

LEXICAL FREQUENCY, TEXTBOOKS AND LEARNING FROM A COGNITIVE PERSPECTIVE. A CORPUS-BASED SAMPLE ANALYSIS OF ELT MATERIALS

RAQUEL CRIADO*
AQUILINO SÁNCHEZ
Universidad de Murcia

ABSTRACT. *There is not full agreement among authors regarding the role of repetition in the learning process, but its positive effects have been emphasized in Cognitive Psychology, Psycholinguistics and SLA. We assume here that repetition is a key factor in language teaching/learning and focus our attention on the lexical component. We analyze the lexis in two well-known ELT textbooks in order to verify up to what point opportunities for lexical repetition are offered. Methodologically, 1) all the vocabulary was examined using Nation's RANGE software to identify its distribution throughout the first three first ranges defined by Nation and to compute the standardised type/token ratio in both textbooks; 2) the repetitive practice vocabulary activities were identified and quantified. Results show that the two course books are similar in the amount of words included and their distribution throughout the ranges studied, but they differ in some other fields, such as the standardised type/token ratio and the emphasis on repetitive practice. Pedagogical and research implications are discussed.*

KEYWORDS. *Cognitive learning, ELT materials, frequency, repetitive practice.*

RESUMEN. *Los autores discrepan sobre el cometido de la repetición en la enseñanza/aprendizaje, pero sus efectos positivos han sido puestos de manifiesto en Psicología Cognitiva, Psicolingüística y ASL. En este trabajo asumimos que la repetición es un factor clave en la enseñanza/aprendizaje de lenguas y nos centramos en el léxico. Analizamos el léxico de dos libros de texto para la enseñanza del inglés como lengua extranjera con el fin de comprobar si ofrecen oportunidades para la repetición léxica. Metodológicamente, (i) primero analizamos el vocabulario de cada libro mediante el programa RANGE de Nation para identificar su distribución en los tres primeros tramos de frecuencia y la ratio estandarizada type/token; (ii) identificamos y cuantificamos todas las actividades prácticas de repetición léxica. Los resultados revelan que ambos libros son muy semejantes en el número total de palabras y su distribución por tramos de frecuencia, pero difieren en otras áreas, como la ratio type/token estandarizada.*

zada y el énfasis en las prácticas repetitivas. En base a estos resultados, se incorporan implicaciones pedagógicas e investigadoras.

PALABRAS CLAVE. Aprendizaje cognitivo, materiales para la Enseñanza de Inglés como Lengua Extranjera, frecuencia, prácticas repetitivas.

1. INTRODUCTION

In this study we particularly focus on vocabulary frequency as offered in textbooks, which we consider to be the main learning and teaching tool in most Foreign Language (FL) contexts. Frequency of practice has been the subject of discussion and divergent views in teaching methods. Indeed, repetition of forms has been one of the most important strategies in language education to foster learning and consolidate knowledge.

Accordingly, this study is premised on the need to carry out close analyses of textbooks regarding the lexical component they offer in terms of the adequacy of their frequency and distribution according to cognitive patterns of learning as well as their related pedagogical means or strategies to facilitate lexical learning (as materialised in the number and kind of activities). Lexical repetition in textbooks may be favoured in different ways: a) by including the adequate amount of items, according to the level, and favouring their occurrence across the course book, so as to offer opportunities for abundant repetition; b) by means of the right kind of activities, i.e., repetitive practice activities, whose procedural strategies demand the repetition of the same lexical items to accomplish the activities. Studies such as these may be relevant on the one hand for textbook authors and teachers, in that they offer relevant information on what these materials really offer from the perspective of cognitive requirements regarding repetition in terms of mere frequency of lexical items and discover whether the materials also include specific activities providing opportunities for repetitive practice. In this way, authors would become aware of whether they need to adjust their own procedures for the pedagogical inclusion of lexis in textbooks and teachers would know whether they should supplement their lexical teaching or not with additional (tailor-made) materials. On the other hand, studies such as the present one may be helpful to researchers so as to raise awareness in the academia about the need to pursue further studies on the optimal degree of frequency for vocabulary learning and about the related pedagogical implications. Overall, then, this type of studies aims at providing sound guidelines for language teaching materials design regarding the treatment of the lexical component.

