

**A COMPUTER-AIDED RE-APPRAISAL OF CO-SELECTION:
IMPLICATIONS FOR INTERLINGUAL AND
INTERCULTURAL COMMUNICATION**

Luis Pérez González

Universidad Europea-CEES, Madrid

Antonia Sánchez Macarro

Universitat de València

ABSTRACT

This article argues that the systemic study of the realisational relationship which holds between “meaning” and “form” could benefit from the appropriation of recent advances in the field of Corpus Linguistics. We take as our point of departure the key systemic contention that the meaning potential of language can be ultimately reduced down to finite system networks of choices, whose most delicate realisations —i.e. the ultimate units conveying meaning— have so far received a relatively small amount of scholarly attention. By contrast, the fact that meaning often spans several lexical units correlating with distinctive grammatical patterns has come to feature increasingly higher in the research agenda of corpus linguists, thus bringing to the fore the importance of the so-called “co-selectional” relationships. This article illustrates the dynamics of this relationship between lexis and grammar by drawing on a corpus-based analysis of semi-idiomatic expressions. Further, it attempts to stake out some sort of middle ground between Systemic Functional and Corpus Linguistics as well as find a common underpinning to the formal realisation of meaning.

**1. INTRODUCTION: THE SYNTAGMATIC AND PARADIGMATIC AXES IN
SYSTEMIC FUNCTIONAL LINGUISTICS**

It is a widely held view that Systemic Functional Linguistics (hereafter SFL) has made, since the earliest formulations of the model, the greatest strands in recognising

“use” and “meaning” as key issues in the study of language. Drawing upon Halliday’s (1961) earlier account of “context” as an interlevel between meaning and situation, Gregory set out to formalise the “recurrent characteristics of user’s use of language in situations” (1967: 185). The ensuing account of context in terms of the well-known “field,” “tenor” and “mode”—related constraints on language constitutes a purely functional response to the question of *how language is structured for use*. But the systemic approach to language has also reinterpreted this issue in semantic terms, and explained *how language is organised to make meanings*. In this sense, it is held (Halliday 1985) that the recurrent contextual categories of field, tenor and mode correlate with three major types of meaning which are activated concurrently. Accordingly, the “experiential,” “interpersonal” and “textual” meanings are bundled together in the plane of language form. In defining these types of meanings, Halliday contends that of all the uses we make of language, which are limitless and changing, language is designed to fulfil three main functions: a function for relating experience, a function for creating interpersonal relationships and a function for organising information.

The communicator’s construal of linguistic meanings in accordance with the contextually relevant constraints proceeds through networks of systems embracing the range of options available at any given moment, at different levels of linguistic organisation. Each system provides the speaker with the opportunity to choose among the constitutive elements of the currently relevant “paradigm” in order to convey the meaning in question. The speaker’s choice engages then in “syntagmatic” relations with the items selected from the contiguous systems or paradigms, all of which constitute instances of the same “rank.” The construal of linguistic meanings may thus be held accountable for the consequential interaction between the paradigmatic axis of choice and the syntagmatic axis of chain at each rank of linguistic organisation.

In a recent publication, Tucker (1996) acknowledges that “systemicists have always insisted on the primacy of choice of paradigmatic relations. If the primacy of choice is an accepted starting point, then *chain* (syntagmatic relations) is to be explained as a consequence of choice” (italics in the original; 1996: 541). However, the prominence given to the paradigmatic axis at the expense of its syntagmatic counterpart is being increasingly challenged by other practitioners of SFL. O’Donnell (1996), for example, contends that the model’s reliance on the system networks of choices—which causes context to be modelled exclusively in paradigmatic terms—neglects other aspects of context and argues for the need “to add a syntagmatic axis to context modelling, taking into account the structural relations between contextual entities” (1996: 16). This paper will tackle SFL’s biased orientation toward the paradigm, taking as its point of departure the assumption that the latter has been detrimental for the systemic study of lexis.

The systemic notation makes it clear that choices made at the discourse-semantic level are realised through the pre-selection of features from the grammatical systems. Accordingly, the choice of an “initiating” move involves, for instance, the pre-selection from the “mood” system of a congruent clause-type which, in its turn, will be realised through the pre-selection of certain actor/process/goal-types at the “transitivity” level. To a certain extent, SFL has concentrated on the projection of the experiential, interpersonal and textual meanings downward to the level of lexico-gram-

mar, such that the lexical items are regarded as paradigmatic choices which fill slots in the ensuing syntagmatic structures. More pointedly, SFL has done so primarily with a view to establish how such choices contribute to the overall cohesion of a text. However, recent developments in the field of Corpus Linguistics have brought into sharp relief the importance of the syntagmatic relations of co-occurrence between contiguous words. Such relationships include traditional notions such as “collocation” and “colligation,” but also others such as “co-selection” whose existence relied so far entirely on the linguist’s intuition. In the remaining sections of this paper, we propose to look at the interface between context and “co-text,” thus balancing the paradigmatic and syntagmatic dimensions and demonstrating that pre-selection also exists at the level of lexical combination.

2. THE STUDY OF LEXIS WITHIN THE MAP OF SYSTEMIC FUNCTIONAL LINGUISTICS

The role of lexis within SFL has undergone several changes since the earliest formulations of the model. It stands to reason that, from a strictly *functional* point of view, the rank of group/phrase consists of one/more words; each of these constituents may be regarded as slots in which lexical items can be fitted, subject to certain obvious restrictions. To give but one example, the modifier slot before the head of a nominal group is typically filled by members of the class of adjective or noun, and the qualifier slot is commonly realised by means of a prepositional phrase. Now, SFL is also *semantic*. Accordingly, it is expected to account for the fact that a grammatical utterance does not always result from the choice of a given word from the right class to fill the slot in question. Alternatively, other additional factors have to be borne in mind when carrying out the selection. In this section we intend to discuss how SFL has traditionally dealt with the functional and semantic aspects of lexical organisation.

