

MULTI-WORD LEXICAL PHENOMENA IN FUNCTIONAL GRAMMAR

Chris Butler

University College of Ripon and York St John, York

1. INTRODUCTION

Theories of grammar, functional theories among them, have long been constructed in accordance with the assumption that the word is a basic unit of the grammar, and that words combine to produce phrases, which in turn combine to give clauses, from which complexes (“sentences”) may be formed by processes of co-ordination and subordination. The process of constructing a piece of language, for the language user as well as the analyst, has thus been seen largely as a matter of “putting words together” according to the grammatical rules of the language, which themselves make use of word, phrase and clause level units.

Most grammars (not least those adopted, or more likely adapted, for use in the teaching of languages) have, of course, recognised the importance of “idioms”, in the sense of stretches of language consisting of more than one orthographic word, whose meaning cannot be transparently derived from the meanings of the individual components. Indeed, idioms present problems for grammars which have attracted a good deal of attention. Until relatively recently, however, the true qualitative and quantitative importance of “multi-word phenomena” in language was not fully appreciated. Progress in this area has been greatly facilitated by the advent of large computer-searchable corpora, and by the ready availability of computer hardware and software which is capable of performing linguistic analysis, albeit of a fairly basic level, on a large scale.

The concept of a grammar as based on the combination of word-level units into hierarchically-structured larger units is favoured by an approach, such as that of Chomsky and other formalists, which divorces the structure of language from the use to which that structure is put in human communication. Functional grammarians,

however, are committed to the view that language structure can only be satisfactorily explained in terms of the communicative functions which languages serve, and the psychological and social conditions of language use. There is now an impressive, and steadily growing, amount of evidence that the hierarchical, word-combining view of language, in which the construction of a stretch of language is seen in terms of open choice from the patterns allowed by the grammar, is, by itself, a rather poor model of language-in-use, requiring supplementation from a rather different model which Sinclair has termed the “idiom principle”¹:

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.

The last clause in this quotation is especially worth noting: the claim is not simply that there are multi-word items which cannot simply be resolved into their components according to the rules of the normal grammar, but that even many of the sequences which could be generated perfectly well by that grammar are, in fact, better seen as prefabricated pieces on which the language user can draw without recourse to the open choice grammar. As we shall see below, this evidence comes not only from corpus-based studies of adult language, but also from work on first and second language acquisition, processing mechanisms and language evolution. Given their commitment to an account of language-in-use, functionalists cannot afford to ignore the implications of this work for the construction of their own grammars.

One of the most striking conclusions to emerge from the corpus-based work alluded to above is the intricate interplay of grammatical and lexical aspects of patterning: words, singly or in sets, have their own sets of grammatical properties, which interact in complex ways with the properties of other words. The challenge to linguistic theory which is posed by this conclusion has been taken up by Tucker, within the framework of Systemic Functional Grammar, in which lexis and grammar are subsumed into a single paradigmatically-oriented “lexicogrammar”, in which meaningful choices can be realised by means of grammatical elements, lexical items, or both.² Tucker illustrates his discussion by means of an extended example involving expressions of the type *I haven't the faintest idea*, which arose in corpus-based investigation of superlative adjectives. Tucker points out that expressions of this kind are “semi-fixed”, and he goes on to show how Systemic Functional Grammar can handle the grammatical and lexical constraints involved (definiteness of the NP, superlative of the adjective, negative polarity, choice of head noun and adjective from restricted lexical sets: *idea/notion, faint/foggy/slight*).³

It would seem, then, that the area of (semi)-prefabrication is one which may be of considerable interest within a functional approach to language. In what follows, I will first review briefly the kinds of evidence which have been adduced for the importance of multi-word phenomena. I will then discuss to what extent, and how, these phenomena have been accounted for so far in Functional Grammar (FG). Finally, I will make some suggestions for future work in this area.

2. THE NATURE AND IMPORTANCE OF MULTI-WORD SEQUENCES

2.1 TYPES OF MULTI-WORD SEQUENCE

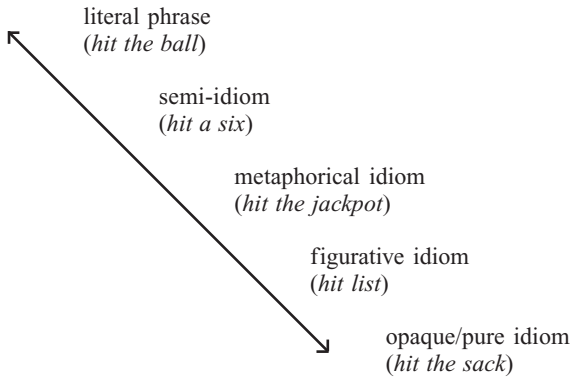
A large number of terms have been used in the literature to refer to what I have called multi-word phenomena: Wray gives a useful list with sources, which includes *chunks*, *sentence builders*, *formulaic speech*, *lexical phrases*, *routine formulae*, *preassembled speech*, *idioms*, *gambits*, *holophrases* and a number of others. It is by no means clear that all these terms are used by their originators to mean exactly the same thing, which makes discussion more difficult, but at least serves to highlight the multi-faceted nature of the phenomena. I have reluctantly adopted yet another term, *multi-word lexical phenomena* (MWLP), taken from Nattinger and DeCarrico, as an attempt at neutrality, though it is singularly inelegant and could even be criticised on the grounds that idioms, routine formulae and the like could themselves be regarded as single “words” in one sense of that very slippery term.⁴

One difference between approaches which is hidden by the plethora of terms is between those MWLPs which constitute structurally and/or functionally complete units and those which do not. Those who have come at the problem from the point of view of language acquisition and learning, or linguistic processing, have concentrated on sequences which, although consisting of more than one “word”, can be recognised as structural units in their own right, with their own semantic and discoursal functions. Much work in corpus linguistics, however, tends towards a rather more inclusive approach, motivated by Sinclair’s injunction to “trust the text”, in the sense of allowing the computer-based analysis to suggest patterns on a mechanically objective basis, and then interpreting the results with as open a mind as possible.⁵ Such work has, for example, demonstrated that multi-word sequences tend to overlap in texts, so that rather than being subject to fixed limits, they flow into one another to form a rich interlocking fabric (see Altenberg’s and Renouf’s work on English and Butler on Spanish).⁶

The former approach is emphasised in Luelsdorff’s definition of *phraseology* as a branch of lexicology:

Phraseology is that branch of language study which examines the properties of fixed turns of speech, i.e. those phrases, clauses, and sentences which are reproduced in their entirety by the speakers of a language. Such turns of speech include quotations, proverbs, idioms, and clichés.⁷

and by the use of terms such as *fixed expression*, in, for example, Alexander’s discussion of “idiomaticity” in the context of the teaching of English to non-native speakers.⁸ The “fixedness” of such expressions can be interpreted both in terms of boundaries, and in relation to the restrictions on alteration of the expressions. As Alexander, in common with other writers on the topic, fully recognises, there is a cline of fixedness, from structures fully generated by the sentence grammar at one end, to total immutability (actually a fairly rare phenomenon) at the other, passing through various intermediate positions on the way. Expressions also differ in their degree of semantic transparency, as Alexander demonstrates with the very neat example reproduced below.⁹



Related to the clines of formal (im)mutability and semantic transparency/opacity is the collocational behaviour (i.e. lexical co-occurrence patterns) of linguistic items. Indeed, expressions which are “fixed” to varying extents can be seen in terms of differing strengths of collocability.

The classification of fixed expressions proposed by Alexander is based on a mixture of form and function.¹⁰ He distinguishes *idioms* (including phrasal verbs, “*tourneurs*” such as *to stand one’s ground*, and irreversible binomial expressions such as *hard and fast* cf. **fast and hard*), from *discourse-structuring devices* (greetings and “gambits”), *proverbs*, *catch phrases* (clichés and slogans), and quotations/allusions.

