

Usage, Blends, and Grammar

MICHAEL BARLOW

Rice University

1. Introduction

The extract in (1)¹ contains a short sample of language in use, an orthographic representation of part of an exchange between two American speakers set in an academic environment.

- (1) a. Why is it chapter 3?
b. Well, that's the thing to talk about.

While the representation in (1) clearly omits a considerable amount of information, extending from the phonetic realisation of the utterances to their discourse context, it does nevertheless provide empirical data relating to an actual speech event, and as an illustration of language usage it will serve as a useful starting point for the discussion of usage-based grammar² presented here. We can take some comfort from the fact that example (1) represents an actual exchange, if small one, between two speakers. However, if we look to the example for insights into usage-based grammar, we see the difficulty of bridging the gap between the particular, the utterance itself, and the general, the grammatical representation that ultimately underpins the

production and comprehension of the utterance—differences between production and comprehension notwithstanding. The strategy adopted here is simply to examine multiple utterances (or sentences) in a corpus, looking for regularities in usage. The relationships drawn between multiple instances of utterances and linguistic generalisations have some parallels in the connection between usage data and the form of a usage-based grammar in that, as we will see, both relationships involve data-driven pattern extraction. Even a cursory inspection of a corpus reveals recurrent regularities in the data, the most obvious of which are collocations or recurrent word associations (Firth 1957, Sinclair 1991).³ We find, for example, that the word *thing*, which appears in (1b), is used very frequently (in both spoken and written discourses) and occurs in a number of collocations such as *the thing is* and *the...thing to do*. With appropriate usage data (i.e., corpus data) at hand and keeping in mind the parallels alluded to above between corpus-based generalisations and usage-based grammar, we can make some progress towards answering a question posed (somewhat rhetorically) in Bolinger (1961: 381):

Is grammar something where speakers produce (i.e. originate) constructions, or where they reach for them from a pre-established inventory, when the occasion presents itself?

Obviously, the answer to this question is not going to simply appear out of thin air once we have the requisite corpus data. Consequently, we have to look for clues to the relationship between frequent collocational patterns and their representation in a usage-based grammar, and then consider the implications for less frequent collocational patterns such as *the thing to talk about* in (1b), which is not a particularly strong collocation.

The prevailing view of syntax is that at its core it comprises a set of rules or constraints which account for the creative use of language, and that secondarily or peripherally, the grammar (or lexicon) contains a list of multi-word idioms and other set phrases. I claim that this is exactly backwards: The main component of grammar instead comprises a large set of redundantly specified schemata, both abstract and lexically-specified, and the role of rules or constraints (or highly abstract schemata) is to provide the glue or mortar to combine these prefabricated chunks.⁴ Exploring this alternative view, I investigate the relation between instances of usage illustrated by the fragment in (1) and a usage-based grammar, paying particular attention to the role of blending as described in Turner and Fauconnier (1995), Fauconnier and Turner (1996), and Turner (1996) *inter alia*. I aim to show that what looks like creativity in language is in many

instances the result of blending of prefabricated units rather than the output of a set of generative rules.

The structure of the paper is as follows: I first examine the consequences of the pervasiveness of collocations for the structure of grammar, turning in Section 3 to an exploration of the nature of syntax in a collocation-rich grammar. The main part of the paper, Section 4, demonstrates the use of corpus data to investigate the extent of blending of prefabricated forms in language production.

2. Collocations and Usage-based Grammar

We can describe how a usage-based grammar might differ from a standard grammatical description by considering the following contrasts. Adopting a generic X-bar syntax framework, one could view the utterances in (1) as bracketed structures or tree-structures in which the main category represents a sentence containing phrasal categories (such as NP, N', etc.) which, in turn, contain or dominate lexical categories (such as N). Only the latter categories are linked to individual lexical items, and therefore the control of connections among words is achieved only indirectly through restrictions on the combination of syntactic categories. From this perspective, syntax is a set of constraints governing the combination of phrasal and lexical categories.⁵ And, as is well-known, the typical data used in this tradition is grounded not in actual usage but rather in intuitions about sentences and specifically judgements of the grammaticality of sentences.

A model of syntax that aims to explain the distribution of syntactic categories, say AP and N' (rather than lexical items themselves), clearly has value as a way of capturing the range of possibilities of association (as famously illustrated by *colorless green ideas*), but it is bound to be unsuccessful in accounting for frequent collocations such as those involving the cooccurrence of particular adjectives (e.g., *broad*) with particular nouns (e.g., *daylight*). In other words, an X-bar framework can only account for relations between syntactic categories and is unable to represent the lexical or collocational dimension of syntagmatic structure. The ubiquity of dependencies among words (as opposed to dependencies among syntactic categories) is self-evident once corpus analyses are undertaken, suggesting the need for some kind of collocational grammar. And while no such grammar has been produced, the extent of collocations is extensively documented in dictionaries and other reference works such as Benson, Benson and Ilson (1997), Kjellmer (1994), and Moon (1998), which provide extensive lists of collocations. Moreover, the degree of attraction (colloc-

ational strength) between any of the words in a particular corpus can be estimated by statistical measures such as mutual information, t-score, cost and entropy.⁶ Collocations are not only a feature of corpora; speakers know about collocations, and collocations come in different strengths in terms of how well routinised or entrenched they are in speakers' linguistic systems.

Does it matter that X-bar syntax fails to capture certain aspects of what it means to know a language? In one sense, no, it does not. If syntax is taken to be an idealised view of those aspects of language structure related to conjectural learnability issues rather than a model of day-to-day language use, then one might say that the details associated with cooccurrence patterns and frequency of occurrence are peripheral and of little or no importance. On the other hand, if we focus on usage and usage-based grammar, we find that it is the idealised syntax which is of secondary importance and plays a background role in speaker's knowledge, as explained in the discussion below.

Using corpora to examine patterns of usage reveals the important role that lexical units larger than a word perform in language production and shows that much of language in use is not creative in the Chomskyan sense, but is based mainly on the use of prefabricated or semi-prefabricated chunks. I argue that the creativity or the expressive dimension to language comes in large part from the modification of prefab structures, rather than the novel combination of lexical categories.

Well-entrenched collocations like *broad daylight* and other examples of fixed and semi-fixed structures such as those shown in (2) are instances of syntagmatic units that can be described in terms of schemata or constructions.⁷

- (2) a. one thing
 b. the thing is
 c. the right thing to do
 d. sort of thing
 e. it is one thing to ... it is another thing to ...

The building blocks of a usage-based grammar are not lexical items dominated by lexical categories, but are form-meaning pairings of differing degrees of complexity and different degrees of specificity. Adopting this view of the units of grammar, we see that, in fact, the building block metaphor is not the best for this approach. (See also Langacker, this volume.) For one thing, the "blocks" have both a formal and a meaning side; and some of the blocks come already partially assembled. Nevertheless,

some sort of "building block model" is necessary to capture the compositional nature of syntax/semantics. The sort of model required, however, must allow for the composition and blending of units large and small.

