In Search of Common Core of Key Vocabulary Among EFL Coursebooks for Secundary Education in Cantabria Using Corpus Linguistics Applications

Elaine Millar

Universidad de Cantabria, Avenida de Los Castros, Santander 35005
Facultad de Educación, Departamento de Filología

ABSTRACT

Nowadays, vocabulary is widely accepted as a crucial component in EFL learning and teaching. Investigators have demonstrated direct links between learners' overall language performance and their vocabulary size. In this paper, we examine the glossaries of three EFL textbooks widely used for the 4th year of Secondary Education in Cantabria to determine to what extent the samples share key vocabulary. To carry out our research, we used the digital resource, Text Inspector, which draws on the corpus English Vocabulary Profile (EVP). A significant degree of lexical irregularity was found among the three samples, with entries ranging from Common European Framework of Reference (CEFR) levels A1 to A2 and significant proportions of vocabulary not listed at all in EVP. Furthermore, the three glossaries have very little vocabulary in common. We argue that this lexical inconsistency can have a negative impact on learners, who are required to take a standardized test of English at A2 level at the end of Secondary Obligatory Education (ESO).

Keywords: vocabulary, EFL textbooks, Secondary Education, Corpus Linguistics, glossaries, English Vocabulary Profile

1. INTRODUCTION

Foreign Language (FL) teaching has evolved significantly over the last century. Research into language acquisition and learning has grown exponentially and, since its creation in the 1960s, Corpus Linguistics has been an important player in this field. In the 21st century, it is now widely accepted that vocabulary, alongside grammar, is a central component of language acquisition and learning, thus, it should be dealt with thoroughly in the FL classroom[1]. This new outlook signifies an important change for FL teaching, which, until recently, had focused primarily on grammatical structures within formal deductive methodologies[2]. Nowadays, FL classes focus on building learners' communicative capacity. As a result, content, context and fluency are key. Vocabulary tends to feature prominently in English as a Foreign Language (EFL) textbooks, often contextualised and accompanied with "word-building" activities[3]. The Common European Framework of Reference for Languages (CEFR) is an important resource for FL materials designers, providing categorical guidelines on the social, grammatical and lexical aspects of a language that a learner knows at each learning stage (from A1 to C2). The CEFR, along with the great advances that have been made in Corpus Linguistics over recent years, mean that publishers now have extremely reliable and up-to-date tools with which to create their coursebooks. However, it is unclear if due attention is paid to lexical content for EFL materials. Some investigators have cast doubts over authors' methods of vocabulary selection and treatment. In previous investigations, Miranda García[4] and Alcaraz Mármol[5] found significant differences in vocabulary selection between textbooks for the same educational level. We find this disconcerting, especially if we consider the wide variety of EFL textbooks that are used in Secondary Education in Cantabria. In the school year 2015/16 there were at least twelve different textbooks, from three different publishers, used for the 4th year of Obligatory Secondary Education (ESO) in public high schools. If this wide range of textbooks differ significantly in lexical content, it is most likely that students across Cantabria are receiving an incongruent lexical input in class.

In a bid to improve EFL teaching and learning, the regional government of Cantabria has introduced a series of standardized exams, in line with the skills and levels of the CEFR. Since 2011, at the end of Obligatory Secondary Education (4th of ESO), students must take an A2 level exam set by the *Escuela Oficial de Idiomas* of Santander. If all students in Cantabria have to sit the same exam at the end of the school year, this raises the question: do the range of textbooks being used have similar and relevant lexical content that will prepare students equally for the exam? Whilst

grammar structures required for A2 level are well known and covered extensively in EFL textbooks[6], our primary area of concern in this study is lexical content. In our opinion, vocabulary is an extremely important linguistic component which, until recently, has been somewhat neglected in the EFL classroom. Corpus Linguistics has allowed vocabulary research to flourish. We now know there are considerable benefits to having a solid vocabulary base in EFL[7] and furthermore, we now have the technological means to select the best vocabulary for each learning stage. This study uses Corpus Linguistics applications to measure the adequacy of the lexical content in EFL textbooks being used for Secondary Education.