Halliday (1961) is known to have set high expectations as to what SFL may one day tell us about the organisation of lexis. As a grammarian, Halliday “would like to turn the whole of linguistic form into grammar, hoping to show that lexis can be defined as the “most delicate grammar” (...) and all the exponents ranged in systems” (1961: 267). Halliday (1966), however, assesses the feasibility of this enterprise and reformulates his initial claims. The theory of lexis is now claimed to be complementary to, but not part of grammatical theory. Drawing upon Firth’s (1957) previous work on “collocations” —i.e. the way words convey meaning by keeping company with other specific words— Halliday lays stress on the parallels existing between grammar and lexis. As intimated above, grammar is organised by means of “systems” (paradigms) and “structures” (syntagms). As far as lexis is concerned, the paradigmatic dimension is represented by “sets” —that is, the systems of conceptually linked words available at any given moment— while the syntagmatic dimension correlates with the notion of “collocations” —i.e. the co-occurrence of certain words within a given syntactic structure. This being so, simple interpretations of these postulates might lead to infer that SFL is mainly concerned with how the lexical fillers of the syntagmatic slots resemble/differ from the paradigmatic choices which realise other syntagmatic structures elsewhere in the text, thus explaining how the relevant field,

tenor and mode constraints result in similar/opposing choices at different compartmentalised slots.

But SFL's predominantly functional stance toward lexis in the sixties evolved and came to exhibit a significantly increased semantic orientation in the following decade. As a matter of fact, SFL's shift of attention to collocation serves as a springboard to reinstate the original approach to lexicon as the most delicate grammar. By the mid-seventies, Halliday & Hasan (1976) assert that "there is no sharp line between grammar and vocabulary: the vocabulary or lexis, is simply the open-ended and most "delicate" aspect of the grammar of a language [and] the distinction between grammatical and lexical is really one of degree" (1976: 281). In doing so, lexicon is considered to be much more than simple grammar slot fillers. More pointedly, it is postulated that "there is only one network of lexico-grammatical options. And as these become more and more specific, they tend more and more to be realized by the choice of a lexical item rather than by the choice of a grammatical structure. But it is still part of a grammatical system" (Halliday 1974; as quoted in Tucker 1996: 534).

Work currently being undertaken under the umbrella of SFL's increasingly semantic orientation seeks to implement the latest directions for the study of lexis (Tucker 1996: 538), mainly by attempting to range all lexico-grammatical choices into systems. Insofar as system networks aim to map the meaning potential of language onto a finite number of choices, the task consists in specifying and classifying the right-most delicate realisations of such networks, i.e. the ultimate units that convey meaning.

While traditional linguistic models would have readily equated such semantic choices with individual lexical items, semantic-oriented systemicists have set out to explore the potential realisations of the aforementioned lexico-grammatical options. In doing so, they cannot ignore the growing body of literature on the phenomenon of collocation, which has received a boost in the last decades thanks to the growing availability of huge computerised corpora of linguistic data. Although the groundbreaking, intuition-based work on collocation capably showed that certain semantic meanings are realised through chains of syntagmatically related items, the latest corpus-based studies have made an important contribution, bringing to the fore the fact that certain meanings are conveyed by multi-word units whose combination responds not only to the principles of "lexical combination," but also of "grammatical colligation." In other words, recent developments in the field of Corpus Linguistics confirm that the paradigmatic lexical choices cannot always be made independently of their position in the syntagmatic chain, thereby acknowledging the existence of units of meaning activated through a single lexico-grammatical choice, as represented in the corresponding system network. In systemic terms, the upshot of the foregoing exposition is that the choice of a particular "process," for instance, allows for the selection of a restricted range of "goals." The increasing awareness that the boundaries of semantic units do not correlate with those of distinct words would hint at the necessity to reappraise the interaction between choice and chain at the level of lexico-grammar. Since "form" is the organisation of material language into meanings (Halliday 1961), the consequential spill of lexical meaning across the boundaries of discreet words along the syntagmatic axis of chain would seem to call for a more rigorous study of the formal manifestations of the most delicate lexico-grammatical choices.

3. CORPUS-BASED INSIGHTS ON THE UNIT OF MEANING

The scholarly interest in lexical co-occurrence arose from Firth's (1957) pathbreaking study of "collocation," which led him to postulate that the meaning of a word is directly contingent on the words with which it commonly associates and that competent language users must be capable of recognising typical and untypical patterns of lexical co-occurrence. This phenomenon was further explored by Halliday (1966) and Sinclair (1966), whose intuition-based characterisations of "collocation" resulted in a preliminary taxonomy of what may be loosely termed as "multi-word units." (Carter and McCarthy 1988). Accordingly, the internal bonds within these lexical patterns have been formalised with reference to a cline of predictability ranging from loosely fixed phrases to idioms and other multi-word lexical items whose collocability is totally predictable.

Generally speaking, these intuition-based studies helped arouse the scholars' awareness that the incongruous selection of specific items from contiguous lexical sets leads to the breaking of the local collocational constraints and expectations. However, the classification that emerged from these efforts proved insufficient for analysts to establish reliably which linguistic stretches constitute a fixed expression and what their internal degree of idiomaticity is.

Drawing upon his extensive research on The Bank of English, a growing computerised corpus held at the University of Birmingham (COBUILD), Sinclair (1991) himself challenges the trustworthiness of intuition-based generalisations. His preliminary corpus-based observations led him to distinguish between "collocations"—where the co-occurring elements are not necessarily adjacent—and other multi-word expressions subject to the "idiom principle," which he considers as important as grammar in the explanation of how meaning arises in text (1991: 110). Before expounding the major thrust of this new stance, let us browse through the concordance from the British National Corpus represented in Table 1 below (edited by Banks 1996). The reader is invited to propose candidates for the search node, which is common to all the concordance lines and has been blanked throughout:

hurried back to the ferrymaster. "Tell me," he to bite her lip. "Drink some of this tea," Maggie hn, aren't you going to say something?" Laura ck to school with their children, Princess Anne he Archbishop of Cape Town, Desmond Tutu, e knuckles white. Minnie poured her some tea, available and have refused to say if they have rejected the claims of the offer document and gister. All teachers in England and Wales were class culture and lived within it. They were not dvice held good for young refugees. They were ves they might know the identity of the driver is ggle one step further. Sisters and brothers are is believed to be the first time a developer has e is quite right. Our campaign has consistently