The connection between usage and usage-based grammar is not a simple one, but we can attempt a basic description as follows: repeated exposure to collocations leads to the entrenchment of collocational patterns and their associated meanings in the grammar.⁸ This process applies not only to strong lexical collocations such as *broad daylight*, but also to looser collocational links between words as in [*it is one thing to X, it is another thing to Y*]. The existence of a cline from strong to weak collocational links among words and the presence of links between collocations (collocations of collocations) are amenable to treatment in a usage-based grammar that, by definition, is structured in a way that reflects input. Additionally, appropriate conditions of language usage are clearly a part of a speaker's knowledge of language and the connections between grammatical units and register, genre, and other types of situational information must also be part of the grammatical representation, but such associations are not pursued here.

3. Combining and Modifying Collocations

If we take the position that chunks represented by schemata are the fundamental units of syntax, then we must face up to a couple of problems. One concerns the modification of chunks. How are fixed or semi-fixed expressions such as *close, but no cigar* modified? If they are unanalysed chunks, then it would not be possible for them to (ever) take modifiers or to allow word substitutions. And it is in solving this problem that something akin to X-bar syntax⁹ may be useful as a guide to the internal structure of chunks, which is a necessary step in the creation of modified structures. A phrase like *close, but no cigar* may be represented in the grammar with a category label for the whole phrase, but with little identification of internal structure. However, a syntactic template may be imposed on the structure if a speaker wants to create a modified structure, as in *close, but no banking cigar*¹⁰ or *close, but no goal*. The characterisation of this procedure in terms of an imposition of a syntactic template is, perhaps, an overly dramatic description. Chunks in the grammar have differing degrees of internal structure and sometimes the internal structure will be quite transparent, as in *broad daylight* and *the thing is* but in other cases, rather opaque, as in *one fell swoop* and *easy does it*. Focussing on usage-based grammar at this level

of detail, we also face the fact that individual speakers will vary, perhaps quite markedly, in the internal structure assigned to different fixed and semi-fixed expressions. Individual variation aside, some kind of analysis of chunks can always be made, but given the fundamental role of collocational units, then inevitably higher order syntactic categories play a secondary role. Further issues relating to the modification of chunks will be explored in detail in Section 4.

The second problem can be stated as follows. We have made the claim that collocations are an important aspect of syntagmatic structure and yet do not fit within traditional syntactic approaches. However, if collocational units are taken to be fundamental units of syntax, then we need a mechanism to account for the combination of these units. Here once again we can turn to abstract schemata to provide the glue allowing combination of syntagmatic units, although the structures involved will have to reflect the organisation of spoken discourse and hence may differ from traditional tree structures, which are perhaps better suited to written discourse. In a similar vein, we should note that collocational units do not necessarily correspond to traditional syntactic categories. And without promising a complete resolution of these issues, we can make some progress in understanding form/meaning composition by examining evidence from corpus data concerning usage and cognitive representations, focussing in particular on the role of blending.

To illustrate these notions, let us look at a couple of examples. In the sentence taken from a spoken corpus shown in (3a), we see the common collocation *the thing to do* forming a part of the utterance. Based on an inductive view of language learning, the frequency of this construction in corpora might lead one to suggest that the internal representation of this string will have a cognitive reality as a set phrase or grammatical unit. On the other hand, we might ask what is to be made of the occurrence of *a worthwhile thing to footnote* in (3b). This latter example seems, on the face of it, to be a good illustration of the need for a generative grammar in which syntactic processes control the combination of lexical categories. Certainly *a worthwhile thing to footnote* is going to be so rare in usage as to count for all intents and purposes as a unique utterance.

- (3) a. And I think probably **the thing to do** is have a group write that
- b. I think at some point that might be **a worthwhile thing to footnote**, ...

There are a variety of ways in which this pair of utterances could be handled. For instance, it might be argued that *the thing to do* is a more or

less prefabricated chunk, whereas *a worthwhile thing to footnote* is created according to a rule-like compositional system. Adopting this approach is equivalent to saying that there are two main ways of constructing sentences: adopting a bottom-up metaphor, we can say that one way is from the word up, and that the other is from the word up, except for those parts for which a prefabricated chunk exists. Little, if any, discussion is given within general syntax to the manner in which the two construction methods can co-exist so as to ensure that a prefabricated chunk such as *the thing to do* meshes with a syntax based on the combination of syntactic categories.¹¹ Disregarding the collocational connections between the chunk and other words in the sentence, we still have the problem of how a grammar in which phrasal nodes dominate either other phrasal nodes or lexical categories can cope with a complex four-word expression like *the thing to do* or with variants such as *the best thing to do*.

Alternatively, some might argue that if a rule-like syntactic system is needed in any case for (3b), then in the interests of parsimony the string in (3a) should also be accounted for by the same rule-like system. To the extent that the utterance in (3a) can be generated by the same kinds of rules as (3b), this seems reasonable. However, this approach is implausible because we are then left with the gulf between, on the one hand, a fully compositional syntax and, on the other hand, usage data that indicates the extensive presence of prefabricated chunks which have unit status on syntactic or semantic/pragmatic grounds.

Let us examine a third possibility, which is that a prefabricated chunk is involved in some way in both (3a) and (3b). This third option, explored in detail below, rests on the notion that some of the apparent creativity in language is in fact the result of merger or modification of lexical prefabricated chunks—a partial creativity based on the re-use or re-purposing of prefabricated structures, rather than complete from-the-bottom-up assembly. That is not to say that a bottom-up assembly based on the combination of lexical items never occurs, only that it is the exception rather than the rule. In other words, I argue for the view that strings such as *a worthwhile thing to footnote* are the result of blending of stored cognitive representations. Thus we can explore the idea that in meeting the needs of a particular communicative situation, there is a merger, or mixing, or blending of form-meaning pairs which leads to an output that may be much more variable than even the full range of underlying or stored cognitive representations, including all entrenched collocations. In the following section we examine the mechanisms of this blending process.

4. The Role of Blending

Blending is a general cognitive process involving the merger of formal and conceptual structures to produce new structures that contain partial projections from the input domains, along with new emergent properties specific to the blend (Turner and Fauconnier 1995, Fauconnier and Turner 1996, and Turner 1996). Metaphors and figurative language can be taken to be one type of blending, but there are many other kinds of blends in literary and everyday language, and in non-linguistic domains such as advertising images. A series of Absolut Vodka advertisements, for example, have for several years been based on visual blends which entice the viewer to marvel at the integration of the shape of the Absolut bottle with a view of a famous landmark or scene, and which, at the same time, invite the retrieval or isolation of one input to the blend: the vodka bottle itself. Absolut London, for instance, appears to be a common photograph of the Prime Minister's official residence, 10 Downing Street. Looking closely, however, you see that the outline of the famous door combined with a strategically placed lantern represents the vodka bottle. And in another example, Absolut Amsterdam, the scene is of three tall, narrow Dutch houses next to a canal, but the middle house, which has the same texture and same facing as its neighbouring houses, has taken on the shape of the vodka bottle. These particular examples illustrate a situation in which the formal blend does not correspond to a conceptual blend. In the Absolut ads, there is no conceptual blending that accompanies the visual blending of the vodka bottle and the Prime Minister's residence; there is no blending that goes beyond the blend in the visual image itself. The ad does not suggest, for example, that the Prime Minister drinks vodka at home by the bottleful.