The structure of this article would be as follows. In section 2 we will offer a review of the importance of frequency and repetition in learning in general and in language learning in particular, with special emphasis on vocabulary learning. Given that this study is based on a corpus-based lexical analysis of two course books, section 3 will offer a review of studies focused on vocabulary as found and distributed in course books. Section 4 will detail the goals of our study, followed by the methodology followed to attain such goals (section 5). In section 6, results and discussion will be included in relation to the

goals previously established. A conclusion will be finally incorporated with the summary of the main ideas in the article and a reflection on pedagogical and research implications.

2. THE IMPORTANCE OF FREQUENCY IN LEARNING IN GENERAL AND IN VOCABULARY LEARNING

The role of practice trials in learning was studied in Newell & Rosenbloom's (1981) *Power Law of Practice*, which defines and models the effect of practice on learning and performance. This is a learning curve defined by a power function. The *Power Law of Practice* is premised on the fact that there is a specific relationship between practice trials and "reaction time", that is, between frequency of trials (practice) and their effect on learning (see Newell & Rosenbloom 1981; DeKeyser 2001). Many scholars in cognitive and psycholinguistic research highlight an additional finding: language acquisition results from abstracting and consolidating patterns detected in language use, in an unconscious way (Atkinson & Shiffrin 1968; Sinclair 1991; Ellis 2002; Tomasello 2003; Schmitt *et al.* 2004; Erman 2007). Detection of usage patterns is closely related to abundant linguistic input and previous exposure to language, that is, abundant iteration of incoming data. Through the frequent exposure to samples of language, the brain ends up consolidating form-function and form-meaning associations, which allow for the elaboration of a linguistic code, similar to the codes acquired by other speakers and used to interact with one another. We also have the experience that in most human activities and learning processes, repetition and practice seems to be the most important ingredient for the consolidation of knowledge and for accuracy in performance.

However, scholars do not fully agree on the impact of frequency in knowledge acquisition. Two distant approaches to language acquisition may be contrasted in this respect. Both assign a radically different role to frequency. The first one (Roeper 2007, among others) assumes that humans are born with a built-in universal grammar in the brain. Within this position, frequency does not have much to do in the learning process; its role is rather secondary and perhaps subsidiary: human beings come to this world with innate structures (a kind of built-in circuitry) that account for learning by means of interaction with the input. Roeper (2007: 23) argues that frequency is a mere quantitative element, and cannot contribute to knowledge increase, since knowledge is different in nature: "Counting instances of X does not obviously change X. Learning inherently involves adding information. But again, the addition of information is quite separate from counting information". The other approach, held by a wide array of scholars (Behrens 2009; Tomasello 2003, among others), is a usage-based approach and favours a different view: linguistic knowledge is constructed by humans through exposure to input, and frequency is the key element in the process of knowledge construction. Truly, the relationship between frequency and language use and acquisition cannot be discarded. The frequency of a specific structure, for example, maintains a tight connection with its complexity. Complex structures are less frequent than more simple ones. The "principle of least effort", in Zipf's words, seems to be based on real linguistic

usage and on the behavior and distribution of linguistic forms in communicative events (Mandelbrot 1965; Zipf 1935, 1949).

Research in psychology and neurolinguistics has proved that repetition, as tightly connected to frequency, exerts a true effect on linguistic learning (Rundus 1971; Kapur *et al.* 1994; Paradis 1994, 2009; Tomasello & Brooks 1999; Rosa & Leow 2004; Ullman 2004; Bybee 2006; Anderson 2010, among others). The physio-biological bases of knowledge condition its nature, as well as its acquisition. Neurologically, repetition of the same action affects the structure of specific synapses or connections among neurons, thus rendering future and similar tasks easier, and allowing for reaching automaticity after several trials; the quantity of trials may vary depending on other accompanying factors. Repetition and practice strengthen neural connections, and makes the task easier to perform, until it becomes automatic. This neurological pattern is reflected in a well-known learning model widely applied in FLT: The stimulus-response model, which requires frequency of repetition and iteration as essential for building the automatic association between the initial stimulus and the related response.

