



The Distribution of the Lexical Component in ELT Coursebooks and its Suitability for Vocabulary Acquisition from a Cognitive Perspective. A Case Study

RAQUEL CRIADO¹

LACELL RESEARCH GROUP – UNIVERSIDAD DE MURCIA

ABSTRACT

The psycholinguistic and neurolinguistic perspective of language acquisition requires some essential conditions in vocabulary acquisition: a) repetitive practice, which allows for data to reach long-term memory, and thus become proceduralised and automatized; b) how relevant the lexical items are regarding the communicative needs of the learners insofar as communicative relevance is linked to frequency in general linguistic usage; c) the potential in vocabulary acquisition, which will necessarily relate to the amount of new lexical items introduced in each one of the units in textbook; d) the way words are taught, i.e. whether aimed at explicit or incidental learning. In order to analyse and evaluate these issues, we will study the lexical items presented in a specific textbook from the point of view of frequency, distribution along the manual, opportunities for rehearsal and repetition (which will depend on frequency), and the nature of the activities centred on vocabulary. The results of this case study will allow us to check whether or not they may stand a comparison against the findings of psycholinguistic and neurolinguistic research on vocabulary acquisition.

KEYWORDS: vocabulary acquisition, ELT, frequency, practice, corpus linguistics, psycholinguistics

RESUMEN

Desde la perspectiva de la psicolingüística y de la neurolingüística, deben darse algunas condiciones para la adquisición léxica: a) práctica repetitiva, que facilita el paso de los datos a la memoria de larga duración, con la consiguiente procedimentalización y automatización; b) el grado de relevancia respecto a las necesidades comunicativas de los hablantes, teniendo en cuenta que la relevancia comunicativa se correlaciona a su vez con la frecuencia de las palabras en el uso general de la lengua; c) el potencial de adquisición léxica, que se relacionará necesariamente con el número de palabras introducidas en cada unidad de los libros de texto; d) la manera como se enseñan las palabras, ya sea explícita o incidentalmente. Para analizar y valorar estos temas, se estudiará el léxico introducido en un

¹ *Address for correspondence:* Raquel Criado, Departamento de Filología Inglesa, Campus de *La Merced*. Universidad de Murcia. 30071 Murcia, Spain. Tel: 00 34 868888791; Fax: 00 34 868883185; e-mail: rcriado@um.es

libro de texto en lo relativo a la frecuencia, distribución, oportunidades que ofrece para la repetición (que dependerá de la frecuencia con que aparecen las palabras) y naturaleza de las actividades centradas en el léxico. Los resultados de este análisis nos permitirán también valorar si el manual se ajusta y en qué medida a las más recientes investigaciones nacidas de la psicolingüística y la neurolingüística en relación con la adquisición de vocabulario.

PALABRAS CLAVE: aprendizaje de vocabulario, enseñanza del inglés como lengua extranjera, frecuencia, práctica, lingüística de corpus, psicolingüística

I. INTRODUCTION

Interest in *vocabulary acquisition* has never diminished throughout the history of language teaching (Howatt, 2004; Kelly, 1969; Sánchez, 1997, 2009; Schmitt, 2000). Textbooks for language teaching have often included lists of lexical items for *explicit learning* by students, as it was the case in the Grammar-Translation Method, Gouin's Method or the Direct Method. Other methods, such as the Audio-lingual or Communicative ones used to introduce vocabulary items within situational or communicative contexts. The teaching and learning of vocabulary lists have been at the core of the teaching / learning process in the classroom for centuries. A typical sequence in the classroom was (i) the explanation of grammatical rules by the teacher, followed by (ii) classroom practices in which the words were first memorised and later used to build the sentences prescribed by the rules. Words were learned and manipulated as single and fully autonomous lexical units.

The statement by Lewis that "Lexis is the core or heart of language but in language teaching has always been the Cinderella" (Lewis, 1993: 89) is not fully true if words are taken as basic and autonomous lexical units. After all, the teaching tradition reveals that learning words has always been one of the main tasks recommended by textbooks. Still, the importance of vocabulary has not always been adequately emphasized, and particularly the nature of words and its contribution and role in the building of meaning has not been correctly evaluated by most teaching methods. In the last two decades though, the importance of vocabulary knowledge has been brought to the forefront, in the field of vocabulary acquisition research and assessment (Laufer & Hulstijn, 2001; Nagy & Scott, 2000; Nation, 2001, 2006; Read, 2000, etc.), in order to improve communicative potential, fluency and accuracy. Research has also contributed to a better understanding of the word and its dependency on context.

The aim of this paper addresses the issue of whether the teaching materials have adapted or not, and how much, to this new dimension in the understanding of vocabulary and its role in language learning. We will approach the problem analysing first the most important theoretical aspects related to *vocabulary acquisition and learning* from a cognitive and pedagogic perspective. Secondly, we will select and analyse a textbook on the following parameters: lexical *frequency* and distribution along the manual (compared to lexical frequency in general English), opportunities for *rehearsal and repetition* (which will depend

on *frequency*), and the nature of the activities centred on vocabulary. The comparison of these data against what is to be expected from a pedagogical and cognitive point of view will shed light on whether the textbook is suitable or not to favour effective *vocabulary acquisition*, that is, proceduralisation and automatism, which is the final goal of knowledge acquisition from a cognitive perspective.

II. SINGLE WORDS VS. 'PREFABS'

Recent interest in vocabulary has given rise to the debate about which is the nature of the semantic unit. Some scholars (Teubert, 2004, 2005) have challenged the widespread and traditional belief that words are the ultimate semantic units. In debates on the nature of the lexical unit the belief is more and more reinforced that collocates and context in general are to be taken into consideration in order to define semantic units. But the assumption that words are not the only units of meaning brings with it some consequences to language teaching and learning. Larger lexical units, if claimed, must first be identified, defined and conveniently presented as an object for teaching and learning.