2. RESEARCH QUESTIONS

This study uses Corpus Linguistics applications to measure the adequacy of lexical content in EFL textbooks for Secondary Education in Cantabria, Spain. We will attempt to answer the following questions:

1. Do the glossaries of three popular EFL textbooks used for 4th ESO focus on the target level (A2) vocabulary?

2. Do the glossaries of three popular EFL textbooks used for 4th ESO share a common core of vocabulary?

3. THEORETICAL FRAMEWORK

3.1 EFL learning and vocabulary

According to Criado & Sánchez[6], educational authorities in Spain did not prioritize foreign languages in public education until the nineteen-seventies, when Spain opened to the rest of Europe. The *Ley General de Educación* (LGE) of 1970 endorsed the Audiolingual Method of FL teaching for Secondary Education. Although this was an improvement from previous teaching approaches (i.e. Grammar Translation), the Audiolingual Method continued to focus heavily on grammatical structures, with vocabulary being taught in an isolated manner. This meant that the influence of context on word meaning was often missed. It is now understood that the absence of context in vocabulary learning can lead to significant breakdowns in communication, and as a consequence, teaching methodologies have evolved to include context dependent vocabulary. Currently, the Council of Europe and the Ministry of Education in Spain recommend the Communicative Approach to FL teaching[6]. With language being treated as tool for communication, emphasis in the classroom shifts from the memorisation of grammatical structures and isolated word lists to the transmission of ideas and meaning between speakers. Richards[8] indicates two distinct periods in the Communicative Approach: The Classic Communicative Approach (from the nineteen-seventies to the nineteen-nineties) and the Current Communicative Approach (from the nineteen-nineties to the present). Some key features of the Current Communicative Approach are:

• Language is acquired when there is significant communication between users.

• Language is learned through collaboration and exchange of ideas.

• Learners need relevant and interesting content.

• To promote fluency, learners need freedom to experiment with the FL.

• Activities should focus on negotiation of meaning acquisition and significant interaction.

Given the fundamentals of the Communicative Approach, it is clear that vocabulary plays a vital role in FL teaching and learning. In many respects this has been reflected in EFL textbooks, however, investigations indicate that publishers tend to come up short when it comes to selection and treatment lexis.

Words are carriers of meaning and messages. A solid knowledge of vocabulary has a direct impact on a learners' comprehension of a foreign language, their fluency and precision[7]. It is estimated that the English language consists of around 54,000 word families, and that the average native controls between 17,000 and 20,000 of these families[5]. This is an immense amount of vocabulary, which is almost impossible to learn in a classroom setting (students would have to learn up to ten word combinations a day, five days a week, over a period of twenty years). Furthermore, learners must come into contact with a word various times before they can internalize it into their vocabulary. Taking these factors into account, it becomes very clear that vocabulary selection for successful EFL teaching is essential. Research has revealed that to reach A2 level, learners should know around 2000 of the 5000 most frequent words in English[7]. Based on this

evidence, we believe that students in 4th of ESO should be taught a similar volume of vocabulary from relevant semantic fields. 2000 vocabulary items may seem low if we consider the amount of hours of EFL instruction learners have received over the course of their studies, however, we argue that one of the fundamental objectives of the Current Communicative Approach is to promote communication between learners, minimizing the dominant role of the teacher. According Chujo[9], learners must understand around 95% of the language presented to them, to ensure meaningful learning based on the Communicative Approach. If learners do not understand more than 5% of the words in a text, the focus of the class can shift from significant communication the teaching and learning of language form.

3.2 Literature Review

In Spain there have been a number of investigations focusing on the lexical content of EFL textbooks. While qualitative investigations analyze the methodological treatment of vocabulary, quantitative studies focus on the selection, distribution and frequency of words. In this section we will summarize the findings from investigations of EFL textbooks in Spain.

Qualitative:

Mancebo Francisco (cited in Jiménez Catalán & Mancebo Francisco[10]) analyzed a selection of primary school EFL textbooks to see if they complied with CEFR recommendations. The study found that the samples followed all CEFR recommendations apart from those related to vocabulary selection, frequency and grading.

Criado and Sánchez[6] examined a sample of seven EFL textbooks from different educational levels in Spain to establish to what extent their activities followed the Communicative Approach (the method officially endorsed by the Ministry of Education in Spain). They concluded that around 50% of the activities complied with the recommended approach, and the remaining activities focused purely on the learning grammar and vocabulary.

