Frequency and Functionality: Two Keys for L2 Coursebooks

GEMA ALCARAZ

LACELL RESEARCH GROUP – UNIVERSIDAD DE MURCIA

ABSTRACT
There is no doubt of the importance of vocabulary in EFL coursebooks. Yet, criteria for the selection of target words remain unclear. Frequency and functionality have been proposed as the two underpinning for selection. However, they are usually incompatible so that materials designers must opt for one or the other. The question is which one should be chosen. Despite the fact that recent corpora studies remark the importance of frequency, it seems that functionality still dominates coursebook vocabulary selection.

KEYWORDS: Corpus, coursebook, frequency, functionality, second language, vocabulary

I. INTRODUCTION
In one way or another, vocabulary has always been present in L2 teaching methodology from the traditional Grammar-Translation to the Communicative Method (Sánchez 1997). In fact, it has been gaining positions since the mid-seventies, so that we can state that vocabulary is nowadays one of the cornerstones in Second Language learning.

1 Address for correspondence: Gema Alcaraz, Departamento de Filología Inglesa, Campus de La Merced. Universidad de Murcia. 30071 Murcia, Spain. e-mail: gemam@um.es

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved. IJES, Special Issue, 2009, pp. 61-72
Materials designers are cognisant of the importance of vocabulary in learning a second language. That is why they lend considerable weight to vocabulary in their textbooks. Normally, vocabulary is presented in different contexts such as dialogues, narrations, songs or other text types. Nonetheless, there are also explicit allusions to certain groups of words which constitute the target vocabulary of the didactic units. Those words are dealt with throughout the unit by means of different activities.

Selecting an appropriate coursebook is one of the challenges both teachers and institutions have to face every year. However, selection is not easy at all. The market is full of offers, some better than others, which complicate the process. Yet, it is the task of teachers and institutions to choose materials in tune of both the learner’s needs and the experts’ advice. What is more, it is the designers’ task to live up to the expectations of authorities, teachers and learners.

Despite the recognized importance of both vocabulary and coursebooks in L2 acquisition, the reasons behind the selection of the latter are not always clear. Thus, the Spanish documents for the regularization of Foreign Language teaching, that is, the so called ‘ Diseño Curricular Base’ and the Education Law (2006), do not provide any specific information about the words the learner should know at the end of his/her Primary and Secondary Education.

As a general rule, these official documents make vague and inconclusive allusions to vocabulary such as “ se evalúa la capacidad de utilizar expresiones y frases para hablar en términos sencillos sobre su familia y otras personas, el tiempo atmosférico, la ropa, libros, juegos y sobre contenidos de las diferentes áreas” (Minimum contents curriculum, 2006: 43095). The present example illustrates the document failure to specify how many and which words are exactly those that students must know related to family, weather or clothes, among other semantic fields. It is here where frequency and functionality should have a role.

We should not expect that an average L2 learner can equal a native speaker, at least in terms of vocabulary size; nor is it necessary for him/her to do so. Most vocabulary researchers agree that around 2000 word families are enough for basic communication (Nation 1990; 2001; Schmitt 2000; Nation and Gu 2007). Taking into account that the English language contains around 54000 word families (Webster 1963), those 2000 word families do not even reach 4% the whole set of English words, which begs the following question: how is it possible that L2 learners are expected to communicate with only 4% of a language, if only in a basic fashion?

The answer is frequency, that is, the number of times a word occurs in general discourse. In fact, knowing the 1000 most frequent words leads to understand up to 78% of words in a non-specialized text. This percentage increases up to 84% with the addition of the second 1000 most frequent words. In this sense, the benefit is much higher than the cost, for with a relatively small number of words (around 4% of the total set) learners can
communicate in a simple but acceptable way. Hence, a relatively small amount of vocabulary can provide the learner with considerably communicative reward.

Nevertheless, frequency presents some limitations which concern the definition of word and word knowledge. In fact, there is still disagreement about what the term ‘word’ means. Traditionally, the word has been defined in orthographic, semantic or formal terms. From an orthographic standpoint, the word is considered a sequence of letters bounded by spaces or punctuation marks (Cruse 1986). However, this definition does not contemplate oral discourse and disregard possible spelling variations. Regarding semantic criteria, the word is considered an indivisible unit of thought (Jackson and Amvela 2001). The problem is disagreement on the unclear definition of ‘unit of thought’. Formal criteria – by which the word is considered “the minimal free form” (Bloomfield 1933: 178) – are not much better. This definition implies mobility and internal stability. Yet, though most words are internally stable, some of them occupy a fixed place in the sentence and cannot be moved, as it is the case of determiners.

