FORMULAIC SEQUENCES IN FUNCTIONAL AND COGNITIVE LINGUISTICS*

Christopher S. Butler
Swansea University

ABSTRACT

This paper attempts to resolve the tension between an approach to language in which lexical items are matched individually with slots in frames provided by the syntax of a language, and one which holds that much of our language consists of recurrent, reusable multiword chunks, with differing degrees of variability, and often with rather ill-defined boundaries. The properties of three formulaic sequences are briefly described, and then four linguistic approaches are examined to determine to what extent they can account for these properties. It is concluded that all the approaches fail to accommodate semantic prosodies which can extend over ill-defined stretches of language. A model is proposed in which associative patterns at different levels of description are linked by constraint satisfaction mechanisms.

KEY WORDS: Formulaic sequence, corpora, functionalism, cognitivism, semantic prosody.

RESUMEN

Este artículo intenta resolver la tensión entre un enfoque en el que las unidades léxicas se asocian con posiciones determinadas dentro de los marcos sintácticos, y otro que mantiene que gran parte del lenguaje está impregnado de piezas complejas recurrentes y reutilizables que muestran distintos grados de variabilidad y cuyos límites no están siempre definidos con claridad. Se describen las propiedades de tres secuencias formulaicas y se examinan cuatro enfoques lingüísticos y su capacidad para dar cuenta de estas propiedades. Se concluye que ninguno de estos enfoques es capaz de acomodar diferencias de prosodia semántica y se propone un modelo alternativo en el que patrones asociativos que funcionan en distintos niveles de descripción se interrelacionan según mecanismos de satisfacción de restricciones.

PALABRAS CLAVE: secuencia formulaica, corpus, funcionalismo, cognitivismo, prosodia semántica.

1. INTRODUCTION: TWO VIEWS OF LINGUISTIC STRUCTURE

The principal aim of this article is to discuss the tension between two views of linguistic structure, and to take some initial steps towards reconciling them. Grammars have often been constructed according to the "open choice" principle,

according to which lexical items are matched individually with slots in frames provided by the syntax of a language. On the other hand, work in corpus linguistics (especially that associated with Sinclair and his colleagues)[1] suggests that this is a grossly oversimplified, indeed potentially misleading view, if we take as our source of data for description the attested productions of native speakers of a language. Such language, as opposed to the neatly packaged constructed sentences of the armchair linguist, consists to a considerable extent of recurrent, reusable multiword chunks, with differing degrees of variability, and often with somewhat ill-defined boundaries, if we take into account not only syntax but also syntagmatic lexical patterning (collocation) and, above all, meaning. There has been much discussion of idioms and other aspects of phraseology in the literature, from a variety of theoretical and applied points of view (Nunberg, Sag and Wasow; Moon; Cowie; Taylor, Ch. 27; among others), but it is corpus analysis which has revealed that actual usage systematically makes use of much subtler devices than are generally discussed by writers of grammars. So pervasive is the use of semi-preconstructed chunks that Sinclair has proposed that the open choice principle on which most grammars are based should be supplemented by an **idiom principle**, stated as follows:

> The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. (Sinclair, *Corpus* 110)

Furthermore, the work of Wray, with roots in psycholinguistic and sociolinguistic modes of explanation, comes to similar conclusions, postulating a key role, in "normal" adult language, child language acquisition, and the language of aphasics and second/foreign language learners, for reusable, holistically stored and processed formulaic sequences. Wray defines a **formulaic sequence** as,

> [A] sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (Wray 9)

All this is, of course, of little consequence for those who largely rely on native speaker intuition as a source of data and follow Chomsky's line, namely that since "linguistic theory is mentalistic,"

> Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of the this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline. (Chomsky 4)

---

* I am grateful to Gordon Tucker and Francisco Gonzálvez García for comments which led to improvements in this paper.

[1] See, for example, Sinclair, *Corpus*, "Search", *Trust*.

But for functionalists, and also for proponents of Cognitive Linguistics, including some types of constructionist approach, the patterns revealed by corpus analysis are, or should be, a serious matter, since these linguists adopt a perspective which involves seeing language as essentially a form of human communication, and a primary aim, at least for some proponents of such approaches, is to explore the structure (and also the effects) of the actual usage events engaged in by speakers and writers, hearers and readers, in their linguistic interactions.[2]

I shall first discuss briefly three formulaic expressions which will be used for illustration in the rest of the article. I shall then look at four approaches, within functional and cognitive linguistics, which have taken the issue of formulaic language seriously, the aim being to assess the extent to which they are able to account for the properties of formulaic constructions. The conclusion from this survey is that the models examined are able to account for only those properties which can be described in terms of the constituent structure of expressions, so that any phenomena which operate over stretches of language not coextensive with such constituents remain unexplained. I then propose an alternative way of approaching formulaic phenomena, based on the concept of syntagmatic associations operating at different levels, unified by constraint satisfaction.

2. THE DATA

The extensive literature on idiomatic, formulaic language shows clearly that there is a cline from totally fixed expressions such as *by and large* at one end (Taylor 543, *inter alios*) to looser collocational patterning at the other. Most formulaic expressions, however, display some degree of variability.

The first formulaic expression we shall look at is *COME a cropper*, with the meaning "have an unexpected, embarrassing, and disastrous failure" (*Collins COBUILD Dictionary*). Detailed examination of the 50 idiomatic occurrences of *cropper* in the British National Corpus, World Edition (henceforth BNC) reveals the following variants in addition to the basic form:

came a 5 & 3 cropper [in a sports match]
come an almighty cropper
came the most appalling cropper
gonna come a right bloody cropper
came a complete cropper
has come a catastrophic cropper
has come the most frightful cropper
had nearly come a nasty cropper
came such a cropper

[2] For discussion of the roles of corpora in functional linguistics, see Butler, "Corpus".

This shows that although the default is to use *COME a cropper*, suggesting a possible fixed unit (except for variations of tense, aspect and person) it is possible to treat the sequence analytically, in order to insert a modifier of the noun which is normally concerned with the size or disastrous nature of the failure.[3] The modifier is normally (i) an ordinary adjective, (ii) an adjective indicating size or disastrous nature premodified by the degree word *most*, or (iii) the predeterminer *such*. A further point to note is that in the default form of the expression, and also with modification just by an adjective or predeterminer, the article is indefinite *a*, an example of **colligation**, the preference of an item for a particular grammatical item. However, with *most* premodifying the adjective, the article changes to the definite *the*, despite the fact that expressions such as *came a most appalling/frightful cropper* would seem to be grammatical.

Syntactically, we could either treat *COME* as, exceptionally, a (non-passivisable) transitive verb, or regard the idiom *COME a cropper* as occupying a slot which would normally be occupied by a single intransitive verb. However, the fact that the NP can be modified suggests that we need a schema with an open slot, into which only a restricted range of items can be inserted, as shown in (1).