, "did you take anyone across the Forth tha , handing it to Laura. Laura took several s . He took a deep breath, and managed to s yesterday. She spoke out after a new surve Israelis to forgive the Nazis for killing her to remove her coat if she felt the room him to resign. However, Tory leader Council shareholders to hold on to their shares whi by the National Union of Teachers to have by their teachers to achieve educational su to learn English, to refrain from speaking to contact Colchester CID on (0206) 76221 to help smash sexists and classist legislat the Government to schedule an archaeologi the Government to undertake a full and pro

<p>President Bush, for the first time, last week e house arrests of main opposition figures and that there's a sucker born every minute, Klein f the nineteenth century other Indian Christians his department to look at physical education. It g what he had done, James took the reins and h he does not wish to go past, and yet is being nced co-driver, what to expect. Evans, 30, then m in July, jumping aboard at the last moment, incident which saw the flag-waving Mongolian</p>	<p>the Iraqis to take matters into their own the military government to restore democrat the suckers of the world to unite in protes that Christianity in India should assume an that all children should be taught to swim. his horse out of Corporation Lane with all on by his rider. His confusion will be indi his Sierra through Grizedale faster than Mc on by a friend who tought she might tell th on by a gaggle of grannies, invading the ri</p>
---	--

Table 1. A concordance sample from the British National Corpus (Banks 1996)

The reader will probably note that each group of concordance lines favours one or more candidate verbs to fill the slot. However, the fact that the missing filler is to be figured out on a cumulative basis will lead him/her to drop some of the so-far candidates until s/he comes up with a verb which proves suitable throughout the whole concordance (i.e. “urge”). The fact that the reader’s range of choices is progressively restricted by the text featuring at both sides of the gap suggests that the semantics of the verb is defined by the totality of its collocational complementation, that is its “co-text.” In other words, once we are provided with relevant surrounding language —the co-text of “urge”— only one lexical item can be selected to convey a given meaning. The “idiom principle” is drawn upon by Sinclair (1994) to postulate that we need to adjust our received perception on the relationship which holds between form and meaning, shifting from the discreet correlation to the indistinguishability of the latter:

- The operation of the idiom principle results in significant co-occurrences and the distribution of meaning across the constitutive elements of the idiomatic expression: “if the words collocate significantly, then to the extent of that significance, their presence is the result of a single choice” (Sinclair 1994: 24).
- The meanings of words selected together, i.e. engaged in co-occurrence patterns, is different from their independent meanings as they become at least partly delexicalised; in Sinclair’s words, a word “will not have its independent meaning in full if it is only part of a choice involving one or more words.” Sinclair illustrates his claim by looking at the traditional assumption that the adjective either adds something to the meaning of a noun or restricts it; according to his analysis, the everyday use of adjectives demonstrates that, in many cases, these elements do not add any distinct or clear meaning load; on the contrary, they are “simply underlining part of the meaning of the noun” (1994: 24), as in the following example of a delexicalised adjective:

(1) *general* trend *general* perception *general* consent
 general drift *general* opinion

- Thirdly, further work undertaken by Sinclair and later by Stubbs (1996) has shown that different forms of a given “lemma” may attract different co-texts.

In other words, each constituent of a lemma is surrounded by different lexical environments which do not normally have the capacity to replace each other. The fact that different collocational selections result in idiosyncratic co-occurring patternings is illustrated by Sinclair (1995) with the co-texts of the constituents of the lemma “eye:”

(2)	<i>eye</i>	Specific meaning only refers to the anatomical organ when talking about injuries or handicaps	turn a blind <i>eye</i> keep an <i>eye</i> on
	<i>eyes</i>	Specific meaning Co-occurs with adjectives such as “blue,” “brown,” “covetous” or “manic”	all <i>eyes</i> will be on rolling their <i>eyes</i>

All in all, the practitioners of Corpus Linguistics in the nineties have been able to substantiate in empirical terms the basics of two notions that had been first proposed by Firth forty years earlier (Palmer 1968: 181). The combination of individual words along the syntagmatic axis of chain revolves either around the relations of “collocation” —that is, the mutual expectancy between words— or the relations of “colligation” —namely, the interrelation of items belonging to specific word classes or grammatical categories.

However, the new computerised data have also allowed corpus analysts to explore and define the notion of “co-selection” as a blend of collocation and colligation. Francis (1993) and (1996) constitute an application of these theoretical developments. Building upon the assumption that traditional intuition-based grammar allows for more freedom in lexical combination than is in fact possible, Francis’ work tackles the co-selection of certain syntactic structures and lexical items/strings of lexical items and argues that the communication of certain meanings requires from the speaker the selection of certain lexical items which are, in their turn, realised through specific grammatical patterns. In short, it is Francis’ contention that each lexical co-occurrence has its unique grammatical profile.

Francis (1993) illustrates this point by looking at a specific grammatical structure which Quirk *et al.* (1985) term as “introductory *it* as object.” The concordance she works with (see Table 2 below) shows that (i) the structure at issue combines with an extremely restricted range of verbs, “find” and “make” being the most frequent thereof; (ii) likewise, when “*it*” is followed by an adjective, the range of co-occurring adjectives is also unique and restricted:

policies because they often	Find	it	difficult	to explain why a r
unemployment, many will	Find	it	harder	than usual to relax i
ist, who, unlike Genscher,	Finds	it	easy	to make friends among
orers in this country, has	Found	it	almost	impossible to get a g
nted by Lord Graham who	Found	it	necessary	to the point of hi
put Gwynned first. But he	Found	it	hard	, too, being wed to John
want to pressure him, and I	Made	it	clear	that I was not angry.
cent. Dealers said the move	Made	it	clear	that the Spanish authori

ut his party's policies, he	Made	it	plain	that Labour planned to
his strategy would at least	Make	it	possible	to conceptualise mo
ss ahead of the match have	Made	it	inevitable	that Frech rugby
inflationary pressures and	Makes	it	less	and less likely that th
against a rape conviction,	Making	it	likely	that he will remain i
area of refurbishment, by	Making	it	almost	impossible for them t

Table 2. A stepped concordance of the structure “introductory *it* as object”
(Francis 1993: 140)

The mechanism of co-selection is expounded by Francis as follows:

These lexical patterns are closely connected with the communicative function of the structure, which is to present a situation in terms of how it is evaluated [...] The ways in which we typically evaluate situations using this particular structure, are stereotyped, showing that our concerns tend very often to be with how difficult or easy life is made for us, how predictable things are, and how well we understand what is going on. (Francis 1993: 141).