Lexical approaches to language teaching, such as the one proposed by Lewis (Lewis, 1993), cope with this problem suggesting that words should be learnt within the context they appear in communication, that is, words should not be learnt in isolation, as it was the case in the Grammar Translation Method, for example. Moreover, lexicalists also claim that we do not only acquire words as isolated items; quite often we memorise what they refer to as 'lexical chunks', that is, two or more words taken as a whole. This is typically the case of idioms, but it also applies to collocates and many other phrases of frequent usage. Learning 'prefabricated' chunks offers many advantages in communicative situations, since speakers retrieve already proceduralised knowledge which does not require any special conscious processing. If chunks are already there, ready for use as encapsulated units, the communicative process gains fluency. From a methodological perspective, pre-existing chunks prove that we do not only store single word units, which are later on processed following the rules of the language. Our mental lexicon stores lexical items in many different patterns and in various complex composites, with different morphological and syntactic implications. As a consequence, single word units cannot be taken as the only lexical items present in our cognitive system, as traditionally assumed.

This intuition is supported by some incipient research. Erman (2007) challenges the view that we only store single words in our mental lexicon. On the basis of the evidence he gathered, he concludes that 'at least 50% of the written and spoken language (and probably more) is made up of prefabricated structures (Erman, 2007: 28). Erman departs from Anderson's ACT theory of the human cognitive system (Anderson, 1983, 2005) regarding the types of memory subsystems in the brain. The detection and duration of pauses when retrieving lexical information is taken as a proof for deciding when a compound lexical unit is automatized (that is, should be considered as proceduralised knowledge) or not. Erman's

conclusion had already been advanced by various authors (Jackendoff, 1997; Schmitt, Grandage & Adolphs, 2004; Sinclair, 1991, etc.), although no experimental evidence accompanied their intuitions.

Moreover, Erman's study seems to confirm that adolescents are less secure in eliciting prefabs, while adults did not pause as much as young people did. This fully agrees with the following hypothesis: the acquisition of prefabs occurs in more advanced stages of linguistic knowledge, be it the mother tongue or a second or foreign language. In other words, the learning of single lexical units (words) may be the first stage before we initiate the learning of prefabs and 'linguistic composites'. This view complements the assumption that *vocabulary acquisition* is 'incremental in nature' (Schmitt, 2000: 117).

III. THE NATURE OF 'WORD'

The concept of 'incremental learning' applied to vocabulary is extremely complex. The complexity and composite nature of what we mean by 'word' suggests that its learning will necessarily be 'progressive' or incremental, since learning all the elements involved, formal and semantic, is more likely to take time and opportunities for practice. When we say 'we have learned a word' we most often mean that we are aware of its 'essential' semantic features, not necessarily all its possible instances of occurrence in communication. The basic contour of a lexical item may be called its 'identity card': its ID allows contrasting that word against other words, even though not all the details are specified. We may therefore affirm that we know the word *table* and ignore some of the meanings of this word in contexts other than the physical nature of 'a flat surface, usually supported by four legs', as for example 'an arrangement of facts and numbers in rows or blocks, especially in printed materials'. In parallelism to that, we can also say that features like 'a table of contents', a 'round-table', or 'the periodic table' enlarge the semantic field of *table*, but their ignorance by specific speakers does not hinder to state that a learner of English, or an infant, 'know' the word *table*. The incremental knowledge of words should be considered as a scale starting with a minimum and a maximum, the minimum including their ID, that is, the essential contrastive features that define a specific word in the linguistic system it belongs to. How advanced a user of a language is in the scale of vocabulary knowledge will define the general and global knowledge of the language. An advanced command of the language implies that the speaker knows many words and most or all of their meanings. This, in its turn, asks for the knowledge of many of the prefabs and collocates a word is involved in.

The dependency of lexical items among themselves derives from the very nature of words: they are always used within a context and all of them collaborate in building the contextual meaning. Word context, on the other hand, is not to be reduced to the co-text, that is, the amount of words immediately accompanying a lexical item. Context implies larger settings as well, both lexical and semantic, within which a specific word is necessarily embedded. The co-text and the larger context may be often decisive for identifying the exact

meaning of lexical items. This is what happens in the acquisition of the mother tongue: we begin learning words with the meaning they have in the specific context in which they are used for the first time. In the case of *table*, for example, the meaning learned for the first time is normally associated to 'a flat surface supported by four legs', the tool on which people seat for dinner, or the tool on which we put things on. Such a link between this word and similar situations and contexts, when noticed repeatedly, reinforces the association, which may end up being proceduralised and finally automatised. The exposure of the learner to other contexts and situations in which the same word may be used with partially or totally different meanings will bring with additional associations and links. Repeated exposure to these novel associations will contribute new meanings, new collocates and new prefabs. This is the kind of incremental knowledge some authors refer to (Bhans & Eldaw, 1993; Schmitt, 1998, 2000).

IV. TEACHING MATERIALS AND EXPLICIT / INCIDENTAL LEARNING

When approaching the learning of vocabulary two options are generally considered: *explicit* and *incidental* learning. *Explicit* learning advocates for a conscious presentation of the information to be learned. It is assumed that being conscious and aware of what we have to learn is more efficient for acquisition. On the other hand, explicit attention consumes a lot of time and this slows down the process. *Incidental* learning advocates usage (meaningful usage, with no explicit information on the words). *Incidental* learning may be less efficient because (i) students may not be able to capture the meaning of all the words they come across while reading or speaking, and (ii) they have fewer chances to come across less frequent words and therefore more difficulties to increase their vocabulary. *Incidental* learning has also some advantages: it makes students rely more heavily on context for discovering the meaning of new items and learning in general proceeds smoothly, more in line with natural language acquisition processes. The exposure to repeated instances of vocabulary use is higher, since fluency in linguistic production is not interrupted by explicit information and there are more opportunities for the (unconscious) proceduralisation of linguistic knowledge.

Classroom practice and most teaching materials combine both options, allowing for explicit and incidental *vocabulary acquisition*. The consensus very often centers on teaching explicitly most usual words, while less frequent items are left to incidental learning (Nagy, 1997; Nation, 2001, 2006). Incidental learning, however, is heavily dependent on usage, as annotated above, and the problem in most classrooms is that opportunities for communicative interaction are poor. The lack of real communicative contexts and contact with native speakers adds further problems to the efficacy of this type of learning. Teaching materials, on the other hand, are not adequate substitutes for real and intensive language use since they are limited by nature and cannot offer the students the linguistic richness needed for incidental learning alone.