(1)	*COME* (*such*) *a*	(adjective)	*cropper*
	the most	adjective of size or seriousness	

Secondly, let us consider an example where the co-occurrence patterns are more complex, and defy packaging into neat bundles. The BNC contains 70 examples of the sequence *bare hands*, and only 3 of the singular *bare hand*. Taking both forms together, the dominant determiners are possessives (58/73 = 79%), almost all other occurrences having a zero determiner (13 = 18%) with only single occurrences of indefinite and definite articles. For the plural form *bare hands*, 62/70 (= 89%) occur within a PP with *with* as the preposition, acting instrumentally, the others being phrases acting as subject, object, complement or *by*-agentive. Expressions centred around *bare hands* show what Sinclair terms a **semantic preference** for verbs indicating force (50%), the proportion rising to 75% if the lexical head of the direct object is also taken into account. There is also a **semantic prosody** of difficulty attached to this construction[4]: not only does the use of force imply difficulty, but there is also often a further indication of this in the surrounding context,

[3] The view of formulaic sequences as analysable constructs is supported by psycholinguistic evidence on the processing of such sequences (for a summary, see Gibbs).

[4] Note that the term **semantic prosody** is being used here with the meaning given to it in Sinclair's work, in which it "expresses something close to the 'function' of the item – it shows how the rest of the item is to be interpreted functionally" (Sinclair, "Search" 87-88). The term has also been used (see e.g. Partington 66ff, also Stubbs, *Words* 65-66, who himself prefers the term **discourse prosody**) to refer to attitudinal features revealed by the list of single words with which a particular word collocates (e.g. CAUSE collocates strongly with words indicating unpleasant events – see Stubbs, "Collocations"). For discussion of different approaches to semantic prosody, see Hunston, "Semantic".

for instance, the conjunction of *bare hands* with more effective means, for example, *using daggers, tridents and bare hands* or *with buckets, shovels and bare hands*; the use of *nothing/little more than* in front of *bare hands*; or *if it meant* (*if it meant digging the hole with my bare hands*). Expressions with *bare hands* are in some ways similar to, but in other ways different from, those with *naked eye* studied by Sinclair ("Search"). Table 1 compares the two types of expression.

TABLE 1: COMPARISON OF *NAKED EYE* AND *BARE HANDS*.

FEATURE	*NAKED EYE*	*BARE HANDS*
Number of noun	Singular	plural
Determiner	definite article	possessive/zero
Preposition before det	yes: mainly *with*, *to*	Yes: mainly *with*
Semantic preference	Visibility	force
Semantic prosody	Difficulty	difficulty

The important difference between the *naked eye* and the *bare hands* examples and those considered earlier is that we have here a semantic prosody, in this case one of difficulty, which is not confined to a constituent with some particular grammatical and/or semantic function but operates over a wider span which cannot be defined in terms of constituency.

In the next four sections, four approaches to formulaic constructions are discussed, in order to determine to what extent they can cope with the examples just presented.

3. CONSTRUCTIONIST APPROACHES

A natural place to start looking for ways of accommodating the corpus findings within a grammar is the set of approaches which are subsumed under the term "constructionist."[5] As pointed out by Croft and Cruse (Ch 9) in the useful though brief guide to these approaches on which the following discussion is based,

[5] I use the term "constructionist approach" to refer to the whole set of approaches which are based on the construction as a pairing of form and meaning, including, for example, the unification-based Construction Grammar of Fillmore and his colleagues (Fillmore, Kay and O'Connor), the model of Goldberg (*Constructions*), the Cognitive Grammar of Langacker and the Radical Construction Grammar of Croft. For reasons of space, I shall concentrate here on the approach of Fillmore Kay and O'Connor, which I shall refer to simply as Construction Grammar, and that of Goldberg.

the Construction Grammar of Fillmore and his colleagues, which, together with Langacker's Cognitive Grammar, was instrumental in stimulating the development of grammars in which the construction is central, came about largely through an attempt to deal with idioms.

Fillmore, Kay and O'Connor, after presenting a classification of idioms, proceed to examine in some detail what they describe as formal, lexically open idioms, illustrating their discussion by means of a detailed analysis of the *let alone* construction, as in *I barely got up in time to* EAT LUNCH, *let alone* COOK BREAKFAST. What this analysis shows is that the *let alone* construction, although similar to other constructions in various ways, has its own set of properties, at syntactic, semantic and pragmatic levels, which are not entirely predictable from more general principles operating at these levels. As Croft and Cruse (240) observe, later work has brought to light many other constructions whose properties cannot be predicted from the individual constituents of the constructions or other constructions in the language under description, so motivating their assignment to the construction as a free-standing theoretical entity.

In the conclusion to their article, Fillmore, Kay and O'Connor (534) say the following:

> It has seemed to us that a large part of a language user's competence is to be described as a repertory of clusters of information including, simultaneously, morphosyntactic patterns, semantic interpretation principles to which these are dedicated, and, in many cases, specific pragmatic functions in whose service they exist.

Construction Grammar thus recognises that there are patterns at the morphosyntactic, semantic and pragmatic levels, all of which need to be brought together in specifying the properties of a construction.[6]

Fillmore, Kay and O'Connor (534) also express the hope that the machinery for dealing with idioms may also be applicable to the grammar as a whole. This, of course, is the basic claim of constructionist approaches. In another key paper, Kay and Fillmore demonstrate that the *What's X doing Y?* (WXDY) construction, as in *What is this scratch doing on the table?*, although having its own unique semantic interpretations and morphosyntactic properties, and so qualifying as a construction in its own right, interacts with other constructions (the VP, the left isolation construction, etc.) to give the final forms in which the WXDY construction can appear.

Constructions containing varying degrees of idiomatic material have also been studied in other variants of construction grammar which arose under the stimulus of the work of Fillmore and his colleagues. Goldberg (*Constructions*), for instance, devotes a whole chapter of her book to the *way* construction, as in *Frank*

---

[6] It should be noted that later work by Fillmore and his colleagues has introduced some changes to the model originally put forward in the *let alone* paper. Space constraints preclude further discussion here.

dug his way out of the prison, in which "possessive + *way*" is obligatory and unpredicatable. She shows that there is a basic interpretation involving the means for creation of a path (as in the example just given), and a less basic interpretation involving manner (e.g. *He seemed to be whistling his way along*), and provides a formalisation of these variants. For instance, the means interpretation involves, at the semantic level, a CREATE-MOVE predicate, with "creator-theme," "createe-way" and "path" as its arguments, while at the syntactic level these are mapped on to a Verb, Subject, Obj_{way} and Oblique respectively. She also discusses semantic constraints on the construction: the verb must represent repeated action or unbounded activity; the motion must be self-propelled and directed. Furthermore, Goldberg explicitly makes use of the concept of simultaneous satisfaction of constraints when she says that "[c]onstructions are combined freely to form actual expressions as long as they are not in conflict" (Goldberg, *Work* 12). Note, however, that this applied to combinations of constructions, rather than to constraints at different levels on particular constructions.

We see, then, that constructionist approaches not only treat idiomatic expressions as an important, rather than a merely peripheral, part of language, but that they also recognise clusters of morphosyntactic, semantic and pragmatic properties which attach to them and differentiate them from other pieces of structure, so motivating the postulation of a separate construction,[7] and also providing a mechanism which, it is claimed, is equally operative within what Chomskyans would call the "core" grammar. We have also seen that a constraint satisfaction mechanism operates in relation to the combination of different constructions, and that Construction Grammar describes the native speaker's competence as consisting of clusters of simultaneously operative syntactic, semantic and pragmatic properties. How, then, might such grammars cope with the three formulaic sequences selected for exemplification here, *COME a cropper* and expressions centred around *bare hands* and *naked eye*?