In order to back up and confirm these findings, we looked at the occurrences of this structure in a 50-million-word version of the COBUILD corpus (Pérez-González & Sánchez Macarro 1996). Taking the foregoing structure as our starting point, we focused our analysis on three other candidate verbs to report the communicator's opinion, i.e. “think,” “consider” and “deem.” In doing so, it was our intention to check out whether the co-selectional restrictions also apply to a whole set of verbs which are regarded as synonyms by traditional grammars. Our scrutiny of the corpus revealed that, despite the conspicuous differences in the rate of occurrence of each of these verbs, the collocational restrictions still apply. The sets of concordances help us establish the following patterns of co-selection:

(3)

<i>find it</i>	difficult easy hard helpful interesting useful	<i>think it</i>	possible probable unlikely	<i>Consider it</i>	appropriate necessary worthwhile	<i>deem it</i>	good better
----------------	---	-----------------	----------------------------------	--------------------	--	----------------	----------------

Again, data suggest that the above co-selection patterns are highly stable and restricted, although these constraints are by no means absolute. “Appropriate” and “worthwhile,” for instance, co-occur also with “think.” The co-occurrence rate per million words is, however, considerably lower in the case of this second verb. Consequently, our analysis confirms that there are not only stereotyped ways in which situations can be evaluated, but also stereotyped ways in which the speaker evaluates by providing his/her opinion. “Deem” is the least often occurring evaluative verb and

introduces a general positive evaluation. The three remaining verbs are more selective, though; “find,” as first suggested by Francis (1993), shows the speaker’s concern with difficult and easy situations, as well as with elements which can prove useful to change the situation in question; “think” seems to be the most congruent choice when the speaker intends to evaluate an event in terms of the likelihood of its taking place. Finally, “consider” seems to be the preferred option when evaluating an element in terms of its appropriacy or relevance to a given situation.

As a matter of fact, accounting for the co-patterning of different forms and senses of lemmas and syntax would seem to have become a key objective for most of the latest pedagogical and reference tools. In broad terms, the methodological attempt to compile a grammar of the typical meanings that human communication encodes and, thereby, to tackle the interface between lexis and grammar has resulted in the redefinition of the boundaries between grammars and dictionaries (Leitner 1993). It has become a widely held assumption that language learners not only need to understand a foreign language, but also to produce it idiomatically. In order to allow communicators to create messages by making use of preformed sequences of words, grammars are to account for co-selectional formulae, whereas dictionaries need to contain a fair amount of information on the patterning of lexical items. This would seem to have become by now a well-attested insight resulting in the introduction —to a greater or lesser extent, and in various ways— of grammar into monolingual pedagogical dictionaries. Recent work, such as Noël et al.’s (1996) project “Contragram,” is intended to help this insight find its way into bilingual dictionaries. Given that co-selectional patterns mediate between the dictionary and the grammar of a language, more grammar information on the source side of a bilingual dictionary could facilitate the learner’s disambiguation task. Likewise, more grammar on the target language side could prevent many ungrammatical or unidiomatic translation.

Francis (1996) own approach to the study of the interface between lexis and grammar has been developed from the basic insight that “co-selectional patterns” —a term which, incidentally, has come to replace “structure” in her writings— incorporate a blend of semantic and grammatical information. Accordingly, verbs which share their respective complementation would be expected to share all or part of their meaning. What is more important, the ensuing patterns often flow into one another, an idea which was first suggested by Winter and his notion of “collocational streams.” The fact that one pattern may flow into the next has led these researchers to reinforce their conviction that it is patterns rather than clauses which should be regarded as the units of analysis at the interface or meeting point of lexis and syntax.

4. CO-SELECTION: SETTING THE BOUNDARIES OF DISTINCT LEXICO-GRAMMATICAL CHOICES

In the foregoing sections, we have argued that the use of computer-held corpora and the implementation of a corpus-driven methodology have proved exhaustive in their treatment of lexico-grammatical co-occurrence phenomena, as well as relevant to the formalisation of the syntagmatic relationships between more or less contiguous items. In the remainder of this paper, we intend to explore the impact that such cor-

pus-based insights may have on the design of lexical system networks within the domain of SFL.

It is obvious that evidence from corpus-based research has been able to describe the relationship between form and meaning more accurately, mainly by demonstrating that meaning units do not map on to individual lexical items and, thereby, that meaning is distributed along the syntagmatic axis between such items and their environments. While these insights are readily applicable to the design of lexical system networks in the case of juxtaposed collocates or co-occurring items, it remains to be seen how we may possibly set the boundaries and formalise the internal structure of other multi-word units whose constituents are never/not always immediately contiguous and/or, to a certain extent, subject to formal variation.

In order to explore this latest point, we took on the analysis of lexical variation within multi-word lexical units which are not totally predictable, that is semi-fixed idiomatic expressions, such as “put all one’s eggs in one basket” or “to be one sandwich short of a picnic.” The aim of our analysis was to ascertain the reliability of a corpus-based analysis of co-text with regard to the (in)variability of such expressions, as attested by the speakers’ everyday utterances. Because of constraints on the length of this article, the remainder of this section will report on the analysis of a specific example, i.e. the expression “a new broom sweeps clean.” Accordingly, Table 3 below represents the concordance provided by the fifty-million-word COBUILD on-line corpus in response to our search query.