The work of Nattinger is concerned with the importance, in the teaching and learning of languages, of *lexical phrases*.¹¹ Lexical phrases are defined as follows:

... multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time.¹²

Rather than conflate structural and functional criteria into a single classification, Nattinger and DeCarrico suggest cross-classifying lexical phrases according to the two types of property. Four structural criteria are recognised as important:

The first has to do with their length and grammatical status; the second, with whether the phrase has a canonical or non-canonical shape; the third, whether the phrase is variable or fixed; and the fourth, whether the phrase is continuous or discontinuous, that is, whether it consists of an unbroken sequence of words or whether it is interrupted by variable lexical fillers.¹³

Various combinations of these properties, seen as graded, define four basic structural types of lexical phrase: *polywords*, which are short phrases acting very like individual lexical items, can be canonical or not, and are fixed and continuous (e.g. *for the most part, by the way*); *institutionalized expressions*, at sentence level, usually canonical, fixed and mostly continuous (these correspond most closely to Alexan-

der's proverbs, catch phrases and quotations, but also include social routines such as *how are you?*, *nice meeting you*); *phrasal constraints*, which are short to medium length phrases, canonical or not, which allow variation of lexical and phrasal categories (e.g. *a ___ ago*, *good ___ (morning, etc, as a greeting)*); and *sentence builders*, which provide a framework for whole sentences, can be canonical or not, allow considerable variation, and can be continuous or discontinuous (e.g. *not only X but also Y*, *the ___er X*, *the ___er Y*).¹⁴

Functionally, lexical phrases are grouped into: *social interaction markers* relating to conversational maintenance and purpose (summoning, nominating a topic, shifting and closing topics, expressing politeness, performing particular types of speech act, etc); *necessary topics* for learners expressing themselves in a foreign language; and *discourse devices* such as logical, spatial and temporal connectors, fluency devices, summarisers, etc.¹⁵

Corpas Pastor, after reviewing previous classifications of phraseological units in Spanish, proposes her own classification¹⁶. At an initial level, these units are divided into those which constitute complete speech acts and those which do not. Within the latter class, she recognises "collocations", which conform to the rules of the language but are fixed, to varying degrees, in terms of norms of usage, and "locutions", which are combinations of words whose meanings are not merely the sum of the meanings of their component items.

Wray proposes that what she calls "formulae" fall into three basic categories according to their social function: the *manipulation* of other people through "pragmatically determined interactional strategies" such as indirectness; the expression of *group membership* through the use of particular ways of saying things which are associated with particular groups, and also through the use of social platitudes as phatic devices; and *fluency and holding the turn*, which may be perceived by language users as more important than continuity of content.¹⁷ I will return to Wray's classification in later discussion of my own work.

As typifying the corpus-based, initially more inclusive and mechanical approach to MWLPs, we may take the work of Sinclair and Renouf, of Altenberg and of Kjellmer on English, and my own work on Spanish. All of these are concerned, from varying perspectives, with the computerised isolation of recurrent sequences of words from corpora of written and/or spoken language, and the interpretation of the resulting strings.

The work of Sinclair and Renouf, anchored in the needs of language teaching as well as in the exploitation of the corpus developed by the Cobuild team at the University of Birmingham, has investigated the occurrence of *collocational frameworks* in which two grammatical words (articles, prepositions, etc) enclose a lexical word (e.g. *a ___ of*, with words such as *pound*, *series*, *sheet*, etc., in the gap), and also the phraseology of core, high frequency words in English.¹⁸

The aim of Altenberg's project is:

... to make a detailed investigation of recurrent word combinations in the London-Lund Corpus (LLC) with special emphasis on collocations and prefabricated expressions that reflect the speech process and various speaker strategies.¹⁹

Altenberg and his colleagues have reported work on recurrent word combinations, discontinuous combinations (collocational frameworks) and recurrent verb-complement constructions.²⁰

Kjellmer's work is concerned with the extraction of a corpus of ready-made phrases from the Brown Corpus of American written English.²¹ Kjellmer has named this corpus the Gothenburg Corpus of Collocations, but his definition of "collocations", as "recurring sequences that have grammatical structure", makes it clear that he is dealing with sequences similar to those discussed earlier.²²

My own work on Spanish has been concerned with the isolation and interpretation of recurrent word sequences and of collocational frameworks.²³ I will refer to this work again later in the context of implications for Functional Grammar.

2.2 EVIDENCE FOR THE IMPORTANCE OF MULTI-WORD LEXICAL PHENOMENA

2.2.1 *Corpus-based studies*

The corpus-based studies reviewed above have demonstrated the pervasive nature of MWLPs, whose frequencies, especially in spoken language are very high.²⁴

2.2.2 *Language acquisition and learning*

Ellis claims a major role for memorisation of sequences in the learning of native and second languages:²⁵

The attainment of fluent comprehension and production, both in native (L1) and second (L2) languages, involves the acquisition of memorized sequences of language. Learning vocabulary involves sequencing the phonological properties of the language: the categorial units, syllable structure, and phonotactic sequences. Learning discourse involves sequencing the lexical units of the language: phrases and collocations. Learning grammar involves abstracting regularities from the stock of known lexical sequences.²⁶

He presents an impressive array of detailed evidence in support of this claim. Similarly Weinert presents a review of the importance of formulaic language in second language acquisition, in which she discusses evidence from a variety of approaches, as well as the theoretical and methodological problems associated with defining and identifying formulaic language.²⁷ Both Peters, concerned largely with L1, and Nattinger and DeCarrico, with a primary interest in L2, suggest a more prominent role for preformed language than had generally been recognised.²⁸ These works, and the many references given in them, should be consulted for further details of the available evidence.

2.2.3 *Processing mechanisms*

Weinert claims that "[t]here is some evidence for the psychological reality of formulaic language in terms of storage and production", and gives a number of references to back this up.²⁹ Ellis presents a detailed discussion of the relationships between the learning and analysis of sequences and short and long term memory components.³⁰ Wray cites work which associates formulaic language with processing in aphasics.³¹

2.2.4 *Language evolution*

Wray makes a persuasive case for the centrality of a holistic system involving preformed sequences in the evolution of language in the human race.³² Her evolution-

ary conclusions, as well as the reason usually advanced for the dual systems of holistic and analytic processing today, are summarised in the following quotation:

The solution to the problem of having too little on-line processing capacity to easily handle analytic language at all times, was to fall back on the older, reliable holistic system for precisely those aspects of communication which it was best at: communicational functions. Holistic processing became the preferred strategy for coping with a range of different problems that could arise in the course of interacting through the medium of language: where the social aspects of the interaction were more important, or where it was necessary to retain fluency whilst constructing a complex utterance, formulaic language was, as it still is, drawn upon to meet the communicational shortfall.³³

3. MULTI-WORD LEXICAL PHENOMENA IN FG

3.1 MWLPs IN RELATION TO THE BASIC TENETS OF FG

As the following quotations demonstrate, FG is strongly committed to providing an account of language which focuses on its communicative use:

The primary aim of natural languages is the establishment of inter-human communication; other aims are either secondary or derived.³⁴

In the functional paradigm, [...] a language is in the first place conceptualized as an instrument for social interaction among human beings, used with the intention of establishing communicative relationships. Within this paradigm one attempts to reveal the instrumentality of language with respect to what people do and achieve with it in social interaction.³⁵

In accordance with this orientation, FG makes a commitment to three fundamental criteria of adequacy:

pragmatic adequacy:

Since a natural language is an instrument used for communicative purposes, there is little point in considering its properties in abstraction from the functional uses to which it is put. The system underlying the construction of linguistic expressions is a functional system. From the very start, it must be studied within the framework of the rules, principles, and strategies which govern its natural communicative use. In other words, the question of how a language is organized cannot be profitably studied in abstraction from the question of why it is organized the way it is, given the communicative functions which it fulfils.