A balance is to be found between *explicit and incidental vocabulary acquisition*. In natural language learning environments, incidental learning is the rule. Learners are not given any explanation of the meaning of the words used and must therefore rely on the context as the only source of information. Context is useful and operative when it is meaningful for the user, that is, when it offers a good amount of already known words around a few lexical items with which the user is not yet familiar. Studies based on word *frequency* counts reveal that the amount of words used in daily communication is not as high as one might consider. According to the *Brown* corpus (Kucera & Francis, 1967), a one million-word corpus of American English, the first one thousand more frequent lemmas cover 72% of the whole text, and the first five thousand lemmas cover up to 88.6% of the total (see Table 1).

First 1,000 lemmas	72%
Second 1,000 lemmas	79.7%
Third 1,000 lemmas	84%
Fourth 1,000 lemmas	86.7%
5000 lemmas	88.6%
6000 lemmas	89.9%

Table 1. Text coverage by the first 6000 lemmas in the Brown Corpus.

There is a heavy tendency in language for a reduced number of words to be used intensively, while the rest of words are progressively less frequent. Reading and understanding authentic written texts require a vocabulary of three to five thousand word families (*word family*: a base word plus its derivatives) (Nation & Waring, 1997). More simple texts may be understood with a far less number of words, somewhere over one thousand. Studies on the distribution of words in linguistic usage conclude that two thousand word families will cover ca. 99% of the basic communicative needs (Schmitt, 2000; Schonell, Meddleton & Shaw, 1956). Others increase the number to three thousand (Adolphs & Schmitt, 2003; McCarthy, 1998). On the basis of these facts it is reasonable to assume that efficiency in *learning vocabulary* is connected (i) to the learning of the most frequent words, and (ii) the amount of the most frequent words to which priority should be given is around three thousand word families. The issue of explicit and incidental teaching of vocabulary may find on those data a useful reference for taking decisions on which words to teach and how to teach them.

V. THE COGNITIVE PERSPECTIVE OF VOCABULARY ACQUISITION

Language acquisition and *vocabulary acquisition* are mutually dependent. After all, words are the formal symbols associated to concepts, and storing and manipulation of concepts are key issues in communication through language. It can be safely stated that the degree of knowledge of a specific language will be directly related to the amount of vocabulary a

speaker knows of that language. It is quite relevant therefore to attend to the issue of how vocabulary is learned and memorised.

Knowledge of the words of a language is the type of knowledge referred to as 'declarative knowledge' (*DEC*). *DEC* opposes 'procedural knowledge' (*PRO*) mainly because the former can be brought to consciousness as often as we want and can be quickly acquired through reflection and conscious cognitive action, while the latter takes more time to consolidate, escapes consciousness and awareness and is quicker in performance. The nature of both types of knowledge may imply different strategies for their acquisition. The nature of both types of knowledge may imply different strategies for their acquisition, such as when we refer to the role of consciousness (explicit) or implicitness (incidental) in learning, for example.. Regarding the consolidation of both types of knowledge, the basic strategy is the same: consolidation is connected to memorisation and memorisation is governed by rehearsal. It is true that *DEC* may require at times only a single stimulus to be acquired (Ullman, 2004), while *PRO* will always result from repeated action triggered by recurrent stimuli. Nevertheless, the consolidation of both *DEC* and *PRO* share a similar need for *repetition* before becoming automatised (Sánchez & Criado, in press). Automatisation is the only condition in skill learning that guarantees fluency of performance. As referred to language, fluency is the necessary condition for establishing meaningful and easy communication among the members of a speaking community.

Vocabulary is declarative knowledge, that is, knowledge about facts. *DEC* is acquired through association. In the case of vocabulary, the acquisition depends on the association of a real thing in the outside world to a concept in our mind. Associations are triggered by stimuli in the neural network (Ullman, 2004). A stimulus may begin at a specific neural node and is transmitted to other neurons by means of neurotransmitters, which result from the release of chemicals that change the electric polarization of the membrane in the neural receptors. The transmission of the electrical signals runs along specific channels, which strengthen under certain conditions. Full consolidation is reached when the same stimulus is able to automatically activate an already shaped channel and produce similar results at the end of the neural circuitry. There is still a long way ahead to understand fully how these initial electrical bits generated by and transmitted through the neural system derive into knowledge. *Psycholinguistics* firstly and *neurolinguistics* in the last decades are contributing a better understanding of the cognitive processes that generate what we refer to as 'knowledge' (Anderson, 2005).

One of the most relevant areas of cognitive processes is how data are accessed, transmitted and memorised. Memory is particularly important in cognitive processes, since it is the device responsible for storing data, keeping them at our disposal and accessing them whenever we need them. Our neural system is known to work with two types of memorisation devices: *short-term memory* and *long-term memory* (Anderson, 2005; Atkinson & Shiffrin, 1968). Data captured are first presented to short-term memory, a kind

of working memory acting as an interface with the outside world. Input entering the working memory flows very quickly and is immediately lost unless it enters long-term memory. It can therefore be stated that our working memory is the main entrance for input data, which is equipped with a filter for evaluating and selecting only the data considered relevant or necessary.

From the point of view of efficiency in *vocabulary acquisition*, what matters is the amount of lexical information entering and consolidating in long-term memory. Neurologists and psycholinguists tell us that long-term memory is activated and strengthened mainly (i) through rehearsal or *repetitive practice* and activation, (ii) when attention is focused on specific data, and (iii) when new data are associated in some way to already consolidated information. The three options are accessible to learners and teachers. *Repetitive practice* has been present all throughout the history of school teaching and there is no doubt on its efficacy as a teaching and learning technique. Awareness, attention and explicit reflection on the data to learn has been the subject of opposed views in teaching methods, but what is known about the biological bases of knowledge acquisition invites us to seriously reconsider the issue and analyse its practical applications in language teaching materials and the classroom (Sánchez & Criado, in press). Associating new data to already memorised data is usually connected to individual learning strategies and admits a good amount of variation.