- | | |
|-----|--|
| 1. | iance for Cinema and TV, used the word disaster.’ TX with a new broom at Number 10 the expectations engendered in June 1990 by the |
| 2. | land BL. DL> 28 June 1992 TX. ADRIAN SHINWELL, president and new broom at the Scottish Conservative and Unionist Association June 2 |
| 3. | Pounds 4 billion. The year say Nick Temple, the company’s new broom chief executive, carry through the most thorough restructurations |
| 4. | supply as it is impossible to get planning permission for new broom cupboards because they are judged too small for health an |
| 5. | at the Leeds headquarters of the 205 strong store chain, his new broom has already swept into some dark corners of the Asda empire. |
| 6. | are not good. Morale in the DTI is low and Heseltine’s new broom has yet to show its bristles. It will remain a low budget de |
| 7. | St, NW 1 71 724 2739 until March 14 TX. XR. Issue 64256 HL. New broom in Bloomsbury Robert Anderson BL. Simon Tail DL. 14 February |
| 8. | in Falkirk and latterly an officer in Northampton, is the new broom in the cosy council corridors. He says the fire fighting has |
| 9. | pecially if you live in the Network Southeast area. The new broom is sweeping through other government departments, too. |
| 10. | the country’s east coast liberal establishment. The new broom is determined to sweep clean. But he never expected the out |
| 11. | thout the food truck, ‘ he said. Issu 8758 HL. Tory new broom must tackle constituency tyrants Letter Scotland BL. DL. 28 |
| 12. | loss in the six months to end December as a result of the new broom ’s policy of Ian Rock, the managing director. About a fifth |

13. could spell Are you taking your pills. Issue 9757 HL. **New broom** ready to sweep out sacred cows Scottish Conservative and
14. leading enthusiastic strength to the hand that pushes the **new broom**. Issue 8757 HL. Why Kane is not able Leading Article
15. Socialists ballast in rough seas ahead. But he is hardly a **new broom**. With this change, M Miterrand has effectively ruled out
16. not Royal Street or Andrassy Avenue are back again. But the **new broom** seems to have swept indiscriminately. The street named after
17. rein in Orrell's jumbos, Leicester meet Rosslyn Park. The **new broom** that swept in with the new season is held by Tony Russ, form
18. 33 p to 356 p and Clinton Cards roared 15 p to 80 p. TX.HL. **New broom** for unilever City BL. DL. 28 January 1992. TX. THE jobs of La
19. for Pickard, coach of world No 1 Stefan Edberg, and the **new broom** he says has swept through the previously musty Lawn Tennis A
20. ious Oxford camp 12 months ago, reckons he has wielded a **new broom** through the Cambridge set up and produced, he reckons, a lea
21. ollecting KUA and Becker's out LTH You got it, Professor. **New broom**. Becker was the old guys' guy. LTH ZG 1 KD 1 I see. And what
22. tablished practice at Century. This is symptomatic of your **new broom**, is it, Director General. LTH CES in order that those with day
23. and the countryside looked as if it had been swept with a **new broom**. The rowns of poplars that lined the driveway were just beginning
24. NPR90- At the same time the fact that there is a **new broom** —Sharon Pratt Dixon's **new broom** sweeping out supposedly the old ways of Marion Barry— this gives members
25. high LTH They're still feeling their way, Alec, and **new brooms** always want to sweep cleaner than the old. Incur their disple
26. ier Tomlinson explained, "is our **new broom**, "and Castle noticed the way
27. open check. You must forgive the **new broom**. I have to learn the ropes, "
28. see how far blow the table the **new broom** was liable to sweep. He said "
29. ly today. I said to myself—the **new broom** may still be sweeping around.
30. r the very devil to work with. A **new broom** sweeps clean and all that. I h
31. I think we might be in for the **new broom** business here. You see what I m
32. eart of what?" "Europea. Oh some **new broom**'s bright idea for a new versio

Table 3. COBUILD concordance for "new broom"

As shown in the concordance, the fifty-million-word COBUILD corpus contained only two occurrences of the foregoing expression as a whole (lines 10 and 25 of the concordance). However, further searches focusing on the individual constituents of the expression at issue unveiled the stability of "new broom" as a fairly invariable chunk within the broader linguistic item. Altogether, more than thirty occurrences of the metaphor "new broom" were spotted in the COBUILD corpus. In all these cases, "new broom" referred either to a person or to a specific policy; additionally, it became evident that "new broom" is occasionally used as a premodifier when referring to people (as in concordance line 3). Despite other isolated instances in which "new broom" itself is premodified by other items, this invariable sequence

usually collocates (on the left side) with the definite article “the” and, not so often, with the indefinite article “a.” As far as the right side of the syntagmatic dimension is concerned, “new broom” collocates mostly with an adjunct or the verb “sweep” followed by an adjunct. Although the right-hand side complementation is substantially more variable than its left-hand side counterpart, it is worthwhile mentioning that the adjunct refers, in both cases, to institutions, organisations or associations in which the new broom is the incumbent of an executive position.

Before proceeding with the report of our findings, we should perhaps draw the reader’s attention to the fact that the relational patterns emerging from this concordance match the results of previous case-analyses. Take, for instance, Sinclair’s (1995) corpus-based analysis of those multi-word linguistic units allowing for a relative and idiosyncratic degree of formal variability. Drawing upon his detailed scrutiny of corpus concordances, Sinclair claims that such semi-idiomatic multi-word units or patterns revolve around a “fixed core structure” which prescribes its more or less variable co-text both to the left and right of the invariable item. Taking as his point of departure the exclusive collocational relationships that the constitutive forms of the lemma “eye” (see example 2 above) engage in, Sinclair notes, for instance, that the singular often collocates with terms such as “blind,” “caught,” “glint” and “naked.” Should we opt to focus on the lexico-grammatical association between “eye” and “naked,” the corpus reveals that the group “the naked eye” constitutes a solid and fixed core unit of meaning resulting from a process of collocation and that the core collocation is able to predict the surrounding verbal environment as we move away from it along the syntagmatic axis of chain, either to the left or the right.