Idioms of the *COME a cropper* kind are syntactically, semantically and pragmatically irregular, and must be listed at the lexical end of the so-called ***construct-i-con***, the total set of constructions in a language, which in some constructionist models (e.g. that of Goldberg) includes individual word-sized lexical items. It would presumably not be difficult to specify conditions on, for example, the need for an indefinite article except in cases where there is a superlative adjective (e.g. *came the most appalling cropper*).

With our *bare hands* and *naked eye* examples, however, things get somewhat trickier. As pointed out in section 2, these expressions act as the core of extended units of meaning whose boundaries are fuzzy rather than discrete, in that the semantic prosody of difficulty which is associated with the idiom may appear not

[7] It should be noted that individual constructionist approaches differ somewhat in the exact nature of what they recognise as a construction. For discussion, see Croft and Cruse (Ch. 10), Goldberg (*Work* Ch. 10), and Gonzálvez-García and Butler.

only in various lexical guises, but also at varying distances and varying positions with respect to the core. This is difficult in a model, such as the various constructionist approaches, in which the boundaries of any particular construction must be clearly specified. In order to account for the *bare hands/naked eye* type of phenomenon, a *sine qua non* of our grammar must be a syntagmatic component which takes account of probabilistically characterised lexical co-occurrence, and this so far appears to be lacking in constructionist approaches, in the sense that the only way of indicating co-occurrence is to specify particular lexical items which are obligatory for the construction.

To summarise, then, constructionist approaches recognise that clusters of morphosyntactic, semantic and pragmatic properties must be brought together in relation to a particular construction, but individual constructions, as studied in constructionist approaches, have precisely defined boundaries, making it difficult to see how the more diffuse constraints involved in semantic prosody could be accommodated.

4. THE COLLOSTRUCTIONAL ANALYSIS

Stefanowitsch and Gries present a model which combines the assumptions of constructionist approaches (in particular, the approach developed in the work of Lakoff (1997) and Goldberg (*Constructions* with corpus-based collocational analysis. This **collostructional** approach follows up on Goldberg's (*Constructions*) demonstration that grammatical constructions provide a meaning of their own, which interacts with that of the lexical items which occur in the construction. For instance, although *hit* does not itself have the concept of transfer as part of its meaning, it can be combined with the ditransitive construction in expressions such as *Pat hit Chris the ball*, where transfer is clearly a part of the overall meaning, which must have come from the construction itself. The aim of collostructional analysis is to explore in detail the associations between constructions and the lexical items which occur in them in corpora. The analysis begins with the isolation of corpus examples of particular constructions, and then examines which lexical items are strongly attracted to, or repelled from, that construction, or rather a particular slot within it, as determined by the results of a Fischer exact test applied to the two-by-two table which contains the single and joint frequencies of the construction and the associated word, or **collexeme**. For instance, Stefanowitsch and Gries investigate the construction *N waiting to happen*, drawing up a table in which the cells represent the frequency of co-occurrence of a particular noun, say *accident*, with *waiting to happen*, the frequency with which *waiting to happen* occurs in the absence of *accident*, and also the frequency of *accident* in the absence of *waiting to happen*, and the frequency of other relevant words (in this case verbs) which have nothing to do with either of the elements under investigation. The value of the Fischer exact probability is then taken as a measure of collostructional strength, the lower the value, the stronger being the bond. In this particular case, the results show that *accident* and *disaster* are by far the most strongly attracted collocates.

Interestingly from the perspective of the present article, Stefanowitsch and Gries also investigate the behaviour of a single word, the verb *cause*, whose attitudinal collocational preferences have been studied in previous work, and a partially variable idiom, the *X think nothing of* V_{gerund} construction. The results for *cause* confirm that it co-occurs strongly with items which have a negative connotation. However, the authors also perform a more detailed analysis which isolates the words most strongly associated with each of the grammatical constructions with which verbal *cause* is associated, namely transitive (*it's progressively caused slight breathing problems*), prepositional dative (*it caused harm to others*) and ditransitive (*I am sorry to have caused you some inconvenience*). Although all three constructions attract collexemes with negative connotations, the transitive occurs only, and the prepositional dative mainly, with external states and events, while the ditransitive occurs chiefly with mental states and experiences. For *X think nothing of* V_{gerund}, the ranked collexemes reveal some verbs which, in appropriate contexts, could denote activities which could be risky or otherwise potentially undesirable.

Stefanowitsch and Gries go on to demonstrate the usefulness of their technique with more abstract constructions: the *into*-causative (e.g. *He tricked me into employing him*), the ditransitive, progressive aspect, the imperative and past tense. More recently, the collostructional technique has been used to study a variety of "alternations": the dative alternation (*John sent Mary the book* vs. *John sent the book to Mary*), active and passive, word order in verb-particle constructions (*John picked up the book* vs. *John picked the book up*), markers of futurity (*will* vs. *be going to*), and alternative ways of indicating possession (*s*-genitive vs. *of* construction) (Gries and Stefanowitsch).

The collostructional technique is undoubtedly of great value in empirical studies of the relationship between constructions and the lexical items which occur in them, making the study of collocation more precise by relating it to particular structures. It is also clear that such studies yield useful data in relation to the attitudinal associations of particular lexical items. However, because the technique deliberately anchors collocates to slots in particular constructions, it will not tell us anything about the context surrounding those constructions and this, we have seen, is crucial for the analysis of expressions such as those centred on *naked eye* or *bare hands*.

5. JACKENDOFF'S PARALLEL ARCHITECTURE MODEL

We have seen that constructionist approaches are based on the concept of constraint satisfaction. Another model based on the same principle, and itself showing strong constructionist tendencies, is the parallel architecture model developed by Jackendoff, whose latest manifestation is the Simpler Syntax model of Culicover and Jackendoff. Although still formalist in the sense of upholding the importance of Universal Grammar (though as a guide to acquisition rather than determining it) and postulating the autonomy of syntax from the other two levels, Culicover and Jackendoff's model rejects four of the key postulates of mainstream generative grammar, substituting for them claims which were long ago made by functionalists and

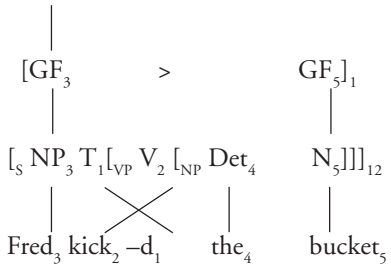
cognitivists. Firstly, as has already been pointed out, the model is constraint-based rather than derivational; secondly, there are no hidden levels of syntax, such as "deep structure"; thirdly, the model rather than being syntactocentric as in mainstream generative grammar, is organised on the principle that conceptual semantics, syntax and phonology all have their own autonomy, and that the representations they generate are mapped on to one another through the simultaneous satisfaction of constraints at the three levels; and fourthly, grammar and the lexicon are seen as a continuum rather than as separate components of the model (Culicover and Jackendoff 14-15).

Idioms and other "syntactic nuts" (the term is from Culicover) were seminal to the development of this model, as they were in constructionist approaches. As Culicover and Jackendoff (25) point out, these aspects of language, regarded by mainstream generativists as "peripheral" rather than belonging to the "core grammar," turn out to be very numerous, and present at least as many problems for children acquiring a language as does the "core." Culicover and Jackendoff therefore pursue the line that a theory of learning which is capable of explaining how the lexicon and the "peripheral" elements of languages can be acquired should, in principle, be applicable to the learning of the "core" too.