Rundus' experimental studies (Rundus, 1971) revealed that the more participants rehearsed an item, the more they remembered it. The results match perfectly well with neural observations (Ullman, 2004). Kapur, Craik, Tulving, Wilson & Brown's findings (1994) reinforce the importance of rehearsal with a new element: attention and awareness. Their experiments reveal that rehearsal is more efficient when it is meaningful and fully conscious and focused explicitly on the data being learned. That fact confirms the well-known experience about the usefulness of sheer mechanical repetition. The efficacy of repetition is due to the structural changes that take place in the neural synapses (or connections among neurons). Repeated connections strengthen the connection, and so the task is rendered easier. When the task becomes so easy that you can perform it with less effort or attention, it is because a certain degree of proceduralisation of the process has been reached (proceduralisation can be complete after the first 16-item block of practice items, as was shown by DeKeyser (1997). At this point in the process, structural changes in the synapses affected apparently cease and become stable. In addition, more practice implies execution that is more efficient. Facts regarding the two types of memory and the consolidation of data may be synthesized in the following way: most of the information which flows through the short-term memory is usually lost, pressed by the permanent flow of incoming data, unless repetitive iteration and/or attention favours its selection to enter long-term memory. Iteration or *repetition*, together with attention, is therefore the habitual mechanism, which guarantees permanence and avoids oblivion in information storing.

Coming back to incidental vocabulary acquisition, it is worth noticing that learning with this method is also dependent on repetition and exposure. The difference from explicit learning lies on the way exposure takes place. Non-reflective exposure is more likely to take more time and predictably more *repetitive practice*, even if it may also be effective in shifting input data from short-term to long-term memory. The advantage is that incidental learning spares explicit attention, which, on the other hand, cannot be focused on every single data incoming our working memory. The complementary character of explicit and incidental vocabulary acquisition is then based on the capabilities of our neural system for accessing and storing input data.

VI. A CASE STUDY: THE LEXICAL COMPONENT IN A TEXTBOOK

I have commented and briefly analysed in the previous sections some of the basic issues in vocabulary acquisition and learning, namely, (i) the importance of the word in the linguistic communicative system, as an item to be taught and learnt; (ii) the vocabulary size needed for engaging in basic communication; (iii) the *frequency* index of the vocabulary learnt (which will seriously affect communicative efficiency); (iv) the need for *repetition* as a necessary condition for proceduralisation; and (v) the activation of explicit and incidental learning, favoured by the activities offered in the manual. In this section we will analyse a specific textbook in order to find out whether the topics mentioned above are positively approached or not, and in which way. A positive approach will depend on the amount and nature of the vocabulary introduced, on their adequacy to the general frequency list of English, and on how activities are designed, so as to favour explicit or incidental learning.

Meara & Jones (1988) claim that ‘vocabulary knowledge is heavily implicated in all practical language skills, and that speakers with a large vocabulary perform better than speakers with a more limited vocabulary’. We fully concur with this statement. Another consensus refers to the amount of word families learners need to know in order to be able to communicate in a second language. As annotated above, the range of 2,000 to 3,000 word families is considered adequate for engaging successfully in basic communication. Even though speakers with higher vocabulary levels will gain in fluency, 3,000 word families is a reasonable reference for measuring basic communicative capabilities in learners. The three base-word vocabulary ranges defined by Nation in 2001 and 2006 (1,000, 2,000, and 3,000 most frequent words of English), against which we will contrast the vocabulary found in the textbook, appears to be therefore a sound indicator of vocabulary knowledge. The task requires that we first identify the vocabulary offered in the textbook. In a second stage, we will find out if the vocabulary in the textbook matches the ‘expected’ vocabulary according to the *frequency* list of general English, and particularly in relation to the first 3,000 most frequent words as identified by Nation (2001, 2006).

The computational tool we will use for counting and comparing vocabulary in the textbook and the general use of English will be RANGE

(<http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>), which classifies the vocabulary of any text into three frequency categories: the first 1,000, the second 1,000 and the third 1,000 most frequent words of general English. Words not included within these first three categories appear as off ranges. The classification of words as tokens (every word form in the text, be it repeated or not), types (different words in the text: *friend* and *friends* are two types) and word families (the headword, its inflected forms and its closely related derived forms) is very relevant for our study. This feature of RANGE refines considerably the information available. The identification of tokens vs. types allows for a contrast between the raw vocabulary input against the new words really introduced in a specific text. From the perspective of vocabulary acquisition we will later check if the textbook complies with the specific conditions governing knowledge acquisition, in particular (i) those regarding ‘opportunities for repetition’, which will depend on the *frequency* of occurrence of lexical items throughout the textbook, and (ii) the presence of activities favouring explicit or incidental learning.

The textbook analysed is *Valid Choice 2*, by Jane Lawrence and Alan Williams, published by Burlington Books (2006). The manual is adapted to the syllabus of the Spanish Bachillerato, Course 2. The methodological approach must therefore adjust to the Communicative Method and to the principles underlying the *Common European Framework of Reference for Languages* (2001). The book is structured in 6 main units, 10 pages each. Some specific sections are also included in the student’s book: a section for ‘exam preparation’, a grammar appendix and a glossary. It is worth noticing that the glossary includes about 500 words only, which are defined as ‘the most frequent words’, with no further specifications. Such a glossary clearly contrasts against the 3,225 types (ca. 2,320 word families) used in the manual (see next section). It should also be assumed that students using *Valid Choice 2* have already used *Valid Choice 1* (which is not the object of analysis in this paper).

VI.1. Word counts in the textbook and word ranges

The textbook contains 25,687 running words (tokens). Out of this total, 3,225 are distinct words (types). This figure amounts to about 2,320 word families. Regarding the word ranges defined by Nation (2001, 2006), only 148 types belong to range 1; 630 to range 2, and 242 to range 3. The same types classified as word families amount to 113 for range 1; 434 for range 2, and 187 for range 3 (see Table 2).