This means that linguistic expressions can be understood properly only when they are considered as functioning in settings, the properties of which are co-determined by the contextual and situational information available to speakers

and addressees. Language does not function in isolation: it is an integrated part of a living human (psychological and social) reality.³⁶

Psychological adequacy:

... such a grammar must also aim at psychological adequacy, in the sense that it must relate as closely as possible to psychological models of linguistic competence and linguistic behaviour.³⁷

Typological adequacy:

... it should be *typologically adequate*, i.e., [...] it should be capable of providing grammars for languages of any type, while at the same time accounting in a systematic way for the similarities and differences between these languages.³⁸

The implications for the study of MWLPs are that FG should be able to provide an account which recognises the importance of (partially or fully) pre-formed sequences in linguistic communication, and describes their functions, as well as being consistent with what is known about the processing of such sequences. The grammar should also be able to accommodate facts about MWLPs in all languages, and to account for similarities and differences between languages in this area.

Some of the most fundamental changes which have been made in FG over the past decade are concerned with the recognition of layers and levels of meaning within the clause. The work of Hengeveld on modality demonstrated the need for a multi-layered model of the clause, which, in addition to recognising predicate and illocutionary layers, distinguishes two functions of the predication: the designation of states of affairs and the representation of the content of a speech event.³⁹ This is the model which is reflected in Dik's account:⁴⁰

clauses	represent	speech acts
propositions	represent	possible facts
predications	represent	states of affairs
predicates	represent	properties/relations and are applied to
terms	represent	entities

Whereas predications represent (cognitive correlates of) states of affairs (SoAs) in some world, propositions represent these states of affairs as a content which can be stated, denied, argued about, believed, doubted, etc. In Hengeveld's work, this model is further elaborated with the recognition that the predication and its constituent predicates and terms constitute a *representational* level of analysis, whereas the propositional and illocutionary layers constitute an *interpersonal* level:

At the representational level a SoA is described in such a way that the addressee is able to understand what real or hypothesized situation is referred to. At the interpersonal level this situation is presented in such a way that the addressee is able to recognize the communicative intention of the speaker.⁴¹

Hengeveld explicitly recognises the clear link here between his proposal and the ideational and interpersonal functions of Halliday's metafunctional model.⁴²

We shall see later that the distinction between representational and interpersonal meaning is of crucial importance in the interpretation of findings relating to MWLPs.

3.2 DIK ON IDIOMS

Dik's account of MWLPs is confined to idioms, defined as follows:

By "idiom" we shall understand any composite linguistic expression, the meaning of which cannot be compositionally derived from the meanings of its parts.⁴³

A brief account is given in Dik's *The Theory of Functional Grammar, Part 1*, and a fuller treatment in an article devoted to idioms, from which the summary below is taken.⁴⁴ Dik also discusses the ways in which idioms can be handled in a FG-based natural language processing system.⁴⁵

Dik's discussion is concerned with "phrasal idioms", in other words multi-word combinations with non-compositional meanings, and shows that some of the problems which these have been claimed to pose for linguistic theory are easily and naturally handled in FG. All the "fully lexical" items of a language are represented in FG as *predicates*, which are embedded in a *predicate frame* giving the form of the predicate, its syntactic class, the number and semantic function of its arguments, any selection restrictions associated with those arguments, and a meaning definition in terms of meaning postulates relating the predicate under definition to other predicates of the language. The predicate frame is the semantic nucleus of the clause, in that the semantics of the simplest type of clause is generated by inserting, into the argument slots of the predicate frame, *terms* referring to entities in some world. More complicated clauses can be constructed by adding *operators* for meanings which are realised grammatically in the language, and *satellites* for those which are realised lexically. As we shall see later, operators and satellites operate at several levels of structure.

The lexicon contains all predicates which cannot be derived by productive rules from other predicates; it must therefore contain idiomatic predicates as wholes. For an idiomatic predicate such as *kick the bucket* in English, a unified meaning can be assigned by a meaning definition equating the meaning of the idiom to that of the predicate *die*:

$$(1) \quad \text{kick}_{\text{V}}(x_1; \langle \text{hum} \rangle(x_1))_{\text{Proc}} (\text{the bucket})_{\text{Pat}} =_{\text{df}} \text{die}_{\text{V}}(x_1)_{\text{Proc}}$$

"to say, of a human being, that he kicked the bucket is to say that he died" = Dik's (9)⁴⁶

A later formulation recognises that animate beings other than humans may be involved, and uses a labelling for semantic function which is more in line with current practice ("Goal" rather than "Patient"):

$$(2) \quad \text{kick}_{\text{V}}(x_1; \langle \text{anim} \rangle(x_1))_{\text{Proc}} (\text{d1}x_1; \text{bucket}_{\text{N}}(x_1))_{[\text{Go}]} \leftrightarrow \text{die}_{\text{V}}(x_1)_{\text{Proc}} = \text{Dik's (40)}^{47}$$

This formulation illustrates the way in which the “frozen” parts of an idiom can be treated as such in FG: the appropriate material is stated as such in the predicate frame itself. Thus the specification of the Goal as “d1” (i.e. definite and singular) prevents this material from being subjected to certain rules which would normally apply in the generation of an NP with a count noun —pluralisation, or making the NP indefinite. Note that the original formulation, spelling the Goal out as *the bucket*, also has the effect of blocking modification of the noun, whereas the later formulation does not restrict the expression in this way. The reason may be that, as Dik points out in discussion of some problems even for the FG approach, some modification may indeed be possible, as in the following example:

- (3) *He finally kicked the dull bucket of his empty life.* = Dik’s (34)b⁴⁸

It is curious, then, that Dik returns to specification as *the bucket* in the second edition of *The Theory of Functional Grammar, Part 1*:

- (4) *kick* [V] (x_1 :<anim>)_{Proc} (*the bucket*)_[Go] ↔ *die* [V] (x_1)_{Proc} = Dik’s (57)⁴⁹

Placing the Goal function in brackets is an *ad hoc* convention for indicating that the entity is a “Pseudo-Goal” rather than a true goal —in other words, it behaves as a Goal with respect to the form and placing of the NP, but not with respect to the assignment of Subject and Object functions, since passivisation, involving the realignment of these functions with the NPs of the clause, is not possible for the idiom, though perfectly acceptable for the literal interpretation:⁵⁰

- (5) *John kicked the bucket.* (lit./idiom) = Dik’s (32)a⁵¹

- (6) *The bucket was kicked by John.* (lit.) = Dik’s (32)b⁵²

One reservation about the formulation given above is perhaps worth expressing: Dik comments that the first argument is given the semantic function Processed, which is the function given to entities which undergo a non-controlled event, because “the idiom does not designate an Action carried out by x_1 , but rather a process that x_1 undergoes”.⁵³ But surely this fact is shown by the assignment of the Processed function to *die* in the meaning definition: assignment of the same function to the first argument of *kick* is inappropriate, since we then have no controlling entity to explain the assignment of the function (Pseudo-)Goal to *the bucket*.

Dik also shows that those idioms which are not fully grammatical according to the normal rules of the grammar again present no problem for FG.⁵⁴ For instance, the expression *spic(k) and span*, meaning “clean and tidy”, contains two items which have no similar use outside the idiom, and so needs to be treated as a complex adjectival predicate with its own entry in the lexicon.