Turner and Fauconnier (1995: 202) discuss the potential disconnection of conceptual and formal blends, noting that formal blends can occur in the absence of conceptual blends (as illustrated above). We will return to this disengagement of concept and form below.

An important motivation for blending is the push to consolidate several events into a single unit. Noting the pressure to integrate conceptual structure, Fauconnier and Turner (1996: 117) give an example of a non-integrated sequence of actions: 'Jack sneezed. The napkin moved. It was on the table. Now it is off the table.' They point out that English allows the same content to be expressed with the form *Jack sneezed the napkin off the table*, which represents an integrated conceptual structure. This integration occurs by conceptual blending in which one input is a [NP V NP PP] structure, representing an integrated caused-motion event of the type 'John threw the ball into the basket,' and a second input, which is the sequence of

actions given above. Thus while the two inputs to the blend are different, they are similar enough to support a partial merger into an integrated conceptual structure expressed through a particular syntactic configuration.¹²

Turner and Fauconnier (1995: 184) state that:

interestingly, and rather unexpectedly, conceptual blends in thought are seldom mirrored by formal blends in language. Language has other means for prompting us to perform conceptual integration.

The motivation for this statement comes from a particular view of language that most linguists would agree with, which is that language only hints at intended meaning. An utterance can be thought of as providing signposts which the language user can take note of, and within a particular context, exploit to retrieve or construct the intended meaning. One illustration of this view comes from the fact that the mere juxtaposition, rather than formal blending, of two words is generally sufficient to lead the hearer to the appropriate meaning. In other words, given the requisite background knowledge, the simple contiguity of *dolphin* and *safe* is enough to guide the hearer to the conceptual integration necessary to understand *dolphin-safe tuna*; similarly, the occurrence of *land yacht* can be seen as an instruction to create an appropriate conceptual blend of spaces associated with *land* and *yacht* (Turner and Fauconnier 1995). The precise relations between the linguistic components of the compound are severely underspecified. Language appears to us to be rich in meaning, but much of that meaning is in the context of the words being used, and, ultimately, of course, it is in the minds of language users rather than in the words themselves.

Are there situations in which conceptual blends are commonly matched by formal blends? The answer depends on where you look. We have already mentioned some noun-noun compounds analysed by Fauconnier and Turner such as *dolphin-safe tuna* and *land yacht*, in which formal blending does not occur. What about grammatical structures? When Fauconnier and Turner talk about blending and grammar, they refer to schemata such as [NP V NP PP] and [NP(x) be NP(y) of NP(z)]¹³ and are thus focussing only on abstract clausal and phrasal constructions, which are also discussed extensively by Goldberg (1995). However, a considerable component of grammar consists of units larger than noun-noun compounds like *land yacht* or *fire station* and which are, at the same time, less abstract than clausal structures such as [NP V NP PP]. I maintain that it is at this intermediate level of granularity that formal blends are much more common than Fauconnier and Turner seem to suggest. In short, I am arguing that if we look to the semi-lexical, semi-syntactic structures such as [*the main/best/first thing is*] and their associated

meanings, we will find the locus of a considerable amount of formal blending.

Let us explore this idea further. As we saw above, Fauconnier and Turner state that one fundamental reason for the occurrence of blends is the need to merge conceptualisations of events to form an integrated structure. At the heart of what I am proposing is a different kind of merger: the blending of general, stored schema-meaning pairs to fit the particular—the current thought or intention. A thought or intention may fit well with a stored representation, in which case the word or phrase can simply be recruited, and presumably, this is what happens in the use of utterances such as *thank you*, *good morning*, etc. Longer and more complex phrases, however, are more likely to be the result of not only concatenative processes, but blending or modification of suitable prefabricated chunks.

This kind of blending is an abstract process not open to inspection, but it is not unreasonable to propose that there is within a grammar a set of form-meaning pairs with different degrees of entrenchment and that in expressing an idea certain chunks can be recruited and perhaps modified or blended. One way to look at this is to view the utterance to be created as a blended space in which stored form-meaning pairs are not simply concatenated, but are merged.

There is an obvious difficulty in investigating this topic that can be referred to as the concrete mixer problem. We have much more of an idea of the nature of the stuff that comes out of the mixer than we do about what goes into it. Blending by its nature obscures the input to the blend; and so finding empirical evidence of blending is difficult. The strategy I pursue here is to look for examples in which the input to the blends is retrievable, and then study the properties of the blending process. The assumption is that the same blending processes are operating whether or not we cannot identify the inputs to the blend. In other words, instead of having a two-part model of grammar in which there is a generative, word-based component supplemented by the listing of some fixed expressions, we have instead a unitary, coherent system in which the speaker can exploit a variety of lexical resources and combine them using concatenative and blending processes.

4.1 Lexical Blends

The fact that we tend to think of words as among the more stable and tangible elements in the universe of linguistic entities suggests that blending within words is rare. A blend, when it does occur, will be obvious—at least as long as the blend remains novel. Thus a word such as

automobilia clearly results from the process of blending. In the production, and probably also in the comprehension, of *automobilia* two separate units are activated: *automobile* and *memorabilia*. It is reasonable to assume that the inputs to the blend are these two words, and that similarity in the form of these words (and perhaps other properties that they have) facilitates the production of a blend when the communicative situation calls for an intimate linking of the meanings associated with the two words.¹⁴

Let us briefly examine another example and consider the blend *digerati*. In this case, the lexical items merged in the blend are harder to discern; one component is probably *literati* with the other being a word something like *digital*, although the precise form remains unclear. And again, in comprehending *digerati* many hearers will activate a 'literati' meaning and a 'digital' meaning.

For some speakers, *automobilia* may actually have the status of a unit with only minimal activation of the associated contributing units, just as *smog* is treated as an independent lexical item (although wordsmiths may not be able to prevent the associated words *smoke* and *fog* from rising into consciousness). The main point of these examples, however, is to show that for many speakers the existence of a lexical blend and typically even the identity of the elements contributing to the blend are obvious. (See Kemmer forthcoming for discussion of the properties of component words that facilitate blending.) The blends are new enough to be identifiable as such and it is the very stability of words which makes the occurrence of lexical blends easy to recognise. The more difficult task to which we now turn is to search for evidence for syntactic blending.

4.2 Syntactic Blends

The constructional character of syntax means that while we may expect blending to occur, or even to be commonplace, evidence of blending is going to be difficult to find, since the units of syntax and the process of syntactic composition itself are hidden from view. In looking at the sentence or utterance—the output of constructional processes—it is difficult if not impossible to discern the elements used in the compositional process. We can see what the words making up the utterance are, and we can usually give a possible tree structure for the utterance, but beyond that, for a given usage event, it is hard to say definitively whether units larger than the word were part of the input. For example, if the output string is series of lexemes represented by $a b x y z$, we can then ask, are the basic units lexical categories associated with $[a][b][x][y][z]$, as commonly supposed; or $[a b y]$ which is merged in some manner with $[x - z]$; or $[a b x w z]$ with v

replacing *w*; or any number of other alternative combinations? It is difficult to say, and yet it is a fundamental question if we wish to understand the nature of syntax.