As for word knowledge, linguists also seem to differ in their perspectives. Thus, Richards (1967) or Nation (1990; 2001) understand word knowledge as a taxonomy where knowing a word involves different types of sub-knowledges referring to form, meaning and use. Others (Palmberg 1987; Melka 1997) conceive word knowledge as a continuum with different degrees of familiarity. The reductionist view nurtures both from the taxonomy and the continuum proposals, whereas authors such as Sánchez (2007) focus on the semantic content as the main parameter in their word knowledge analyses. Hence, Sánchez (ibid.) represents word knowledge as set of aspects which are organized into constellations. Put another way, word features bear a relationship which can be relatively more or less closer. This is what Sánchez called ‘lexical constellations’.

Another important drawback in frequency is lack of psychological validity. In other words, vocabulary introduced by frequency lists is organized alphabetically, so that no semantic or psychological association exists between the different words. If found, it is just fruit of coincidence. This fact does not help vocabulary acquisition at all, and, what is more, it may somehow hinder the learning process. By contrast, it has been shown that learning words which are semantically related – organized into semantic groups – not only favours but also reinforces acquisition (Aitchison 1996; Haycraft 1993; Meara 1983; Craik and Lockhart 1972).

We have observed that frequency alone is not enough for vocabulary selection. Functionality is also necessary. Nation and Gu state that “in terms of usefulness, all words are not created equal” (Nation and Gu 2007: 20). ‘Usefulness’ here refers to communicative usefulness. In fact, functionality involves the learners’ communicative needs. It relies on questions such as ‘which words must students know in order to talk about people, things and events in the place where they study and live?”; ‘which words must the student know in order to respond to routine directions and commands?”; ‘which words are required for
certain classroom experiences?’; ‘which words are needed in connection with the students’ particular academic interests?’ (Allen 1983: 108).

These four questions seem to cover the most important learners’ needs. Thus, the first one would allow the learner to put the new language to use. The second and the third one help learners to manage classroom situations. Finally, the fourth question considers the profile diversification which is found nowadays in relation to the learning of a second language. In other words, the vocabulary that is needed by a lawyer is partly different from economists’ vocabulary or coaches’ vocabulary. This is what we call ‘English for specific purposes’.

Therefore, it seems that adequate vocabulary selection for L2 materials should rely both on frequency and functionality. Yet, these two factors do not always match, that is, there some words which should be known from a functional point of view, but they do not appear among the most frequent ones in general discourse. In this sense, many words as demanded by students’ communicative needs do not belong to the previously mentioned threshold of 2000 word families.

The debate is on the table: whether materials designers should opt for one or the other is something that warrants further research. On one hand, there is no doubt that frequency and the role of corpora are gaining currency in Second Language teaching, at least in theory. On the other hand, it seems that functionality has pride of place in coursebooks, given the predominance of communicative methods.

II. METHOD
II.1. Questions
The present study answers the following questions:
• Which are the frequency levels of vocabulary in the coursebook?
• Which are the frequency levels of vocabulary in each didactic unit of the coursebook?
• To what an extent is the 2000 most frequent words in General English represented in the coursebook?

II.2 Instruments and Procedure
II.2.1. Coursebook
The coursebook chosen for the analysis is entitled Bugs 3. It aims at young learners in their third year of Primary Education. There are two main reasons why this textbook was selected as the object of our analysis. First, it is widely used at numerous schools in the Region of Murcia. It is very well known among teachers, students and parents due to its positive effect on the children’s learning process. The present course is built upon eight main units, plus an introduction. The set is completed with two special sections: Christmas and Easter. The units are delivered as follows: the first term comprehends the introduction and units 1, 2 and 3.
The second part of the academic year develops units 4 from 6, whereas units 7 and 8 are displayed in the last term.

*Bugs* 3 is claimed to fall within the Communicative Method. Vocabulary is presented mostly, though not exclusively, in context by means of dialogues, stories or songs. Those are complemented with different activities where vocabulary learning is one of the aims. Units are built upon the topics of animals, clothes, the body, food, sports, routines and holidays.

**II.2.2. Range**

The coursebook analysis was carried out by means of the RANGE programme. Among other things, this computer program allows you to classify the vocabulary of any text into frequency categories. RANGE includes three different lists: list one and two (hereafter List 1 and List 2) contain the first and second most frequent words in General English, whereas list three (L3) presents words that are not found among the 2000 most frequent words but are usually found in upper secondary and university texts. The words that do not fit in any of the three lists appear in the “not found” category.