The fact that idioms show a range of degrees of “frozenness” in terms of the grammatical operations they allow constitutes a problem for all grammatical models. Dik is, however, able to demonstrate that some of the restrictions commonly found

can be incorporated into the predicate frame.⁵⁵ One such frequent restriction is on passivisation: Dik observes that if a clause containing the idiom *kick the bucket* is passivised, the idiomatic reading is lost, whereas *bury the hatchet* is perhaps more tolerant of passivisation:

- (7) *The bucket was kicked by John.* (only literal) = Dik's (26)a⁵⁶
- (8) *The hatchet was buried by the contesting parties* (possibly idiomatic)
= Dik's (26)b⁵⁷

In FG, the active/passive contrast is handled in terms of the assignment of the Subject syntactic function to different items in the semantic structure of the clause. The FG solution to the difference between the two idioms in (7) and (8) above is therefore to block Subject assignment to the term representing *the bucket* in the idiom *kick the bucket*, but not to the term representing *the hatchet* in *bury the hatchet*.

Finally, Dik discusses the problems which arise if we assume that the literal interpretation of an idiomatic expression plays no part in the semantics of the idiom itself.⁵⁸ An example is provided by (3), repeated for convenience below:

- (3) *He finally kicked the dull bucket of his empty life.* = Dik's (34)b

Dik suggests that the non-idiomatic interpretation of the elements of the expression (in the above case, the predicates *kick*, *bucket*) should still be accessible from the idiom.

3.3 MOUTAOUAKIL ON IDIOMS

Moutaouakil takes up the issues related to examples such as (3) above, and discusses possible solutions to the problem of integrating literal and idiomatic meanings.⁵⁹ More generally, he is concerned with the processes of idiomatisation, metaphorisation and de-idiomatisation, and how they may be accounted for within FG.

As an illustration of the problem under discussion, Moutaouakil cites a picture, in a French magazine, of an official luncheon party, in which a woman, dancing on the table, has just put her foot in one of the dishes. A member of the party comments as in (9) below:

- (9) *Elle a mis le pied dans le plat !* = Moutaouakil's (3)⁶⁰

She has put the foot in the dish.
"She has put her foot in the dish."
= "She has made a mistake."

As Moutaouakil points out, (9) is intended, in the context in which it is used, to have both the literal and the idiomatic interpretations. As usual, the idiomatic interpretation is blocked if certain types of change are made (e.g. substitution of *assiette* for *plat* or *posé* for *mis*).

Moutaouakil observes that phenomena of this type are important to linguistic theory for a number of reasons:

they are very productive devices, of importance to the expression of, for example, mitigation and irony;
 they appear to be widespread in languages of different types;
 they involve the deliberate use of a kind of ambiguity, as part of the speaker's discourse strategy;
 they are connected with the phenomenon of metaphor;
 they involve various levels of linguistic description: syntax, semantics, pragmatics, lexicon, logical reasoning;
 they present problems for translation;
 resolution of the problems associated with them would enrich the FG model in fundamental ways.

Moutaouakil goes on to examine the process by which an expression with a literal and a metaphorical meaning may achieve the status of an idiom. *Idiomatisation* begins with the two meanings, the metaphorical one being derivable from the literal one by logical inferential processes of the kind proposed in the Functional Logic component of Dik's model of the Natural Language User.⁶¹ The next stage is the *lexicalisation* of the metaphor through the process of markedness shift.⁶² According to Dik's account of markedness shift an initially marked metaphorical way of expressing a meaning becomes progressively de-marked, so that it eventually becomes the most usual way of saying what is meant, largely or wholly replacing an original literal way of expressing the meaning.

Moutaouakil recognises that Dik's proposals, as reviewed above, are sufficient to account for the properties of idioms which arise by the mechanisms summarised above, but then goes on to suggest how the further problem of *de-idiomatisation* (i.e. the revival of the literal meaning of an idiomatic string, to account for examples such as (9)) can be handled in FG. He defines *de-idiomatisation* as "any process having as a consequence the fact that an idiom loses (all or some of) its idiom features".⁶³ *De-idiomatisation* can be complete or partial, and may occur as a result of diachronic processes, or synchronically in the use made of an expression by a speaker. Moutaouakil identifies two kinds of factors which provide mechanisms for synchronic *de-idiomatisation*: contextual and structural. Contextual *de-idiomatisation* occurs when the idiom is used in a situation where both its literal and idiomatic meanings are applicable, as in (9) above. *De-idiomatisation* may also occur through the incorporation, into the structure of the expression, of a lexical item which is appropriate to the literal interpretation. Example (3) given earlier, taken from Dik, is an instance of this type. Moutaouakil also points out that in partially "unfrozen" idioms, the idiomatic meaning is an accepted, uncancellable part of the inherent meaning of the expression, rather than being derived from the literal meaning by any inferential mechanism.⁶⁴

Moutaouakil⁶⁵ goes on to propose an account of partial *de-idiomatisation* within a FG framework. He first characterises *de-idiomatized* idioms in terms of intended, rather than accidental, ambiguity: they are part of the speaker's communicative planning. He then proposes that since the literal and the idiomatic meanings are both "basic features of the literally understood semantic content", both should be repre-

sented in the underlying clause structure.⁶⁶ His first solution to this problem is to provide two separate structures, linked by a double arrow signifying that the structures refer to the same linguistic expression. The first structure codes the idiomatic meaning by means of a predicate frame in which the idiomatised term slots are lexically filled, and is provided with a meaning definition which reflects the idiomatic interpretation. The second structure has the term slots unfilled, and is provided with a meaning definition reflecting the literal interpretation. Later, Moutaouakil collapses the two structures in such a way that their common parts appear only once. His examples are taken from Standard Modern Arabic, and are rather hard to follow for the reader who has no knowledge of that language; the principle, however, is clear enough.

3.4 EVIDENCE FROM CORPUS STUDIES: NON-IDIOM MWLPs

I now turn to work on corpora which has extended our knowledge of MWLPs beyond the boundaries of idioms as traditionally defined. The work of Altenberg and his colleagues on English, and my own work on Spanish, has demonstrated the occurrence, in corpora, of frequently repeated sequences which constitute a single meaning choice, even though their structures can be generated by the open-choice grammar and their interpretations are semantically transparent. Table 1 lists the most common combinations of four or more words in the London-Lund Corpus of spoken English, as given by Altenberg.⁶⁷ The frequency of each combination in the corpus is also given.

at the end of the	13	in the middle of the	7
at the beginning of the	11	and so on and so forth	7
thank you very much indeed	11	and all the rest of it	7
as a matter of fact	10	going to be able to	6
but on the other hand	10	I don't know how many	6
it seems to me that	10	and I said well I	6
from the point of view of	9	and at the same time	6
in the House of Commons	8	the the the the the	6
and that sort of thing	8	that would be very nice	5
as far as I know	7	at the bottom of the	5

Table 1: The most frequently occurring sequences of four or more words in the London-Lund Corpus of spoken English (taken from Altenberg 1990:136)

Similar lists of frequent sequences of 4 and 5 words in the spoken Spanish component of the Corpus de Referencia are shown in Tables 2 and 3^{68, 69}:

la verdad es que	200	no no no no	172
que pasa es que	126	lo que pasa que	24
lo que pasa es	124	qué es lo que	124
sí sí sí sí	122	eso es lo que	103
a la hora de	94	que es lo que	93
que yo creo que	75	el punto de vista	72
yo creo que es	72	desde el punto de	71

yo creo que no	70	que a lo mejor	70
a mí me parece	67	yo no sé si	65
y yo creo que	60	no yo creo que	56
en el caso de	55	buenas noches buenas noches*	55
a lo largo de	55	lo que hay que	54
yo creo que el	54	y por lo tanto	53
gol gol gol gol	53	los medios de comunicación	52
de la de la	51	pero yo creo que	51
que se va a	51	vamos a ver si	50

Table 2: The most frequent sequences of four words in the spoken component of the Referencia Corpus of Spanish (taken from Butler 1997:67)

* This sequence is almost certainly spread over two speaker turns.

lo que pasa es que	124	no no no no no	90
desde el punto de vista	70	gol gol gol gol gol	48
sí sí sí sí sí	42	lo que ocurre es que	42
a mí me parece que	32	no tiene nada que ver	31
buenas tardes hola buenas tardes	31	y la verdad es que	28
en el sentido de que	28	lo que hay que hacer	26
la verdad es que no	25	el punto de vista de	25
buenas noches hola buenas noches*	25	hay que tener en cuenta	23
el congreso de los diputados	23	que hay que hacer es	21
y eso es lo que	20		

Table 3: The most frequent sequences of five words in the spoken component of the spoken component of the Corpus de Referencia (taken from Butler 1997:67)

* This sequence is almost certainly spread over two speaker turns.