Within the Simpler Syntax model, parallel structure accounts are given for illustrative idiomatic constructions. (2) below shows the lexical entry for the idiom *KICK the bucket* meaning DIE, while (3) shows how this can be integrated into a sentence. Both are taken from Culicover and Jackendoff (225, example (64) a and b).

(2) $[\text{DIE (X)}]_2 \leftrightarrow [_{\text{VP}}\ \text{V}_2\ [_{\text{NP}}\ \text{Det}_4\ \text{N}_5]]_2 \leftrightarrow \text{kick}_2\ \text{the}_4\ \text{bucket}_5$

(3) $[\text{PAST}\ ([\text{DIE}\ (\text{FRED}_3)]_2]_1$

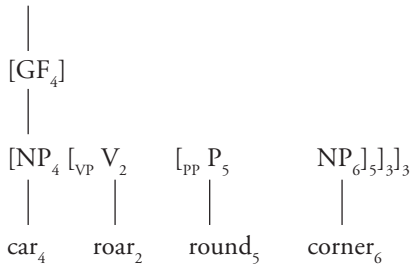


In (3), the conceptual semantic element is co-indexed with the whole construction but also with the head verb KICK. GF represents the grammatical function tier of the model (including the ranking of the two arguments), T represents tense, and again the various elements of the syntactic structure are co-indexed not only with elements of conceptual semantic structure but also with the final phonological structure, represented here orthographically.

Culicover and Jackendoff (226-227) are also able to formalise the structure of other VP constructions such as the sound + motion construction, as in the corpus example (4), the structure of which, modelled on a similar example given by Culicover and Jackendoff (227, example (65)b), is shown in (5).

(4) ... a car roared round the corner, ... (BNC CDT 48)

(5) [GO ([CAR]$_4^{á}$), [ROUND ([CORNER]$_6$)]$_5$; [ROAR] (á)]$_2$]$_3$



The conceptual semantics here is exactly the same as for the version shown in (6):

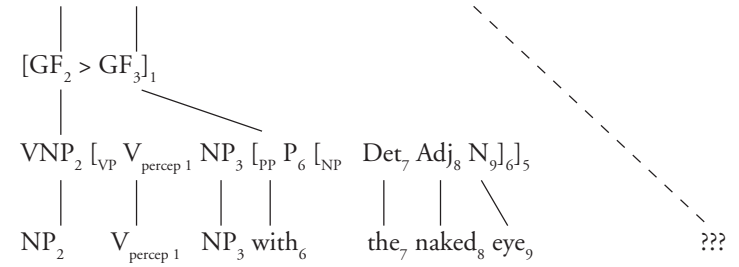
(6) ... a car went round the corner, roaring ...

The mapping on to the syntax, however, is clearly different in the two cases, and (5) shows how it occurs for the sound + motion construction. The symbol á indicates that the two arguments are bound within the conceptual structure, and again subscripts show the matching of the three levels.

It remains to be seen whether such formalisations can be extended to cover the wider range of types of semi-fixed expression revealed by corpus analysis. The specification of examples such as *COME a cropper* should also prove possible, though it might be harder to formulate an adequate conceptual semantic representation than for *KICK the bucket*, which can be paraphrased in terms of a single conceptual entity DIE. Furthermore, the mapping on to the syntax would have to be more complex, in order to account for the various structural possibilities found in the corpus.

Turning now to constructions of the kind typified by the sequences centred around *naked eye* and *bare hands*, we might attempt a mapping of conceptual structure on to syntactic structure as in (7), which shows the situation for expressions of the form *NP$_1$ SEE NP$_2$ with the naked eye*.

(7) [PERCEIVE ([A]$_2$ [B]$_3$)]$_4$; [UNAIDED]$_5$; [DIFFICULT]$_4$]$_1$



It would be possible to formulate similar structures for the alternative pattern exemplified by *visible to the naked eye*. However, the analysis runs into the same problems as that offered by constructionist approaches. One, as we have seen, is

concerned with mapping the meaning of difficulty on to some specific item or structure, since we have shown that there is a whole range of possible realisations of this semantic prosody, including the grammatical category negativity, lexical adverbs (e.g. *barely*, *hardly*, *just*), and also a variety of more subtle, and less easily categorised, means. The other problem is the inherently probabilistic nature of the choices made in the generation of this structure, again something which is not discussed in the context of the Simpler Syntax model.

6. SYSTEMIC FUNCTIONAL GRAMMAR

At the end of the previous section, I mentioned the choices made in the generation of a particular type of multiword sequence which can be considered as an extended unit of meaning. Systemic Functional Grammar (henceforth SFG) is the only grammatical theory which places at its generative heart the choices (or, to put it more neutrally, the paradigmatic oppositions) which are available to users of a language. It is therefore possible, within this framework, to think about alternative strategies for the expression of particular meanings. This gives us some cause to hope that the theoretical apparatus of SFG will be appropriate for modelling the things that speakers actually say, in relation to what they could have said but did not, and this is clearly of particular importance when we are looking at conventionalised ways of expressing particular meanings. Furthermore, SFG fully recognises the probabilistic nature of the choices made.[8]

A further advantage of SFG for our purposes is the fact that from the very inception of its precursor, Scale and Category Grammar (Halliday, "Categories"), syntagmatic lexical association, or collocation, as described initially by Firth (196), has been recognised as one dimension of linguistic patterning. Halliday ("Some") treats lexis and grammar as distinct though related kinds of linguistic patterning, collocational patterning along the syntagmatic axis being superimposed on the syntagmatic patterns of the grammar. However, in another paper published in the same year, Halliday ("Lexis" 62) suggests that it might be useful to see paradigmatic description as the central, underlying core of the grammar, if it could be shown that structural description could be derived from it. Thus was born Systemic Functional Grammar, where "systemic" refers to the technical device of the system, as a means for representing related options. As this idea took root and developed, it came to be seen that both grammar and lexis are instruments for realising meaningful choices, but that they differ in their degree of specificity. We see here, then, the genesis of the idea, shared by cognitively-oriented theories, that grammatical and lexical elements of language are not separate, but form a continuum, the **lexicogrammar**, in which lexis is seen as "most delicate grammar," where "delicate" is to be understood

[8] For discussion of this concept within SFG see Tucker, "Systemic"; "Between"; "Exposure".

as "detailed, more specific." In other words, the options towards the left-hand end of a network of systemic oppositions (that is, a set of systems related by dependency) tend to be realised grammatically, while the most detailed options, to the far right of such a network, tend to be realised lexically. At the point where this paradigmatic approach to the relationship between grammar and lexis was formulated, syntagmatic relations began to recede in importance within the theory, only to be revived relatively recently, in the work of Tucker[9].

Tucker ("Grammarians," "Getting, "Extending," "Systemic, "Between," "Sorry") has attempted to show how system networks can be formulated, together with rules for the realisation of the various options, in such a way that collocational phenomena, including the limited variability of semi-fixed idiomatic constructions, can be accounted for. Tucker's aim is to achieve a reconciliation of the "lexis as most delicate grammar" approach of Halliday and the lexis-driven approach of Sinclair, "by showing how systemic functional grammar can model the relationship between lexically and grammatically realized meanings in a unified manner, where grammar and lexis are interdependent" (Tucker, "Grammarians" 148). The key point from which Tucker's argument emerges is that seeing lexical items merely as realisations of very specific, delicate choices in system networks is misleading, in that the choice of (options leading to) specific lexical items, or groupings of such items, can in turn have an effect on further grammatical choices. This approach fits nicely with the demonstration, by corpus linguists, of the intimate association between grammatical and lexical patterning.

