Abstract
This paper aims to shed light on the type and nature of vocabulary included in the input given to Bachillerato 2nd Form students. We will analyse the vocabulary of a textbook used in a high school so as to see whether it includes or not the most frequent words listed in Nation’s frequency lists. This will allow us to know if our students are exposed to the most frequent and therefore most useful type of input. Teachers and researchers should question whether the type of vocabulary taught is the one that students need and whether our Bachillerato students reach what is theoretically considered to be a suitable level of vocabulary for a foreign student. The creation of a corpus using the textbook and its analysis using Lexical Frequency Profile developed by Laufer and Nation will allow us to draw conclusions on the type of language exposure that students are receiving.

Keywords.
Frequency, input, vocabulary, corpus.

I. Introduction.

Despite the fact that Spanish students spend at least eight years of their lives studying English as a compulsory subject within the Spanish school system, there exists a striking difference between native speakers and non-native speakers in terms of the amount of lexis they have acquired during this period. Most students will never realise that their skills in the foreign language are too poor to perform well in most communicative contexts. Only those who devote their lives to the study of the language or those who use English as a necessary tool in their working environment will become aware of the fact that in spite of studying so many years the knowledge acquired is not sufficient to enable them as efficient language users. The Spanish education system and the royal decrees that regulate education at this level establish certain objectives that students have to reach by the end of Bachillerato. These objectives include the critical understanding and interpretation of oral texts, written texts and visuals used in habitual communicative situations and by mass media. Students are also supposed to read different texts in the foreign language as a source of information, knowledge or enjoyment and leisure both comprehensively and at the same time in an autonomous way. Even university authorities prepare the university entrance exam taking for granted the fact that most students at this point of their academic experience should by this time have reached a level at which they are able to use the language autonomously. This means that teachers and textbooks should provide students with the necessary tools by which they are able to fulfil the established objectives. However, it seems that the majority of students do not reach such expected standards and that there is not a general concern on the nature of either the input provided within classrooms or the level actually achieved by students.
There are many research studies that have pointed out the importance of vocabulary for the development of all the linguistic skills. Meara and Glyn (1988: 80) claimed that “vocabulary knowledge is heavily implicated in all practical language skills, and that in general speakers with a large vocabulary perform better than speakers with a more limited vocabulary”. This is something that learners realise and leads them to complain about a lack of vocabulary that prevents them from understanding to the full both native and non-native speakers of English and from communicating and transmitting their feelings and ideas with precision. At this terminal stage of education teachers and researchers should question whether the type of vocabulary taught is the one that students need to reach such objectives and whether or not our Bachillerato students reach what is theoretically considered to be a suitable level of vocabulary for a foreign student. No matter whether they are good or bad students, in most cases students would find themselves unable to read a newspaper article without being aided at the end of their time in Bachillerato studies. There may be many different aspects to be taken into account as being responsible for this situation and vocabulary must definitely be one of them.

I.2. Aims

The main objective of this research study is to analyse the vocabulary included in the textbook so as to see whether it includes or not the most frequent words listed in Nation’s frequency lists. This will allow us to know if our students are exposed to the most frequent and therefore most useful type of input. This goal is related to this research question:

- Does the vocabulary input that students are exposed to include the most frequent words of English as listed in Nation’s frequency lists?

I.3. The role of Vocabulary in Foreign Language Learning.

There are many studies that reinforce the key role of vocabulary size on reading and writing skills, especially with native speakers. Laufer (1998: 256) when analysing the vocabulary of Israeli students, established that a good vocabulary size was important for reading comprehension and for fluency in speech. Goulden et al. (1990: 342) claimed that “measures of vocabulary size- particularly size of academic vocabulary- are important indicators of the ability of second language learners to achieve academic success”. Anderson and Freebody (1981) cited in Read (2000: 76) found out that there was a high correlation between a good performance in a test of vocabulary and reading comprehension. However, if we follow Laufer and Nation (2001: 7) “for vocabulary knowledge to play a significant role in everyday language use, words knowledge is not enough. This knowledge must be fluently available, if fluent conversation or reading is expected.” Thus, when dealing with foreign language vocabulary learning, many different facets of the construct must be taken into account.