Fernández Orío[11] assessed the representation of knowledge dimensions in vocabulary activities in two EFL textbooks for 4th of ESO. The study found significant inconsistencies between the two samples with regard to distribution and attention paid to lexical knowledge dimensions. It was concluded that this could have a negative impact on the development of the learners' lexical competence.

Quantitative:

Miranda García[4] analyzed the lexical content of sixteen EFL textbooks used at the then *BUP* level (equivalent to the current *Bachillerato*). The aim of the investigation was to determine the size of the common lexical core between the samples. An irregular lexical distribution was found, indicating that publishers adopted disparate criteria for vocabulary selection, if any at all.

Jiménez Catalán and Mancebo Francisco[10] analyzed the lexical content of two textbooks from different educational stages (6th of Primary and 4th of ESO) to ascertain if a systematic approach had been taken to vocabulary input. They found important discrepancies between the two samples with regard to quantity of vocabulary, frequency of vocabulary occurrence, and distribution of word types.

Alcaraz Mármol[3] compared the lexical input of two textbooks from 3rd of Primary against the British National Corpus (BNC) High Frequency Word List. Much of the vocabulary found in the samples did not feature on the BNC High Frequency Word List. This was attributed to the difference in interests and communication needs between children and adults, leading to a divergent vocabulary selection.

In general, previous studies have come to a similar conclusion: publishers of EFL textbooks still have yet to establish a systematic criterion for the selection and treatment of vocabulary. Three main problem areas have been identified: adequacy of vocabulary selection, quantity and distribution of vocabulary throughout a textbook; methodological treatment and recycling of vocabulary. Our research focuses on vocabulary selection for the glossaries of EFL textbooks. We consider vocabulary selection a very important criterion, especially if we take into account the diversity of teaching materials used in secondary schools in Cantabria.

4.   RESEARCH METHODOLOGY

4.1   Materials

We selected three of the most commonly used textbooks in public high schools in Cantabria: *Real English ESO 4*[12], *English Alive 4*[13] and *Voices 4*[14]. Each publication has been designed specifically for the 4th year of Secondary Education in Spain by leading EFL publishers. They all consist of a student's book and workbook with nine learning units. At the back of each workbook is a glossary with Spanish definitions of key vocabulary from each unit. All of the publications dedicate a considerable amount of activities to vocabulary with sections such as *Word Bank*, *Language Builder* and *Vocabulary Plus*. In this study, we focus on vocabulary selection, i.e., words that the publishers have considered key and, therefore, included in their glossaries.

4.2   Instruments

We used the online application Text Inspector (www.textinspector.com) to analyze the glossaries of the three sample textbooks, categorizing the vocabulary from each into CEFR levels. Text Inspector draws on the English Vocabulary Profile (EVP), which is part of the English Profile project (http://www.englishprofile.org/wordlists), a global research program dedicated to creating a database of the grammar and vocabulary EFL learners know at each CEFR level.

We used the Cambridge English KET Vocabulary List as a point of reference for A2 vocabulary, comparing it with our three sample glossaries. It has been argued that, in order to ensure successful learning, students require a vocabulary that corresponds with their needs and interests[9]. We believe that the KET Vocabulary List is an appropriate starting point for 4th of ESO. The KET Vocabulary List is based closely on the semantic fields identified in Threshold 1990[15]. It is updated periodically, drawing on the Cambridge Learners Corpus and English Vocabulary Profile.

Given that we have decided to analyze the glossaries of the three textbooks, we chose the "type" as the lexical unit for our investigation. We define a type as a single lexical unit and its derivated forms. The "lemma" was also an appropriate option; however, we could not use this due to practical restraints.

4.3   Procedure

We There are three phases to our investigation. Firstly, we calculated the total number of words in each glossary and the KET Vocabulary list. The purpose of this was to see if each glossary is similar in size to the KET Vocabulary list. We also needed these figures to analyse data later on in the study. Secondly, using Text Inspector, we classified the content of each glossary into CEFR levels. Our aim was to assess the appropriateness of the vocabulary chosen by each publisher in relation to the target level (A2). Finally, we determined the common core of vocabulary between the three glossaries.