An example of the lexical conditioning of further grammatical choices (Tucker, "Grammarians" 161-162) is that when we traverse what is known as the transitivity network, which specifies types of process in the clause, we may first select the category of "mental process" (as opposed to a "material" process of doing and happening, or a "relational" process of being, having and the like), and then choose among the more delicate possibilities afforded by the English language, resulting in a lexical verb such as *love*, *like*, *know*, or *remember*, to give just a few examples. But these verbs differ in their complementation patterns: for instance, *like* can take a *to*-infinitive clause as complement, but *dislike* cannot.

Tucker ("Grammarians") goes on to examine in detail the expression *I haven't the faintest idea*. He points out that this expression is interesting in a number of ways: it is "a stretch of lexical organisation that is not coextensive with the grammatical unit 'word'" (164); it is semi-fixed rather than completely invariant; selecting it also means co-selecting from other systems in the grammar; the range of variation is bound up with choices at other points in the lexicogrammar, for instance in polarity (positive/negative); it involves what Halliday calls **grammatical**

[9] It should be noted here that Tucker works within the 'Cardiff grammar' variant of SFG, rather than the 'Sydney grammar' formulated by Halliday and his colleagues. The two approaches have a great deal in common, but also display some important differences (for discussion see Fawcett; Butler, *Structure* Parts 1 and 2).

metaphor, in that it is not the "congruent" form for expression of the particular meaning conveyed; and it involves collocational sets of lexical items (*faintest*, *slightest*, *foggiest*, etc.).

The first problem is the nature of the process involved. On the surface, the expression has a relational process (*have*), but there is clear semantic parallelism with *I don't know*, which is a mental process of cognition. The two classifications have different implications for the grammar of an example such as *I haven't the faintest idea where she is*: in the mental process analysis, we have a process consisting of the material *haven't the faintest idea* (parallel to *(not) know*), with *I* as the clause participant labelled "Cognizant" and *where she is* as "Phenomenon"; with the surface relational clause analysis, the process is *haven't*, *I* is the so-called "Carrier" and *the faintest idea* "Possessed," but such clauses do not normally have a position for the part of the expression realised as *where she is*. Tucker therefore opts for the mental process analysis, treating *the faintest idea* syntactically as a "main verb completing complement," which is also the category used for the particles of phrasal verbs[10].

The fixed and semi-fixed elements in *I haven't the faintest idea* are modelled through the mechanism of preselection of choices within the grammar, operationalised as the allocation of 100% probability to particular choices in the generation of the clause. For instance, in order to rule out **I have the faintest idea*, the choice of a negative rather than a positive clause, in the system of polarity, is set to 100%. At the appropriate points in the selection of systemic features, the verb *have* is preselected, the possibilities for the head of the NP (or, as systemicists call it, nominal group) *the faintest idea* are restricted to a small set of nouns also including *notion* and *clue*, and the choice of modifier for this noun is restricted to a second small set including *foggiest*, *slightest* and perhaps one or two more adjectives. Tucker ("Grammarians") gives full details of the mechanisms involved in each case.

In more recent work, Tucker has explored certain aspects of his proposals in more detail and applied them to other types of semi-fixed expression. In Tucker ("Getting") the issue of classification in terms of basic process type, together with the consequences of the available options, is addressed in relation to the expression *I can't get my head around it*. Tucker ("Between") concentrates on demonstrating that "the full range of phraseological expres-sions and their variants can be modelled systemically and functionally, without recourse to the undesired treatment of these phenomena at a separate level of description" (974), using *you've got hold of the wrong end of the stick* and *my lips are sealed* as his main examples. In Tucker ("Sorry") the emphasis is on formulaic aspects of speech act realisation, exemplified from apologies involving *sorry*, and again demonstrating the ability of the standard mechanisms of SFG to account for the variations encountered in a corpus.

Of particular interest is the attempt, in Tucker ("Systemic"), to account for collocation in general, rather than only semi-fixed expressions, in terms of the normal

[10] The current Cardiff grammar no longer uses this term, replacing it by Main Verb Extension.

apparatus of SFG. Tucker takes as an example the discussion of the noun *gaze* and its collocates in Hunston (*Corpora* 69ff). His strategy is to see collocation in terms of the probabilistic narrowing down of options for elements within the functional structure of the clause, and in this respect it is very like what is done in collostructional analysis[11]. According to Matthiessen and Martin's concept of "nuclear relations" within the clause, each clausal or phrasal unit sets up a set of relationships among its elements. For instance, the Subject, which is by default nominal, and the Complement(s), by default nominal or adjectival, represent participants in the process which is realised by the main verbal element, and there may also be adjuncts which represent the circumstances under which the process occurs. These elements are then candidates for housing items which frequently co-occur. In the case of the noun *gaze*, estimation of collocational strength by means of the t-score shows that the high frequency collocates include possessives (*his*, *her*, *my*). This, Tucker observes, is a reflection of the fact that the NP (nominal group) whose head is *gaze* is semantically related, in terms of Halliday's concept of grammatical metaphor, to a clause with *gaze* as the main verb, realising a process which in this case normally takes a human Subject/Agent. A further frequent collocate is *under*, and concordances show that this occurs in structures such as *under the gaze of a handsome young curate*, where the possessive relationship is expressed by means of a prepositional structure with *of*. Estimation of collocational relationships using the Mutual Information (MI) score, which tends to emphasise less frequent, but tightly bound collocates, reveals that two such items are *avert* and *averts*, giving a structure *avert* + possessive + *gaze* which is semantically related, again through grammatical metaphor, to *look away*, *turn* + possessive + *gaze/eyes/face away*. Once again, we can model these relationships through preselection: selection of the delicate transitivity options leading to the lexical item *avert* will limit the Complement to a small range of options realised through the lexical items *gaze*, *face*, *eyes*, together with a probabilistic preference for a possessive determiner. Further MI collocates include *unblinking*, *unseeing*, *unfocused*, *baleful*, *unwavering*, *watchful*, which are modifiers of the head noun: once again, the corpus data can help us to restrict the lexical classes of modifier in a probabilistic fashion.

Let us now look briefly at how Tucker's proposals might handle our three exemplificatory sequences. In generating COME *a cropper* and its variants, the SFG approach would presumably start with an already fairly delicate subdivision of material (doing/happening) processes leading to the lexically-realised area of failing, and then provide further subdivisions concerned with the failure being disastrous, with yet more delicate options showing the possibilities for variation, as sketched in Figure 1, with realisation rules in Table 1. The functional relationship between COME and *a cropper* can be handled, in the Cardiff grammar, by means of the element Main Verb Extension (see footnote 5 above).

[11] A very similar approach is also taken in Butler ("Matter"), in relation to the verbs GIVE and TAKE, though the investigation is also pursued beyond the limits of the arguments of the verbs, to examine other aspects of collocation.

