



Words as “Lexical Units” in Learning/Teaching Vocabulary

MOISÉS ALMELA
AQUILINO SÁNCHEZ*
Universidad de Murcia

ABSTRACT

One of the genuine contributions of theoretical linguistics to the interdisciplinary field of applied linguistics is to elucidate the nature of *what should be taught* and *how it should be taught*. Traditionally, the input supplied in vocabulary teaching has consisted either of word lists (most often) or of words-in-context (more recently). In the first case, words are treated as self-contained receptacles of meaning, and in the second case, they are considered as nodes of semantic relationships. However, recent directions in corpus-driven lexicology are exploring the gulf between the concept of a “word” and that of a “semantic unit”. The main purpose of this paper is to update some implications of this discussion for one of the applied disciplines, namely FL/L2 vocabulary teaching and learning.

KEYWORDS: second language teaching/learning, lexical unit, word, corpus linguistics, collocation, lexicology.

* *Address for correspondence:* Moisés Almela, Departamento de Filología Inglesa, Campus de La Merced 30071, Murcia, Spain. Tel: 00 34 968363212; e-mail: moisesal@um.es. Aquilino Sánchez, Departamento de Filología Inglesa, Campus de La Merced 30071, Murcia, Spain. Tel: 00 34 968363175; Fax: 00 34 968363185; e-mail: asanchez@um.es.

I. INTRODUCTORY NOTES

“Lexis is the core or heart of language but in language teaching has always been de Cinderella”, states M. Lewis (Lewis, 1993: 89). This situation of “neglect” is changing lately: vocabulary attracts more and more the attention of scholars and is the subject of numerous research projects (Laufer & Hulstijn 2001; Nagy & Scott, 2000; Nation, 2001; Read, 2000, etc.), especially in the field of vocabulary acquisition and assessment.

Paradoxically, the current “lexicalist turn” in linguistics –both theoretical and applied– has coincided with a questioning of the very foundations of lexicology. The increasing interest in vocabulary has given rise to a lively debate about the nature and structure of a semantic unit, and some scholars have challenged the assumption that words qualify as such kind of units (Teubert, 2004, 2005).

The above paradox can be resolved insofar as we accept that there is a clear-cut borderline between the “lexicalist” and the “phraseologist” tenets. In the last resort, there is no contradiction between the idea of lexicalism and the traditional modular approaches to language structure. Strictly speaking, the notion of lexicalism does not exclude words from functioning as self-contained lexical units, and this notion in turn favours a naïf concept of language as a construct made out of elements functioning as “building blocks” of the linguistic system, a system made out of individual elements which combine with each other to build sentences and ultimately to generate discourse. Within this perspective the word gains a high degree of autonomy and tends to be considered a unit which can be manipulated with ease and can easily become an objective element for study and analysis, or, more recently, for processing by computers. In short: it is not the lexical but, more specifically, the phraseological bias that is diametrically opposed to the grammatical one and is able to complement it —in syntax, this has already become evident in the conflict between projectionist and constructionist approaches.

The proponents of a phraseological/idiomatic approach to language argue that the conception of a word as a lexical unit is grounded in spelling rather than in semantics. The fact that words are presented as separate units in the written language has consolidated the idea that they function as independent units in discourse. However, orthography has a tricky relationship with language structure. We know that a space between letters is not necessarily a delimitation of a semantic unit. The orthographic definition of a lexical unit is not free from difficulties when attempting to explain compound words, formally presented as various - hyphenated or not- units (*the White House; lower-case letter*).

The equation word-lexical unit comes more clearly into crisis when confronted with grammar. The composition of units of (lexical) meaning is heteromorphous not only with spelling but also with morphology. Morphological units are not always simple constructs (*shopkeeper*) and quite often a bunch of morphological units, e.g. a phrase, is used to designate a single concept (*as a matter of fact*). Nor are the limits of phraseological patterns necessarily coincident with those of syntactical units. According to Biber et al. (2004), "lexical bundles" may overlap clauses or phrases (e.g. *If you look at...*), without necessarily forming grammatically (syntactically) complete units.

Besides, the independent use of words in communicative events is more than questionable, since their full power for meaning is only displayed in discourse, that is, in the company of other words. For instance, from the mere selection of the single word *strong* we cannot predict whether it describes a physical or a psychological quality (compare *strong coffee* with *strong personality*).

On the other side, however, there are also strong arguments in the literature for an underlying structure of meaning inside the word (see section III). The argumentation is multiple and comes from both minimalist and maximalist approaches to lexical semantics. The relationships among the diverse senses of a word often show sufficient analogies to be traced back to a single representation. Besides, the predictability and regularity shown in the devices of sense extension underpin the treatment of polysemy as part of the language system. Thus, the fact that the actual word senses are variable is not sufficient to discard the unit-status of the word insofar as such variability is restricted and structured.

All in all, it is evident that any statement of the word as a unit of meaning requires renewed and sophisticated argumentation. Rather than taking the traditional assumptions for granted, these should be updated and subjected to revision in the light of new evidence.