Some of the above sequences (quite apart from those which simply constitute repetitions for emphasis or through hesitation) are grammatically anomalous: for instance, *thank you very much indeed* has no Subject, where one would normally be expected; *and so on and so forth* does not correspond to the normal rules for the use of *so*. The majority, however, can quite easily be generated by the normal rules of the grammar. Furthermore, although some are not entirely transparent semantically (for instance the idiomatic interpretation *on the other hand*, although related metaphorically to its literal interpretation, cannot easily be deduced from it), the majority are. In other words, while some of the frequent sequences in spoken English and Spanish have at least some of the properties traditionally associated with idioms, most do not. In an earlier article, I make some suggestions for the treatment of such sequences in FG, which are summarised and further developed in what follows.⁷⁰

3.4.1 Formal properties

Firstly, note that there are groups of sequences containing some common material, but with variation in what precedes and/or follows: examples are given in diagrammatic form in Table 4.

at	the	end	of	the	
at	the	beginning	of	the	
at	the	bottom	of	the	
in	the	middle	of	the	
y		yo	creo	que	
pero		yo	creo	que	
no(,)	yo	creo	que		
	yo	creo	que	el	
	yo	creo	que	es	
	yo	creo	que	no	
	lo	que	pasa	es	que
	lo	que	ocurre	es	que
y	la verdad			es	que
	la verdad			es	que no
	que	es	lo	que	
	qué	es	lo	que	
y	eso	es	lo	que	

Table 4: Examples of common core content in English and Spanish sequences

Secondly we find that almost all frequent sequences begin with a function word rather than a lexical word: conjunctions, articles, pronouns, prepositions are all represented in the English and Spanish examples. This suggests that these expressions are largely located structurally within the nominal and prepositional phrases of the two languages, in some cases optionally preceded by a conjunction. Indeed, we can simplify this picture even further: both nominal and prepositional phrases correspond to term structures in FG. We may therefore propose that the basic principles underlying the semantic structure of the clause in FG can be maintained, in that there is still a basic distinction between predicates and terms, but that the structure of terms inserted into argument and satellite positions in the clause will be formed in accordance with some combination, as yet not understood, of grammatical and phraseological mechanisms.

3.5.2 *Functional properties*

Reference to Tables 1-3 will show that the frequent sequences of 4 or more words in English and Spanish, setting aside those which consist of repetitions for emphasis or through hesitation, represent three major types of meaning: representational, interpersonal and a third category concerned with information management. Each of these is dealt with in turn below.

3.5.2.1 Representational meanings

Rather few of the frequent sequences in English or Spanish encode representational meanings, and those which do tend to reflect one of the following:

common topics in the corpus under analysis: *in the House of Commons*, *los medios de comunicación* (“the media”), *el Congreso de los Diputados* (“the Chamber of Deputies”);
 expressions of time and place (*at the end /beginning/bottom of the, in the middle of the, and at the same time, a lo largo de* (“along, throughout”), *a la hora de* (“when it comes to”).

3.5.2.2 Interpersonal meanings

Most of the frequent sequences represent meanings concerned with speaker-addressee interaction, the expression of modalities of various kinds, and speaker-oriented comments:

the expression of thanks or appreciation in English: *thank you very much indeed, that would be very nice*;
 speaker agreement or disagreement in Spanish: *sí, sí, sí, sí, sí* (“yes, yes, yes, yes, yes”), *no, no, no, no, no* (“no, no, no, no, no”);
 speaker assessment of probability: *it seems to me that, as far as I know, as a matter of fact, (y) la verdad es que (no)* (“(and) the truth is that”, with or without a negative), *a mí me parece que* (“it seems to me that”), *que a lo mejor* (“that probably”), *yo no sé si* (“I don’t know if/whether”), and all the sequences with *yo creo que* (“I believe/think that”);
 obligation: *(lo que hay que hacer (es))* (“what needs to be done (is)”), *hay que tener en cuenta* (“it’s necessary to take into account”);
 point of view (*from the point of view of*, the overlapping components of *desde el punto de vista de* (“from the point of view of”));
 other meanings connected with speaker comment: *and I said well I, vamos a ver si* (“let’s/we’ll see if/whether”), *no tiene nada que ver* (which would be followed by *con* —“has nothing to do with”).

In principle, mechanisms for the analysis of such meanings are available within FG. Various proposals for the analysis of illocution have been made in the literature.⁷¹ There is also an extensive FG literature on aspects of modality.⁷² Rather than attempt to deal with every kind of meaning exemplified above, I will concentrate here on one sequence in Spanish, in order to illustrate the possibilities, and indeed the complexities, which are involved.

Person/number	Simple form	Reinforced form	Interrupted
1 singular	157	31	3
2 singular	-	1	-
polite 2 singular	1	2	-
3 singular	-	-	-
1 plural	3	1	-
2 plural	-	-	-

polite 2 plural	-	-	-
3 plural	-	-	-

Table 5: Frequencies of (a mí) me parece que, (a ti) te parece que, etc., in the spoken part of the Corpus de Referencia

Consider the sequence *a mí me parece que* (“it seems to me that”) in Spanish. This is formed according to completely regular syntactic rules, which allow the “redundant” double expression of the indirect object pronoun (*a mí* and *me*), so giving a more emphatic, reinforced version of the simpler form *me parece que*. There are parallel syntactically well-formed possibilities for other pronouns and for full noun phrases: (*a él/ella*) *le parece que* (“it seems to him/her that”), *a Juan le parece que* (“it seems to John that”), (*a usted*) *le parece que* (“it seems to you (polite) that”), and so on. If, however, we look at the actual use of these expressions in corpora of spoken Spanish, we note a very strong skewing of frequencies. From the sequences found in the spoken component of the Corpus de Referencia, listed in Table 5, it can be seen that the first person form (*a mí*) *me parece que* occur with very much higher frequency than the others. Note that there was not a single example, in a corpus of about a million words, of a sequence of the form *a* + 3rd person pronoun/full NP + *le parece que*, or of any 2nd or 3rd person plural forms. In other words, although the productive pattern is there to be used for persons other than first singular, and indeed is so used on rare occasions, it is overwhelmingly the expression of the speaker’s own view, rather than questioning that of the addressee or reporting on that of a third party, for which this construction with *parece* is used.

Note also that in three cases the sequence *a mí me parece que* is interrupted by other material:

- (10) ... *a mí este mallorquín me parece que* ...
 ... it seems to me that this Majorcan ...
- (11) ... *a mí eso me parece que* ...
 ... it seems to me that that ...
- (12) ... *a mí, en principio, me parece que* ...
 ... it seems to me, in principle, that ...

In the first two cases, we have preposing of the subject of the *que* clause: compare *a mí me parece que este mallorquín/eso...* In the third case the interrupting material is an adverbial. These examples remind us once again that expressions such as this are not entirely fixed and immutable. Furthermore, the fact that the initial *a mí* can be omitted, and that the sequence can be interrupted only immediately after this constituent, shows that we are still dealing here with the internal structure of the sequence as generated by the normal syntactic rules of the language.