So much for the relevance of the “collocation” phenomena to the formalisation of a multi-word lexico-grammatical choice. It is then turn to move on to the implications of “colligation,” a phenomenon which allows the researcher to ascertain which (variable) elements from the co-text tend to co-occur with the (invariable) core unit on a regular basis. Despite their lexical variability, those items engaged in colligational relations with “the naked eye” belong to a restricted set of grammatical categories. In other words, the invariable core unit grounded in collocational patterns colligate with a range of grammatically equivalent elements that contribute to define the boundaries of the ultimate lexico-grammatical choice.

The joint action of collocation and colligation helps sketch the co-selectional profile of a multi-word unit. Taking up Sinclair’s example, the collocational structure “the naked eye” does not only co-select the prepositions “with” and “to”; as we move away from the former, it also attracts verbal elements or adjectives revolving around the semantic concept of “visibility.” Our discussion is summarised in Figure 1 below.

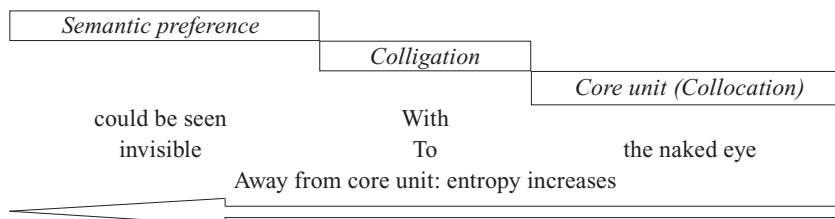


Figure 1. The lexico-grammatical co-selection profile of “the naked eye”

In Sinclair's view, it is the grammatical and semantic consistencies that arise from such co-selectional phenomena that warrant the analyst's approach to the overall linguistic structure as a single unit of meaning resulting from a single lexico-grammatical choice despite their formal variability. The linguist's failure to detect any more co-selectional consistency or regularity whatsoever at either side of the so-far delimited structure would then call for the treatment of the latter as a linguistic item on its own, or —putting it in SFL's terms— as the most delicate realisation in a lexical system network. As Sinclair himself explains:

Information theory can be used on consolidated corpus data to calculate the entropy (chaos) at each word space, moving away from the core word. When it reaches a high value there must be an item boundary, because the co-text is not longer predicted by the core word (Sinclair 1995).

These emerging units convey a distinctive “semantic prosody” (Louw 1993, Sinclair 1994, Stubbs 1995), a term which acknowledges the fact that the habitual collocates of the core “naked eye” —more specifically, their semantic constituency— are capable of colouring it. In the case of Sinclair's example, the co-selection of the unit with verbs/adjectives related to the notion of “visibility” activates a semantic prosody which suggests “difficulty.”

As far as our own example is concerned, the internal structure of “new broom” would seem to revolve around a system of choices which opens immediately after the word space that follows this invariable core unit. It is at this point that the need to choose between the adjunct and “sweep + adjunct” arises. Our scrutiny of the concordance could detect no further regularity beyond this point.

Data suggest that the semantic prosody emerging from these co-selectional pattern is one of profound changes taking place within a particular professional domain as a result of a new appointment. This claim is corroborated by those concordance lines where “new broom” engages in atypical patterns of co-selection, for instance, as a premodifier with reference to things or as part of an adjunct (lines 4 and 23, respectively). In other words, atypical patterns of co-selection remove the distinctive semantic prosody which characterises the multi-word unit under scrutiny and reinstates the relevance of the meaning of each individual constituent. Figure 2 below represents in graphic form the lexico-grammatical co-selection profile corresponding to the core unit “new broom.”

For want of more exhaustive work along this pathbreaking line of research led by corpus linguists, we may put forward the interim contention that computer-aided insights are particularly relevant to the description of multi-word units with shades of idiomatic meaning. In doing so, they assist researchers with their attempt to formalise the most delicate lexico-grammatical realisations of language and, thereby, speed up the journey towards the design of the system networks that range the former. The phenomenon of co-selection arrives at a clearer understanding of certain co-occurrence patterns that traditional intuition-based dictionaries and grammars do not always encompass. At any rate, such multi-word units consist of a core item responding to an invariable realisation throughout a concordance. As we move away from the core, the range of variation among the collocates increases until the predictability

falls fairly dramatically. It is precisely at this point that a boundary between multi-word units of meaning responding to a single lexico-grammatical choice falls.

	<i>Collocation</i>	<i>Colligation</i>	<i>Semantic preference</i>
<i>Core</i>			
the new broom		at	institutions companies political party
the new broom sweep around/through			institutions companies' boards political party sport associations

Semantic Preference:Incumbent of executive position

Semantic Prosody:Make important changes

Figure 2. The lexico-grammatical co-selection profile of “new broom”

5. CO-SELECTION: IMPLICATIONS FOR INTERLINGUAL AND INTERCULTURAL COMMUNICATION

The fact that co-occurrence patterns allow individuals to predict what lexical items and/or grammatical classes are to be found in the vicinity of the current one has led a number of researchers from different backgrounds to lay particular emphasis on the hypothesis that human beings understand/produce language “in chunks.” This being so, it is contended that language is stored redundantly, i.e. not always as individual items but rather as complex prefabricated sequences of speech that are often retrieved from memory as such on the grounds that they would seem to be activated by a single lexico-grammatical choice. The chunk-oriented approach to the analysis of co-occurrence patterns has concerned itself with the pedagogical implications of this linguistic phenomenon, mainly with the ways in which co-occurrence can be best exploited for language learning purposes.

Nattinger (1988), for instance, regards language use by native speakers of a language as a “compositional process” whereby preassembled phrases—that is, multi-word units responding to a single lexico-grammatical choice on the system network of choices—are stitched together. Accordingly, he proposes a number of reasons for deliberately drawing the learner’s attention to the phenomenon of co-occurrence: (i) they provide raw material for later analysis and segmentation; (ii) they allow the learner to avoid incongruities of register; (iii) finally, they lead to fluency, as learners are not required to concentrate on each individual words. In postulating these premises, Nattinger (1988) gives prominence to the syntagmatic dimension of lexis vis-à-vis its paradigmatic counterpart, thus implying that the regularity of certain paradigmatic choices on the part of the speaker result in the stability of the ensuing syntagmatic

patterns, thus warranting the relevance of co-occurrence phenomena to the process of second language acquisition.