The outcomes of such revision should have implications for applied linguistics, where the debate about the nature of a semantic unit has still not been consistently incorporated. Although the question about what constitutes a semantic unit is far from being resolved, teachers and learners of language have most often a simplified view of the issue. Most often words are taken as lexical units 'as a matter of fact'. Such an assumption is based on a long tradition of linguistic beliefs, well rooted in the mind of the speakers and clearly favoured by some methods of teaching/learning languages or practices habitual in the classroom.

The history of vocabulary teaching has been centred on the teaching of words as isolated or de-contextualized items. It has been clearly so in the traditional *Grammar Translation Method*, but also in other methods not so strongly based on grammar, as might be

the case of the *Direct Method* and some approaches heavily based on teaching through reading and memorization of dialogues, closer to real linguistic usage. The teaching and learning of vocabulary lists has been one of the pillars in the classroom for centuries. The explanation of grammatical rules by the teacher was followed by classroom practices in which the words learned were combined in order to build the kind of sentences required by the rules. There is not explicit information on the nature of the words being learned or taught, but the way teachers and students present them helps in consolidating the perception that they are fully autonomous.

In this paper, our main purpose is to bring EFL research in line with current issues in lexical semantics. More precisely, we shall discuss some of the implications which collocational research has for the understanding of vocabulary learning processes and the design of teaching methods.

II. COLLOCATION AND VOCABULARY

The relationship between collocational input and lexical knowledge has been predominantly approached from a word-centred perspective. By and large, the word keeps being widely accepted as the main unit of lexico-semantic analysis in linguistics, and consequently, it is also presumed the default unit of vocabulary teaching-learning. Normally, the role of collocational data is limited to facilitating the process of learning word meaning(s). For instance, part of the meaning of the node *mesa* (in its ‘furniture’ sense) can be inferred from collocates such as *silla*, *comer*, *sentar(se)*, etc. The underlying assumption is that vocabulary knowledge consists in knowing words, and that the knowledge of a word in turn can be improved by reference to contexts of use. Thus, the substantial difference between this strategy and the use of word lists does not reside in the shift of unit, but in the addition of a meaningful environment to the unit in question, i.e. the word.

This approach is not without its limitations. The potential of a collocate for giving information about the meaning of the node is overstated insofar as the ambiguity of the collocate itself is neglected. Where the node and the collocate are realizations of ambiguous words, the analysis of word meaning basing on collocation is at risk of ultimately causing a sort of vicious circle, whereby the construction of a meaning for a node *n* depends on the interpretation of a collocate *c* whose reading is in turn relative to the node. The informativeness of the collocate is often determined by its actual sense, but this sense in turn may depend on a syntagmatic reference to the node.

Let us consider the collocation *abnormal cell*, whose components are polysemous when analysed as single words. The fact that the noun *cell* carries here the feature 'living being' is not obvious from the adjectival collocate. The use of *abnormal* in the attributive position does not fully predict a noun with a 'living being' meaning —e.g. compare the foregoing collocation with *abnormal gain/loss*. The implications should be pondered over: if the feature 'living being' is predictable from the entire collocation but not from any of the collocates taken individually or separately (in divergent usages), it follows that, strictly speaking, *abnormal* is not a decisive clue to the meaning of *cell*. It is rather the case that the feature 'living being' is a semantic property of the multi-word pattern taken as a unit. Where the component words of a collocation are interdependent, it might be wise to promote the holistic learning of the usage pattern.

This begs the question of why collocations in language teaching should be ascribed the status of combinations if they are used as items in the discourse. Related to this is also the question of why words should be treated as the building blocks of vocabulary despite the fact that they cannot determine their own actual reading in language use, whereas there are other units whose actual senses are subject to minimal variation from text to text. In the last resort, the answer will depend on whether lexical competence is conceived of as primarily a communicative skill or not (see section IV). In this sense, the option for either words or collocations as the basis of vocabulary teaching will be informed by linguistic theory. On the assumption that lexical knowledge does not form an autonomous system but is determined by *functions* of language, the decision to teach words as the main vocabulary items is not fully adequate.

From this premise, it is no surprise that recent advances in corpus linguistics mark a departure from the word-centred approach. It is suggested that vocabulary teaching should be inspired by a revised notion of what constitutes a lexical unit (Teubert, 2004). Several corpus linguists have been preoccupied with distinguishing between words and units of meaning (Ooi, 1998; Sinclair, 1998; Stubbs, 2002; Teubert, 2005; Tognini-Bonelli, 2002). The concept of an **extended lexical item** (ELI) has implications both for the *structure of the lexicon* and for the *scope of the phrasicon*. Regarding the first point, the ELI implies that the paradigm of lexical unit consists of a network of interdependence links among co-occurring words. A further implication for lexicology is the thesis that any paradigmatic relation between two or more words is contingent on the syntagmatic framework determined by a higher-level unit of meaning (for instance, *day* means the opposite of *night* in the usage pattern *during the day* but not in *a gap of X days*).

As regards the extension of the phraseological realm, the postulate of an ELI involves the ultimate comparability of collocations and idioms. Inherent in this view is the claim that the realm of multi-word units has been understated in the mainstream models (Almela, 2006; Teubert, 2004; Tognini-Bonelli, 2001). The stock of idioms stored in the real vocabulary of a language largely outnumbers the stock of idioms described in standard linguistic research and in reference works, notably dictionaries.