The foregoing discussion suggests that we need to build in two types of information about *a mí me parece que* and the simpler *me parece que*. On the one hand, they can be generated by the normal open-choice syntactic rules, and the possibilities for

interruption of the larger sequence are those which would be predicted by these rules. On the other hand, the whole sequence acts in a unitary fashion as far as its meaning is concerned: it acts as a modifier of the propositional content of the clause which, syntactically, follows it, and should therefore be represented in the underlying structure of that content as a subjective, epistemological modality.⁷³

We saw earlier that according to mainstream FG theory, a meaning should be treated in terms of an *operator* if it is realised grammatically, but as a *satellite* if realised lexically. Hengeveld illustrates the two possibilities in relation to example (13) below:

(13) *It seems that it is possible that he can cure blindness.* = Hengeveld's (49)⁷⁴

For a language in which the modality (which Hengeveld labels “quotative”) is grammatically realised, the underlying structure would involve a Quotative operator (as well as an operator for the objective modality represented in (11) by *it is possible*). For English, however, as the quotative modality is represented lexically, Hengeveld proposes the following structure⁷⁵:

(14) DECL (S) (A) (X_i: [Seem_v (X_j): Pres e_i: Possible_A (e_j: [Can_v Cure_vinf (x_j: p3(x_j))_{Ag} (x_k: blindness (x_k))_{Go}] (e_j)_o] (e_i)] (X_j)_o] (X_i))

By analogy, we might propose a structure such as (16) for the example from the Corpus de Referencia given below as (15):

(15) ... *a mí me parece que es impresionante.*
... it seems to me that it is impressive.

(16) DECL (S) (A) (X_i: [Parecer_v (X_j: Pres e_i: [Impresionante_A (x_j: p3(x_j))_o] (e_i)] (X_j)_o (p1(x₂))_{RecExp} (X_i))

(16), then, corresponds to the formal properties of the sentence: it allows the generation of the correct structure, and is related to the simpler form *me parece que* by the presence or absence of the first person argument. However, it ignores the fact that (*a mí*) *me parece que* as a whole realises a meaning of (emphatic or less emphatic) speaker qualification.

In order to represent this fact, we need a second analysis, in which the sequence is seen as modifying the propositional content corresponding to *es impresionante*. There is, however, a problem with such a formulation, which is not specific to this particular area of the grammar, but strikes at the heart of the operator/satellite distinction. This problem has been recognised by Vet, in his analysis of *je crois* (“I believe, think”), *je sais* (“I know”) and epistemic uses of *devoir* (“must, have to”) in French.⁷⁶ Vet demonstrates that the first two of these elements do not behave in the same way as other persons and tenses of the verbs, but must lie outside the proposition, acting as modal modifiers. Epistemic *devoir* is likewise analysed as an extra-propositional modifier. But as Vet points out, according to the classical model we cannot readily analyse these elements as satellites, since these are normally adverbial in nature.⁷⁷

Neither, however, do they correspond to what are normally classified as operators, since they are not realised by some grammatical morpheme (affix, or whatever). Vet concludes:

the category of modal expressions can be realized by a great variety of formal means. Suffixes are one of them, but similar values can also be expressed by a kind of matrix clause (*je crois* “I believe”) or an epistemic auxiliary verb (*devoir_E*) which are outside the underlying proposition and do not function as the predicate of the predication. They cannot be regarded as adverbial satellites, and consequently have to be regarded as operators (or modifiers).⁷⁸

Similarly, in our Spanish example it would not be possible, within the classical model, to treat (*a mí me parece que* as either a satellite (as it is not adverbial) or as a normal operator (as it contains lexical material). The same argument would apply to *it seems to me that*, which Altenberg’s work demonstrated to be a frequent sequence in English (see Table 1).

It seems, then, that we need to broaden the concept of operator (or, alternatively, that of satellite). If we do this, however, the rationale for the distinction between operators and satellites is seriously weakened. Furthermore, the problem is exacerbated by the fact that the relationship between grammatical and lexical realisations of a particular type of meaning cannot easily be captured in the current mainstream model of FG, because no connection is made, within the theory, between operators and corresponding satellites. This means, for example, that a past time meaning realised by tense cannot be systematically related to the more specific meaning provided by a past time adverbial in the same clause.

A proposal which might, in principle, solve the problem is made by Nuyts, who suggests that operators represent the less specific, and satellites the more specific, meanings within a particular semantic domain, and that

... one might wish to have some deeper level in the grammar at which each qualification can be treated as the coherent conceptual phenomenon it appears to be, irrespective of its varying expression forms.⁷⁹

This proposal remains to be worked out, but would be a welcome addition to FG if it meant that we could make links between similar meanings, either within a language or across languages.

3.5.2.3 Information management meanings

One group of frequent sequences remains to be discussed: those involved in the management of information in the text. These are of several kinds:

sequences in English which have the function of indicating that further information which could have been given is being assimilated to that which has already been provided: *and that sort of thing, and so on and so forth, and all the rest of it*; sequences in Spanish which give focus to a particular element of the message: *lo que pasa/ocurre (es) que* (lit. “what happens is that”, often best translated as “the

thing is (that)”), and the forms with *es lo que*: (*y eso es lo que* (“(and) that’s what ...”), *que es lo que* (“which is what ...”), *qué es lo que* “what is it that ...?/what it is that”);

the reformulating device *en el sentido de que* (“in the sense that”) in Spanish; the Spanish sequence *y por lo tanto* (“and so”) which acts as a connective between clauses or sentences;

we can perhaps include here *en el caso de* (“in the case of”), which is often used to pick out one particular situation in contrast to others.

Aspects of the meaning of some of these expressions could perhaps be handled in FG in connection with the assignment of the Focus pragmatic function. Others will have to await the development of a more detailed model of information management than is yet available. This is clearly an area where Systemic Functional Grammar is at a much more advanced stage than FG: compare, for example, the minimal treatment of “connectors” in *The Theory of Functional Grammar, Part 2*⁸⁰ with the much more extensive account given in Halliday’s *An Introduction to Functional Grammar*.⁸¹

3.6 CONCLUDING REMARKS

I hope, in this article, to have succeeded in demonstrating the importance which should be attached to the incorporation of multi-word lexical phenomena into the theoretical perspective of functional approaches to language. This is an area which is only just beginning to be explored within the FG framework. Initial indications are that we may be able, to some extent at least, to accommodate phraseological phenomena within the basic framework of FG, but that this will require developments in a number of areas.

Firstly, since many frequently used multi-word sequences have interpersonal functions, in introducing the addressee’s opinion, signalling point of view, modal meanings, certain types of speech act, and so on, we need a more fine-grained account of the interpersonal level in FG than is currently available. Secondly, since a further function of multi-word sequences is as information management devices, we also need a more detailed account of certain types of discourse-related phenomena. The recent appearance of a volume of articles dedicated to pragmatics and discourse in FG,⁸² and also the inclusion of a chapter on discourse in *The Theory of Functional Grammar, Part 2*, suggests that we may not have to wait too long for an appropriate discourse model. Thirdly, since the vast majority of multi-word sequences consist of noun phrases, either alone or within prepositional phrases, and optionally preceded by a conjunction, it would seem that many phraseological phenomena are located within terms, suggesting that we might initially concentrate on term structure in our efforts to formulate a model which takes account of both open-choice grammar and the idiom principle.