It is precisely this field of scholarly inquiry that would seem to have most recurrently capitalised on the increasing attention devoted to the pre-assembling of language. Cowie's (1988) notion of "pre-fabricated routines," for instance, refers to her claim that children first extract multi-word units from the flow of speech to which they are exposed and store them as such. Even though they will later be able to segment them, children retain these multi-word chunks both as a whole and in term of its constitutive elements. In Cowie's opinion, such units can be divided into two major groups. On the one hand, there is the idiomatic or "semantically specialized" multi-word expressions which are subject to different degrees of internal variation; broadly speaking, they correlate with the first three types of collocational units proposed by Nattinger (see above). On the other hand, there is the "pragmatically specialized" multi-word expressions, which span a great diversity of grammatical patterns and correlate with the four last classes put forward by Nattinger. At any rate, Cowie concludes that it is reasonable to create a methodology for vocabulary teaching which acknowledges the fact that both native speakers and learners of a language are predisposed to store and reuse units as much as, if not more than, to generate them from scratch.

The importance of the realisation of certain communicative choices or units of meaning through complex lexico-grammatical patterns has also been attested to by different studies on interlingual translation. Translation studies are particularly interested in analysing how such multi-word units are handled when rendered into a different language and, more specifically, in discussing whether they can be regarded as the ultimate response to the search of translation scholars for the basic "unit of translation."

Shuttleworth & Cowie (1997) report on the development of this search. It would thus seem that it is a commonly held view that "unit of translation" refers to "the linguistic level at which the Source Language is recodified in the Target Language" (1997: 192). The focus of the debate lies in whether units of translation should be elements of linguistic form or content. Early views on this matter suggest that units of thought, units of meaning and units of translation should be synonymous, although more recent approaches to this issue define the "unit of translation" as "the smallest unit of Source language which has an equivalent in the Target Language." In many cases, translators are recommended to deal with certain multi-word structures as units of translation as long as its parts taken individually are untranslatable, i.e. no equivalents can be found for them in the target language. Accordingly, "not only the word *generally*, but also the expression *by and large*, although it is made up of three words, would be treated as a single unit" (1997: 192).

From a strictly systemic approach to the analysis of language, Bell (1981) argues that the process whereby a translator renders a Source Language text into its Target counterpart requires the linguistic input to pass through a "frequent lexis store" and a "frequent structure store." The first store is the equivalent of a physical terminology database and constitutes "an instant "look-up" facility for lexical items, both words and idioms" (1981: 47). For its part, the second store contains frequently occurring structures in their entirety, with "direct access to phrases and sentences nearly as rapid as it is for individual words" (1981: 48). Again, recurring multi-word co-selectional units are tackled as units of meaning which by-pass any lexical or syntac-

tic parsing mechanism of linguistic processing. In other words, they are dealt with in the same way as single-word units.

Bringing the relevance of and co-occurrence phenomena to interlingual translation into a sharper focus, Baker (1992) and Hervey *et al.* (1995) distinguish between “collocations,” on the one hand, and “idioms” and “fixed expressions,” on the other. As far as the first member of the dichotomy is concerned, it is particularly important to be aware of the contrastive co-occurrence patterns in the Source and Target languages. Accordingly, the translator’s failure to realise the implications of the latter often results in important translation pitfalls, such as the misinterpretation of the meaning of a Source-language collocation, particularly in those cases when the collocation pattern is a culture-specific one. Putting this issue in Baker’s words,

When the translation of a word or a stretch of language is criticized as being inaccurate or inappropriate in a given context, the criticism may refer to the translator’s inability to recognize a collocational pattern with a unique meaning different from the sum of the meaning of its individual elements. (Baker, 1992: 53).

As far as the translation of idioms is concerned, one of the main difficulties for translators seems to be the identification of idioms as such, especially in those cases where there is no equivalent idiomatic construction in the target language. It is often the case that a Source language unit is translated into a Target language unit located at a different rank of the linguistic system —e.g. when a word is translated by a phrase; as a matter of fact, it is widely known that trainee translators find it easier to translate by shifting upwards —i.e. from a ST group to a TT clause (Taylor 1993). But regardless of the equivalence of the lack of the latter across languages, the spotting of the multi-word structure as a single unit of meaning is crucial to the accomplishment of a good translation.

6. CONCLUSION: IMPLICATIONS FOR SFL

Corpus-based insights into the actual linguistic behaviour of individuals have proved to be enormously useful to challenge the speakers’ and researchers’ intuitions on language organisation as well as to provide them with precise information regarding the relative occurrence rate of certain items or patterns. More specifically, they allow us to identify multi-word units of meaning consisting of a core constitutive item surrounded by a number of stable co-selections, thus attaching a specific “semantic prosody” (Stubbs 1995) to such linguistic patterns. In so doing, they lend support to the semantic-oriented slant of SFL and its search for a number of rigorously formalised system networks of lexico-grammatical choices, these being the right-most delicate realisations of the latter —i.e. the ultimate units that convey meaning.

On the one hand, the fact that each linguistic item tends to occur within its idiosyncratic co-text as well as to be realised in accordance with its particular grammatical profile may help researchers to gain a deeper insight into the mechanisms whereby speakers produce cohesive texts. Once the spontaneous principles of cohesion have

been thus apprehended, the linguist/teacher may (i) arouse the language learners' awareness of and compliance with the required cohesion standards for a given language and register; (ii) pin down certain deviancies from such conventional trends whereby the speaker/writer might be attempting to accomplish certain communicative purposes, and (iii) distinguish between the cohesion patterns which characterise each text-type and, thereby, relate the latter to the relevant set of contextual/co-textual constraints. On the other hand, the fact that certain co-selection patterns are necessary to achieve a specific semantic prosody may help the linguist to argue that the ironic, puzzling or challenging connotations of the co-selection at issue result from the speaker's opting for a marked or incongruent choice instead of the semantic preferences favoured by the core item of the multi-word unit (Louw 1993).