Traditionally, the concept of a collocation has been neatly distinguished from that of an idiom (or a fixed expression) basing on the allegedly compositional structure of the former (Liang 1991). However, the definition of collocation as a standardized lexical combination has also drawn great criticism for lack of empirical adequacy. Penadés Martínez (2001) has remarked that the mainstream definitions of collocation are unable to yield a clear-cut category word co-occurrence, distinct from idioms, when applied to actual data. Almela (2006) has contended that the borderline between collocations and idioms does not reside as much in the different nature of these structures as in the different scale or proportion of their cohesive devices.

The new tendency of linguistic theory to widen the scope of idiomatic language has been synchronized with the increasing importance attached to formulaic language and chunking in the field of EFL. Many specialists recommend that teaching procedures be based solidly on “pre-fabricated” language, chunks, and routines (Granger, 1998; Lewis, 1993). Indeed, the new models of the lexical unit call for a new approach to the relationship between collocational input and vocabulary learning. Rather than focusing on the semantic information which the collocates give about the node word, the attention has been turned towards the choice of the collocation itself considered as a whole. That is to say, there has been a shift of the unit status from the word to the pattern. Accordingly, there are intrinsic rather than extrinsic motivations for promoting exposure to collocational data. Instead of conceiving collocations as useful for learning vocabulary, they are deemed to constitute themselves the vocabulary items. The idea is not that collocations should help the learner to acquire word meanings but that collocations should substitute for the words as the target lexical items in the foreign/second language.

In sum, the discussion on the concept of a lexical unit in lexicology manifests itself in the debate between word-centred and collocation-centred approaches to vocabulary teaching. In what follows, we shall comment on some arguments for and against each of these two approaches.

III. COLLOCATION AND WORD MEANING

Context-dependency of word senses is one of the main arguments for an idiom principle. The postulate of an ELI has been associated with a revision of mainstream lexicographic practice. The standard dictionary micro-structure evinces that the usages displayed in the first part of the entry, before the idioms, represent "free" senses or autonomous meanings. Contrary to the fixed expressions placed at the end of the entry, the allegedly "free" senses are not explicitly assigned any co-textual restraint in the form of lexical or lexicogrammatical structure. Accordingly, Sinclair (1991) has criticized lexicographic tradition for implicitly assuming that any occurrence of a word could signal any one of its meanings, which would make communication impossible. Indeed, the practice of relegating fixed expressions to the end of a lexical entry is insufficient for capturing the correlations between lexical meaning and word co-occurrences.

As regards language teaching, the context-dependency of word senses raises the question of how useful it is to learn a given word sense without its corresponding co-textual correlates. Without mastering the patterns of sense-context coordination, the chances of engaging successfully in communication are seriously hampered. It might be counter-argued that the selection of sense in a polysemous word follows naturally from the operation of common sense knowledge. However, this is not always the case. Many co-textual restrictions on sense activation are highly idiosyncratic and difficult to predict from either encyclopaedic knowledge or L1 competence. A case in point is the collocation *previous conviction*. In principle, two readings ('prejudice' and 'criminal record') can be assigned to this collocation, based on a modular knowledge of the lexicon and the syntactical rules. The selection of one or other reading depends on which of the two homonyms (*conviction* = 'firm belief' / 'guilty verdict') is deemed to underlie the form *conviction* in the aforementioned collocation. Is the speaker free to select the word sense in whatever way (s)he pleases? Not quite. Phraseology reduces the range of meaning to just the second reading. Note that the knowledge of lexis and syntax as separate modules does not suffice for the learner to predict the meaning of *previous conviction*. The 'guilty' sense of *conviction* in this collocation is not determined by either the word meaning or the grammatical rules; rather, it is determined by the idiosyncratic formation of an upper-level (multi-word) unit. In cases like this, the meaning of the collocation should be learned in toto; or put differently, the 'guilty' sense of *conviction* should be learned alongside its co-textual correlates. These correlations play an important role in precluding communication breakdowns. They prevent the message from being decoded in a different way than the one intended by the speaker.

Notwithstanding, there are arguments in the literature for the concept of “word meaning” and against the model of an extended item. Four of such arguments are commented and counter-argued below. Firstly, many authors have made the case for the existence of default (prior) word meanings, i.e. senses that tend to be activated in absence of a specific phraseological pattern or a lexicogrammatical unit. Telegraphic speech, where content words are accumulated without forming any upper-level structural (language) unit, provides beginners with an effective communicative strategy and constitutes one of the initial steps in the development of linguistic competence. After all, individual words are capable of fixing a referent in actual contexts. This is especially true for nouns, e.g. a sign saying *hospital* on the top of a building, or a sign with an arrow next to the word *exit* fixed on a door.

Secondly, there are semantic features in the word that are not affected by variations in the textual environment. For instance, the predicative function of ‘intensification’ remains invariable in the adjective *strong*; the aforementioned feature does not co-vary with the different co-texts in which *strong* is used, say, the nouns *personality*, *man*, *coffee*, *tea*, *argument*. Thus, ‘intensification’ seems to be part of the autonomous word meaning of this adjective. From minimalist approaches, it has been contended that sense variation is a matter of pragmatics, not of semantics (Ruhl, 1989). In structural lexicology, the same remark has motivated the distinction between the concept of “meaning”, on the one hand, and of “sense” or actual reading, on the other (Casas Gómez, 2002). Thus, the *meaning* of ‘intensification’ is realized in multiple *senses* of the adjective *strong*. The different interpretation of this word across collocations such as *strong argument*, *strong personality*, *strong coffee*, etc., can be explained as the actualization of a single feature in multifarious contexts. The variation would be located at the level of actual use, not of lexical structure.