5.   RESULTS & DISCUSSION

5.1   Total number of words in each glossary

English Alive 4 and *Voices 4* contain glossaries similar in size to the KET Vocabulary list, whereas *Real English ESO 4* has a much smaller glossary (Table 1). Based on this result, we would hesitate to assume that this means the glossary in *Real English ESO 4* is less adequate than the other two textbooks. It has been recommended that teachers introduce six or seven new words per hour of instruction[5], so one could indeed argue that *Real English ESO 4* offers a more manageable wordlist for 4th of ESO. In any case, the size of the glossaries alone cannot establish their appropriateness for the objective learning level. To do this, we must further analyse the contents of the three glossaries.

Table 1. Total number of types in each glossary

	KET Vocabulary List	Real English 4	English Alive 4	Voices 4
No of types	1449	502	1016	1061

5.2   Content of each glossary classified into CEFR levels

Figure 1. shows the variation of the types listed in the KET Vocabulary List and our three samples. The KET Vocabulary List consists mainly of vocabulary from levels A1 and A2. 0.7% of the types in this list are B2 or higher, and 2.2% are

not listed in EVP. This differs significantly from our three sample glossaries, which contain a mixture of vocabulary from all CEFR levels. Furthermore, our three samples contain much higher proportions of vocabulary not listed by EVP. In table 2, we can see that 61.95% of the types in *Real English ESO 4* are higher than A2 or not listed. This figure stands at 68.01% for *English Alive 4* and 65.79% for *Voices 4.*
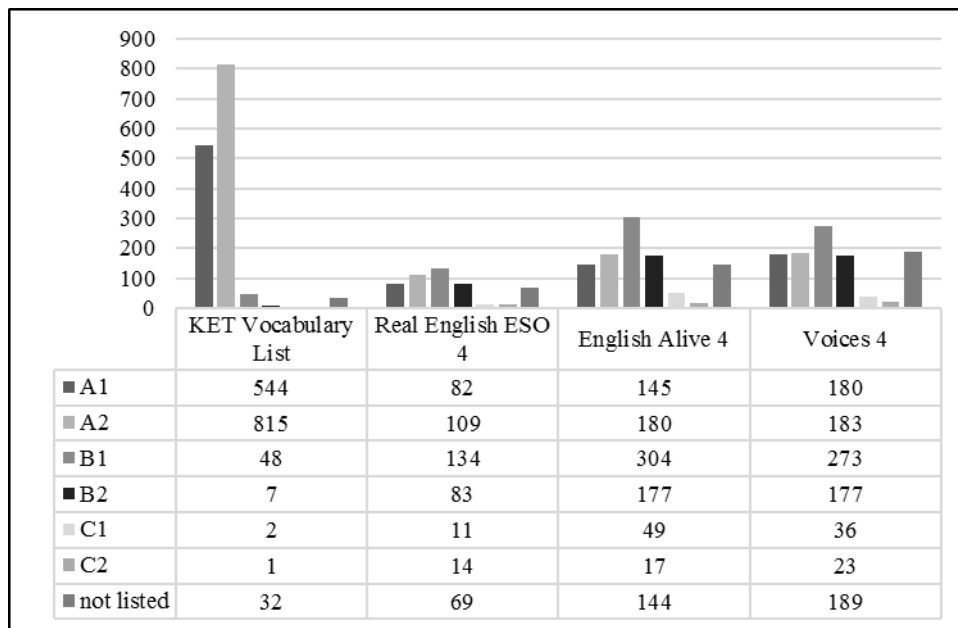

	KET Vocabulary List	Real English ESO 4	English Alive 4	Voices 4
■ A1	544	82	145	180
■ A2	815	109	180	183
■ B1	48	134	304	273
■ B2	7	83	177	177
■ C1	2	11	49	36
■ C2	1	14	17	23
■ not listed	32	69	144	189

Figure 1. Types classified into CEFR levels

Table 2. Percentages of types A1-A2, B1-C2 and not listed

	KET Vocabulary List	**Real English 4**	**English Alive 4**	**Voices 4**
A1-A2	93,79%	38,05%	31,99%	34,21%
B1-C2 & not listed	6,21%	61,95%	68,01%	65,79%