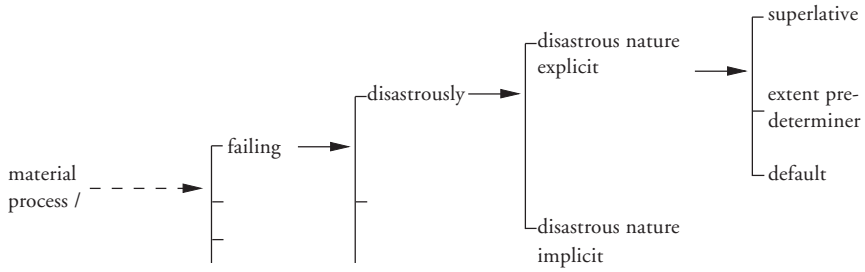


Figure 1: Tentative proposal for options leading to variants of COME *a cropper*.

TABLE 2: REALISATIONS FOR FEATURES IN FIGURE 1.

SYSTEMIC FEATURE	REALISATION
disastrous nature explicit/superlative	COME *the most* + adj showing seriousness + *cropper*
disastrous nature explicit/extent predeterminer	COME *such a cropper*
disastrous nature explicit/default	COME *a/an* + adj showing seriousness + *cropper*
disastrous nature implicit	COME *a cropper*

The generation of structure for expressions such as SEE *with the naked eye* would start with the selection of a mental process, narrowed down to one of perception, and preselection would determine the probabilistically stated range of options within what the Cardiff grammar calls "qualities of Situation" (as opposed to the "qualities of Thing" which act to modify nominally-realised entities). In particular, there would be a range of such qualities relating to unaided perceiving, and an appropriate arrangement of interconnected systemic options would allow the selection of a main verb such as *see*, *discern*, *perceive* to predetermine a prepositional structure with *with*, while the selection of a quality-of-Thing related to perception, such as *visible*, would predetermine a prepositional structure with *to*. The complement of the preposition would have its determiner set to high probability for *the*, with possessive determiners as options of lower probability, and the head noun set to *eye*.

How, then, does the SFG approach fare overall? Its paradigmatically based nature is attractive in that it allows us to model the choices available to the speaker, and the concept of delicacy of choice makes possible the progressive subdivision of these choices to reflect very detailed meanings, with preselection providing a mechanism for restricting the final choice of lexical sets of individual lexical items. However, underneath this apparent strength lies what many grammarians would see as a weakness: what, we may ask, is the justification for positing particular subdivisions of choice? Proposals such as that in Figure 1 seem *ad hoc*, tailored to the requirements of what we want to generate, without any independent justification. It may be that this is inevita-

ble if we want to model such delicate choices, and Tucker ("Motivating") has pointed out that it is probably futile to look for "reactances" within the grammar in order to distinguish between similar lexical items. He also shows an awareness of the problems when he remarks that "at this level of delicacy lexicogrammatical description enters into uncharted territory, extending well beyond the range of broad grammatical phenomena that can lend their names to feature labelling" (Tucker, "Between" 958).

A further problem is that Tucker's approach to collocation takes us back to the situation we encountered with constructionist approaches in general and Culicover and Jackendoff's model, namely that co-occurrence ends up subsumed under structural relations between elements with fixed boundaries, as evidenced by the treatment of *gaze* discussed earlier. It is therefore difficult to see how Tucker would deal with semantic prosodies which, as we have seen, extend over stretches of language which are not always easy to delimit structurally. Tucker ("Between" 963) mentions, for example, Stubbs' (*Words* 45) demonstration that the verb *cause* tends to associate with words indicating unpleasant things such as *damage*, *death*, *disease*. But again, the discussion is confined to cases where the collocational relationship can easily be subsumed under a structural one: the nouns with unpleasant connotations are generally found as heads of the Complement of *cause*. With multiword expressions such as those centred around *naked eye* or *bare hands*, the situation is more complex. It will be interesting to see if further work within the Cardiff grammar framework will be able to resolve this issue.

7. TOWARDS A UNIFIED ACCOUNT: RE-ASSESSING THE BASIS OF FORMULAIC LANGUAGE

We cannot, I think, escape the conclusion that although each of the approaches discussed here goes some considerable way towards accounting for the properties of idiomatic multiword expressions, all of them suffer from one crucial disadvantage, namely that they have not yet shown that they can provide a mechanism through which semantic prosodies that are not contained within fixed constituents can be accommodated. For this, we need to revert to the concept of collocation as a separate level of lexical syntagmatic constraint, which only sometimes maps isomorphically on to grammatically well-defined structures. How important this is in the long run will depend on whether the more diffuse realisations of semantic prosody turn out to be a frequent and systematic part of how we communicate, as is suggested by Sinclair's work, or just a fairly marginal phenomenon.

We have seen that Sinclair's claim is that semi-preconstructed strings of words are stored and selected as single items. We have also seen that the same basic idea underlies Wray's model of formulaic language. Wray's model of language acquisition operates on the principle of "needs only analysis," according to which language acquirers operate with the largest possible unit, until such time as the input reveals the potential of parts of the unit to be recombinable (Wray 130-132, 138). Some sequences will remain unanalysed even in the adult. This idea is also consonant with the proposals of usage-based grammars, such as Cognitive Grammar and some kinds

of constructionist model[12] which postulate that the language user's grammar sediments out of very large numbers of individually experienced language events, from which generalisations are progressively made. For functionally-minded linguists, these proposals make a great deal of sense, in that one of the central tenets of functionalism is that language acquisition is based on the linguistic input and on a set of cognitive capacities and dispositions which act on it (the "constructivist" approach, as opposed to the "nativist," Universal Grammar approach of the Chomskyan school).

There is, however, another way to look at formulaic language, which although appearing to take the opposite view to Sinclair and Wray, can, I think, actually be seen as compatible with their work. One of the most striking characteristics of formulaic sequences, illustrated not only in this brief presentation but throughout the literature on this area, is just how few such sequences are truly fixed: expressions of the type *by and large* are very much the exception rather than the rule. Variability is endemic to formulaic language, but it is restricted, controlled variability, and one of our tasks as linguists is to tease out the constraints. This is important, because it means that at least for the many language users whose productions find their way into corpora such as the Bank of English and the BNC, most formulaic sequences do have an internal structure which can be accessed and exploited if the communicative need arises (note that this formulation accords with the Needs Only Principle). If we are to take account of these variations, as well as of the default pattern, we must recognise that most formulaic sequences are indeed made up of components, which can often be modified.

But if this is the case, how do formulaic sequences differ from non-formulaic strings? The answer, surely, is that they realise strong **associations** between particular components, in terms of collocation and colligation, which may operate at a very specific level (i.e. between specific words, whether lexical or grammatical in function) or at a more general level (i.e. with a semantic or syntactic class of items).

In order to examine associations between specific lexical words (**collocations**), we shall use three indicators of collocational strength. The z-score is a measure of how often two items collocate within a given distance (in this case, as in many studies, a distance of 5 words on either side of the node word has been used), as compared with how often they would be expected to co-occur merely on the basis of their overall frequencies in the corpus. The Mutual Information (MI) score likewise uses observed and expected co-occurrence frequencies to compute a measure of association. Both z and MI give too much weight to rare words, and the Log Likelihood (LL) statistics corrects this bias, and is the measure of choice in many studies.[13] Table 3 shows z, MI and LL scores for some pairs of words we have used in previous parts of our discussion, as calculated by WordSmith Tools.[14]

---

[12] See Bybee and Hopper, *Frequency*; also the introduction to Barlow and Kemmer and the papers in that collection.

[13] See for example those in the excellent advanced resource book for corpus linguistics by McEnery, Xiao and Tono.