Thirdly, there are analogies at the level of the actual senses themselves. For instance, the ‘computer device’ sense of *mouse* is said to origin from visual comparisons with the referents of the ‘animal’ sense of *mouse*. Research on polysemy has revealed the existence of regular –hence predictable– mechanisms of sense extension or conceptual shift. This has motivated the notion of systematic polysemy, which has been developed mainly in cognitive linguistics.

Fourthly, it has been shown that certain words are able to trigger more or less coherent and structured representations or categories. The usage of the word *bird* is subject to stark variation across individual speakers and situations. The denoted animal may or may not have feathers, and it may or may not fly (e.g. the denotata of *penguin*). However, in spite of this variability, there is very little doubt that virtually the whole language community will agree

on deciding that *sparrow* is an exemplar of the category defined by the noun *bird*. Thus the semantic potential of a word keeps a relative stability. The same applies to the specialized use of certain terms. Recently, expert astronomers decided in a meeting that Pluto should stop being listed as a *planet*. Nevertheless, the other eight planets of the solar system have maintained their status in the astronomy jargon. This indicates that the category meant by *planet* has a relative stability: a part of its definition and its membership can be subject to discussion, while the rest is taken for granted. Only a part of the extensional range (one from nine) was subtracted.

However, the above four arguments for word meaning can be counteracted by the following remarks. Firstly, prior word meanings have a very limited validity. Some words do not lend themselves readily to a hierarchy of sense activation, because none of their senses is either much more frequent or conceptually salient than the others. For instance, there are no objective features to establish which sense of *basin* is the primary one. Without recourse to etymology, it is difficult to determine which senses of this word derive conceptually from other senses. Moreover, in the case of words for which a prior meaning can be identified, it should be noted that the default meaning is operative only in specific situations and can be suspended at any time by a usage pattern which activates a non-prior sense. Thus, the 'door' sense of *exit* can be regarded as a default meaning, in that it does not require any lexicogrammatical environment for being activated. Yet, this independence of *exit* (= 'door') from any syntagmatic context is balanced by a strong dependence on the extra-linguistic context or situation. The 'door' sense of *exit* does not require any specific collocational environment for its activation, but to balance, it requires a specific extra-textual scene.

Secondly, the existence of constant semantic properties in the word (i.e. lexemic features) is recognized by the model of an ELI. The distinction between a *meaning* or *sense*, on the one hand, and a *meaning-component* or *seme*, on the other, is crucial for an adequate description of lexical meaning. The feature 'intensification' is virtually invariable in the adjective *strong*, but the meaning communicated always involves more layers of meaning than the purely lexemic. Thus, the collocational pattern *NUMERAL-strong crowd/mob* expresses an estimation of the number of people in a group. The actual sense of a word in a particular textual environment often includes semantic features that are contributed not by the word but by the usage pattern.

To explain this, Almela (2006) devised a tripartite classification of semantic features according to their distribution. *Lexemic* features are inherent in the word and remain invariable across usage variation; *specialized* features are carried by the word but activated by

the collocations, hence they are variable; finally, the *prosodic* layer consists of features that are carried by the collocations, not by the words. Thus, only one of the three layers of semantic features that can be identified in a stretch of text constitute a genuinely autonomous contribution from the individual words.

This taxonomy can be illustrated with the collocation of *strong/strength/strengthen* with *case*. Here, the meaning-component ‘intensification’ is a lexemic feature of *strong*; ‘facts and arguments’ is a specialized feature of *case*; and ‘opinion’ is a prosodic feature of the entire collocation. These conclusions have been reached after comparing the respective meanings activated by *case* and *strong/strength/strengthen* both individually and in conjunction with one another. Thus, the intensifying function is invariably attached to the selection of *strong/strength/strengthen* across variegated lexical environments such as *coffee*, *argument*, or *case*; the content ‘facts and arguments’ is attributed to *case* contingently on some distributions (e.g. *a good case for X*) but not on others (e.g. *in the case of X*); finally, the semantic domain ‘opinion’ forms part of the prosodic layer, in that it is predictable from collocations such as *strengthen your case* or *strong case*, but it is not necessarily expressed by separate usages of these collocates. Thus, the semantic domain ‘opinion’ is a function of the multi-word pattern considered as a whole.

This multiplicity of semantic layers detracts from the importance of the autonomous semantic features in the word. Such features exist, but the role they play in communication is limited. The reason why the senses of *strong* are not fully context-independent is not that the word lacks context-independent semes but that every actual sense incorporates some or other kind of context-dependent feature. The actual reading of the adjective *strong* is made up of more semantic traits than just the lexemic feature ‘intensification’.

The main implication for language teaching/learning is that the knowledge of autonomous word meaning (the lexemic features) has little impact on the development of communicative competence. For the learner to use the word appropriately in communication or to assign it the correct sense, (s)he needs to express/decode semantic features that reside in the collocation and not in the word. The combination of word meanings forms only a subset of the lexical meaning communicated by means of word usage. Words keep some semantic features of their own, but their actual senses are rarely independent from distribution patterns.