3. RESEARCH ON THE VOCABULARY INCLUDED IN LANGUAGE TEACHING MATERIALS

Research in the field of vocabulary learning has been scarce until recently; even more so in the field of vocabulary teaching. The situation seems to be changing in the last two or three decades, and a few studies have already engaged in the analysis of textbooks from the point of view of their lexical component. Research and publications by Nation (2001) have been decisive for promoting studies on vocabulary and textbook teaching. Milton (2009) published a thorough and comprehensive study on vocabulary acquisition in which, among other issues, he discusses the role of textbooks for teaching and learning vocabulary. We call the attention on two groups of studies. Firstly, those studies centred on the actual amount of vocabulary and vocabulary coverage in teaching materials. Some of those studies examine the associative relationship of two factors as well, textbook vocabulary and teacher's lexical production –his/her oral input in the classroom– with students' vocabulary gains. The second group consists of those studies focused on the rate of vocabulary learning.

As to the first group of studies, Davis & Face (2006) explored the vocabulary coverage of Spanish textbooks comparing the words they included with corpora frequency lists. Their conclusion is that “for whatever N number of vocabulary words a textbook includes, only 10-50% of those are among the N most frequent lemma in the language. For example, if a textbook presents 2000 vocabulary words, only 10-50% of those words are among the most frequently used 2000 lemma in the language” (Davies & Face 2006: 142). Jiménez and Mancebo (2008) carried out an investigation on the vocabulary input found in EFL textbooks for primary and secondary education, restricted to the first most frequent fifty words in each textbook. They also found important differences in the words presented by different textbooks. Donzelli (2007)

studied the acquisition of lexical items relative to their frequency in textbooks and the teacher's speech at the end of the school year of a specific course, in 9-year-old children. Her results showed that "the words that can only be found in the course-book scored a lower number of hits than the ones available in both types of lexical input –the book and the teacher's speech– which were better acquired by the subjects" (p. 123). Laufer & Rozovski-Roitblat (2011) did not study teacher's oral production but the effects on vocabulary learning of the written input from the textbook and task types. The task types were occasionally focused on form when learners used dictionaries (T+F) and word focused exercises when reading texts (T+F_s). Their results of two tests –passive recall and passive recognition–showed on the one hand that "starting with 4 occurrences, T+F_s fared/behaved better than T+F"; on the other, "task type effect was superior to the effect of word occurrence in recall only (2 word exercises fared better than 6-7 occurrences in text)" (Laufer and Rozovski-Roitblat 2011: 391). Finally, Criado and Sánchez (2012) compared the effects on vocabulary performance of vocabulary input in a textbook and in the teacher's oral discourse in an undergraduate classroom. Correlations between gain (post-test scores minus pre-test scores) and frequency of lexical units in the teacher's oral input, in the textbook, and in the sum of both types of frequency were low and not statistically significant, but positive.

Concerning the second group of studies, Scholfield (1991) analysed the rate at which new vocabulary items are introduced in course books. He found relevant differences in the textbooks investigated. Ito and Bauman (1995) carried out one of the first investigations on the rate of vocabulary acquisition by Japanese college learners of English in the classroom environment. The students tested learnt only one word per hour. Waring and Takaki (2003) conducted a study on the rate of vocabulary learning from reading a graded reader. They arrived at similar conclusions as Ito and Bauman: "the subjects learned one new word from an hour of reading" (Waring and Takaki 2003: 148). Alcaraz's (2011) students reached the average of 3.6 words per class hour all along a three month period. Though the results of these studies on the rate of learning are not fully conclusive either, they will provide a useful point of reference to determine whether the two ELT textbooks analysed in this work offer *realistic* lexical learning opportunities for the students.

4. GOALS

The goal of this study is to ascertain the opportunities for lexical learning offered to students by two ELT course books from the point of view of the frequency of lexical items and repetitive activities focused on vocabulary. For that purpose we will compute and compare in both course books: (i) the distribution of lexical items per RANGE, in order to determine whether such a distribution adjusts to expectations regarding the teaching level that course books aim at; (ii) the amount of target types included and its adequacy to the students' learning potential, on the basis of the findings from i); (iii) the *standardised type/token ratio (STTR)*, so as to find out the average amount of opportunities for lexical input (and repetition) offered; (iv) the amount of repetitive

practice activities included in order to contribute to vocabulary learning. We should emphasize that by “repetitive practice” we do not mean the mechanical drills from the Audiolingual method alone, but all activities involving repetition of lexical items, whether via mechanical, meaningful or communicative practice.