Reflection on the language shows us that even if we knew all the grammatical rules of English we would never be able to use them without words. Vocabulary is the basic tool for shaping and transmitting meaning. A learner may learn that the first person pronoun in English is ‘I’, but if he does not know the denotation of this function word, he will never be capable of using it in real language usage. On the contrary, a learner who wants to buy something in a shop will probably be able to do it even if he only knows some words, such as bread, even if he is unable to use it, to build a sentence or even if he ignores the pragmatical issues that are involved in a commercial exchange. The reason why we teach English is to
provide our students with the resources to understand and transmit ideas in another language. Thus, vocabulary should have a central role in language learning. Years ago vocabulary was learnt by heart using long decontextualised lists and it was taught explicitly. With the arrival of communicative teaching methods vocabulary started to be taught implicitly and contextualised. No matter how vocabulary is taught it always deserves a central role in the teaching and learning practice.

The fact that vocabulary has a central role in the learning of a foreign language makes it necessary to select meticulously the words that are to be taught and included in the input that our students are going to be exposed to. It needs to be taken into account that differently to native speakers foreign language learners only have a limited exposure to the language and this language should include the most frequent forms. Studying the most frequent words allows our students to have more possibilities to use the vocabulary they have learnt in real life situations and gives them higher text coverage. According to Carrol, Davies and Richman (1971) cited in Read (2000: 158) “2000 words represented 81% of the total”. This means that students knowing the most frequent 2000 words would be able to understand 79.7% of any text. We can find these figures in Nation and Waring (1977) cited in Schmitt and McCarthy (1997: 9) as can be seen in the following chart.

<table>
<thead>
<tr>
<th>Vocabulary Size</th>
<th>Text Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>15851</td>
<td>97.8%</td>
</tr>
<tr>
<td>6,000</td>
<td>89.9%</td>
</tr>
<tr>
<td>5,000</td>
<td>88.7%</td>
</tr>
<tr>
<td>4,000</td>
<td>86.8%</td>
</tr>
<tr>
<td>3,000</td>
<td>84.0%</td>
</tr>
<tr>
<td>2,000</td>
<td>79.7%</td>
</tr>
<tr>
<td>1,000</td>
<td>72.0%</td>
</tr>
</tbody>
</table>

The late development of corpus and frequency studies has reinforced the idea that there are certain words that are more useful than others. Nation (2002: 11-12) offers a categorisation of words based on frequency and communicative dimension that distinguishes among high-frequency, academic, technical and low-frequency words. The main purpose of Bachillerato is to provide students with a general knowledge of the language and thus allow them to communicate and understand language in real life situations. Therefore, it makes sense that high-frequency words should be preferred while more specialised ones are to be postponed, leaving specialisation for those who need it at a university level. The problem is to determine exactly how many of these high-frequency words a learner needs to know. Laufer (1998) expected Israeli high school students to have learned between 3,500 and 4,000 word families, while a native speaker of the same age was supposed to know between 18,000 and 20,000 word families. In the same study it is established that a threshold level for minimal comprehension is to be found in 3,000 word families and one of 5,000 word families is needed to be capable of reading for pleasure. Therefore, teachers in Spain should expect their students to reach at least the 3,000 threshold level. While in native speakers their vocabulary size will highly depend on the level of education and the amount of reading in a foreign language context the level of vocabulary reached will depend on the type and amount of input...
that the learner is exposed to. Nation (2002: 97) affirms “instruction should be directed towards the high-frequency words of the language” and “the benefits of knowing high-frequency vocabulary compensate for the time and effort required for direct vocabulary instruction”.

I.4. Other research studies

There isn’t much research done on the type of vocabulary included in textbooks. Therefore, our revision of research studies will include not only those which have directly dealt with this issue but others which have been important for elaboration of this research study.

We must necessarily begin by mentioning a very similar study; the one elaborated by Meara, Lightbown and Halter (1997) entitled Classrooms as Lexical Environments. In this paper the researchers analysed a corpus based on the classroom talk of different teachers of English as a foreign language. One of the basic assumptions behind this study is that “a large number of unusual words would be indicative of a rich lexical environment; whereas the absence or extreme rarity of such words would indicate that the classroom vocabulary was poor” Meara, Lightbown and Halter (1997: 28). Their interest on the language produced by teachers within the classroom came from the idea that most of the words that students learn come from the comprehensible input they are exposed to and from the assumption that “a word which occurs only once or twice is not likely to be learned”, Meara, Lightbown and Halter (1997: 35) Despite the fact that other studies had analysed the input found on textbooks and on teachers’ talk, none of them analysed it in terms of frequency. Using short pieces of classroom talk for the creation of a corpus from which frequency lists were created. However, the amount of teacher talk analysed was quite limited as it consisted of ten samples of just over 30 minutes of class each time. For the elaboration of the frequency counts they counted lemmas and they acknowledged that this had the effect of reducing ordinary words and increasing the amount of unusual ones. The results showed at first sight that classrooms were not rich lexical environments. However, taking into account the limited sample studies and that the samples belonged to intensive classes this perception cannot be generalised as we are unaware of other types of input that students may be exposed to.