Let us review the characteristics of language described above in Section 3. On the one hand, we have a standard view of syntax in which syntactic composition makes reference only to syntactic categories, not lexical items. This results in a grammar well able to account for some of the creative aspects of language, but not the collocational patterns that are found in language usage. On the other hand, the frequency of occurrence of some collocations suggests that on psychological grounds these collocations have a unit status within the grammar; similarly, the semantic or pragmatic idiosyncrasy of some collocations offers no alternative other than to assume that they are listed whole in the grammar. The presence in the grammar of numerous prefabricated units means that it is unclear what the appropriate building blocks of syntax are; hence the nature of syntactic composition must be re-examined.

Two further characteristics are worthy of note. One is the absence of a marked division between strong and weak collocations, which makes a creative/fixed dichotomy difficult to maintain. The highly frequent shorter collocations would be most likely to be inserted as a whole chunk, with modification or blending being possible, but less likely. And the looser (and longer) the collocation, the more likely it will be that compositional, blending processes will play a role.

The second, related, characteristic is that collocations sometimes appear with novel lexemes embedded in them, as in *a good thing to footnote*. Thus it is just as necessary to account for the novel items in fixed phrases as for the occurrence of fixed phrases themselves.

These properties suggest that some sort of blending of schema-meaning pairs may play a major role in syntactic composition, and so we will now turn to look at blending in more detail.

4.3 Evidence of Blending

As we have noted, while it is difficult to see the result of blending processes in general, we should still be able to find some evidence of blending in syntax. One possible indicator of blending is the occurrence of idiosyncratic combinations of syntactic categories. In many cases the blending of chunks will result in structures which are analysable by X-bar syntax and are therefore unremarkable. However, if phrasal chunks are blended, we might expect to find some cases in which the parts don't follow the expected syntactic patterns. Let us look at an example. The phrase *very much* has

both a typically prenominal adjectival use (along with *very many*) and an adverbial use, as illustrated in the examples in (4a) and (4b).

- (4) a. the result is not going to be of very much interest ...
 b. I agree very much with what you said.

How then do we analyse the use of *very much* in (5)?

- (5) So thanks very much to my Steering Committee.

The adverbial use of *very much* appears to have become attached to the nominal *thanks*, and this odd combination of syntactic categories may indicate that this now fixed construction originated as a blend. Historically, the inputs to the blend were perhaps something like *many thanks* (or *thanks for X*) and *thank you very much*.

Other examples of odd word-class combinations noted by Moon (1998: 81-82) as expressions "that cannot be parsed according to normal syntactic rules" include *for free* and *the back of beyond*. There may be several ways to explain the occurrence of forms such as these, and here I simply mention them without further comment as possible candidates for blended structures.

Further evidence of blending may be gleaned from historical changes in syntactic structure and here we focus on a single change in the use of the verb *claim*. The examples in (6) taken from the OED show the typical use of *claim* in the eighteenth century in which the verb subcategorises for a noun direct object.

- (6) a. Both sides claimed the victory (1722)
 b. These instances of kindness claim my most grateful acknowledgments. (1775)
 c. Heroines of such a cast may claim our admiration. (1776)
 d. Much learned dust involves the combatants; each claiming Truth.(1784)

Examples illustrating the use of the verb *assert* from around the same period in (7) show that, unlike *claim*, *assert* can occur with a *that*-complement.

examples in (13) the set expression is again given in its full form and then followed by a conjoined phrase that provides the appropriate connections to the current situation. It is the formal juxtaposition along with some cohesive links that invite a conceptual integration of two events, and in these two examples there are also links between the fat lady in the set expression and individuals in the current situation being described.

- (13) a. "However," Wagner says, "there is a basic problem. You know, **it isn't over until the fat lady sings** ... but in this athletic crowd, there are no fat ladies." (NAN 94/95)
- b. "**It's not over until the fat lady sings** and there's no music coming out of Blanca's office yet," Harrison said Monday. (NYT 94)

In the following examples in (14), the fat lady does not have an actual counterpart in the situation described, but she acts within an imaginary space in such a way as to convey information about the current situation. There is no formal blending in these examples; all the action described in the sentences occurs in a constructed "opera world."

- (14) a. But **it's not all over 'til the fat lady sings** and she's not even on stage yet. (Times/ST 96)
- b. Of course, **it is not over until the fat lady sings** and she has yet to approach the piano. (Times/ST 96)
- c. Well, Grandpa, **it's not over until the fat lady sings**, and she hasn't been called back for an encore. (SJM 91)

The remaining examples, (15)-(20), are similar in overall form to the sentences given above, but in these cases there is also some evidence of formal blending. For instance, the first sentence in (15a) contains an inserted comment and in the following sentence the fat lady acts as a protagonist in the world of Big Ten football. In (15b) and (15c), the noun phrase *opera* replaces the *it* in *it ain't over*. The explicit mention of opera in these examples is motivated by the need to evoke the opera frame so that the additional comment (e.g., *I think I heard her walking to the microphone*) can be interpreted by the hearer. The result in all these cases is a rich conceptual blend in which aspects of the opera world, crucially including a fat lady, are merged with aspects of the situation being described.

- (7) a. Because a Council of the other Side asserted it was coming down. (1712)
- b. As they confidently assert that the first inhabitants of their Island were fairies, so do they maintain that these little people have still their residence among them. (1726-31)
- c. No man should have even a colour to assert that I received a compensation. (1765)
- d. He asserts that he not only invented polyphonic music, or counterpoint, but the polyplectrum or spinet, ... (1782)

The historical evidence shows that *claim* never occurs with a *that*-complement early in the eighteenth century, but when we look at the use of *claim* in the nineteenth century, we see a change in behaviour that could be interpreted as the result of a blend of schemata associated with *claim* and *assert*. The sentences from the nineteenth century in (8) show that *claim* began to appear with a *that*-complement.

- (8) a. He claimed that his word should be law. (1850)
- b. Watt claimed that Hornblower...was an infringer upon his patents. (1878)
- c. I claim that we are before them in the matter of uncapping machines [for honeycombs]. (1886)
- d. He was afraid to bet and crawfished out of the issue by claiming that he didn't drink. (1888)

Another phenomenon that may be an indicator of blending is the double copula (*is is*), as illustrated in (9).

- (9) a. And my question is, is he going to do that?
- b. She said, all it is, is a bunch of riddles.
- c. Then what we would take a look at is, is this thing scorable by this kind of a rubric?
- d. What essentially it is, is Jim and Don want us to fix the discussion on that item.
- e. So the thing is, is that's the kind of level of comparison that you get, like it or like it not, at the fourth grade level.
- f. My point is, is that their objection is a red herring.

The examples in (9a)-(9d) are syntactically unremarkable; each instance of *is* plays a distinct role in each clause. However, the double copula in (9e) and (9f) is more unusual and the second *is* seems redundant, although there are different opinions on the status of these constructions. In an extensive discussion of the source of these forms, Tuggy (1996: 715) states that “parallelism with legitimate *is is* structures” may be one source for these sentences. Tuggy suggests that source of these double copula sentences is a blend or “intersection” of utterances of the type shown in (10a) and (10b).¹⁵

- (10) a. So the thing is, it's not a diagnostic.
 b. Yes, the important thing is that the long informational would be way too much for them.