In Figure 2, we have displayed the proportion of each glossary containing vocabulary from B2 level and higher, along with that of vocabulary not listed on the CEFR scale according to EVP. In *Real English ESO 4* 35.26% of types are either above B1 or not listed at all. In *English Alive 4* this figure stands at 38.09% and in *Voices 4* at 40.05%. We would argue that the three samples have a disconcertingly high proportion of vocabulary straying from the target level.
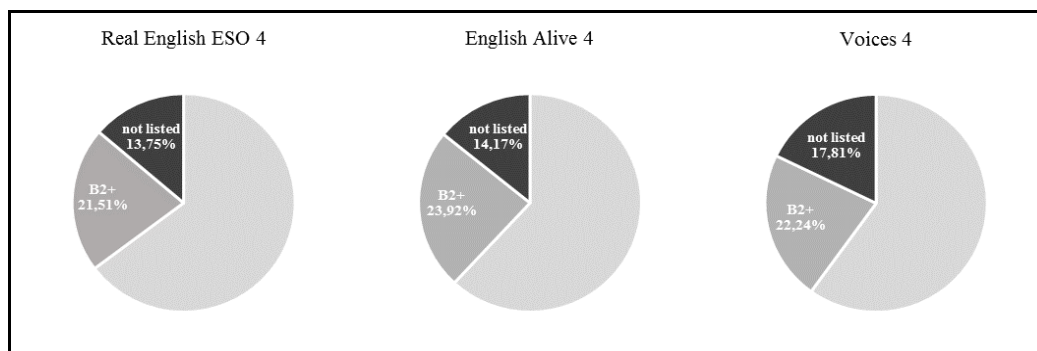

Figure 2. Proportion of vocabulary shared with KET Vocavulary list

In answer to our first research question, we have found that the glossaries from our three sample textbooks do not focus primarily on A2 vocabulary. Instead, it seems that each publisher has followed an unclear criterion for vocabulary selection. In our opinion, the most troubling finding is the quantity of vocabulary in each glossary not listed by EVP.

5.3 Common core of key vocabulary

We have discovered that each glossary contains a wide variety of vocabulary from the CEFR scale and significant proportions of unlisted vocabulary, but we would also like to ascertain if the three samples share common vocabulary. This matter is of particular importance because all students of 4th ESO in Cantabria have to take a standardized exam of English at A2 level. In Figure 3 we compare each sample individually with the KET Vocabulary List. 38.34% of the *Real English ESO 4* glossary features vocabulary from the KET Vocabulary List. This figure stands at 31.90% for *English Alive 4*, and 45.09% for *Voices 4*. In answer to our second research question, we have found a common core of forty-seven types between the three samples, which we have listed by their CEFR level in Table 3.
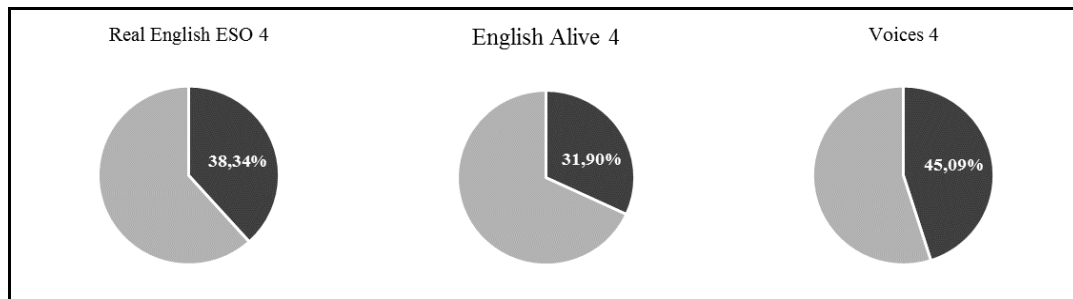

Figure 2. Proportion of vocabulary higher than B1 or not listed in EVP.

Table 3. Common core of vocabulary among all samples according to CEFR level.