There are, however, some thorny problems to be resolved. Firstly, we have seen that a satisfying analysis of MWLPs with modal meanings presents a serious challenge to the distinction between grammatically-realised meanings handled through operators, and lexically-realised meanings dealt with through the insertion of satellites. Further development of the ideas put forward programmatically by Nuyts might lead to a resolution of these problems. Secondly, although FG is, as we have seen, committed to the achievement of psychological adequacy, there is as yet very little

work which takes the findings of psycholinguists (or, for that matter, neurolinguists, language acquisition specialists, etc.) and uses them to reshape the FG model. The further study of multi-word lexical phenomena could provide an important arena for tackling these fundamental issues.

Notes

1. John McH. Sinclair, *Corpus, Concordance, Collocation* (Oxford: Oxford University Press, 1991) 110.
2. Gordon H. Tucker, "So Grammarians Haven't the Faintest Idea: Reconciling Lexis-Oriented and Grammar-Oriented Approaches to Language", in *Functional Descriptions: Theory in Practice*, ed. Ruqaiya Hasan, Carmel Cloran, and David G. Butt (Amsterdam & Philadelphia: John Benjamins Publishing Company, 1996) 146-78.
3. Tucker, "So Grammarians Haven't the Faintest Idea" 164.
4. James R. Nattinger and Jeanette S. DeCarrico, *Lexical Phrases and Language Teaching* (Oxford: Oxford University Press, 1992) 1.
5. John M. Sinclair, "Trust the Text", in *Advances in Systemic Linguistics*, ed. Martin Davies, and Louise Ravelli (London: Pinter, 1992) 5-19.
6. Bengt Altenberg, "Speech as Linear Composition", in *Proceedings from the Fourth Nordic Conference for English Studies, Helsingør, May 11-13, 1989. Vol. 1*, ed. Graham Caie, et al. (Copenhagen: Department of English, University of Copenhagen, 1990) 133-43; Antoinette Renouf, "What Do You Think of That: A Pilot Study of the Phraseology of the Core Words of English", in *New Directions in English Language Corpora: Methodology, Results, Software Developments*, ed. Gerhard Leitner (Berlin and New York: Mouton de Gruyter, 1992) 301-17; Christopher S. Butler, "Repeated Word Combinations in Spoken and Written Text: Some Implications for Functional Grammar", in *A Fund of Ideas: Recent Developments in Functional Grammar*, ed. Chris S. Butler, et al. (Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), 1997) 60-77.
7. P.A. Luelsdorff, *What is a Cliché?* LAUT Papers. 85 (Trier: LAUT, 1981) 1, as cited in Richard J. Alexander, "Problems in Understanding and Teaching Idiomaticity in English", in *Wortschatzarbeit. Anglistik & Englischunterricht Vol. 32*, ed. Hans-Jürgen Diller, et al. (Heldelberg: Carl Winter Universitätsverlag, 1987) 107.
8. Richard J. Alexander, "Fixed Expressions in English: Reference Books and the Teacher", *English Language Teaching Journal* 38 (1984): 127-34; Alexander, "Problems" 105-22; Richard J. Alexander, "Fixed Expressions, Idioms and Collocations Revisited", in *Beyond Words. Papers from the Annual Meeting of the British Association for Applied Linguistics, University of Exeter, September 1988. British Studies in Applied Linguistics 4*, ed. Paul Meara (London: British Association of Applied Linguistics in association with Centre for Information on Language Teaching and Research, 1989) 15-24.
9. Alexander, "Problems" 110.
10. Alexander, "Problems" 108; Alexander, "Fixed Expressions, Idioms and Collocations Revisited" 17. See also the discussion in Ronald A. Carter, *Vocabulary: Applied Linguistic Perspectives* (London: Unwin Hyman, 1987) 58-62.
11. James R. Nattinger, "A Lexical Phrase Grammar for ESL", *TESOL Quarterly* 14 (1980): 337-44; James R. Nattinger, "Some Trends in Vocabulary Teaching", in *Vocabulary and Language Teaching*, ed. Ronald Carter, and Michael McCarthy (London and New York: Longman, 1988) 62-82; Nattinger and DeCarrico.
12. Nattinger and DeCarrico 1.
13. Nattinger and DeCarrico 38.
14. Nattinger and DeCarrico 38ff.

15. Nattinger and DeCarrico 59ff.
16. Gloria Corpas Pastor, *Manual de fraseología española* (Madrid: Gredos, 1996) 50-52.
17. Alison Wray, "Protolanguage as a Holistic System for Social Interaction", to appear in *Language and Communication* 18.1 (1998).
18. John M. Sinclair and Antoinette Renouf, "A Lexical Syllabus for Language Learning", in *Vocabulary and Language Teaching*, ed. Ronald Carter, and Michael McCarthy (London and New York: Longman, 1988) 140-60; Antoinette Renouf and John McH. Sinclair, "Collocational Frameworks in English", in *English Corpus Linguistics. Studies in Honour of Jan Svartvik*, ed. Karin Aijmer, and Bengt Altenberg (London and New York: Longman, 1991) 128-43; Renouf, "What Do You Think of That".
19. Bengt Altenberg and Mats Eeg-Olofsson, "Phraseology in Spoken English: Presentation of a Project", in *Theory and Practice in Corpus Linguistics*, ed. Jan Aarts, and Willem Meijs (Amsterdam and Atlanta, GA: Rodopi, 1990) 6-7.
20. Altenberg, "Speech as Linear Composition"; Mats Eeg-Olofsson and Bengt Altenberg, "Discontinuous Recurrent Word Combinations in the London-Lund Corpus", in *Creating and Using English Language Corpora. Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*, ed. Udo Fries, Gunnel Tottie, and Peter Schneider (Amsterdam: Rodopi, 1994) 63-77; Bengt Altenberg, "Recurrent Verb-Complement Constructions in the London-Lund Corpus", in *English Language Corpora: Design, Analysis and Exploitation. Papers from the Thirteenth International Conference on English Language Research on Computerized Corpora, Nijmegen 1992*, ed. Jan Aarts, Pieter de Haan, and Nelleke Oostdijk (Amsterdam: Rodopi, 1993) 227-45.
21. See especially Göran Kjellmer, "Aspects of English Collocation", in *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora*, ed. Willem Meijs (Amsterdam: Rodopi, 1987) 133-40; Göran Kjellmer, "A Mint of Phrases", in *English Corpus Linguistics. Studies in Honour of Jan Svartvik*, ed. Karin Aijmer, and Bengt Altenberg (London and New York: Longman, 1991) 111-27.
22. Kjellmer, "A Mint of Phrases" 116.
23. Butler, "Repeated Word Combinations"; Christopher S. Butler, "Collocational Frameworks in Spanish", *International Journal of Corpus Linguistics* (1998).
24. Butler, "Repeated Word Combinations"; Butler, "Collocational Frameworks in Spanish."
25. Nick C. Ellis, "Sequencing in SLA: Phonological Memory, Chunking and Points of Order", *Studies in Second Language Acquisition* 18 (1996): 91-126; Nick C. Ellis, "The Epignesis of Language: Acquisition as a Sequence Learning Problem", in *Evolving Models of Language. Papers from the Annual Meeting of the British Association for Applied Linguistics Held at the University of Wales, Swansea, September 1996*, ed. Ann Ryan, and Alison Wray (Clevedon: Multilingual Matters, 1997) 41-57; Nick C. Ellis, "Memory for Language", in *Cognition and Second Language Instruction*, ed. P. Robinson (Cambridge: Cambridge University Press, forthcoming).
26. Ellis, "Sequencing in SLA" 93.
27. Regina Weinert, "The Role of Formulaic Language in Second Language Acquisition: A Review", *Applied Linguistics* 16.2 (1995): 180-205.
28. A. M. Peters, *The Units of Language Acquisition* (Cambridge: Cambridge University Press, 1983); Nattinger and DeCarrico.
29. Weinert, "The Role of Formulaic Language" 185.
30. Ellis, "Sequencing in SLA."
31. Wray, "Protolanguage."
32. Wray, "Protolanguage."
33. Wray, "Protolanguage" page number to be inserted when article is published.
34. Simon C. Dik, "On the Notion 'Functional Explanation'," *Belgian Journal of Linguistics* 1 (1986): 21.