5. METHODOLOGY (MATERIALS AND PROCEDURE)

To extract the relevant quantitative data in both textbooks we have relied on RANGE (<http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>), a computational tool designed by Nation. RANGE counts and classifies vocabulary in three frequency categories (at present up to 14)¹: the first 1,000, the second 1,000 and the third 1,000 most frequent words of general English. These frequency categories are called RANGES as well, so that the word RANGE in the present study is taken as each one of the 1,000 word groups sequentially determined in the frequency list considered. In our study, words not included within the first three most frequent ranges (i.e., the most frequent 3,000 words) appear as off ranges.

RANGE reads each one of the books selected and previously digitized and classifies the words within each range as tokens (every word form in the text, be it repeated or not), types (different words in the text: *friend* and *friends* are two types) and word families (the headword, its inflected forms and its closely related derived forms), which is relevant for our study.

The textbooks analyzed are two very-well known course books from the ELT international market and which are used in Spanish educational settings such as *Escuelas Oficiales de Idiomas* (Official Schools of Languages): *New English File Elementary. Student's Book* (NEF henceforth), by Clive Oxenden, Christina Latham-Koenig and Paul Seligson, published in 2004 by Oxford University Press and *Face2Face Elementary Student's Book* (F2F henceforth), by Chris Redston and Gillie Cunningham, published in 2005 by Cambridge University Press. Both textbooks are similar in goals –the attainment of the A2 level following the *Common European Framework of Reference for Languages* (CEFR, 2001), and methodological approach –Communicative Language Teaching Approach (CLT).

F2F is structured in 12 units, 8 pages each, totalling 96 pages. NEF includes 9 units, 12 pages each, with a total of 108 pages. Both course books include sections on the four skills and grammar, pronunciation and vocabulary in each unit. Additional sections on Grammar, Pronunciation and Vocabulary can be found at the end of both teaching materials. For the purpose of this analysis, only the “core” units have been processed and studied.

6. RESULTS AND DISCUSSION

In this section we will report and discuss the results of the figures concerning the amount of types, tokens, estimated lemmas, word families, the STTR (section 6.1.), the

lexical distribution of words per range in each of the two textbooks (section 6.2). This analysis will be complemented by the comparison of the number of practice activities involving repetition in the two course books (section 6.3).

6.1. *Types, tokens, estimated lemmas, word families and STTR*

F2F contains 36,721 running words (tokens). Out of this total, 2,509 are distinct words (types). Word families amount to 1,719². Types included are distributed as follows:

- 1,207 types (over 845 estimated lemmas³) belong to Range 1 (first thousand most frequent words);
- 365 types (around 255 estimated lemmas) belong to Range 2 (second thousand most frequent words),
- 136 types (ca. 95 estimated lemmas) belong to Range 3 (third thousand most frequent words).

In addition to that, there are still 801 types “off-ranges”, beyond the three thousand most frequent words of English (Table 1).

TABLE 1. *Tokens, types and word families for Ranges 1, 2 and 3 in F2F.*

WORD range	TOKENS/%	TYPES/%	FAMILIES
one	31,752/86.47	1,207/48.11	691
two	1,886/5.14	365/14.55	278
three	552/1.50	136/5.42	109
off-ranges	2,531/6.89	801/31.93	(estimated) 641⁴
Total	36,721	2,509	1,719

As can be seen in Table 2, *NEF* contains 30,855 words (tokens), and 2,841 types (ca. 1,988 estimated lemmas). Word families amount to 1,825. Types included are distributed as follows:

- 1,297 types (over 907 estimated lemmas) belong to Range 1 (first thousand most frequent words);
- 429 types (around 300 estimated lemmas) belong to Range 2 (second thousand most frequent words),
- 185 types (ca. 129 estimated lemmas) belong to Range 3 (third thousand most frequent words).

In addition to that, there are still 930 *off-range* types, beyond the three thousand most frequent words of English (Table 2).

TABLE 2. *Tokens, types and word families for Ranges 1, 2 and 3 in NEF.*

WORD range	TOKENS/%	TYPES/%	FAMILIES
one	26,102/84.60	1,297/45.65	700
two	1,876/6.08	429/15.10	334
three	677/2.19	185/6.51	150
off-ranges	2,200/7.13	930/32.73	(estimated) 641
Total	30,855	2,841	1,825

Both textbooks are close in the total of word families included: 1,825 in *NEF* and 1,719 in *F2F*. The estimated lemmas would amount to ca. 1,988 (*NEF*) and 1,756 (*F2F*). It should be taken into account that only ca. 1,000 lemmas (Range 1) are the expected learning target at the elementary level (A1 and A2, according to the *CEFR*, 2001). Range 2 and Range 3 would correspond to the intermediate and advanced level respectively.