A1				A2				B1			
AFTER	GOOD	MISS	UP	ADVERTISEMENT		BREAK	HIGH	OUT	ADMIT	LAND	SENSE	TOUCH
CAN	HAVE	OF	WELL	AGREE		ENGINE	HOLD	SOMEONE	ADVERT	LAW	SUCCEED	WASTE
FIRST	IN	ON	WITH	APPOINTMENT		EXPLAIN	MIND	TRIP	**B2**			
FLAT	LOVE	TAKE	WORK	AROUND		FIT	MOVE		CURRENT			
GO	MAKE	THE	YOU	AWAY		HEAVY	OIL					

6. CONCLUSIONS

The aim of this study was to assess the lexical content of EFL textbooks being used for the 4th year of Secondary Education in Cantabria. We found a significant degree of lexical inconsistency among the three samples. Firstly, each glossary contained large amounts of vocabulary unrelated to the target level A2. We also found a large amount of vocabulary unlisted in the EVP. This indicates that the publishers of these textbooks have not considered the relevant corpus linked to the CEFR scale when selecting vocabulary. Secondly, the three glossaries shared a very small core of vocabulary. These results are disconcerting when we consider that all of the students will have to take the same A2 exam at the end of the school year. We feel it is unfair that learners may receive an uneven lexical input in the secondary classroom because this could have a negative impact on the learning outcomes of each group. There is a possibility that some students will be better prepared than others for the A2 exam, depending on the textbook they have used in class. Whilst we acknowledge that secondary school students can learn a great deal of English from sources other than the textbook (and that this is a very important, positive part of the EFL learning process), we argue that the textbook in many cases lies at the core of the 4th of ESO class, and its lexical content should promote communicative competency and fluency. The glossaries in these textbooks should contain comparable and tangible vocabulary, relevant to the learners' educational level. We believe that a common core of key vocabulary is necessary for all students in this educational stage to ensure the same learning conditions for all students in Cantabria.

REFERENCES

[1] Larrotta, Clarena. "Second language vocabulary learning and teaching: still a hot topic." Journal of Adult Education 40(1), (2011).

[2] Viña Rouco, Mar. "The teaching of foreign languages in Europe: a historical perspective on foreign language teaching in Spain." CAUCE, Revista de Filología y su Didáctica 25, 255-280 (2002).

[3] Alcaraz Mármol, Gema. "Vocabulary input in EFL textbooks: frequency levels." A Survey of Corpus-based Research, 771-783 (2009).

[4] Miranda García, Antonio. "The vocabulary of the English coursebooks: an analysis." Revista Española de Lingüística Aplicada 6, 111-117 (1990).

[5] Alcaraz Mármol, Gema. "Vocabulary input in classroom materials: two EFL coursebooks used in Spanish schools." Revista Española de Lingüística Aplicada 24, (2011).

[6] Criado, Raquel, and Sánchez, Aquilino. "English language teaching in Spain: do textbooks comply with the official methodological regulations? a sample analysis." International Journal of English Studies 9(1) (2009).

[7] Milton, James. "The development of vocabulary breadth across the CEFR levels." Communicative Proficiency and Linguistic Development: Intersections between SLA and language testing research 211-232 (2010).

[8] Richards, Jack C. [Communicative language teaching today] SEAMEO Regional Language Centre (2005).

[9] Chujo, Kiyomi. "Measuring vocabulary levels of English textbooks and tests using a BNC lemmatised high frequency word list." Language and Computers 51(1), 231-249 (2004).

[10] Jiménez Catalán, Rosa María & Mancebo Francisco, Rocío. "Vocabulary input in EFL textbooks." Revista española de lingüística aplicada 21, 147-16 (2008).

[11] Fernández Orío, Susana. "Vocabulary knowledge dimensions in EFL textbooks." Encuentro: revista de investigación e innovación en la clase de idiomas 23, 30-37 (2014).

[12] Marks, L. & Addison C. [Real English Workbook: ESO 4], Burlington Books, (2010).

[13] Wetz, B. Halliwell, H. & Tims, N. [English Alive! Workbook 4] Oxford University Press, (2007).

[14] Bilsborough, K. & Bilsborough, S. [Voices Workbook 4] MacMillan (2009).

[15] Van Ek, J.A. & Trim, J.L.M. [Threshold 1990] Council of Europe, Cambridge University Press (1991).