Examples like (9e) and (9f) tend to be treated as errors arising out of similarities in structure, and are thus seen as the syntactic equivalent of slips of the tongue; here we will avoid passing judgement on the well-formedness of these sentences and simply take them to be evidence not only that blending occurs in language production, but that such a blend can itself become entrenched as a new constructional form.

The unusual combination of syntactic categories in *thanks very much* and *for free*, the change in the subcategorisation of *claim*, and the repetition of the word *is* in *the thing is, is* are worthy of further discussion and analysis, but here I simply put them forward as examples in which the operation of a blending process can be discerned.

4.4 The Fat Lady

Let us turn now to a more extensive corpus-based analysis of blending and examine set expressions such as *it ain't over 'til the fat lady sings* and *make hay while the sun shines*. The strategy employed is to examine usage data in corpora to find out how these set expressions “surface” in use in order to uncover evidence of blending and provide insight into the role of formal and conceptual blending in the production of these utterances.¹⁶ The obvious advantage of working with set expressions, assuming that they truly are set expressions, is that any blending that occurs will be apparent and therefore we can use these expressions as a test bed to determine the extent and nature of blending in general and to assess the likelihood that semi-fixed expressions such as *a good thing to do* take part in blending processes. We find cases ranging from unchanged fixed expressions to minimally changed

expressions to expressions that deviate markedly from the original, but which can still be seen as modifications of the fixed expression.

Let us start by examining the phrase *it ain't over till the fat lady sings*. How is this phrase actually used in an utterance? Is the phrase plucked from memory and inserted whole, or is it sometimes one input to a blended structure? Since we are discussing blending, it will be no surprise to learn that we will concentrate on its role in blended structures. When we talk about the phrase as being an input to a blend, we have to consider both the formal and conceptual aspects of this blend. As we will see in the examples below, the use of this phrase is complex and is often, but not always, involved in both formal and conceptual blending. The main focus of the present paper is on formal blending, but, as will become obvious, this phrase, in particular, seems to serve as a catalyst for conceptual blending, a full description of which is not attempted here. The numerous sentences presented below provide an extensive sampling of the use of the phrase and give a good indication of the nature of formal and conceptual blending involved.

The set phrase occurs in its full canonical form fairly frequently, as illustrated in (11).

- (11) **"It ain't over till the fat lady sings,"** he said.
(Times/ST 96)

The reference to a fat lady singing suggests the conclusion or end to some process, but it also evokes a frame or series of events (typically described with reference to an opera or other performances) that lead up to the finale. In the examples below there are several cases in which the conceptual blending between the finale of the opera and the situation being described is not matched by an intimate formal blending. For example, the set phrase may be followed by a conjoined phrase in which links are established to the current situation, as in (12), where the set phrase is simply conjoined with a description of a celebration that ended with a performance by Kate Smith (a fat lady singer).

- (12) **It ain't over 'til the fat lady sings** and Columbus's 14th annual Fourth of July celebration, Red, White and Boom, ended just that way. (NAN 94/95)

In some of the examples there is not only the association of a fat lady with the end of a process, but, by extension, the presence of the fat lady is taken to be a necessary condition for the conclusion of the event. In the

- (15) a. **It will not be over**, as sports folk say, **until the fat lady sings**. Though at this rate, if the fat lady is knowledgeable about the Big Ten and handy with a blue pencil, she might want to stick around. (LATWP 95)
- b. "It's been said the opera **isn't over until the fat lady sings**," Gephardt said. "Last Saturday in Michigan I think I heard her walking to the microphone." (AP 88)
- c. "The opera **ain't over 'till the fat lady sings**," it said. "And we're not going to let her sing." (AP 90)

The example in (16) is similar to those in (15) except that the subject is *competition*, or rather *no competition*, and here we see some evidence of a simple formal blend in which a new noun phrase is placed into the schema.

- (16) No competition is **over until the fat lady sings**. The current status of the battle over the megaplex location suggests she is only in rehearsal. (NYT 95)

Example (17) is complex, but interesting, and while it would be an exaggeration to say that the process of constructing this utterance is open to view, it is not hard to get a good sense of how it was constructed. The speaker, DeHihns, picks up on the previous speaker's description of a project as being dead and integrates that notion in a blend with the fat lady chunk to yield *Nothing is dead until the fat lady sings*. And then, in addition, there is a link between the metaphorical fat lady and an actual person.

- (17) Asked at a news conference if the project is now effectively dead, DeHihns replied, "Nothing is dead **until the fat lady sings**, as they say. LuJuana is not fat, but she will make the final decision." (AP 90)

The following sentence is so intricately blended that it is perhaps indicative of a carefully constructed sentence of the type typically found in a journalistic style of expression. If all the examples were of this kind, one might be justified in considering blends to be an interesting aspect of a particular genre, but most of the examples do not give a sense of being contrived in this way.

- (18) In the federal government, it's never **over until the fat lady** rereads the fine print, and she moves her lips. (LATWP 95)

The form of the fat lady set phrase probably has a less direct influence on the examples listed in (19) and (20). The phrase evokes an opera scene and this opera world is transformed or blended with the current situation to produce some kind of blend involving the *fat lady* and perhaps singing, plus material indicating the described situation. These examples do not involve a merging of chunks of the sort we saw in (16) and (17).

- (19) a. The **fat lady** may not be treading the Portuguese doorstep as readily as it seems. (Times/ST 96)
- b. In any event, **the fat lady** was not ready to **sing**. (Times/ST 96)
- c. "The **fat lady isn't singing** yet, but she may be warming up," South Carolina state chairman Chris Verenes said (AP 88)
- d. ... Census Bureau director for seven months, stressed that the tally isn't finished and "**the fat lady has not sung**." (AP 90)
- e. they resumed normal business with the resilient message that **the fat lady** is still a long way from breaking into **song**. (Times/ST 97)
- f. And **the fat lady** is warming up. (Times/ST 95)

In the following the fat lady is a signal of the end of a process and this may be further interpreted in terms of consequences, possibly negative consequences.

- (20) a. "They incurred the debt, and it's time for **the fat lady to sing**." (AP 88)
- b. "Believe me, when **the fat lady sings**, the 1991 deficit will be higher than it was in 1990," says James C. Miller, co-chairman (AP 90)
- c. Sen. Ed Davis said that it is time for the Los Angeles police chief to step down. "I think **the fat lady has sung**," said Davis, R-Northridge. (SJM 91)
- d. The **fat lady** burst into **song** far too early last Sunday, when referee Pat McEnaney ... (Times/ST 96)

From the examples in this section, we see that the saying *it ain't over till the fat lady sings* participates in a variety of blends and also evokes a rich semantic field which leads to complex and colourful conceptual blends. The reality of syntactic blending appears quite clearly in the corpus data. All parts of the set expression seem to be available for modification or substitution under the appropriate circumstances and there were even examples in the corpus in which *fat lady* was modified, as in *it ain't over 'til the fat guy sings*.

4.5 Make Hay

Let us turn now to the somewhat less spectacular case of *make hay while the sun shines*. This chunk is one example of a large inventory of well-known sayings, which can be considered to be a part of a native speaker's knowledge of English. In this section we investigate how *make hay while the sun shines* is used, looking once again for indications of the nature and extent of formal blending.