Thirdly, it must be conceded that there are conceptual analogies among the various senses of a word, but the importance of such analogies is relative to the language function under consideration. Admittedly, each extension of word meaning is historically and so to say “phylogenetically” dependent on a primary sense, but the role of such relations in achieving

effective communication is questionable. Arguably, the knowledge of conceptual analogies among senses of a word is more important for imitating native-like encyclopaedic knowledge than for using the target language effectively. This point will be explained in some more detail in section IV.

Fourthly, it is true that many words are able to convey more or less stable semantic categories, but it is no less true that the actual sense of a collocation is subject to less variation than the readings of each of the collocates, i.e. the component words. For example, the occurrence of the word *difference* by itself does not indicate us whether its own denotatum is a qualitative or quantitative variation between two things, or a conflict between two people or institutions, etc. However, if we encounter the collocational pattern *resolve/settle their differences*, we know that the meaning expressed is almost invariably ‘strife’; and if we encounter the expression *split the difference*, we know that the meaning conveyed is a ‘(quantitative) variation between two prices or monetary amounts’. In short, the actual interpretation of a collocation from one text to another is susceptible to considerably less variation than the actual senses of each of their component words taken separately.

Of course, it could be counter-argued that collocations are cohesive simply because co-occurring words tend to reinforce each others’ senses. However, the specificity of text meaning and discourse structure is not sufficient to explain the semantic stability of collocations. The various senses of two or more collocates can often be combined in multiple ways, giving rise to several readings. If the meaning of the whole lexical combination was the result of ad hoc selections of senses with the only restriction of text coherence, collocations would be almost as ambiguous as words, but that is not the case. For example, basing on separate usages of *take* and *picture*, their combination could produce various meanings such as ‘make a photograph’, or ‘grab a photograph’, or ‘grab a painting’, or ‘obtain a picture’. Of all these interpretations, the collocation *take a picture* selects just one. This semantic stability cannot be explained on the grounds of text-semantic factors alone, because the senses of *take* and *picture* could be selected and combined in many other ways so as to make sense and achieve coherence. Thus, the monosemy of a collocation such as *take a picture* is at least in part a function of processes that are typical of the formation of lexical units, namely co-selection, repetition/recurrence in discourse, and intertextual bonding.

Besides, the correlation between “monosemous” words and limited collocability casts a shadow of doubt over the monosemy of the words in question. Apparently, the noun *incidence* has a meaning of its own. It systematically refers to the ‘frequency with which something occurs, typically something bad (a disease, problem, etc.)’. However, a closer look

at corpus data will prompt the question of whether *incidence* is monosemous only to the extent that its lexicogrammatical environments are predictable. Among the collocates of *incidence*, two semantic classes or sets abound, namely (*i*) those which denote ‘disease’ or other kinds of catastrophe, and (*ii*) those which denote changes as captured in statistical data. To the first group of collocates belong nouns such as *cancer*, *disease*, *leukaemia*, *violence*, *crime*, *diabetes*, *poverty*, *abuse*, *infection*, *rape*, etc. In the second group, we find adjectival, verbal, and nominal collocates such as *great*, *grow*, *reduce*, *rise*, *low*, *high*, *increase*, or *decrease*, among others (data and calculations from the Bank of English).

A third group of collocates consists of words indirectly attracted to the node *incidence* via the collocates from group (*i*). Such is the case of *coronary*, *skin*, *lung*, or *childhood*, which are not directly related to the meaning of *incidence* but co-occur with it as a result of association with its collocates, especially those which denote ‘disease’ or other ‘harmful phenomena’. Thus, the collocation of *coronary* with *incidence* is a function of two collocational patterns: first, the phrase pattern *coronary heart/artery disease*, and second, the lexical collocation of *disease* with *incidence*. Likewise, the collocational pattern *skin/lung cancer* underlies the indirect attraction between *incidence* and *skin/lung*; and the attraction between *childhood* and *incidence* is not immediately motivated by their respective meanings but is underlain by the phrase *childhood abuse*. Thus, the heterogeneity of this third group of collocates cannot be attributed to unpredictability but to indirect attraction. That is to say, this group of collocates is a by-product of the first group.

All in all, the above data indicates an isomorphism between the collocational profile or distributional behaviour, on the one hand, and the componential analysis of the meaning of *incidence*, on the other. The two main groups of (direct) collocates represent respectively one of the semantic features ‘frequency/amount’ and ‘harmful phenomenon’. Precisely, these features play an essential role in defining the meaning of *incidence* and distinguishing it from lexically related nodes. Hence, there are strong arguments to conclude that knowing the collocations of *incidence* involves knowing the meaning of this word. In this case, the lexical competence not only overlaps with collocational knowledge but seems to be almost coincident with it.