[14] See <http://www.lexically.net/wordsmith>.

TABLE 3: COLLOCATIONAL STRENGTH FOR SOME COMBINATIONS OF WORD FORMS IN THE BNC.

Node	Collocate	z	MI	LL
come	cropper	17.21	7.76	91525.84
came	cropper	25.14	4.43	63951.66
naked	eye	84.28	9.18	5885.67
bare	hands	21.06	6.75	14466.85

Any z value over 3.27 is significant at the p >= 0.001 level, and any value of LL over 10.83 is also significant at this level. It has been suggested that a level of 3 or more for MI indicates important collocation (Hunston, *Corpora* 71). So we know that *cropper* is statistically a highly significant collocate of both *come* and *came*, and similarly *eye* is a very significant collocate of *naked*, and *hands* of *bare*.

We have also shown three further kinds of association in these expressions, backing up the large amount of already existing work on the idiom principle. Firstly, there are clear instances of **colligation**, i.e. relationships with grammatical words or categories: *come* and *cropper* associate with the indefinite article (except in certain well-defined syntactic contexts), *naked eye* with the definite article, *bare hands* with possessives; both *naked eye* and *bare hands* associate with prepositions in front of the determiner, especially *with* (in both cases) and *to* (in the case of *naked eye*). Secondly a combination of lexical words may have a **semantic preference** for a particular semantic class of items: *naked eye* goes with words indicating visibility, *bare hands* with words indicating force. Thirdly, we have **semantic prosodies**: *naked eye* and *bare hands* very often have some indicator, in the textual environment, of difficulty. All these properties can be handled in terms of associations between items, whether specific or more general.

Another way of putting this is that individual words, and combinations of words, are **primed** for occurrence in particular environments: this is the basis of Hoey's theory of **lexical priming**. In fact, Hoey demonstrates not only that words and word combinations are associated with each other and with particular semantic areas, but also that they may be primed for occurrence in particular textual positions, such as the beginning or end of a sentence or paragraph.

As I mentioned earlier, I would want to argue (and so would Hoey —pers. comm.) that this "bottom-up" view, in which words associate in specifiable ways and may even take particular positions in the text, is perfectly compatible with the top-down view espoused by Wray. The association between items in adult language means that the child acquiring a language will come across strings such as [kʌməkɹɔpə], [kəɪməkɹɔpə], [wɪðɪzbɛəkhændz], and so on, and if these are sufficiently frequent they may become entrenched as single units, at least until such time as the child comes across examples where parts of these strings are combined in other ways, e.g. [kəɪməkjudʒkɹɔpə] or [[wɪðmaɪbɛəhændz], at which point the original units may

be analysed into components, though still only to the extent justified by the new data. For some speakers, some strings may never get reanalysed, even into adulthood, because they are so frequent and examples of recombination so rare. But this is exactly what is to be expected on the basis of the well-documented relationship between frequency and the entrenchment of items as units.[15]

In order to see the principle of association in action, let us return to the use of *naked eye*. Corpus analysis has shown clearly that this expression is, statistically speaking, most likely to be used in a syntagmatic discourse context which contains the following:

– reference to the concept of visibility
– reference to unaided vision
– indications of the difficulty of seeing

These principles are clearly operative in examples such as the following:

(8) Keen-sighted people can distinguish them both with the naked eye. I am not sure that I can do so, but with even times 7 binoculars they are clear enough. (BNC EAW 1241-1242)

The choice of *naked eye* brings with it a further association —notice that it is a probabilistic association rather than an absolute one, but one with very high probability— the presence of the definite article *the* in front of *naked eye*. The precise way in which *the naked eye* will be used depends on whether the speaker wants to express the process of seeing, or the quality of visibility. In (8), the speaker chooses the first option, selecting the verb *distinguish*. This, in turn, leads to another association in the form of the embedding of *naked eye* in a prepositional phrase which has *with* as its preposition: neither *to* nor agentive *by* will fit here. The (again probabilistic) association with an indication of difficulty is fulfilled through the use of *keen-sighted* (in implicit comparison with having normal or poor eyesight) and *with even 7 times binoculars*, a comparison with a more adequate way of viewing the stars.

Importantly, however, these are not the only kinds of association which are operative in the construction of the speaker's utterance. For instance, *distinguish* normally requires an object, here *them both*. Furthermore, the constituent order rules, combined with the topic-focus structure required to convey the speaker's assessment of the informational importance of constituents, determine the order in which *distinguish* and the elements associated with it occur: in this case Subject-Aux-Main Verb-Object-Instrumental Adjunct. Here, of course, we are talking about the syntax of English, including that associated with particular lexical items such as *distinguish* (or the classes of which they are members).

[15] For discussion from a functional-cognitive linguistic perspective, see, for example, Bybee and Hopper, *Frequency* and Bybee, and from a psycholinguistic perspective, Ellis.

Such an account in terms of association would not necessarily entail the abandonment of the kinds of theoretical proposal discussed in sections 4-7: rather, we would need to reformulate their claims in terms of the systematic syntagmatic association of particular (types of) entity at particular levels of description, as well as mechanisms for the simultaneous satisfaction of constraints at the different levels. For instance, semantic structures can be seen as constellations composed of elements of particular semantic types which regularly associate with each other: indeed, structures involving, for example, an Agent, a semantic predicate and a Patient are clusters of precisely this type. Syntactic structures, such as "NP Aux V NP PP" in English, can similarly be seen as associations of syntactically defined elements. Morphological structures are composed of strongly associated morphemes in generally fixed arrangements, while phonological structures are made up of phonemes in particular association patterns, such as word-initial /str/ in English but not Spanish, or word-initial /mr/ in Russian but not in English. This is not, of course, saying anything in the least revolutionary: on the contrary, syntagmatic structure has been, as already been pointed out, the major organising principle of grammars for a long time. What has not, however, been so prevalent is the concept of the elements in such structures attracting or repelling one another in what I have called association patterns, sometimes absolute, but sometimes probabilistic[16]. It is this concept which allows us to bring in collocational relationships as just one more kind of association, which must be combined with associations at other levels if a natural-sounding utterance is to be formulated.

Further exploration of these tentative suggestions must await future research. I hope to have shown, however, that there is reason to hope that we can resolve the tensions between the complex and subtle patterning revealed by corpus analysis, on the one hand, and the more cut-and-dried patterns enshrined in grammars, on the other.

WORKS CITED

Bank of English. <http://www.titania.bham.ac.uk>.

Barlow, Michael, and Suzanne Kemmer, eds. *Usage Based Models of Language*. Stanford: CSLI, 2000.

British National Corpus. World Edition. <http://www.natcorp.ox.ac.uk>.

Butler, Christopher S. "Corpus Studies and Functional Linguistic Theories." *Functions of Language* 11.2 (2004): 147-186.

——. "A Matter of *GIVE* and *TAKE*: Corpus Linguistics and the Predicate Frame." *Revista Canaria de Estudios Ingleses* 42 (2001): 55-78.

[16] However, see Bybee and Hopper (*Frequency* 14) for the claim that "the kind of constituency normally studied by syntacticians also has its source in language use and frequency of co-occurrence."

——. *Structure and Function: A Guide to Three Major Structural-Functional Theories. Part 1: Approaches to the Simplex Clause*. Amsterdam: John Benjamins, 2003.