One way of interpreting research done in this area of study is to divide it into two main groups. On the one hand we find scholars that are mainly concerned with the development and validation of tools to measure vocabulary size. On the other, we find those who use the developed tools to measure the vocabulary of particular groups of students. We must start by commenting on those concerned with the creation and validation of different tests. Schmitt et al. (2001) used a range of analysing techniques to validate two new versions of the Vocabulary Levels Test, which has been widely used as a diagnosis test in its original form. Their conclusion was that native speakers performed well in these new versions, that individual items worked well and that the test supplies highly scalable vocabulary. Furthermore, their factor analysis proved that the test is unidimensional and their personal interviews indicated that the test reflects testees’ lexical knowledge. Goulden et al (1990) also discuss all the aspects that researchers need to take into account when developing tools to measure receptive vocabulary. They are basically concerned with native speakers, but their conclusions can be applied to a second language context. In this article they conclude that there are less than 58,000 base words to draw on for a vocabulary test, that the average vocabulary size of an educated native speaker is of 17,000 words and that as native speakers learn vocabulary at a relatively slow rate, direct vocabulary teaching is a feasible position. Nation and Waring (1977) in their article ‘Vocabulary size, text coverage and word lists’ also
deal with the theoretical foundations needed to do research on the area answering questions such as how many words there are in English or how many words a language user needs to do basic things in the foreign language. Another interesting piece of work to be mentioned here is the one developed by J. Charles Alderson (2007) in which the researcher studies the frequency of the words included in the Vocabulary Size Placement Test developed by Meara and Buxton (1987/1988). This is a yes/no vocabulary test and test takers are supposed to distinguish real from pseudo words. In order to study the frequency of the words included in the original test, some educated native speakers were given the task of making a judgement on the frequency they thought that each word belonged to. The results proved that there was a limited correlation between the frequencies obtained by corpus studies and those proposed by the judges using different methods of analysis. The results showed that “tests that aim to estimate vocabulary size on the basis of word frequency data are unlikely to be enhanced by incorporating subjective judgements of word frequency” Alderson (2007: 404). This corroborated once more the leading role of corpus linguistics in the study and development of vocabulary tests.

We mentioned previously that there were a group of researchers that applied different tests to measure the vocabulary size of varied groups of students. Pérez (2005) used Schmitt et al. version 1 of the Vocabulary Levels Test to measure the vocabulary size of students of 5th year English Philology in the University of Granada. She found out that her students were not “completely at ease at the 5,000-word level and scored poorly at the 10,000-word level”. Therefore, it is necessary to adopt effective remedial measures to extend students’ vocabularies. She also proves the influence of cognates on receptive vocabulary in the academic level. Hazenberg and Hulstijn (1996) carried out a similar measure of vocabulary to students with a native language other than Dutch, so as to see how many words of the Dutch language they needed to know to be able to perform well during the 1st university year. Their findings pointed towards a required knowledge of ten thousand base words for those entering university. They also argue that the selection of which words are to be learned must be done taking into account both frequency counts and students’ needs. Cameron (2002) presented another measurement of vocabulary but this time it was carried out in a context of English as an additional language. Therefore, the testees were students living in England whose mother tongue was not English. In the author’s words “the most important outcome of this study is the finding that the receptive vocabulary of EAL students who have been educated through English for 10 years have gaps in the most frequent words and serious problems at the 5K level.” This is something we should take into account when establishing the levels to be reached for students of English as a foreign language, as their contact with the language is less extensive. Laufer and Nation (2001) developed a measure of vocabulary size related also to speed of recognition. The aim of the study was to know how vocabulary size is related to speed of meaning recognition. They found out that they were moderately correlated, except for the less frequent vocabulary. They conclude that “increase in speed lags behind increase in size”, Laufer and Nation (2001: 7). They also found out that students respond more slowly to those words that are infrequent and that native and non-native speakers showed different patterns of response. The last study that included in its structure a measure of vocabulary size that we would like to mention here is one developed by Batia Laufer. In her article “The Development of Passive and Active Vocabulary in a Second Language: Same or Different?” she not only measured passive vocabulary size but also the active one. The sample includes graduates of Israeli high schools. The tool used to measure passive vocabulary size is once more Nation’s Levels Test and it showed that students’ vocabulary increased by 1600 word families in one year of school instruction in a school where students study English 5 hours a week. She concluded that “a large number of words can indeed be learnt even when the
learning context is not natural” Laufer (1998: 265). Active vocabulary did not evolve so much in comparison to passive vocabulary.