IV. LEARNING TO COMMUNICATE VS. LEARNING TO SYMBOLIZE

For obvious reasons, the question of what is the “language”, and what is linguistic competence, has a direct impact on the question of what is taught in the foreign/second

language course. Before deciding what should be the operative unit of vocabulary teaching, it is essential to be clear about what it actually means to be “lexically competent” in a language different than the L1. The answer may depend on whether we conceive of the lexicon as basically a set of cognitive resources or as a component of communicative skills. If the aim of FL/L2 vocabulary teaching is to promote the learner’s construction of mental representations which match those of the native speakers, the word –even if it is polysemous– can be consolidated as a suitable unit, for the reasons we shall explain below. Nonetheless, if the chief goal is to assist the student in engaging successfully in communicative events using the L2, then the ELI emerges as a more appropriate unit than any polysemous word.

It could be counter-argued that both goals are complementary: if you learn the representations of the world that are constructed in the L2 lexical system, you will learn the meanings that are communicated in that language. However, it should be pointed out that the cognitive and the communicative subsystems of linguistic semantics are not functionally coupled with one another, as was argued by Feilke (1996) and Almela (2006). An example of this is the use of the words eng. *cab* and sp. *taxis* and *cabina*. The noun *cab* in English does not trigger the same conceptual representations as *taxis* and *cabina* in Spanish: *cab* has undergone a metonymic shift by which it refers both (*i*) to the ‘front part of a vehicle, where the driver sits’, and (*ii*) to a specific type of vehicle, namely a ‘car used for public transportation in return of money’; in Spanish, the noun *taxis* has only sense (*ii*), whereas *cabina* can be assigned sense (*i*) but not (*ii*). Hence, the conceptual content of *cab* in English must be different than that of *taxis* and *cabina* in Spanish. At the cognitive level, there is no possible equivalence between the semantic representation attached to eng. *cab* and that of sp. *taxis*, because the two senses of *cab* derive from a common conceptual entity that differs from the meaning of *taxis*. However, at the communicative level, there is virtually a one-to-one correspondence between eng. *take/call a cab* and sp. *coger/llamar (a) un taxi*.

A word of caution is in place here: the distinction between the cognitive and the communicative dimension in linguistic semantics should not be mistaken for the distinction between the “concept” and its “denotatum”, or between language-dependent construal, on the one hand, and referential function, on the other. Neither the cognitive nor the communicative categories are to be confused with the referents themselves; in both dimensions, the categories are “constructed pigeonholes”, not objects or ontological entities. In fact, translation practice has demonstrated that referential equivalence does not presuppose semantic equivalence. Two or more expressions, be they from the same language or not, may be able to designate the same referent without categorizing it in the same way.

For instance, the idiomatic expression sp. *orden del día* has no lexical equivalence in English, even though it can be referentially equivalent with some usages of eng. *agenda*. In the examples below, the denotata of the collocational pattern eng. ...*item on the agenda* (1-4) could possibly coincide with those of sp. ...*punto del/en el orden del día* (5-8). Does it mean that these patterns are equivalent at the communicative level? Not quite. The examples 24-27 are an indication that the pattern ...*item on the agenda* can be used to denote a programme. That is, the pattern oscillates between the more abstract sense of ‘plan’ and the more concrete sense of a ‘list of items to be discussed at a meeting’. The proof of this oscillation in the meaning is that the use of *top* (adj.) to pre-modify *item* expresses ‘priority/importance’ rather than ‘position on a list’. The combination *top item on the agenda* does not refer to the first item (on top of the list) to be discussed at a meeting; instead, it refers to the main task to be done. In fact, the content ‘list of items to be discussed at a meeting’ is not lexicalized in English, given that there is no single formal pattern to convey this content systematically. This content shares the *signifiant* with other contents. The expressions that convey the message of ‘list of items...’ can also convey other senses.

- (1) *The second item on the agenda for a meeting in April 1963 was the export of arms to Iraq. The third item was the export of large diameterI breathed a little easier.*
 - (2) *I breathed a little easier. But I doubt if anybody was prepared for the next item on the agenda.*
 - (3) *But there is still no agreement on where to hold the substantive talks, and that will be the first item on the agenda.*
 - (4) *I could concentrate on the next item on the agenda. Panto at the Civic Theatre in Halifax.*
- (Bank of English)
- (5) *Dentro de este mismo punto del orden del día, fue ratificado el Documento de Adaptación del Plan Estratégico de Protección al Consumidor*
 - (6) *Seguidamente y en el cuarto punto del Orden del día los asistentes unánimemente ratifican en su cargo de Presidente a D. Mario Romerales quien acepta el cargo.*
 - (7) *Lamolda como moderador y Leandro Sequeiros como secretario, se pasó a discutir el punto fundamental del orden del día*

- (8) *Como primer punto del orden del día aparece la exposición del edil Leonardo Vinci para referirse al extinto ex edil Prof.*

(Corpus Cumbre)

- (9) *Argentina's integration with foreign markets will be a top item on the agenda when President Bush comes here tomorrow.*

- (10) *The shock result will be the top item on the agenda at the EU Foreign Ministers meeting in Luxembourg.*

- (11) *So far, the top item on the agenda is that old stalwart, across-the-board tax cuts. These have almost mythological status among Republicans,*

- (12) *But I think North Korea was clearly the top item on the agenda.*

(Bank of English)

The above example indicates that a semantic category at the communicative level is not the referent itself but a linguistic category that shows *a stable (sufficiently predictable) behaviour in the discourse*. In contrast, a semantic category is cognitively pertinent if it proves able to establish and organize the (conceptual) links among the (mental) representations of multifarious objects. Hence, one of the essential differences between the two dimensions is the relationship with monosemy, polysemy, and ambiguity. In principle, the property of being ambiguous detracts from communicative effectiveness, whereas polysemy does not diminish the cognitive potential; in fact, the systematic polysemy of a word seems to increase its potential for conceptual representations, in that the attributes of a single category are able to generate a network of further categories. Thus, we can conclude that monosemy is a fundamental property of communicatively relevant language units and a subsidiary or incidental property of cognitively relevant ones. By contrast, polysemy is a primary property of cognitively relevant semantic units. Generally, the unit "word" is appropriate for approaching the cognitive or conceptual aspects of linguistic semantics, while the unit "collocation" is more adequate for the study of communicative aspects.