WORD range	TOKENS/%	TYPES/%	FAMILIES
(1) 1000	8296/ 32.30	148/ 4.59	113
(2) 2000	1970/ 7.67	630/ 19.53	434
(3) 3000	636/ 2.48	242/ 7.50	187
off ranges	14785/ 57.56	2205/68.37	?????
Total	25687	3225	734

Table 2. Tokens, types and word families for Ranges 1, 2 and 3 in *Valid Choice 2*.

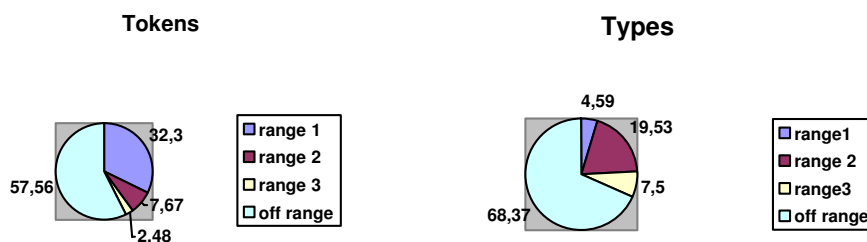
The first striking feature refers to the amount of distinct words used in the textbook: *Valid Choice 2* contains 3,225 types (ca. 2,320 word families²). The class hours during the academic year amount to 100. This fact implies that if students are to learn all the words included, they should learn 32 new types per hour, almost 100 per week, or 400 a month. It must be added to that the consolidation of the words already introduced in previous sessions. Such expectations exceed by far the most optimistic views on vocabulary acquisition and learning. A popular method for learning English, *Maurer Method*, heavily biased by propaganda interests, advertises the efficacy of its materials with the slogan 'Learn the most frequent words of English... in 20 weeks'. Maurer counts on the learning of 7 new words per day, that is, 42 words per week. The prospects by Maurer double the results based on experimental research: Ito (1995) concluded in an experimental study with Japanese students that they learned only 3 new words per day, that is, 20/22 per week. Our textbook lies too far away from expectations. You may argue that textbooks should not only offer the words supposed to be acquired by the learners. Specific communicative events and situations require the use of low frequency contextual vocabulary which must not necessarily be a learning target. In *Valid Choice 2*, out of the 3,225 words (types) introduced, 1,345 occur only once, and 528 occur twice in the texts and exercises. Instances of words occurring once or twice give a total of 1,873. It could be assumed that instances of low occurrence (under three occurrences) can hardly be considered candidates for memorisation and could be excluded. The exclusion of words occurring once and twice would lower the amount of words for acquisition to 1,352, half of the total of types in the textbook. Still, learning 13.5 types per day (1,352 in 100 hours) is far from what experimental studies predict as adequate and within the acquisition potential of learners.

Regarding the three word ranges specified by Nation, 8,296 tokens are included within *range 1* (32.3% of the total). The figure seems reasonable in terms of percentage, but it only covers 148 types (4.59%) and 113 word families (Graphic 1). The unbalance between tokens and types is due to the high frequency of a few lexical items in range 1, which does not contradict *per se* the normal distribution of words in texts. The problem is though that 852 types of range 1 do not occur in the textbook. It is hard to assume that students have

² RANGE does not calculate word families of lexical items outside the three ranges. The figure mentioned here is the result of a probabilistic projection based on the proportion of word families vs. types in ranges 1, 2 and 3.

already fully consolidated those 852 high frequency items. It would have been more reasonable to reinforce their acquisition in level 2, perhaps with occasional occurrences.

Valid Choice 2 is targeted at students who have already completed *Valid Choice 1*, which is designed for initial B1 level. Thus, *Valid Choice 2* is aimed at reaching the consolidation of level B1 (as prescribed by the Spanish official syllabus for secondary education). Accordingly, it should pay attention to the vocabulary of range 2 (second 1,000 most frequent words) and particularly to range 3 (third 1,000 most frequent words). Tokens in *range 2* in the textbook are actually 1,970 (7.67% of the total). They include 630 types (19.53%), and 434 word families (see Graphic 1³).



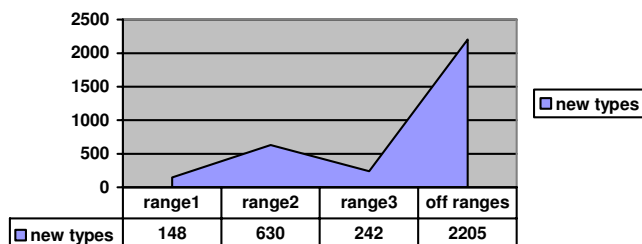
Graphic 1. Types and tokens in *Valid Choice 2*.

These figures need some comments. The amount of tokens of range 2 is too low if compared to the total in the textbook, and more specifically if compared to the items included within range 1 (as the percentage clearly shows). The amount of types in range 2, however, is significantly higher: 630 (19.3%); this is also the case for word families. The relative lack of balance in the amount of tokens and types regarding the total of lexical items in the textbook implies serious negative consequences (as it was the case in range 1 vocabulary, but in the opposite direction). It means that the textbook introduces a reasonable amount of range 2 types (630/1,000), but their frequency is too low to favour effective consolidation or proceduralisation, since students will find each new item only three times along the textbook (the average that results from dividing 1970 (tokens) by 630 (types)).

Range 3 accounts for 636 tokens in the textbook (2.48%). If the textbook is to reach the completion of level B1, this percentage lies far away from expectations, since this level (defined for 'independent users' in the *Common European Framework*) requires a fluent communicative use of English in daily life, very much in line with the third 1,000 words included in range 3 plus the 2,000 words from the previous ranges 1 and 2. Accordingly, the types pertaining to range 3 in *Valid Choice 2* should equal at least the amount of words included in range 2; in any case, the vocabulary learned should follow a steadily ascending

³ Given that the graphics have been performed with the Spanish version of Word, there appears a dot for decimals instead of a comma in all the figures in each graphic.

line from range 1 to range 3. In *Valid Choice 2*, the ascending curve for new vocabulary breaks off in range 2, falls down even more in range 3 and ascends abruptly in 'off ranges' (Graphic 2):