As it was mentioned before not much research has been done on the type and nature of the vocabulary to which students are exposed but this is certainly a very interesting area of research that should be paid attention to.

II. Methodology

II.1. Corpus and Frequency Lists.

A corpus in linguistics and lexicography is a body of texts, utterances, or other specimens considered more or less representative of a language, usually stored as an electronic database. Currently, computer corpora may store many millions of running words, whose features can be analyzed by means of tagging (the addition of identifying and classifying tags to words and other formations) and the use of concordancing programs. Corpus linguistics studies data in any such corpus …’ T.MeA. (The Oxford Companion to the English Language, 1992). Sinclair defines a corpus as “a collection of naturally occurring language text, chosen to characterize a state or variety of a language.” (Corpus Concordance, Collocation, OUP, 1991). A corpus must be planned and based on a sound basis. Therefore, it must be elaborated bearing clearly in mind the reason why it was created. Furthermore, a corpus must be representative. This means that it must give its users a picture of how language is used within a particular variety of language. Depending on the goals behind the corpus elaboration different texts can be included or taken into account. A well-built corpus is very useful for a variety of purposes. It can be used within a teaching and learning context as a tool for the analysis and study of a foreign language. Within the field of research it can be a useful means for the analysis of the language produced by students or the input students are exposed to.

This section will be devoted to the explanation of the criteria that lies behind the elaboration of the corpus used in this research project. The basic assumption is similar to the one that guided Meara, Lightbown and Halter (1997) for the elaboration of their corpus of teacher talk; namely the idea that vocabulary is learned from the input we are exposed to. Obviously it would be extremely difficult to analyse all the input in English that students are in contact with, basically because we cannot know if contact with the language exists outside the classroom and the forms that this contact could take. In this particular case we focused on the textbook used in class and we assume that most of the input given by the teachers will be very often related to the one presented in the book. Besides, the sample of students is formed by normal Spanish high school students who rarely have contact with the language outside the classroom. Therefore, it can be said that the text book is one of the major sources of input that students have contact with and as a result it is the major source of vocabulary learning. Obviously, including a word in a textbook does not mean that students will learn it automatically but its recurrent presence gives students a chance to get to know it. Most often more than one occurrence is necessary for a word to be learned; the more a word is part of the input given in class the higher the possibilities of noticing and learning it.

Elaborating a corpus means making a decision on many different aspects, particularly on how the text is going to be processed. Bearing in mind previous considerations some elements were eliminated from the corpus. Firstly, proper nouns and function words were deleted. Following Nation (2001) these are the function words that were erased from the original text.
### CHART 2:
*Function Words. Nation (2001: 430)*