Furthermore, RANGE distinguishes among three different units of estimation: tokens, types and families. A token is defined as “*every word form in a spoken or written text*” (Nation, 2001: 7), so that each time a word occurs is counted. The number of tokens in a textbook may give us an idea about the amount of input the learners are exposed to in raw terms – which falls out of the purpose of the present study. A family consists of a headword, its inflected forms and its closely related derived forms. Adopting a family as counting unit presents the serious problem of deciding which forms should be included and which should not. What is more, we cast doubt on the idea that young learners know the word *bedroom* once they have learned *bed*. They might guess that *bedroom* is related to *bed* and even what it means, but we must distinguish between guessing ability and learning.

Regarding our study, we have adopted the type as counting unit. A type is any different form that appears in a text. It implies that words such as *walk* and *walked* are counted as two different items. At first sight, this may not be adequate, since a person that has learned the word *walk* possibly also knows *walked*, or at least can identify the latter with the former. However, we have to think that we deal with low levels and young learners. We can not assume that children knowing *walk* also may know *walked* or *walking*. Finally, though RANGE registers any kind of word, our analysis focuses exclusively on content words, that is, nouns, verbs, adjectives and adverbs.

**III. RESULTS**

**III.1. Vocabulary Frequency: the whole book**

After the coursebook was analysed by RANGE, we obtained four frequency lists: List 1 containing items from the 1000 most frequent words in General English, List 2 with that vocabulary pertaining to the second 1000 most frequent words, a third list (L3) with words
of lower frequency but commonly present in upper secondary education and also university texts; finally a set of words which do not belong to any of those categories, labelled the ‘not found words’ (hereafter NF). It is important to remember that our main aim is to observe the weight of the coursebook vocabulary within List 1 and List 2 against the rest of words outside those lists. Thus, for convenience’s sake, we have grouped words into three categories. We have maintained List 1 and List 2 and have put List 3 and NF words together into only one set, what we have called the ‘unusual words’ (hereafter UNW).

Figure 1 shows how the outcomes point out a clear predominance of List 1 with almost 56% (55.91%) words belonging to this frequency rank. This is followed by UNW words with 28.13% and finally by the least represented, List 2 with just 15.95% presence in the course. Focusing on target vocabulary (those words that are especially highlighted to be learned), up to 40% (39.8%) partake the UNW category, closely followed by List 1 with 33.4%. Again, List 2 is the least remarkable, containing the lowest number of target words: 59 items; put it another way, only 11.3% of target vocabulary belongs to List 2.

![Figure 1. Percentages of representation of frequency lists](image)

Regarding parts of speech, we find 227 nouns, 78 verbs, 54 adjectives and 17 adverbs. More than half the nouns, 72% verbs, almost 60% adjectives and the whole of adverbs are part of List 1. However, the proportion amongst all of them differs from the one that is usually given in authentic texts. Proportionally, authentic texts contain 3 nouns every 2 verbs and 1 adjective. In this case, the proportion is dramatically inferior: less than 1 verb and less than 1 adjective every 3 nouns. Thus, presence of nouns is remarkably over the prototypical distribution.

### III.2. Vocabulary Frequency: individual units

In addition to the overall analysis of the textbook, we should also afford insight into each didactic unit. We have omitted the especial sections of Introduction, Christmas and Easter, and just focused on the main course units from 1 to 8.
As with the results of the book as a whole, the predominance of List 1 is manifested in every unit. Its presence oscillates between almost 80% in unit 5 down to 53.21% in unit 3. By contrast, not always the weight of UNW word list is higher than List 2. If we organize the units according to the three terms of the academic year, we can see that the first term offers a higher presence of UNW words (from 17.15% in unit 1 to 31.2% in unit 3) than List 2 words (ranging from 10.10% in unit 2 to 15.6% in unit 3). Yet, tables turn in the second term. Despite the light advantage of UNW list in unit 4, regarding units 5 and 6 List 2 words are clearly superior in number with 11.76% and 12.62% versus 8.4% and 8.7% of UNW. Finally, the third term shows again the hegemony of UNW (20.87% and 15.03%) over List 2 (14.78% and 10.46%).

Figure 2 shows the behaviour of each word list along the eight units. The most regular path is represented by List 2, with soft upward and downward moves. Evolution of List 1 and UNW is opposite, with marked dips of the former in units 3 and 7, which coincide with the UNW peaks.

