they still must have structure, and this paper argues for what must be the minimal structural requirements for idioms in any theory of syntax[17].

5.3 Fixed Expressions and the Boundary of Idiomaticity

At the beginning of this study, idioms were placed in a category of fixed expressions which also included things like clichés, famous quotes, and even proper names. Jackendoff (1997) proposes that once idioms are admitted into the lexicon, the other fixed expressions can follow (though this does not mean that all fixed expressions behave the way idioms do). Schenk (1995) suggests that proverbs and collocations are idiomatic. So then what exactly counts as an idiom?

O'Grady (1998) suggests that there is no reason to set a rigid boundary between idioms and non-idioms. I agree with O'Grady. When confronted with fixed expressions like clichés or proverbs, they often seem to have idiomatic qualities, and no clear line presents itself as to when one can't be another. The criteria of non-compositionality does not always give clear delineation, either. Take the proverb, "every rose has its thorn." On the one hand, its meaning is quite compositional – it actually refers to roses having thorns. But when used, it metaphorically applies to non-rose situations, meaning something like every beautiful or good thing has its downsides. It is unclear exactly how

---

[17] Though they are certainly not always agreed on.

26

to classify this. Certainly a cliché could be idiomatic, or an idiom clichéd, but even the

criteria for idioms of compositionality sometimes only yields a muddy distinction.

5.4 Idiom Breaking

Another issue on the boundary of idiomaticity is that of idiom breaking. Idiom breaking

is a very common kind of word play in which a conventional idiom is altered (Alm-

Arvius, forthcoming). In Section 3.4.1 I alluded to a reason why "He kicked the

proverbial bucket" is not an exception to the rule that the NP in such an opaque idiom can

not take adjectival modification. That is because this is an example of idiom breaking.

The word *proverbial* is not so much modifying the noun *bucket* (though syntactically, it

is in such a position) as calling attention to the fact that the bucket does not exist. That is

why it can exist there, where other adjectives can not. The idiom is broken, so to speak,

by this modifier. Indeed *proverbial* can be inserted as an NP modifier into most any

idiom, transparent or opaque, as a means of idiom breaking:

> (50)   He kicked the *proverbial* bucket.
> (51)   That loudmouth spilled the *proverbial* beans.

This is not the only way idiom breaking can occur. Oftentimes when modification occurs

somewhere it can not normally exist in an idiom, idiom breaking turns out to be the case.

Nicolas (1995) cites the following examples:

> (52)   Many people were eager to jump on the *horse-drawn Reagan* bandwagon.
> (53)   Bruce, a shark , found it [a role in the film *Jaws*] a part he could really
>   sink his *three rows of* teeth into (Nicolas (1995, 40-41); my italics).

These are examples where the actual individual word of the idiom is modified, which I

had previously asserted can not be done. This is only possible through the word play that

is idiom breaking. Idiom breaking draws attention to the idomaticity of the idiom, thus breaking it and allowing normally impossible modification to occur.[18]

5.5 Cognition Studies and Idioms

One area this study does not address is that of cognitive linguistics. This field is currently being used to address many of the tough questions the study of idioms presents. For instance, how do speakers recognize that a phrase needs to be computed idiomatically? Or how exactly do idioms get processed, and is it different from how normal phrases are handled? The answers to these questions may be forthcoming, though as of now it may be that nothing definitive has been proved[19]. What these studies tell us may point to definitive answers on what the structure of idioms might actually be, why they behave as they do, and even perhaps why they exist in the first place.

6.0 Conclusion

I began by arguing that idioms are part of the lexicon. The sheer numbers of idioms which seem to exist indicates that they represent a significant aspect of language behavior. Placing idioms in the lexicon makes sense because as far as phrases and usage go, idioms are nothing special. They can't be written off as abnormalities, because it is evident that they are quite normal. Moreover, placing them within the lexicon allowed them to have internal linguistic structure, which Section 3 demonstrated that they do in fact have by examining their behavior as fixed expressions. We also saw that not only do

---

[18] For further evidence of idiom breaking, simply scan newspaper headlines – examples will come out of the proverbial woodwork, so to speak.
[19] For discussion of several developments in this field see d'Arcais (1993) and Van de Voort and Vonk (1995).

idioms have structure, but that structure is related to the structure of their non-idiomatic or actual meaning. This led to a division between transparent idioms, whose syntactic structure is the same as its non-idiomatic meaning, and opaque idioms, whose structure was disjoint from that of its non-idiomatic meaning. We saw that an idiom can behave syntactically only in ways that its literal counterpart could behave. The Continuity Constraint then allows us to take a closer look at the idiomatic structure which was starting to come into view. Tests of the constraint show that idioms must comprise a continuous chain of head to dependent relations.

This is by no means a complete picture of idioms. Much is yet to be learned, and it is important to remember that, as Gibbs (1993) points out, "Idiomatic language is remarkably complex and each phrase demands its own analysis in terms of its syntactic, semantic, pragmatic, and conceptual properties." This analysis is valuable, as it presents a minimal structure that seems to be common to all idioms. And the study of idioms can tell us increasingly more about the nature of fixed expressions, figurative language, and finally, of language itself.

References

Abeillé, Anne. 1995. The Flexibility of French Idioms: A Representation with Lexicalized Tree Adjoining Grammar. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 15-42. Hillsdale, NJ: Lawrence Erlbaum Associates
Alm-Arvius, Christina. Forthcoming. *Figures of Speech*. *Ms*. Stockholm University.
Collins, V.H. 1957. *A Book of English Idioms*. London: Longmans, Green and Co.
D'Arcais, Giovanni B. Flores. 1993. The Comprehension and Semantic Interpretation of Idioms. In *Idioms: Processing, Structure, and Interpretation*, ed. Cristina Cacciari and Patrizia Tabossi, 79-98. Hillsdale, NJ: Lawrence Erlbaum Associates.
DiSciullo, Anna-Maria and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, Mass.: MIT Press.

Fellbaum, Christiane. 1993. The Determiner in English Idioms. In *Idioms: Processing, Structure, and Interpretation*, ed. Cristina Cacciari and Patrizia Tabossi, 271-296. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, Mass.: Harvard University Press.

Gibbs, Raymond W. 1993. Why Idioms Are Not Dead Metaphors. In *Idioms: Processing, Structure, and Interpretation*, ed. Cristina Cacciari and Patrizia Tabossi, 57-78. Hillsdale, NJ: Lawrence Erlbaum Associates. A

Jackendoff, Ray. 1985. The Boundaries of the Lexicon. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 133-166. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, Mass.: MIT Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: Chicago University Press.

Makkai, Adam. 1972. *Idiom Structure in English*. The Hague: Mouton.

Nicolas, Tim. 1995. Semantics of Idiom Modification. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 233-252. Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Grady, William. 1998. The Syntax of Idioms. In *Natural Language and Linguistic Theory* 16: 279-312.

Schenk, Andre. 1995. The Syntactic Behavior of Idioms. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 253-272. Hillsdale, NJ: Lawrence Erlbaum Associates.

Van de Voort, Marlies E.C. and Wietske Vonk. 1995. You Don't Die Immediately When You Kick an Empty Bucket: A Processing View on Semantic and Syntactic Characteristics of Idioms. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 283-300. Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Gestel, Frank. 1995. En Bloc Insertion. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 75-96. Hillsdale, NJ: Lawrence Erlbaum Associates.

Zeevat, Henk. 1995. Idiomatic Blocking and the Elsewhere Principle. In *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, Andre Schenk, and Rob Schreuder, 301-316. Hillsdale, NJ: Lawrence Erlbaum Associates.