Graphic 2: New types along ranges

In doing so, the book runs into a serious unbalance, which affects negatively the communicative potential of the vocabulary learned. From a pedagogical perspective, we should expect that *Valid Choice 2* (i) reinforces what has been learnt in *Valid Choice 1*, and should thus introduce new lexical items, which are proportionate to the learning potential of the students and to the ascending *frequency* line of general English. The new words should mostly appear first in range 2, and smoothly increase in range 3 (a higher level). A more advanced level (in that case 'off ranges' – presumably B2) is not the goal of this textbook and should consequently be poorly represented. This is not the case here: *Valid Choice 2* offers a strikingly high number of types above range 3: 2,205. The unbalance comes clearly into light in terms of percentage: the new items not included in ranges 1, 2 and 3 take 68.37% of the types detected in the book, against only 19.53% in range 2, and 7.50% in range 3 (apparently the closest to the goals pursued by the manual). A sound distribution would ask just for the opposite: 68.37% of the new items (the highest figure) should belong to ranges 2 and 3, while the off-ranges interval should take lower percentages. Range 1 should be granted a moderate representation for consolidation purposes.

We must therefore conclude that *Valid Choice 2* is clearly unbalanced regarding

- (i) the amount of vocabulary offered;
- (ii) the distribution of vocabulary throughout the three ranges described by Nation (2001, 2006);
- (iii) the *frequency* of the vocabulary included, which is too low and will not favour proceduralisation and automatisisation (both require more opportunities for *repetition*);
- (iv) the amount of words the students are expected to learn, which reach a level well above the more optimistic studies in the field.

VI.2. Vocabulary frequency in presentation texts and activities

Textbooks are typically structured in two main sections: a first section with texts through which vocabulary and grammar relevant for the lesson are introduced in context, and a second section with activities, which aim at practicing the linguistic elements and grammar, selected as the main goal of the unit and introduced in the first section. We will analyse only the distribution of vocabulary in each one of those sections.

Valid Choice 2 deviates significantly from pedagogically based expectations. The section with the texts should abound in new types, while the section with the activities should increase the amount of tokens in relation to types. The reason is obvious: presentation texts are specially selected to introduce new vocabulary, they are supposed to include repetition of words only occasionally. On the other hand, the section with activities is specifically designed to practice words and grammatical structures as a means to consolidate acquisition. Table 3 illustrates quantities in the section with texts and in the section with activities:

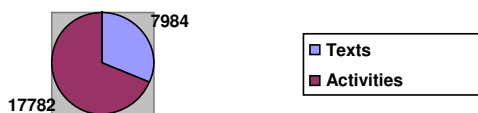
SECTION WITH TEXTS ONLY:			
WORD range	TOKENS/%	TYPES/%	FAMILIES
(1) 1000	2821/35.33	130/ 6.79	101
(2) 2000	568/ 7.11	350/18.28	283
(3) 3000	205/ 2.57	133/ 6.95	109
off ranges	4390/54.98	1302/67.99	(not specified)
Total	7984	1915	493
Section with ACTIVITIES only:			
WORD range	TOKENS/%	TYPES/%	FAMILIES
(1) 1000	5520/31.04	131/ 5.14	104
(2) 2000	1386/ 7.79	473/18.56	342
(3) 3000	438/ 2.46	180/ 7.06	144
off ranges	10438/58.70	1765/69.24	(not specified)
Total	17782	2549	590

Table 3. Tokens, types and word families in the text and activity sections from *Valid Choice 2*.

From the analysis of these data, several facts stand out:

Fact 1: the total number of lexical items in the activity section only doubles the one in the text section. The opportunities for repetition are low: each token introduced in the first section will be repeated only twice on average. This can be easily observed in Graphic 3, which includes the total number of words in the text and activity sections.

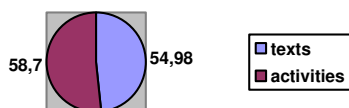
Total of words



Graphic 3. Total number of words in the text and activity sections from *Valid Choice 2*.

Fact 2: the amount of words outside the word ranges is too high in both sections: it takes 54.98% and 58.70% in the first and second section respectively, that is, more than half of the words are above the 3,000 more frequent words threshold (see Graphic 4).

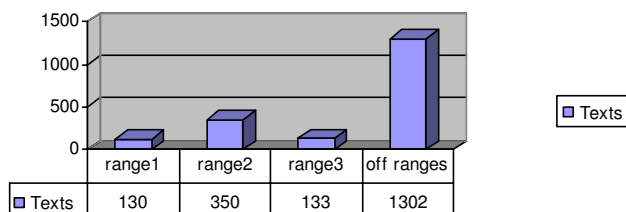
Words off ranges



Graphic 4. Words off ranges in the text and activity sections from *Valid Choice 2*.

Fact 3: As can be seen in Graphic 5 below, the section with texts reveals a strongly marked unbalance in the new words introduced in ranges 1, 2 and 3 versus the rest of words outside these ranges.

New types per range / off ranges in the text section



Graphic 5. New types per range and off ranges in the text section from *Valid Choice 2*.

Range 1 counts with only 130 types, range 2 includes 350 and range 3 only 133; 1,302 types fall outside these ranges (67.99%). Figures are similar in percentage for the activity section, reinforcing the ‘functional’ unbalance between both sections: the ideal proportion

would ask for a significantly higher number of types in the activity section. Higher *frequency* of occurrence favours acquisition because it grants more opportunities for *repetition* and hence for proceduralisation. However, tokens in the activity section total only 17,782 words, just 2.2 times more than in the text section (7,984). The opportunities for repetition are very poor indeed.

The conclusion is necessarily negative regarding the lexical distribution in each one of the sections. The opportunities for repetitive practice in the activity section are very low and this fact distorts the functional expectations of the text and activity sections (introducing new material and practising respectively). The textbook does not offer teachers and students the expected and necessary opportunities for automatising vocabulary. Moreover, (i) as indicated in section III.1., the amount of lexical items introduced exceeds by far the potential of learners for vocabulary acquisition on the one hand and the *rate* of vocabulary acquisition on the other, (ii) the lexical items introduced do not keep in line with the frequency lists; too many of them (1,302 out of a total of 1,915 in the text section) are not included in the three most frequent word ranges. This means that students will be primed to learn words of poor potential for communication in the level prescribed (B1).