We can also analyse didactic units from a qualitative perspective. A qualitative analysis answers the third question of our study regarding representativeness in the textbook of the 2000 most frequent words in General English. Figure 3 presents the classification of target vocabulary in each unit according to frequency. Thus, unit 1 combines school materials with pets; unit two focuses on clothes and family, whereas unit 3 mostly deals with body parts and wild animals; units 4, 5 and 6 are devoted to food, sports and everyday routines, respectively; summer holidays and weather are seen in unit 7, and finally, unit eight presents a traditional tale, where royal characters, cultural public places and several physical, mental and social features are boarded.
<table>
<thead>
<tr>
<th>LIST 1</th>
<th>LIST 2</th>
<th>UNW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black, chair, dog, eye, fish, green, head, leg, rule, white</td>
<td>Bird, brown, desk, mouth, orange, wing, yellow</td>
<td>Ant, antennae, grasshopper, hamster, notebook, schoolbag, spider, tortoise</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Blue, brother, dad, dress, game, go, make, mum, party, red, shoes, sister, trainers, wear</td>
<td>Antie, coat, hat, sock</td>
<td>Cousin, jeans, costume, lantern, pumpkin, t-shirt, uncle, witch</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Big, body, eat, fish, great, live, long, sleep, small</td>
<td>Bird, ear, mouth, sea, tail, tooth</td>
<td>Tongue, crocodile, giraffe, gorilla, grasslands, jungle, leopard, lion, snake, tiger, whale, whisker, zebra</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Egg, fish, five, four, glass, milk, one, six, three, two</td>
<td>Bread, cheese, chicken, chip, fruit, juice, knife, plate, vegetables</td>
<td>Sausages, spoon, cereal, ice cream, macaroni, rice, salad</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Fifty, football, forty, jump, pass, play, run, score, thirty, throw, twenty, walk</td>
<td>Bike, goal, ride, swim</td>
<td>Tennis, basketball, rollerblade, skate, skateboard</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Bed, dinner, get dressed, go, lunch, school</td>
<td>Breakfast, lunch, shower, stretch</td>
<td>Shampoo, soap, splash, munch, zip</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Bear, cold, country, have, land, park, summer, sunny</td>
<td>Camera, cloudy, cream, hat, mountain, raining, snow</td>
<td>Beach, continent, lake, pole, seal, towel, fox, ocean, penguin, polar, reindeer, suitcase, sunglasses, swimming pool, swimsuit, trunks, whale</td>
</tr>
<tr>
<td>LIST 1</td>
<td>LIST 2</td>
<td>UNW</td>
</tr>
<tr>
<td>Beautiful, cat, English, father, give, help, king, man, marry, match, old, poor, speak, young</td>
<td>Boot, clever, coat, hat, queen, rich, river, theatre</td>
<td>Crown, museum, prince, princess, rabbit, donkey, handsome, zoo</td>
</tr>
</tbody>
</table>

Figure 3 Target vocabulary from units 1 to 8 in Bugs 3

All those semantic fields are closely related to the communicative environment of young learners. As the main aim of a textbook is to fulfil the students’ needs as List 2 users, the frequency criteria is placed in the background, and is taken over by functionality – which at the end of the day, represents the ultimate goal of linguistic communication. This functional character does not necessarily have to contrast with General English in all cases. As we can observe from the chart, a significant number of target words belong to the first 1000 most frequent items, and some of them – though in a more modest proportion- are among the second 1000 most frequent vocabulary.

Nonetheless, the proportion of the three lists varies from unit to unit, according to the semantic fields that predominate in each of them. Due to space limitations, we will comment...
on just four out of the eight units in the coursebook. We consider those four units most representative of children’s communicative context, and, at the same time, they are the ones that better represent that contrast between the functional and the frequent.

The selected units are unit 3 (wild animals), unit 4 (food), unit 5 (sports) and unit 7 (holidays). All target words in unit 4, except numbers one to six, are directly related to food. Thirteen of those twenty food words are not among the 1000 most frequent ones. Although words such as sausages, spoon or ice cream are classified as unusual vocabulary, they are highly functional in the learners’ communicative context. Regarding unit 5, all sports but football cannot be found within the most frequent words. Thus, basketball, tennis, skateboard or rollerblade - foci of that unit- belong to a low frequency rank in General English. A similar case is found in unit 3, with the whole set of wild animals - crocodile, giraffe, gorilla, leopard, lion, snake, tiger, whale, zebra – labelled as unusual. Finally, unit devoted to summer holidays also offers a great number of its target words from the UNW category. In fact, central terms to summer holidays such as beach, sunglasses, trunks or suitcase fall out of the most frequent General English, but are paramount to the semantic field of unit 7. Those examples show how the communicative context of children and therefore their linguistic functions do not usually coincide with those in General English corpora.