35. Simon C. Dik, *The Theory of Functional Grammar. Part 1: The Structure of the Clause*, 2nd ed. Kees Hengeveld (Berlin & New York: Mouton de Gruyter, 1997) 3.
36. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 6.
37. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 13.
38. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 14.
39. Kees Hengeveld, "Clause Structure and Modality in Functional Grammar", in *Ins and Outs of the Predication*, ed. Johan van der Auwera, and Louis Goossens (Dordrecht and Providence RI: Foris Publications, 1987) 53-66; Kees Hengeveld, "Illocution, Mood and Modality in a Functional Grammar of Spanish", *Journal of Semantics* 6 (1988): 227-69.
40. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 50, 55.
41. Kees Hengeveld, "Layers and Operators in Functional Grammar", *Journal of Linguistics* 25 (1989): 128.
42. Michael A. K. Halliday, *An Introduction to Functional Grammar*, 2nd (London: Edward Arnold, 1994).
43. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 103.
44. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 103-04; Simon C. Dik, "Idioms in a Functional Grammar" (Amsterdam: University of Amsterdam, 1988), reprinted in *Computational Lexicology and Lexicography: Special Issue Dedicated to Bernard Quemada, I. Linguistica Computazionale VI*, ed. L. Cignoni, and C. Peters (Pisa: Giardini, 1992) 241-65. All further page references are to the 1992 version.
45. Simon C. Dik, *Functional Grammar in Prolog: An Integrated Implementation for English, French, and Dutch*, Natural Language Processing.2 (Berlin and New York: Mouton de Gruyter, 1992) 48, 203.
46. Dik, "Idioms" 246.
47. Dik, *The Theory of Functional Grammar, Part 1*, 1st edn. 87.
48. Dik, "Idioms" 258.
49. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 103.
50. Rolandt Tweehuysen, "Verbale Suffigering (s-Vorm) in Het Zweeds", Diss. (Amsterdam: University of Amsterdam, 1988).
51. Dik, "Idioms" 256.
52. Dik, "Idioms" 256.
53. Dik, "Idioms" 246.
54. Dik, "Idioms" 246-8.
55. Dik, "Idioms" 253ff.
56. Dik, "Idioms" 253.
57. Dik, "Idioms" 253.
58. Dik, "Idioms" 257ff.
59. Ahmed Moutaouakil, "Discourse Ambiguity in Functional Grammar: Idioms and de-Idiomatized Idioms", in *A Fund of Ideas: Recent Developments in Functional Grammar*, ed. Chris S. Butler, et al. (Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), 1997) 78-94.
60. Moutaouakil, "Discourse Ambiguity in Functional Grammar" 79.
61. Simon C. Dik, "FG*CM*NLU: Functional Grammar Computational Model of the Natural Language User", in *Functional Grammar and the Computer*, ed. John H. Connolly, and Simon C. Dik (Dordrecht and Providence RI: Foris Publications, 1989) 1-28.
62. Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 44-47.
63. Moutaouakil, "Discourse Ambiguity in Functional Grammar" 87.
64. Moutaouakil, "Discourse Ambiguity in Functional Grammar" 89-90.
65. Moutaouakil, "Discourse Ambiguity in Functional Grammar" 90ff.
66. Moutaouakil, "Discourse Ambiguity in Functional Grammar" 92.
67. Altenberg, "Speech as Linear Composition".
68. Butler, "Repeated Word Combinations" 67.

69. The Corpus de Referencia was compiled by Francisco Marcos Marín at the Universidad Autónoma de Madrid. The generally larger frequencies as compared with those for the English corpus are due to differences in corpus size.
70. Butler, "Repeated Word Combinations".
71. See especially: Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 299-307; Simon C. Dik, *The Theory of Functional Grammar, Part 2: Complex and Derived Constructions*, 2nd, ed. Kees Hengeveld (Berlin & New York: Mouton de Gruyter, 1997) 229-56; Hengeveld, "Layers and Operators"; Kees Hengeveld, "The Hierarchical Structure of Utterances", in *Layers and Levels of Representation in Language Theory*, ed. Jan Nuyts, A. Machtelt Bolkestein, and Co Vet (Amsterdam and Philadelphia: John Benjamins Publishing Company, 1990) 1-23; Ahmed Moutaouakil, *Towards and Adequate Representation of Illocutionary Forces in Functional Grammar*, Working Papers in Functional Grammar.10 (Amsterdam: University of Amsterdam, 1986); Ahmed Moutaouakil, *On Representing Implicated Illocutionary Force: Grammar or Logic?* Working Papers in Functional Grammar.40 (Amsterdam: University of Amsterdam, 1991); A. Machtelt Bolkestein, "Limits to Layering: Locatability and Other Problems", in *Layered Structure and Reference in a Functional Perspective*, ed. Michael Fortescue, Peter Harder, and Lars Kristoffersen (Amsterdam and Philadelphia: John Benjamins Publishing Company, 1992) 387-407; Rodie Risselada, *Illocutionary Function and Functional Illocution*, Working Papers in Functional Grammar. 34 (Amsterdam: University of Amsterdam, 1990); Co Vet, "Modal Verbs in the Layered Clause Structure: Speech Acts and Truth Value", in *A Fund of Ideas: Recent Developments in Functional Grammar*, ed. Chris S. Butler, et al. (Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), 1997) 19-28.
72. Hengeveld, "Clause Structure and Modality"; Hengeveld, "Illocution, Mood and Modality"; Hengeveld, "Layers and Operators"; Hengeveld, "The Hierarchical Structure of Utterances"; Kees Hengeveld, "Mood and Modality", in *Morphology: A Handbook on Inflection and Word Formation*, ed. Geert Booij, Christian Lehmann, and Joachim Mugdan (Berlin and New York: Mouton de Gruyter, forthcoming); Dik, *The Theory of Functional Grammar, Part 1*, 2nd edn. 241-42, 295-99; Jan Nuyts, *Aspects of a Cognitive-Pragmatic Theory of Language* (Amsterdam and Philadelphia: John Benjamins Publishing Company, 1992); Jan Nuyts, "How Do You think", in *A Fund of Ideas: Recent Developments in Functional Grammar*, ed. Chris S. Butler, et al. (Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), 1997) 3-18.
73. See, for example, Hengeveld, "Illocution, Mood and Modality"; Hengeveld, "Layers and Operators".
74. Hengeveld, "Illocution, Mood and Modality" 245.
75. p3 is used here as a shorthand notation for the third person pronoun, and p1 will be used for the first person pronoun. For a more detailed specification of pronouns in FG, see Dik *The Theory of Functional Grammar, Part 1*, 2nd edn. 152-153). Note also that the overall approach to illocution exemplified here is different from that advocated in Dik's *The Theory of Functional Grammar*.
76. Vet, "Modal Verbs in the Layered Clause Structure".
77. Vet, "Modal Verbs in the Layered Clause Structure" 25.
78. Vet, "Modal Verbs in the Layered Clause Structure" 27.
79. Nuyts, *Aspects of a Cognitive-Pragmatic Theory of Language* 224-5.
80. Dik, *The Theory of Functional Grammar*, Vol. 2: Complex and Derived Constructions 440-41.
81. Halliday, *An Introduction to Functional Grammar* 49, 323 ff.
82. John H. Connolly, et al., *Discourse and Pragmatics in Functional Grammar*, Functional Grammar Series. 18 (Berlin and New York: Mouton de Gruyter, 1996).