It is our belief that SFL can make interesting and crucial contributions to the description of multi-word units of meaning arising from a single lexico-grammatical choice. In that corpus-based insights often challenge the speakers' intuitions, there is now plenty of evidence available attesting to the fact that usage is not always mirrored in the speaker's consciousness of language. In other words, the speaker's consciousness of language cannot be derived solely from exposure to actual usage. SFL aims to model the choices available for speakers, such that the currently activated options are studied not only in terms of their own implications, but also set against the remaining constituents of the paradigm —i.e. the range of choices that could have also been selected at that particular point in the unfolding of the communicative event. Consequently, it would be highly interesting to look at the structure of multi-word units of meaning from a systemic functional standpoint and see whether any alternative choice might have resulted in a different semantic prosody. Furthermore, it would be extremely useful to explore the ways in which the modelling of the system networks of choices could be exploited for practical purposes, thus giving more prominence to the rôle of language in selected fields of human communication such as literature, advertising as well as psychiatric or forensic practice.

Works Cited

- Baker, M. *In Other Words*. London and New York: Routledge, 1992.
- Banks, P. "Contextual Dependency and Lexical Sets." *International Journal of Corpus Linguistics* 1.2 (1996): 75-98.
- Bell, R. T. *Translation and Translating: Theory and Practice*. London and New York: Longman, 1981.
- Carter, R. and M. McCarthy. *Vocabulary and Language Teaching*. London and New York: Longman, 1988.
- Cowie, A.P. "Stable and Creative Aspects of Vocabulary Use." *Vocabulary and Language Teaching*. Ed. R. Carter and M. McCarthy. London and New York: Longman, 1988. 126-139.
- Firth, J.R. See Palmer.
- Francis, G. "A Corpus-driven Approach to Grammar: Principles, Methods and Examples." *Text and Technology*. Ed. M. Baker, G. Francis and E. Tognini-Bonelli. Amsterdam and Philadelphia: John Benjamins, 1993. 137-156.

- Francis, G. "Grammar and Lexical Patterning." Paper delivered at the 11th World Congress of Applied Linguistics AILA'96, Jyväskylä (Finland), August, 1996.
- Gregory, M. J. "Aspects of Varieties Differentiation." *Journal of Linguistics* 3 (1967): 177-198.
- Halliday, M.A.K. "Categories of the Theory of Grammar." *Word* 17.3 (1961): 241-292.
- "Lexis a Linguistic Level." *In Memory of J. R. Firth*. Ed. Bazell, Catford, Halliday & Robins. London and New York: Longman, 1966. 148-162.
- *Introduction to Functional Grammar*. London: Edward Arnold, 1985.
- & R. Hasan. *Cohesion in English*. London and New York, 1976.
- Hervey, S., I. Higgins and L. Haywood. *Thinking Spanish Translation*. London and New York: Routledge, 1995.
- Leitner, G. "Where To Begin or Start?: Aspectual Verbs in Dictionaries." *Data, Description, Discourse*. Ed. M. Hoey. London: HarperCollins, 1993. 50-63.
- Louw, S. "Irony in the Text or Insincerity in the Writer?: The Diagnostic Potential of Semantic Prosodies." *Text and Technology*. Ed. M. Baker, G. Francis and E. Tognini-Bonelli. Amsterdam and Philadelphia: John Benjamins, 1993. 157-176.
- Nattinger, J. "Some Current Trends in Vocabulary Teaching." *Vocabulary and Language Teaching*. Ed. R. Carter and M. McCarthy. London and New York: Longman, 1988. 62-82.
- Noël, D., B. Defranq and F. Devos. "How Much Grammar Is There/Can There Be/Should There Be in Bilingual Dictionaries?" Paper delivered at the "International Translation Studies Conference. Unity in Diversity?" Dublin, May 1996.
- O'Donnell, M. "The Syntagmatic Nature of Context." Paper delivered at the "VIII Euro-International Systemic Functional Workshop," Nottingham, July 1996.
- Palmer, F. R., ed. *Selected Papers of J. R. Firth 1952-1959*. London and New York: Longman, 1966.
- Pérez-González, L. and A. Sánchez-Macarro. "From Context to Co-text: A Computer-aided Reappraisal of the Linguistic Units for Encoding Meaning." Paper delivered at the "VIII Euro-International Systemic Functional Workshop," Nottingham, July 1996.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.
- Shuttleworth, M. and M. Cowie. *Dictionary of Translation Studies*. Manchester: St. Jerome, 1997.
- Sinclair, J. McH. "Beginning the Study of Texts." *In Memory of J. R. Firth*. Ed. Bazell, Catford, Halliday and Robins. London and New York: Longman, 1966. 410-430.
- *Corpus, Concordance, Collocation*. Oxford: Oxford UP, 1991.
- "Trust the Text." *Advances in Written Text Analysis*. Ed. M. Coulthard. London and New York: Routledge, 1994. 12-25.
- "The Search for a Linguistic Item." Paper delivered at ILASH. Sheffield, May 1995.
- Stubbs, M. "Corpus Evidence for Norms of Lexical Collocation." *Principle and Practice in Applied Linguistics*. Ed. G. Cook and B. Seidlhofer. Oxford: Oxford UP, 1995. 245-256.
- *Text and Corpus Analysis*. Oxford: Blackwell, 1996.

- Taylor, C. "Systemic Linguistics and Translation." *Occasional Papers in Systemic Linguistics* 7 (1993): 87-103.
- Tucker, G. "Cultural Classification and System Networks: A Systemic Functional Approach to Lexical Semantics." *Meaning and Form: Systemic Functional Interpretations. Meaning and Choice in Language: Studies for Michael Halliday*. Ed. M. Berry, C. Butler, R. Fawcett and G. Huang. Norwood, New Jersey: Ablex, 1996. 533-566.
- Winter, E. O. (1979) *Towards a Contextual Grammar of English*. London: George Allen & Unwin.