IV. DISCUSSION

Despite bourgeoning interest on frequency, there is a clear predominance of functionality regarding vocabulary selection in coursebooks. In fact, the present study is not an isolated case. Others such as Vassiliu (2001) or Donzelli (2007) also reached similar conclusions in their analyses. Both found general predominance of List 1 followed by almost one third of ‘unusual’ vocabulary. As it happens with our study, the second 1000 most frequent words were the least represented in textbooks. From a qualitative point of view, Vassiliu and Donzelli’s studies are also similar to the present work. Target vocabulary was also related to sports, food or wild animals. For instance, terms found in our analysis such as basketball, tennis or zebra were also present in Vassiliu and Donzelli.

Given all these outcomes, we can state that there is a kind of ‘coursebook convention’ (Bell and Gower 1998) by which designers adopt the learners’ communicative needs – and therefore functionality – as their basis. Herein lies the high percentage of ‘unusual words’ (UNW). If frequency were to be given more relevance in coursebooks, an increase of List 2 words would be necessary. In principle, there would be no reason not to do so. Spanish EFL students in their third year of Primary Education are supposed to have been learning English for six years, including their three years of kindergarten. Thus, it is expected that a considerable portion of List 1 vocabulary has been learned, or at least dealt with. In this sense, more presence of List 2 in the analysed coursebook would not be unjustified, and frequency would acquire more prominence.
However, if frequency became the designers’ main criterion, words such as *accuse*, *committee* or *faith* – which belong to the semantic fields of Law, politics or religion – should be targeted. The question is whether those words and similar vocabulary is normally handled by average L2 students, especially young learners. In fact, it is not only a question of functionality. Words like the ones above and many others from List 1 and List 2 go beyond young learners’ comprehension and practicality. In other words, this type of vocabulary does not fit students’ cognitive and social scope.

Hence, vocabulary needs to cater for the learners’ needs, and adapt to the communicative context of the target audience. In this sense, textbooks must be built upon familiar topics for the students according to their functionality. That is why EFL courses aim to cover part of what is called the ‘basic area’ of the English language. This basic area is closely related to the so called ‘visitor’s wing’, which tries to adapt to the requirements of the average learner. Thus, it is this ‘wing’ which is partly responsible for the predominance of functionality and, consequently, the considerable amount of unusual words in coursebooks.

Therefore, coursebook design must primarily take into consideration learners’ features and needs. These features and needs are closely related to the terms ‘teachability’ and ‘learnability’. The former can be defined as the easiness or difficulty to teach something, in this case vocabulary. In the same line, the latter refers to the more or less difficulty to learn vocabulary.

The reasons for the low or high degree of ‘teachability’ and ‘learnability’ may rely on the word itself, but also on the learner and even the teacher, for they constitute the three hallmarks in the process of L2 vocabulary acquisition. Thus, word ‘teachability’ or learnability may be modulated by the learner’s age and cognitive status, or even the teacher’s personal and material teaching resources. Moreover, the fact that there is a L1 correspondent of that word also helps acquisition. In this sense, the word *cat* is presumably more learnable by a native speaker of Spanish than the word *burka*, which is an inexistent concept. Not only must the student learn the form-meaning link but also the concept itself, which increases the learning burden (Lakoff 1987).

V. CONCLUSION
Despite the business that may be found behind, L2 materials design is a great responsibility and a noble professional track. In fact, the textbook has become one of the essential elements in the EFL classroom. Among coursebook contents, vocabulary is one of the areas which have acquired more relevance in the last decades. The two main criteria for its selection are frequency and functionality. Ideally, selection would rely on these two elements in similar proportion.

However, these two factors are usually incompatible, and one must unavoidably prime over the other. The question is which one of these two should. It seems that
functionality dominates vocabulary selection over frequency. Thus, frequency does not seem to cater for the learners’ communicative needs, despite the growing interest in corpora and its application to the EFL classroom. In other words, frequency has modest pedagogical usefulness. Therefore, it is functionality which dominates the fieldwork of EFL materials design.

REFERENCES