VI.3. Explicit vs. incidental vocabulary learning activities

Vocabulary knowledge is necessary for language fluency (Anderson & Freebody, 1981; Goulden, Nation & Read, 1990; Laufer, 1998; Laufer & Nation, 2001; Read, 2000). As explained in section III, explicit and incidental learning activities are both important for vocabulary acquisition. *Explicit learning* is important because it attracts the attention of students and so it triggers the transfer of data from short-term to long-term memory; *incidental learning* is also relevant because it favours lexical proceduralisation, more slowly indeed but adding the advantage of contextualisation and more realistic communicative contexts. As for lexical acquisition, a textbook may therefore offer explicit or incidental opportunities depending on the kind of activities included. Explicit vocabulary learning will be the goal of activities in which the students' attention is directly drawn on to specific words or phrases by means of various strategies such as the ones suggested by the following instructions:

Match the synonyms below.

Complete the sentences with a suitable adjective from the list above.

Incidental *vocabulary acquisition* will be triggered by activities in which students are involved in language use, like reading, writing, speaking or listening, or whenever they must engage in exercises centred on the reception, interpretation, reshaping and transmission of meaning, as shown in the following instructions:

Skim the text and find out what Shakira's greatest challenge was.

Are the sentences below true or false? Find evidence in the text to support your answers.

A careful analysis of the activities in *Valid Choice 2* offers the following results (Table 4):

Unit	Activities favouring explicit vocabulary learning	Activities favouring incidental vocabulary acquisition	Other activities	Total of activities per unit
1	10	7	12	29
2	12	8	11	31
3	14	9	8	31
4	11	5	15	31
5	10	6	14	30
6	11	7	13	31
TOTAL	68	42	73	183

Table 4. Vocabulary and other activities in *Valid Choice 2*.

Graphic 6 below illustrates those figures:

Types of activities



Graphic 6. Number of activities per type of learning in *Valid Choice 2*

From a global point of view, the amount of explicit vocabulary learning activities is reasonably high, since it reaches 1/3 of all the activities in the book. Activities that favour incidental vocabulary acquisition are also high: they cover 23% of all the activities. It must be added to this the fact that the proportion of explicit and incidental vocabulary activities is homogeneously distributed all along the units. We must therefore conclude that from the point of view of the amount of exercises devoted to *vocabulary learning*, *Valid Choice 2* is on the right track to reach the expected goals in the field of lexical acquisition.

VI.4. Semantic fields covered by the textbook vocabulary

It has already been mentioned that there is a clear unbalance between the words included within the most frequent 3,000 thousand lexical items and the rest. This is the case as well among the most frequent words in the textbook. A comparison of the first 50 words in a

corpus of English (*LACELL* corpus⁴, a 20 million-word corpus) and *Valid Choice 2* is indicative of the unbalance: only 24 are common to both lists, that is, 1/3 of the total. The rest (2/3) in the textbook do not adjust to the frequency rank in the corpus (as shown in Table 5).

<i>Valid Choice 2</i> (50 first most common words)	Lacell corpus (20 million words)	Shared words between <i>Valid Choice 2</i> and <i>Lacell Corpus</i> (22)
people - use - film - students - school - find - answer - look - time - write - correct - more - were - paragraph - word - page - person - complete - choose - films - sentence - following - form - make - English - using - friends - get - go - listen - read - see - help - new - say - think - got - our - expressions - like - most - too - part - unit - information - must - true - know - language	time - people - like - new - think - know - get - see - way - work - right - go - years - - make - good - year - going - got - say - take - used - day - use - come - little - world - must - want - life - need - long - home - put - part - things - might - man - look - course - - house - great - old - women - children - number - government - different - give - place - mean -	<i>again - film - get - go - got - help - know - like - look - make - must - new - people - right - say - see - sentence - think - time - use - want - year</i>

Table 5. A comparison between the 50 most common words in *Valid Choice 2* and *Lacell Corpus*.

There are convincing reasons for the disparity in the rank of some words found in the general use of language and in the textbook (*students, school, correct...* in *Valid Choice 2*), but not so much for others (*work, take, day, good...*). Some of the most frequent words in *Valid Choice 2* do not fit the general scale of frequency in general English (*suitable, step, complete, form...*).

The *frequency* of specific words vs. other possible words is no doubt connected to the topics dealt with in the texts presented. Textbooks are subject to important constraints, since the topics you may come across will be limited by the reduced amount of presentation texts you can include in each lesson (perhaps two or three per lesson). This will affect the resulting frequency list in the textbook as a whole. The first most frequent 50 words in *Valid Choice 2* reveal the importance of the school setting and the emphasis on grammatical terminology:

words, people, students, school, sentences, paragraph, word, page, sentence, friends, expressions, unit, language, verb, questions, grammar, name, work, vocabulary, internet, task, exam, summary, connectors, events, letter, appendix, ...

⁴ The *Lacell Corpus* is a balanced 20 million-word English corpus compiled by the *LACELL* Research Group (E020-02) at the University of Murcia, Spain.

Such a list does not match the frequency rank of a general frequency list of English. Still, the school environment makes a frequent use of them and therefore vocabulary constraints of this kind must be taken into account in an overall evaluation of the vocabulary needed in the classroom context.

VII. CONCLUSION

The use of corpora in language teaching has been encouraged by many authors, both theoreticians and language teaching practitioners (Aston, 2000; Granger, 2003; Johns, 1994; Johns & King, 1991; Renouf, 1997; Tribble & Jones, 1989). However, in this case study, the comparison of the results shown above -from a textbook for teaching English as a foreign language- with some basic data derived from corpus-based research do not allow for an optimistic conclusion. The material analysed here does not seem to comply with some fundamental principles governing vocabulary acquisition.