<table>
<thead>
<tr>
<th>ADVERBIAL PARTICLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Again, ago, almost, already, also, always, anywhere, back, else, even, ever, everywhere, far, hence, here, hither, how, however, near, nearby, nearly, never, not, now, nowhere, often, only, quite, rather, sometimes, somewhere, soon, still, then, thence, there, therefore, thither, thus, today, tomorrow, too, underneath, very, when, whence, where, whither, why, yes, yesterday, yet.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AUXILIARY VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Am, are, aren’t, be, been, being, can, can’t, could, couldn’t, do, didn’t, do does, doesn’t, doing, done, don’t, get, gets, getting, got, had, hadn’t, has, hasn’t, have, haven’t, having, he’d, he’ll, he’s, I’d, I’ll, I’m, is, I’ve, isn’t, it’s, may, might, must, mustn’t, ought, oughtn’t, shall, shan’t, she’d, she’ll, she’s, should, shouldn’t, that’s, they’d, they’ll, they’re, was, wasn’t, we’d, we’ll, were, we’re, weren’t, we’ve, will, won’t, would, wouldn’t, you’d, you’ll, you’re, you’ve.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PREPOSITIONS/CONJUNCTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>About, above, after, along, although, among, and, around, as, at, before, below, beneath, beside, between, beyond, but, by, down, during, except, for, from, if, in, into, near, nor, of, off, on, or, out, over, round, since, so, than, that, though, through, till, to, towards, under, unless, until, up, whereas, while, with, within, without.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DETERMINERS/PRONOUNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, all, an, another, any, anybody, anything, both, each, either, enough, every, everybody, everyone, everything, few, fewer, he, her, hers, herself, him, himself, his, I, it, its, itself, less, many, me, mine, more, most, much, my, myself, neither, no, nobody, none, no-one, nothing, other, others, our, ours, ourselves, she, some, somebody, someone, something, such, that, the, their, theirs, them, themselves, these, they, this, those, us, we, what, which, who, whom, whose, you, your, yours, yourself, yourselves.</td>
</tr>
</tbody>
</table>

Apart from all these words we also deleted numbers, single letters indicative of the order of activities and other symbols typical of textbooks.

It also needs to be mentioned that for practical reasons and due to the limited scope of this research project the corpus elaboration did not include the workbook or University entrance exams that may have been used in class. Obviously their inclusion would have given us a much wider vision of the input students are exposed to. Yet, the textbook is the principal source of comprehensible input for students. From the textbook, even the listening activities transcripts have been included. This means that all the input available in the textbook has been taken into account for the elaboration of the corpus. The resulting corpus has 26,887 words and 3,219 types. As we can observe the reduction of words when they are counted in types is considerable. This means that similarly to teacher talk or baby talk the type of language students are exposed to is quite repetitive. Our objective will be to see which words are most frequent in the book and to investigate the frequency level they belong to. The initial frequency count listed the 26,887 words taking into account their frequency order but this list was later lemmatised. We mentioned previously that for this study we would not take into account words but lemmas. Lemmatisation was done following the definition given by Nation (2001: 7), “a headword and some of its inflected and reduced forms”. In this new frequency list the headword was not included unless it appeared in the original corpus, this allows us to realise which forms of the lemma really occur or are used, as very often a word does not appear in the whole book in its base form but in inflected ones. In order not to change the
frequency figures after lemmatising the list, the different frequencies for the varied inflected forms of a lemma were added. This will allow us to apprehend which words are most often included in the textbook, the frequency level they belong to and the form in which they are presented to students.

This research project does not take into account how these words are presented to students. This does not mean that we are unaware of the fact that focus at a particular time on a particular word, even if it is not among the most frequent ones, may imply that students learn it. This is not neglected and it is something that must be borne in mind when drawing conclusions. However, our objective is the study of the type of input included in a textbook and to know if it matches the expectations that teachers have for students. In other words, we want to know if we give students the type of input they need to reach the level of 3,000 most frequent words that is required for autonomous functioning in the foreign language.

In this section, we would like to comment on the methodological aspects that were taken into consideration for the elaboration of the corpus. It was mentioned that the corpus was elaborated taking into account what was found in the textbook, as it was the major source of input within the classroom. Once the texts to be included were selected these texts were digitalised using a normal scanner. After all the images of the textbook were captured they were processed using OCR software. The resulting text was saved in format txt and carefully checked so as to correct any mistakes that could have appeared in the process. After checking the texts and creating a single txt document with all of them, they were loaded using Monoconc Pro. This was the software used for the elaboration of the frequency lists. For the elaboration of the frequency lists, and due to the fact that our objective was not function words, Nation’s list of functions words was loaded into the software. This allowed us to elaborate a list of frequent words without including function words. The frequency list elaborated contained all the words of the corpus and they were arranged in alphabetical order to facilitate the lemmatisation task. For the lemmatisation the list was copied in an excel document and the list was reduced. It needs to be mentioned that only those words that appeared in the textbook appear here. Therefore, if the head form of the word eat does not appear in the textbook it is not included in the list. The same happens if any of its inflectional forms are not present in the textbook. This allows us to acquaint ourselves with how the different lemmas are presented to students. Thus, our main aim when lemmatising the list is not to include all the forms of a lemma but to see how many times a particular lemma in its different forms appear in the textbook. The frequencies of the different forms were added so as to obtain the frequency of the lemma within the input given to students. The analysis of the type of lemmas included in the frequency lists can help us to gain knowledge of common semantic fields students are in constant contact with.