Let us consider the simple representation of a schema for this basic idiom given in (21). Here the formal part of the idiom is represented as an unanalysed phrase labelled only as a sentence, S, with the approximate literal and figurative meanings rendered in capital letters. This is not to say that this particular expression even in its idiomatic interpretation is completely unanalysed, but here we are merely representing the fact that the phrase is known as a unit and has a particular meaning as a unit.

- (21) S [make hay while the sun shines]
- | | |
|----------------|-----------------------|
| MAKE HAY WHILE | TAKE ADVANTAGE OF |
| THE SUN SHINES | FAVOURABLE CONDITIONS |

The relation between the string *make hay while the sun shines* and its meaning is here represented in isolation, but presumably it is a part of a network of "take advantage" forms and meanings. The literal meaning of the *make hay* expression is shown (on the left) in (21) since the existence of a temporal component to the meaning seems to feed into many of the blends.

Insight into the process of utterance planning, starting with an intention to express an idea and moving on to the production of language, is beyond the reach of corpus analysis and here I simply want to argue that the process, whatever its nature, must be analysed in terms of chunks and blending rather than lexical categories and an X-bar style of syntactic

composition. For illustrative purposes we can give the following account. We know that the output is a form-meaning pair, which is a blend of two or more input form-meaning pairs. One of the inputs is the string *make hay while the sun shines* with its associated meaning. The other input relates to the speaker's intention to express a particular situation in which "X is taking advantage of Y" and we can assume that in this input there are potentially also some words or chunks associated with X and Y.

In the large corpora that were searched, there were, in fact, no instances of the *make hay while the sun shines* schema in its imperative form, although the examples in (22) are very similar and include the canonical form as an embedded clause.

- (22) a. "We have got to **make hay while the sun shines**," he said. (Times/ST 95)
- b. "I've got to **make hay while the sun shines**, so to speak." He is behind in harvesting his barley, he said ... (NYT 95)
- c. **Make hay while the sun shines** was the message going out from Merrill Lynch... (Times/ST 97)
- d. Long-shot Oscar nominees often try to **make hay while the sun shines**, lining up as many projects as possible between the announcement of their nomination and probable disappointment on Oscar night. (NYT 95)

These sentences are, of all the examples, most akin to the basic schema. We can see the process of blending or at least the results of blending in these utterances, with the information relevant to the current situation typically being introduced in the portion of the discourse preceding *make hay*, thereby giving information about who or what is taking advantage of current favourable conditions.

The example in (22c) shows a nominalisation of the fixed expression which is used sentence-initially to introduce a topic, with amplification of the topic following in the subsequent discourse. In (22d), as in most of the other sentences, novel material precedes the set expression, but in this particular example, information about the current situation is also given in an appended participial phrase. One could imagine a more intimate blending along the lines of "Oscar nominees try to make hay while the world waits for the results of voting," but this may have been avoided because of constraints on information flow in the discourse and/or because the

description of a time slice between the two identified events does not fit so well with a *while* phrase.

It is interesting to consider that even though the basic schema corresponding to *make hay while the sun shines* appears to native speakers to be well-entrenched, the expression in its canonical form is rare. It is reasonable to suppose, however, that the schema may have been introduced and firmly established during childhood and that while the schema is not much used directly in the language, its presence in the grammar is maintained and even reinforced by instances such as those above. In other words, from the speaker's point of view, the string is used as a prefabricated unit to structure the discourse, but does not emerge intact. And from the hearer's perspective, *make hay while the sun shines* is evoked more than it is actually heard.

Minor modifications consisting of changes to tense/aspect marking can be seen in the examples in (23) below, which are otherwise similar to the sentences in (22).

- (23) a. Could it be that Raymond Blanc, with his recently acquired three Michelin stars, is **making hay while the sun shines**? (Times/ST 95)
- b. ...Gary Lineker, the former Tottenham and England striker lately **making hay while the sun shone** with Grampus 8 in Japan. (Times/ST 95)

These examples differ only in minor ways from those in (22), but they are listed separately because they indicate more clearly than the previous examples that there must be some kind of internal analysis of the set expression. An X-bar analysis or labelling of some sort must be applied to the idiomatic string in order to modify the canonical form to fit the current circumstances. Some analysis of the string, indicated by bracketing in (24), is needed to motivate the modifications of the fixed expression that occur in (22) and (23).

- (24) νp [make [hay] [while [the sun [shines]]]]

More typically, the expression is altered or blended to a more intimate degree to package both the 'take advantage' meaning of the idiom with elements of the actual situation being described. The range of examples in (25)-(28) below illustrate a "progressive obscuring" of the original form of

the idiom by the introduction of increasingly larger chunks relevant to the particular situation.¹⁷

The first group of examples in (25) all contain a temporal phrase containing *while* and have a similar phrasal structure to the original; some of the examples also retain a reference to metaphorical weather conditions, but no reference to the sun.

- (25) a. The Chiefs aren't the only NFL team that hasn't **made hay while** their **quarterback shined**. (NYT 95)
- b. Now is the season of savers' discontent, though borrowers should **make hay while** the warmer climate remains. (Times/ST 95)
- c. Hewden Stuart's plant hire business **made hay while** the rest of the construction industry suffered from a drought. (Times/ST 95)
- d. ... American cable companies are **making hay while** British Telecom fights to protect its market share with one hand tied behind its back. (Times/ST 95)
- e. Republicans strained in their attempts to **make political hay while** Democrats were obsessive in their defensive blocking. (NYT 95)
- f. the big food groups are opening selected stores 24 hours a day in a bid to **make hay while** consumer confidence continues to improve. (Times/ST 96)
- g. **While** Mr Soros apologises, the ladies of Beardstown are **making hay**. (Times/ST 95)

In these examples the favourable conditions expressed idiomatically by *while the sun shines* are replaced by a description of the actual event such as, from (25f), *while consumer confidence continues to improve*. Thus there is a more intricate blend, with information from the current situation being added in both the subject position and in the temporal phrase.

All the examples in (26) contain *make hay* plus a temporal phrase, but the temporal phrase in these sentences is not introduced by *while*.

- (26) a. Opponents of the Tory Right have **made hay since** the Conservative leadership election last month. (Times/ST 95)
- b. Leeds **made hay in the first half** ... (Times/ST 95)

- c. **During his life**, they **made hay**. No piece detailing the marriage of the millionaire chairman of Hanson and the former model Miss Tucker was complete without carping references to the four decades separating them. (Times/ST 95)
- d. ...packaging companies have a **small window in the cycle** in which to **make hay**. (Times/ST 95)
- e. one of the areas where the group has **made hay in recent years**. (Times/ST 95)
- f. ...This field is wide open to Labour. If Tony Blair cannot **make hay** in such **political sunshine**, how will he fare when winter comes? (Times/ST 96)

Still other examples do not have any temporal component at all and instead focus simply on the meaning 'take advantage.' In these cases there is some suggestion of a more complex blending based not only on *make hay while the sun shines* but also on expressions such as *make money* (27) and *make runs* (at cricket) (28).