What are the implications of this dualism for vocabulary teaching? Our hypothesis is that, by laying more emphasis on units from one or other level, it is possible for the teacher to concentrate on developing either symbolization or communicative skills in the learner. "Word-centred" vocabulary teaching is suitable for helping the learner to construct the cognitive contents (representations) that are typical of the L2. This is because the unit "word" has a high potential for structuring the relationships among multiple conceptual entities. The

basic meaning or core sense of a word conceals the potential for generating further (derived) senses by means of conceptual shifting.

For instance, the basic sense of *agenda*, ie ‘things to be done’, contains the features that are mapped onto more specific senses of the same word: (i) ‘political programme, i.e. things that are planned to be done by a government’; (ii) ‘schedule, i.e. list of tasks and the times at which each of them should be done; (iii) ‘list of items (points or issues) to be discussed at a meeting’. In the three cases, the secondary sense results from applying the meaning ‘things to be done’ in specific domains (‘decisions’, ‘timing’, ‘discussion’, etc.). Insofar as a set of senses can be traced back to a primary meaning, the unit “word” can be attributed a potential for structuring the relationships among multiple conceptual entities. Since there is no isomorphism between the conceptual derivations activated in different languages, learning the lexicon of a FL/L2 requires learning a different way of establishing connections among the mental representations of world entities. For example, the word *agenda* in Spanish develops the senses (i) and (ii) above, but not sense (iii). The polysemy of a word, i.e. the network of senses derived from a common conceptual entity, cannot be predicted from world knowledge or L1-knowledge.

By contrast, “collocation-centred” vocabulary teaching is especially appropriate for helping the learner to participate successfully in communicative events using the L2/FL. This does not mean that the unit “word” cannot function in communication, but in this respect, the ELI proves more efficient, because it does not require any recourse to disambiguating processes. In sum, the production or reception of a verbal message basing on the combination of word senses is likely to require more effort than the production/reception of the same message basing on the retrieval/recognition of ELIs. The underpinnings for this claim will be discussed in the next section.

IV. LANGUAGE ECONOMY

Above, it has been suggested that “economy” could be one of the advantages of collocation-centred over word-centred vocabulary teaching. At first sight, this statement could appear to be at odds with the observation that the ELI is structurally more complex than the unit “word”, more so if the predictability of meaning requires more than two co-occurring words. A case in point is the collocational pattern (*top/bottom/lower*) *left/right hand corner/side*. Note that the combination *right/left hand* is not enough to predict the meaning ‘lateral location in space’, since both the sequences *left hand* and *right hand* can form part of other

collocations conveying the meaning ‘body part at the end of someone’s arms’ (eg *my/her/his... left/right hand*). The composition of an ELI often spans more than two words. Besides, the extended item involves a series of idiosyncratic constraints both on paradigmatic lexical sets (e.g. *settle/*calm/*soothe + PERS. PRON. + differences with*) and on grammatical features (e.g. *settle/resolve + PERS. PRON. + differences with*, but **settle/resolve + PERS. PRON. + difference with*).

The question, then, is not whether the storage and representation of an ELI requires higher or lower costs than each one-word entry. The relative simplicity of the internal structure of the word, compared to the ELI, is not at stake. Admittedly, the ELI means an increase in “representational complexity”, but to balance, it also means a decrease in “processing/interpreting complexity”. Thus, the crucial question is whether the internal complexity of the ELI is “cost-effective”, i.e. whether it saves more efforts than the costs involved. To answer this question, comparisons must be drawn between (i) the cost of increasing storage capacity and structural complexity, on the one hand, (ii) and the reduction of the number of operations involved in the production and interpretation of text, on the other. The efficiency of the chunking strategy depends on proving that the growth in storage capacity is worth the “miniaturization” of processing efforts.

The economy factor in formulaic/prefabricated language has been explored by previous literature in the field of applied linguistics (see Lewis, 1993, among others). To quote Nation (2001: 320), “the main advantage of chunking is reduced processing time. That is, speed”, whereas “the main disadvantage of chunking is storage”. The correlation between prefabs and processing speed is consistent with the widely admitted remark that collocational knowledge is a decisive factor in developing fluency. Some psycholinguistic evidence for this can be read in Moon (1998: 30-31), who echoes the finding that “processing speed is linked not so much to the gross measure of information processed as to the *number of highest-level units that must be treated serially*” (emphasis added). This underpins the general applicability of Sinclair’s idiom principle, whose operation can be described as making “fewer and larger choices” (Sinclair, 1991: 113).