Two points are to be stressed in this respect. Firstly, the textbook should have taken into account the most frequent words recorded in frequency lists based on English corpora. The words offered and presented to the students in this coursebook as a goal to be reached in the field of vocabulary acquisition are not in line with the *frequency* list of general English. Put it another way, there are striking mismatches between the words selected in *Valid Choice 2* and the most frequent words from English frequency lists. Secondly, *rehearsal and repetition* are necessary for consolidating vocabulary acquisition, which is a particularly relevant principle widely acknowledged in psycholinguistics, neurolinguistics and the teaching tradition. This means that the lexical items the students should learn are to be often found in the textbook, especially in the practice activities designed for vocabulary learning. We know that textbooks are limited in size and they cannot offer all the possible communicative situations the students may find in real life. Accordingly, the amount of vocabulary included in the manuals is necessarily lower than what should be ideally required. Still, these limitations applicable in specific areas or communicative situations should not severely affect the overall selection of the words included. But, as recently mentioned, *Valid Choice 2* reveals a clear unbalance in all the aspects of vocabulary selection if this is compared to the expected frequency list of general English.

Thus, the conclusion is that *Valid Choice 2* does not seem to take into account some of the most basic issues affecting *vocabulary acquisition*, both from the point of view of which words should be learnt first and the conditions which govern vocabulary acquisition. On the one hand, corpora are the best source to define the words more efficient in linguistic communication, but results based on corpora do not seem to have been considered by the authors of *Valid Choice 2*. On the other hand, research and data on the cognitive processes underlying knowledge and language acquisition ask for frequent rehearsal and repetitive practices in order to consolidate learning.

REFERENCES

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, vol. 24:4, 425-438.
- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (2005). *Cognitive Psychology and its Implications*. (6th ed.). New York: Worth Publishers.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*. Newark, DE: International Reading Association, pp. 77-117.
- Aston, G. (2000). *Learning English with the British National Corpus*. In M. P. Battaner, C. López, C. (Eds.), *VI Jornada de Corpus Lingüística. Corpus lingüística i ensenyament de llengües*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 25-40.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. Spende & J. Spence (Eds.), *The Psychology and Learning of Motivation 2*. New York: Academic Press, pp. 89-195.
- Bhans, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, vol. 21:1, 101-114.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Available on http://www.coe.int/T/DG4/Linguistic/CADRE_EN.asp
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19:2, 195-211.
- Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12:1, 25-53.
- Goulden, R., Nation, P. & Read, J. (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics*, vol. 11, 341-363.
- Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37:3, 538-545.
- Howatt, A.P.R. with Widdowson, H. G. (2004). *A History of English Language Teaching*, Oxford: Oxford University Press.
- Ito, T. (1995). *The Acquisition of Basic Vocabulary by College-Age Japanese Students in an Intensive EAP Program*. Paper given at the *Annual Meeting of the Japan Association of College English Teachers*, Tokyo, September.
- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Boston: The MIT Press.
- Johns, T. & King, P. (Eds.) (1991). *Classroom concordancing*. Birmingham: Centre for English Language Studies, University of Birmingham.
- Johns, T. (1994). From printout to handout. Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives from Pedagogical Grammar*. Cambridge: Cambridge University Press, pp. 293-313.
- Kapur, S., Craik, F. I. M., Tulving, E., Wilson, A. A. & Brown, G. M. 1994. Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. In *Proceedings of the National Academy of Sciences, USA*, 91, 2008-2011.
- Kelly, L. G. (1969). *25 Centuries of Language Teaching*. NY: Newbury House Publishers.
- Kucera, H. & Francis, W. N. (1967). *The Brown University Standard Corpus of Present-Day American English* (Brown Corpus). Providence: Brown University.
- Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of Task-Induced Involvement. *Applied Linguistics*, vol. 22:1, 1-26.

- Laufer, B. & Nation, P. (2001). Passive Vocabulary Size and Speed of Meaning Recognition. *EUROSLA Yearbook 1*, 7-28.
- Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language: Same or Different? *Applied Linguistics*, vol. 19:2, 225- 271.
- Lawrence, J. & Williams, A. (2006). *Valid Choice 2*. Burlington Books.
- Lewis, M. (1993). *The Lexical Approach*. Hove: LTP.
- McCarthy, M. (1998). *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Meara, P. & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied Linguistics in Society*. CILT, London, pp. 80-87.
- Nagy, W. E. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 64-83.
- 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). Mahwah, NJ: Erlbaum, pp. 269-284.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How Large a Vocabulary is Needed for Reading and Listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, vol. 63:1, 59-82.
- Nation, I. S. P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 6-19.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Renouf, A. (1997). Teaching Corpus Linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles, Gerry (Eds.), *Teaching and Language Corpora*. London; New York : Longman, pp. 255-266.
- Rundus, D. J. (1971). Analysis of rehearsal processes in free recall. *Experimental Psychology*, 88, 63-77.
- Sánchez, A. & Criado, R. (in press). Cognitive underpinnings of repetitive practice in the learning of EFL. *Proceedings of the 27th AESLA International Conference*.
- Sánchez, A. (1997). *Los métodos en la enseñanza de idiomas. Evolución histórica y análisis didáctico*, Madrid: SGEL, S.A.
- Sánchez, A. (2009). *La enseñanza de idiomas en los últimos cien años. Métodos y enfoques*. Madrid: SGEL, S. A.
- Sánchez, A. (2007). ‘Y realmente, ¿qué significa conocer o saber una palabra?, in Martínez Cabeza, M., Neil McLaren & Luis Quereda. Eds. *Estudios en honor de Rafael Fente Gómez*. Granada: Universidad de Granada. pp. 179-195.
- Schmitt, N. & McCarthy, M. Eds. (1997). *Vocabulary: Description, Acquisition, Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N. (1998). Quantifying word associations’ responses: What is native like? *System*, 26:3, 389-401.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic Sequences*. Amsterdam: John Benjamins, pp. 127-152.
- Schonell, F. J., Meddleton, I. G. & Shaw, B. A. (1956). *A Study of the Oral Vocabulary of Adults*. Brisbane: University of Queensland Press.

- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- 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.
- Tribble, C. & Jones, G. (1989). *Concordances in the Classroom: A Resource Book for Teachers*. London: Longman.
- Ullman, M. T. (2004). Contributions of Memory Circuits to Language: the Declarative / Procedural Model. *Cognition*, vol. 92, 231-270.