The frequency lists were also processed in order to find out not only which words are the most frequent ones but also the frequency range they belong to. In order to find out the lexical frequency that the different lemmas belong to they were processed using the classical Vocabulary Profiles v. 2.6. This software gave us the information of how many words belong to the different frequency levels; establishing the number of families found in the list per level and that of types and tokens that exist. This computer programme helped us to get to know the type of vocabulary that was being provided to students. This study of the frequency range that words in the textbook belong to will help us see if there is a mapping between the input students receive and the type of vocabulary knowledge that they have.
II.2. Sample

The corpus was elaborated using the textbook Valid edited by Burlington and as we mentioned previously only the textbook and the listening transcriptions were included in it.

III. Results

It has already been explained how the corpus using the textbook was elaborated and now we would like to analyse the results of such corpus. In order to see whether the proper input was being provided to students the frequency lists elaborated with content words were introduced in the software Frequency Profiles. This allowed us to quickly and efficiently analyse the content of the whole word lists. Before explaining the results of the analysis of the corpus it needs to be reminded that research studies done about the relationship between reading and vocabulary size predict that a minimum knowledge of words belonging to the K3 category to be able to reach minimal comprehension. Therefore, if the main objective of teachers and publishers is to help students to accomplish the goal of autonomous reading comprehension, the input provided must include plenty of words belonging to K3 category and also a balanced proportion of K1 and K2 groups. Students cannot be expected to understand words belonging to this category if they are not constantly exposed to it.

The results showed that 54.99% of the words included in the textbook belong to the group of K1 words. This means that most of the input that students receive belongs to this frequency level and therefore it cannot be expected that their understanding of words in other levels is high. We are aware that words belonging to this level are obviously the most frequent ones and therefore it is reasonable that their presence in the textbook is more frequent than those belonging to other categories. However, it cannot be forgotten that students require recurrent exposure to vocabulary in order to learn it and to favour future recall. The analysis of the input provided in the textbook would make us predict that students will perform better in the 1,000 level if we passed them Nation’s Word Level Test than in any other. In the following graph the percentages of the words classified in different levels can be seen.

GRAPH 1:

FREQUENCY LEVELS INCLUDED IN THE TEXTBOOK

54.99% of a total of 1684 tokens corresponds to a total amount of 926 tokens and 902 types belonging to the 1,000 word level. The 2,000 words level is less than a half of the latter and it
represents only 19.89% of the total amount of the content words in the textbook. This percentage includes 335 tokens and 333 types. 9.92% of the words belong to the academic word list. Finally, 225 types and 256 tokens are categorised as being off list; 167 tokens and 164 types, of the words included in the textbook belong to Nation’s 15.20% of the total. Frequency Profile software does not classify words into the K3 level and this is the reason why a manual check of the words that appear as off list would be required to determine whether these words belong to the K3 level or not. This is something that should be done in the future but even if all of them belonged to the K3 level we can see that their presence in the textbook is not enough to foster extensive vocabulary learning at this level. Nevertheless, it should be pointed out that despite the presence of K1 words representing the majority, the distribution of the rest of levels is a logical one. It can be observed that the next level in terms of number of words included in the textbook is K2 and academic words are the ones that are less present in the corpus analysed.

The role of teaching techniques cannot be neglected and we cannot forget that despite the fact that word presence of a particular level may be scarce; focus on a particular lemma can imply its learning. In order to determine how and which words are taught an analysis of those words that belong to the section of vocabulary in the textbook would be required, as they are the ones that receive more attention from teachers and students. Nevertheless, the presence of the K1 level in the textbook is so outstanding that even if no special attention is paid to these words they would be a significant part of students input.

The analysis of words lists also show us that those words that are more frequently used in the textbook include many words related to the metalinguistic function of language. It is understandable that in a textbook there is a high presence of words that are related to the explanation of grammar and other linguistic aspects. Therefore, it is not striking that the most frequently used lemma is ‘word’ appearing a total of 192 times throughout the textbook. Other lemmas such as ‘sentence’, ‘expression’ or ‘adjective’ are also often repeated. Apart from metalinguistic language a high presence of lemmas related to the headings of activities is readily observed, words such as ‘write’, ‘complete’ or ‘choose’. The fact that this type of lemmas has a prominent presence in the textbook reduces the input that students are exposed to even more. Students are usually already familiar with this type of language because it is one that is included in textbooks year after year. The last thing we would like to mention in this section is that out of 1,221 lemmas included in the textbook 579 have recurrences between 4 and 2. The fact that 47.42% of the content words included in the textbook are part of input just between 4 and 2 times reduces the chances of word learning. It cannot be forgotten that the more a lemma is presented to students the higher the possibility of the student learning it. The rest of the percentages can be seen in chart 3.