- (27) a. ...one of the areas where the group has **made hay** in recent years. (Times/ST 95)
- b. ...France, which has seen her neighbour **making hay** on the back of a weak currency (Times/ST 95)
- c. ...one of the areas where the group has **made hay** in recent years. (Times/ST 95)
- d. In the meantime, the banks continue to **make hay**. (Times/ST 97)
- e. ...those companies with money to spend will **make hay**. (Times/ST 97)
- (28) a. However, they survived to **make hay** against far less experienced operators. Butcher hit 20 fours and a six in his... (Times/ST 96)
- b. Slater will **make hay** against county bowling if he conquers his impetuosity on green-tinged mornings... (Times/ST 96)

In some examples, there is an idea of taking advantage politically and, in fact, *make political hay* is a collocation. This usage, illustrated below in (29) and (32), seems particularly common in the American corpora.

- (29) a. A rabble-rousing Scottish parliament could **make hay with** Tony Blair's refusal to turn the clock back. (Times/ST 97)
- b. Still, Labour is hardly in a position to **make hay**. (Times/ST 97)
- c. And they accused Dole of using a critical foreign policy issue to "**make political hay**," as White House press secretary Mike McCurry put it. (LATWP 95)
- d. Old Labour **makes hay** in Brussels (Times/ST 96)

The types of sentences illustrated in (27)-(29) are so far removed from the source expression and are perhaps being influenced by these other phrases (*make money*, *make runs*, etc.) to the extent that it is not at all clear that the expression *make hay while the sun shines* plays a role. A schema covering examples (27)-(29) associated with a simple 'take advantage' meaning is given in (30).

- (30) VP [make [hay] [ADJUNCT]]

The many examples listed above are interesting in themselves, but it is worth reiterating that the main point to be drawn from these sentences is that blending is clearly a part of sentence construction. If we find widespread blending in which an idiom such as *make hay while the sun shines* combines with other form-meaning pairs, then we have no reason not to expect that blending also occurs in those situations in which the input to the blend is harder to identify.

On a final note on *make hay*, we see that in American usage (as exemplified by the examples from the New York Times and other U.S. newspapers) there is much less focus on the temporal aspect, less focus on the weather, and more of a focus on the process of making hay from or out of something. In other words, the *hay* part of *make hay* in American usage is much more likely to be seen as a product and part of an almost industrial process. The example in (31) illustrates very well this manufacturing view of making hay.

- (31) The GOP intends to **make hay with** whatever **sunshine** the committee provides. (AP 89)

Other typical uses from American newspapers are given in (32). Some of these sentences include the phrases *make hay out of* or *make hay of*,

which are totally absent from the British examples and may be influenced by a schema along the lines of *make wine out of grapes*. Here again it is not clear whether there is, in these cases, a strong connection with the full idiomatic expression *make hay while the sun shines*, or whether these are now independent constructions, linked only tenuously to the full idiom. The example above in (31) might be taken as an indication that there is some connection between the shorter *make hay* construction and *make hay while the sun shines*. On the other hand, many of following examples in (32) are associated with the accrual of a political advantage mentioned above—an overtone that is not a part of the longer form of the fixed expression.

- (32) a. But his political adviser, Lee Atwater, warns that rivals who try to “**make hay**” at Bush’s expense “may find it backfires.” (ACLDCI)
- b. The Democrats, who generally oppose term limits, did their best to **make hay of** the Republicans’ multiple measures on term limits, ... (NYT 95)
- c. And not only are respected public figures calling for an amnesty; some politicians have even **made hay of** their Stasi ties. (LATWP95)
- d. ...now find themselves subjected to the repugnant spectacle of their national leaders angling to **make hay out of** the tragedy. (NYT 95)
- e. they realize that the president could **make political hay out of** any Democratic effort to repudiate Gramm-Rudman. (ACLDCI)

The above discussion shows the potential of using corpora of different speech communities to elucidate the differences between the linguistic systems of those communities. A closer look at the American vs. British corpora might reveal some very subtle differences between the usages of otherwise very similar-looking expressions.

In this section we have seen a broad range of examples involving the use of two well-known idioms and on the basis of these results we can say that blending is an important constructional process. The details of the blending process, as well as cross-dialectal variation, provide scope for deeper investigation. Another issue worth pursuing is potential differences between blending processes associated with specific expressions. For example, while the corpus examples showed that both of the sayings we investigated were involved in formal and conceptual blending, there were

clearly differences between the two. The *make hay* saying appears to take part in formal blends to a greater extent than the *fat lady* expression, but it does not evoke a general hay-making conceptual domain that matches the rich opera/performance world associated with *the fat lady*.

5. Conclusion

In this paper we have adopted a perspective on a corpus as a record of language usage events and explored the view that corpus analysis can lead to insights into the nature of usage-based grammar. One characteristic of corpus data is the ubiquity of collocations of various kinds, and it is reasonable to assume that these frequent word combinations have some degree of unit status or cognitive reality in a usage-based grammar. The question explored here is the source of strings such as *a worthwhile thing to footnote* which clearly cannot be analysed as some sort of chunk, but which may be the result of a blending process, with one of the inputs being a construction or some kind of more or less fixed expression such as *a good thing to do*. In order to approach this topic we looked at the way that the fixed expressions *it ain't over 'til the fat lady sings* and *make hay while the sun shines* "surfaced" in usage and the results demonstrate that while the processes involved are complex, it is reasonable to conclude that formal (and conceptual) blending of stored units plays an important role in accounting for the creative aspects of language in use.

Notes

I am grateful to Suzanne Kemmer for comments on earlier drafts of this paper.

1. The usage examples that are not explicitly identified all come from the Corpus of Spoken Professional American English (CPSAE) (Barlow 1998), a 2 million word corpus based on spoken transcripts. To find examples of the idioms discussed, hundreds of millions of words of corpus data were searched. The data sources used in addition to CPSAE were the following:

NYT	<i>New York Times</i>
Times/ST	<i>The Times</i> and <i>Sunday Times</i> (of London)
AP	Associated Press
NAN	North American News (major U.S. newspapers)
LATWP	<i>Los Angeles Times</i> and <i>Washington Post</i>

ACLDCI Brown Corpus and *Wall Street Journal*

Each year of *The Times/Sunday Times* contains around 40 million words. Much of the data was accessed at the LDC Online site (www ldc upenn edu ldc online index html)

See Langacker (1987, 1988, this volume) for discussion of the relation of grammar and usage.

There is a considerable terminology for recurrent word patterns: collocations, idioms, sayings, fixed expressions, prefabricated units, chunks, lexical phrases. In this paper the distinctions among types of recurrent word patterns are not relevant and hence these terms are used more or less interchangeably. See Moon (1998) for discussion of terminology in this area.

The view of grammar as consisting of a stored set of language units or routines has been proposed by researchers working in a variety of paradigms. See, for example, Aijmer (1996), Bybee and Scheibman (forthcoming), Croft (1995), Haiman (1994), Hopper (1987), Langacker (1987), Peters (1983), and Sinclair (1991). Sinclair (1987: 319-320) makes a distinction between the open-choice principle and the idiom principle. The open-choice principle refers to the choices in terms of words and structures available to speakers in constructing utterances. Sinclair notes (1987:320) that "the open choice principle does not provide substantial enough restraints" and proposes, in addition, the idiom principle: "the language user has available [...] a large number of semi-preconstructed phrases that constitute single choices, even though they appear to be analysable into segments."