——. *Structure and Function: A Guide to Three Major Structural-Functional Theories. Part 2: From Clause to Discourse and Beyond*. Amsterdam: John Benjamins, 2003.

Bybee, Joan. "From Usage to Grammar: The Mind's Response to Repetition." *Language* 82.4 (2006): 711-733.

Bybee, Joan and Paul Hopper, eds. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 2001.

——. "Introduction to Frequency and the Emergence of Linguistic Structure." *Frequency and the Emergence of Linguistic Structure*. Ed. J. Bybee and P. Hopper. Amsterdam: John Benjamins, 2001. 1-24.

Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge: MIT P, 1965.

Collins COBUILD English Language Dictionary. London: Harper Collins, 1987.

Cowie, Anthony P. ed. *Phraseology*. Oxford: Oxford UP, 1998.

Croft, William. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford UP, 2001.

Croft, William, and D. Alan Cruse. *Cognitive Linguistics*. Cambridge UP, 2004.

Culicover, Peter W. *Syntactic Nuts*. Oxford: Oxford UP, 1999.

Culicover, Peter W., and Ray Jackendoff. *Simpler Syntax*. Oxford: Oxford UP, 2005.

Ellis, Nick. "Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition." *Studies in Second Language Acquisition* 24 (2002): 143-188.

Fawcett, Robin P. *A Theory of Syntax for Systemic Functional Linguistics*. Amsterdam: John Benjamins, 2000.

Fillmore, Charles J., Paul Kay, and Mary Catherine O'Connor. "Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*." *Language* 64.3: 501-538, 1988.

Firth, John R. *Papers in Linguistics: 1934-1951*. Oxford: Oxford UP, 1957.

Gibbs, Raymond W., Jr. "Idioms and Formulaic Language." *The Oxford Handbook of Cognitive Linguistics*. Ed. D. Geeraerts and H. Cuyckens. Oxford: Oxford UP, 2007. 697-725.

Goldberg, Adele E. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford UP, 2006.

——. *Constructions: A Construction Grammar approach to Argument Structure*. Chicago: U of Chicago P, 1995.

Gonzálvez-García, Francisco and Christopher S. Butler "Mapping Functional-cognitive Space." *Annual Review of Cognitive Linguistics* 4 (2006): 39-96.

Gries, Stefan Th. and Anatol Stefanowitsch. "Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations'." *International Journal of Corpus Linguistics* 9.1 (2004): 97-129.

Halliday, Michael A.K. "Categories of the Theory of Grammar." *Word* 17.3 (1961): 241-292.

——. "Lexis as a Linguistic Level." *In Memory of J. R. Firth*. Ed. C.E. Bazell, J.C. Catford and M.A.K. Halliday. London: Longman, 1966. 148-162.

——. "Some Notes on 'Deep' Grammar." *Journal of Linguistics* 2.1 (1966): 57-67.

HOEY, Michael. *Lexical Priming: A New Theory of Words and Language*. London: Routledge, 2005.

HUNSTON, Susan. *Corpora in Applied Linguistics*. Cambridge: Cambridge UP, 2002.

——. "Semantic Prosody Revisited." *International Journal of Corpus Linguistics* 12.2 (2007): 249-268.

JACKENDOFF, Ray. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford UP, 2002.

KAY, Paul, and Charles J. FILLMORE. "Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction." *Language* 75.1 (1999): 1-33.

LAKOFF, George. *Women, Fire, and Dangerous Things*. Chicago: U of Chicago P, 1987.

LANGACKER, Ronald W. *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford: Stanford UP, 1987.

MARTIN, James R. *English Text: System and Structure*. Amsterdam: John Benjamins, 1992.

MATTHIESSEN, Christian M.I.M. 1992. "Interpreting the Textual Metafunction." *Advances in Systemic Linguistics: Recent Theory and Practice*. Ed. M. Davies and L. Ravelli. London: Pinter. 37-82.

MCENERY, Tony, Richard XIAO, and Yukio TONO. *Corpus-Based Language Studies: An Advanced Resource Book*. Abingdon: Routledge, 2006.

MOON, Rosamund. *Fixed Expressions and Idioms in English: A Corpus-Based Approach.* Oxford: Clarendon, 1998.

NUNBERG, Geoffrey, Ivan A. SAG, and Thomas WASOW. "Idioms." *Language* 70 (1994): 491-538.

PARTINGTON, Alan. "'Utterly Content in Each Other's Company': Semantic Prosody and Semantic Preference." *International Journal of Corpus Linguistics* 9.1 (2004): 131-156.

SINCLAIR, John M. *Corpus, Concordance, Collocation*. Oxford: Oxford UP, 1991.

——. "The Search for Units of Meaning." *Textus* 9.1 (1996): 75-106. Reprinted in Sinclair, 2004,. 24-48.

——. *Trust the Text: Language, Corpus and Discourse*. London: Routledge, 2004.

STEFANOWITSCH, Anatol, and Stefan Th. GRIES. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8.2 (2003): 209-243.

STUBBS, Michael. "Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies." *Functions of Language* 2.1 (1995): 23-55.

——. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell, 2001.

TAYLOR, John R. *Cognitive Grammar*. Oxford: Oxford UP, 2002.

TUCKER, Gordon H. "Exposure, Expectations and Probabilities: Implications for Language Learning." *Advances in Language and Education*. Ed. A. McCabe, M. O'Donnell, and R. Whittaker. London: Continuum, 2007. 239-253.

——. "Extending the Lexicogrammar: Towards a More Comprehensive Account of Extraclausal, Partially Clausal and Non-Clausal Expressions in Spoken Discourse." *Language Sciences* 27 (2005): 679-709.

——. "'Getting Our Heads around It': Semantic and Syntactic Tension in the Transitivity Analysis of Metaphorically-derived Multi-word Verbs." *Studi Italiani di Linguistica Teorica e Applicata* 30.2 (2001): 303-313.

——. "So Grammarians Haven't the Faintest Idea: Reconciling Lexis-oriented and Grammar-oriented Approaches to Language." *Functional Descriptions: Theory in Practice*. Ed. R. Hasan, C. Cloran, and D.G. Butt. Amsterdam: John Benjamins, 1996. 145-178.

——. "Between Lexis and Grammar: Towards a Systemic Functional Approach to Phraseology." *Continuing Discourse on Language: A Functional Perspective*. Vol. 2. Ed. C. Matthiessen, R. Hasan and J. Webster. London: Equinox, 2007. 954-977.

——. "Motivating 'Lexical' Systems in a Systemic Functional Lexicogrammar." Plenary Lecture Given at the 19th European Systemic Functional Linguistics Conference and Workshop. 23-25 July 2007. Saarbrücken, Germany.

——. "'Sorry to Muddy the Waters': Accounting for Speech Act Formulae and Formulaic Variation in a Systemic Functional Model of Language." *Functional Perspectives on Grammar and Discourse: In Honour of Angela Downing*. Ed. C.S. Butler, R. Hidalgo Downing, and J. Lavid. Amsterdam: John Benjamins, 2007. 395-418.

——. "Systemic In*corpora*tion: On the Relationship Between Corpus and Systemic Functional Grammar." *System and Corpus: Exploring Connections*. Ed. G. Thompson and S. Hunston. London: Equinox, 2006. 81-102.

Wray, Alison. *Formulaic Language and the Lexicon*. Cambridge: Cambridge UP, 2001.