**CHART 3:**

<table>
<thead>
<tr>
<th>REPETITIONS</th>
<th>NUMBER OF LEMMAS</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 100 or 100</td>
<td>5</td>
<td>0.409%</td>
</tr>
<tr>
<td>Between 99 and 10</td>
<td>306</td>
<td>25.06%</td>
</tr>
<tr>
<td>Between 9 and 4</td>
<td>329</td>
<td>26.94%</td>
</tr>
<tr>
<td>Between 4 and 2</td>
<td>579</td>
<td>47.42%</td>
</tr>
</tbody>
</table>

Word repetition is something that must be paid attention to as there are several studies that prove that it has a significant role for incidental vocabulary learning. Webb (2007: 47) explains that “research investigating the effects of reading on vocabulary acquisition has found that both L1 learners … and L2 learners… may incidentally gain knowledge of
meaning and form through reading.” Repetition has been proved to affect learning. Horst et al. (1988) cited in Webb (2007: 48) suggest that large learning gains are likely to occur for words which were repeated eight or more times. Acquisition of words which were repeated fewer than eight times was much less predictable with some cases of negative gains found between the pre- and post-tests. Webb (2007: 60) not only found that repetition affected word learning but also that the higher the number of encounters with a word the deeper the knowledge of such a word was in the different tests provided in the sample. Following these learning theories we can conclude that 47.42% of the words included in the textbook have few possibilities to be learnt by students. In contrast a 52.40% of the lemmas included in the textbook have enough recurrence to make it possible for students to learn them incidentally. This percentage is reasonable; however, the nature of the words included in the text needs to be taken into account. It was mentioned previously that most of the words belong to the K1 level, which means that students are in contact with an input that does not give them the tools that they need to develop autonomous reading capacity for authentic texts.

IV. Conclusion

After analysing the frequency lists obtained from the corpus elaborated we can conclude that while our Bachillerato students are supposed to know words belonging to the K3 level the input with which they are in contact belongs in its majority to the K1 level. Using this information we can predict that their vocabulary size would not reach the K3 level if we passed a test to these students and that they would have real problems to complete a task as the one proposed in the university entrance exam here in Murcia. Besides, the richness of the vocabulary included in the textbook is fairly limited as it contains many words that are not repeated as many times as it is necessary to foster learning. Furthermore, from the analysis of the most frequent words we can observe that these lists contain many words related to the metalinguistic function and which will rarely be useful for them in real life communication.

V. Further research.

This research study shows that there is a need for further research on the type of vocabulary presented to students and that this analysis should be done in a very exhaustive way if we want to help our students to get the most of the years they are in contact with the target language.

This research project analysed the vocabulary included in a textbook, obviously in order to be able to generalise other textbooks should be analysed. Furthermore, apart from simply analysing the textbook it would be interesting to create a much more complete corpus that included workbooks and the university entrance exams that are very often worked in class. This would give us a wider idea of the type of input students are in contact with. It would also show whether other publishers include a more balanced distribution of word levels and if this distribution tends to be the same at different academic levels. It would also be of extreme importance to analyse which words receive a special treatment and are included in specific vocabulary activities and which ones are included as part of texts and listening activities. This together with a much deeper testing of students vocabulary size would permit us to know if students end being familiar with those words that are taught explicitly in the textbook.
VI. Acknowledgments.

I gratefully acknowledge the collaboration of teachers working in IES Sangonera la Verde for their help when passing the exams and the collaboration of students who agreed to sit the exams. Thank you to my tutor in this PhD course, Aquilino Sánchez, for his patience and invaluable advice in the elaboration of this research project. Also thank you to all those teachers who had something to teach to make this work easier.

VII. Bibliography


Goulden, R., Nation, P. and Read, J., (1990),’How Large Can a Receptive Vocabulary Be?’, *Applied Linguistics*, vol. 11, pp.341-363