Sag and Wasow (1999: 416) state: "Most contemporary syntactic theories have preserved the most important innovations of the Standard Theory, namely, syntactic features, recursive phrase structure, and some sort of semantic component."

See Church et al. (1991) for discussion of t-score and mutual information, Kita et al. (1994) for their "cost" measure, and Manning and Schütze (1999) for an account of entropy and other measures.

Different kinds of description of schemata are presented in Langacker (1987); Barlow and Kemmer (1994), Barlow (1996), Fillmore et al. (1988), among others.

See, for example, Bybee (1999), Bybee and Scheibman (forthcoming), Croft (1995), and Haiman (1994).

9. I assume that X-bar syntax, which is organised to some extent on semantic grounds, is similar to the more abstract forms of schemata proposed here. Jackendoff (1994: 22, quoted in Goldberg 1996: 8) states "one might want to view the 'core rules' of phrase structure for a language as maximally underspecified constructional idioms." See also Langacker (this volume).
10. Example taken from the *New York Times*, Oct. 22, 1999.
11. Sag and Wasow (1999: 265-269) discuss some idioms such as *take advantage* and *kick the bucket* and in their HPSG representations it is, in fact, quite clear how the idiom is combined with other syntactic components. However, one problem with the HPSG analysis is that these phrases cannot be modified. Thus it seems that the grammar would not be able to generate *take full advantage* or *kick the proverbial bucket* except by listing them as completely separate from their non-modified equivalents. But more seriously, the extent to which collocational expressions must be accounted for within a realistic grammar is not acknowledged. However, the view of grammar as a collection of "signs" in HPSG is promising and suggests that the theory could potentially handle rich collocational connections.
12. There are also meta-blending processes of various kinds, such as the proposal by Kemmer and Verhagen (1994) that causative structures are based on other more basic structures such as transitive clauses.
13. [NP V NP PP] is from Fauconnier and Turner (1996: 117) and [NP(x) be NP(y) of NP(z)] is from Turner (1996: 104).
14. Turner and Fauconnier (1995) mention lexical blends such as *chunnel*. See Kemmer (forthcoming) for a schema-based account of lexical blends at both the conceptual and formal level.
15. According to Tuggy's terminology, the example in (10a) is a Focus formula (Tuggy 1996: 724) and example (10b) illustrates a one-"be" construction (Tuggy 1996: 713). The examples in (9) and (10) are taken from CPSAE.
16. The examples used in this section are all from written corpora, although some of the examples are quotes. We will have to wait for the creation of very large spoken corpora to fully determine the prevalence of blends based on fixed expressions.
17. The reasons underlying the choice of relatively unaltered idiom versus a relatively obscured idiom remain to be elucidated. A more integrated or

highly blended structure is denser in terms of information content, and this may be appropriate for some situations. In other situations the relatively unadulterated, unblended idiom may be useful as a way of introducing a general proposition, the particular details of which can then be fleshed out as the discourse develops.

References

- Aijmer, Karin. 1996. *Conversational Routines in English: Convention and Creativity*. Harlow, Essex: Addison Wesley Longman.
- Barlow, Michael. 1996. Corpora for theory and practice. *International Journal of Corpus Linguistics* 1(1), 1-37.
- Barlow, M. 1998. *Corpus of Spoken Professional American English*. Houston: Athelstan.
- Barlow, Michael and Suzanne Kemmer. 1994. A schema-based approach to grammatical description. In Roberta Corrigan, Gregory Iverson and Susan Lima (eds.), *The Reality of Linguistic Rules*. Amsterdam: Benjamins.
- Benson, Morton, Evelyn Benson, and Robert Ilson. 1997. *The BBI Dictionary of English Word Combinations*. Amsterdam: Benjamins.
- Bolinger, Dwight. 1961. Syntactic blends and other matters. *Language* 37(3), 366-381.
- Bybee, Joan. 1999. The emergent lexicon. Ms. University of New Mexico.
- Bybee, Joan and Joanne Scheibman. Forthcoming. The effect of usage on degrees of constituency: The case of *don't* in English. In Susana Cumming (ed.), *Constituency and Discourse*. Amsterdam: Benjamins.
- Church, Kenneth, William Gale, Patrick Hanks, and Donald Hindle. 1991. Using statistics in lexical analysis. In Uri Zernik (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Coft, William. 1995. Intonation units and grammatical structure. *Linguistics* 33, 839-882.
- Connier, Gilles and Mark Turner. 1996. Blending as a central process of grammar. In Adele E. Goldberg (ed.), *Conceptual Structure, Discourse, and Language*, 113-130. Stanford: Center for the Study of Language and Information (CSLI).
- Emore, Charles J., Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501-38.
- Firth, J. R. 1957. A synopsis of linguistic theory. *Studies in Linguistic Analysis*, 1-32. Special volume. Philological Society.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele E. 1996. Jackendoff and construction-based grammar. *Cognitive Linguistics*, 7(1), 3-19.

- Haiman, John. 1994. Ritualization and development of language. In William Pagliuca (ed.), *Perspectives on Grammaticalization*, 3-28. Amsterdam: Benjamins.
- Hopper, Paul J. 1987. Emergent grammar. *BLS* 13, 139-157. Berkeley: BLS.
- Jackendoff, Ray. 1994. The boundaries of the lexicon. Unpublished manuscript. Brandeis University.
- Kemmer, Suzanne E. Forthcoming. Lexical blends. In Hubert Cuyckens, Thomas Berg, Rene Dirven, Klaus-Uwe Panther (eds.), *Motivation in Language: Studies in Honour of Günter Radden*. Amsterdam: John Benjamins.
- Kemmer, Suzanne and Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5, 115-156.
- Kita, Kenji, Yasuhiko Kato, Takashi Omoto and Yoneo Yano. 1994. Automatically extracting collocations from corpora for language learning. In Andrew Wilson and Anthony McEnery (eds.), *Corpora in Language Education and Research*. UCREL Technical Papers, Vol. 4, 53-64. Lancaster: Department of Linguistics, Lancaster University.
- Kjellmer, Goran. 1994. *A Dictionary of English Collocations: Based on the Brown Corpus*. Oxford: Oxford University Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, Vol. 1: *Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1988. A usage-based model. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics* (Current Issues in Linguistic Theory 50), 127-61. Amsterdam: Benjamins.
- Langacker, Ronald W. This volume. A dynamic usage-based model. In Michael Barlow and Suzanne E. Kemmer (eds.), *Usage-Based Models of Language*. Stanford: CSLI.
- Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: The MIT Press.
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Peters, Ann M. 1983. *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Sag, Ivan A. and Thomas Wasow. 1999. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI.
- Sinclair, John M. 1987. Collocation: A progress report. In Ross Steele and Terry Threadgold (eds.), *Language Topics: Essays in Honour of Michael Halliday*, Vol. II, 319-331. Amsterdam: Benjamins.
- Sinclair, John M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Tuggy, David. 1996. The thing is is that people talk that way. The question is is Why? In Eugene H. Casad (ed.), *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*, 713-752. Berlin: Mouton de Gruyter.
- Turner, Mark. 1996. *The Literary Mind*. New York and Oxford: Oxford University Press.
- Turner, Mark and Gilles Fauconnier. 1995. Conceptual integration and formal expression. *Metaphor and Symbolic Activity* 10(3), 183-203.