Normally, the contribution of prefabs to fluency is attributed to the avoidance of rule based processing. However, corpus linguists have emphasized another important factor: the avoidance of sense disambiguation processes. Sinclair (1998: 10) has remarked that a very simple sentence such as *The cat sat on the mat* can give rise to 41,310,000 possible combinations of its elements. For Sinclair, one of the arguments against the concept of “word sense” is the inconsistency between the intricacy of disambiguation and the apparent ease and

effortlessness with which native speakers engage in verbal communication. Sinclair (1998) and Teubert (2005) think that the multiplicity of potential sense combinations is not an objective property of text structure but a fictitious methodological construction which results from adopting the word as the default lexical unit.

Precisely, one of the disadvantages of basing vocabulary teaching on word meanings is that it creates the need to select from many thousands of sense combinations in every stretch of text. This means that, when confronted with actual communicative events, the learner has to carry out multiple processes such as pragmatic and logical inferences, semantic-syntagmatic operations, etc., in order to decide which combination of word senses is the more coherent one. This operational complexity can be drastically minimized if the more stable and cohesive word co-occurrences have been learned as wholes. To quote Teubert and Čermáková (2004: 151), there is no need to choose among four senses for *friendly* and eight for *fire* if we know that the expression *friendly fire* has a single meaning. One of the advantages of learning collocations instead of words is that the retrieval/recognition of the former makes processing considerably simpler and faster.

V. CONCLUSIONS

This paper has explored some implications of corpus findings for FL/L2 vocabulary teaching. A close examination of corpus data supports the hypothesis that words are co-selected in chunks (Sinclair, 1991). In turn, this suggests that the monosemous sequences in discourse result from complex/extended lexical choices overlapping word boundaries, rather than being the product of successive word sense selections. These corpus-linguistic insights into the structure of text and vocabulary have invigorated the controversy about the lexical unit. The debate has important implications for the applied branches of linguistics, because it affects the discussion of which language units should constitute the main learning target and teaching object. An emerging issue in EFL is whether the new empirical findings from corpora are sufficient to substantiate the need for establishing a new unit of meaning in vocabulary teaching.

Two opposite stances on this issue have been compared throughout the paper. The word-centred perspective assumes that learning vocabulary consists of knowing various aspects of the word. On this view, the role of collocational knowledge is ancillary to the unit “word”. By contrast, the sceptics about word meaning suggest that words should be learned as components of higher-level lexical units (ELIs). This means that lexical competence is not

equated with knowing various aspects of words but with knowing the multi-word patterns which words enter.

After discussing some underpinnings for one and the other perspective, we have conjectured that they can be complementary in one respect: they promote the development of different functions and aspects of lexical competence. Idiomatic patterning constitutes the most efficient language level for promoting fluency and facilitating communicative success in the FL/L2. Meanwhile, the polysemy of words remains available for learning the way in which the foreign/second language community categorizes the world.

VI. REFERENCES

- Almela, M. (2006). *From words to lexical units. A corpus-driven account of collocation and idiomatic patterning in English and English-Spanish*. Frankfurt a. M.: Peter Lang.
- Biber, D., Conrad, S. & Cortes, V. (2004). 'If you look at...': Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 401-435.
- Casas Gómez, M. (2002). *Los niveles del significar*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Feilke, H. (1996). *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt a. M.: Suhrkamp.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (Ed.), *Phraseology. Theory, Analysis and Applications*. Oxford: Clarendon, pp. 145-160.
- Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of Task-Induced Involvement. *Applied Linguistics*, 22:1, 1-26.
- Lewis, M. (1993). *The Lexical Approach*. Hove: LTP.
- Liang, S. Q. (1991). À propos du dictionnaire français-chinois des collocations françaises. *Cahiers de Lexicologie*, 59:2: 151-167.
- Moon, R. (1998). *Fixed expressions and idioms in English. A corpus-based approach*. Oxford: Clarendon.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269-284). Mahwah, NJ: Erlbaum.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

- Ooi, V. B. Y. (1998). *Computer corpus lexicography*. Edinburgh: Edinburgh University Press.
- Penadés Martínez, I. (2001). ¿Colocaciones o locuciones verbales? *Lingüística Española Actual*, 23:1, 57-88.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Ruhl, C. (1989). *On Monosemy. A Study in Linguistic Semantics*. Albany, NY: State University of New York.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics*. Amsterdam, Philadelphia: John Benjamins, pp. 1-24.
- Stubbs, M. (2002). *Words and phrases. Corpus studies of lexical semantics*. Oxford, Malden, Melbourne, Berlin: Blackwell.
- Teubert, W. (2004). Units of meaning, parallel corpora, and their implications for language teaching. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam: Rodopi, pp. 171-189.
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10:1, 1-13.
- Teubert, W. & Čermáková, A. (2004). Directions in corpus linguistics. In M. A. K. Halliday, Teubert, W., Yallop, C. & Čermáková, A., *Lexicology and corpus linguistics. An Introduction*. London, New York: Continuum, pp. 113-165.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian. Towards a corpus-driven approach. In B. Altenberg & S. Granger (Eds.), *Lexis in contrast. Corpus-based approaches*. Amsterdam/Philadelphia: John Benjamins, pp. 73-95.