The number of tokens (total running words) and types (total of different or not repeated words) for each one of the ranges is partially different in both course books (see Tables 1 and 2 above). *F2F* offers more tokens (+5,866) than *NEF*, but contrary to expectations, *NEF* includes more types (+332) than *F2F*, and hence the estimated lemmas in *NEF* are also higher in number (+232).

In this respect, it is possible to establish a coefficient of repetition of words in a text by calculating the *type/token ratio*, that is, the relationship between the number of types and tokens in both textbooks. The type/token ratio measures the lexical diversity and variation of a text and reveals the proportion of raw vocabulary input versus the new words introduced in a text. Two texts with a similar amount of tokens and types are equally varied in lexical variation and richness, while two texts with a similar number of tokens and a different number of types show differences in lexical richness and variation. In the case of the textbooks analysed in this article, both are different in their respective amount of tokens and types. In order to homogenize these figures so that we could compare in a more precise way how really different both textbooks are in relation to their type/token ratio, we calculated the *STTR*, following *WordSmith Tools* (Scott, 2008). The resulting ratio (Fig. 1) shows which one of them offers greater lexical richness.

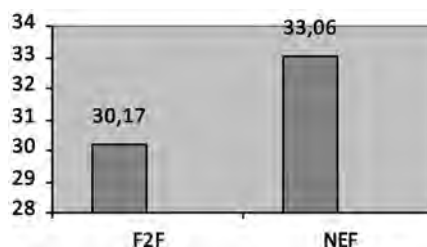


Figure 1⁵. *STTR in F2F and NEF.*

We regard the STTR as an exponent of the opportunities students are given for being exposed to or repeat the same item. As an average, the textbook with the highest STTR will allow for fewer repetitions of the types included. Therefore, students using *F2F* will have more opportunities for the repetition of words than students using *NEF*.

In addition, the lexical component of both textbooks may also be analysed from the angle of their possible learning implications in specific educational systems. For instance, Spanish *Bachillerato* includes an average of 100 teaching hours a year. The average lexical learning load for students following these textbooks in Spanish *Bachillerato* would then rise to ca. 19 new lemmas per hour in the case of *NEF* and 17 in the case of *F2F*, following the amount of estimated lemmas indicated above for each course book. As stated in section 3, experimental studies by Ito and Bauman (1995), Waring and Takaki (2003) and Alcaraz (2011) offer a less optimistic prospect (1 word per hour in the first two studies and 3.6 words in the third study). Accordingly, *F2F* and *NEF* include an amount of vocabulary items far beyond the learning potential of regular students.

6.2. *Lexical distribution of words per range*

The selection and subsequent pedagogical administration of lexical items faces a challenge related to the size of a word list and the frequency of the items included in it. The challenge derives from the law of normal distribution of words in texts. As mentioned in section 2, Zipf's findings (1935, 1949) in word distribution, hand in hand with corpus linguistics studies, predict that the amount of items that appear in a text and their frequency adjusts to a specific pattern or model. Such a model may somehow distort or interfere with the pedagogical goals defined and set for each level.

How does this model work? It predicts that a few sets of tokens are extremely frequent in discourse, while many others occur much less frequently, and about 50% occur only once. In addition to that, the increase in new types follows a parabolic curve, that is, the larger the text in terms of tokens, the fewer the types (see Sánchez & Cantos 1997). This property of word distribution in texts should attract the attention of textbook writers. Adding new words will obviously require the inclusion of more text, but the larger the amounts of texts added, the fewer the types there will be. This means that the inclusion of, say, 1,000 new types with a reasonable frequency of occurrence will require the occurrence of many other additional types besides the one thousand types targeted at. Pedagogical manipulation of texts and intervention in the selection and presentation of words may somehow readjust such a distributional model, but will not change the essentials of this inherent property of word distribution.

Let us examine how the property of normal distributional pattern of words is present in the two ELT textbooks analysed here with respect to a) the STTR, and b) the number of types and tokens per range included in each course book.

As indicated in section 6.1, it should be remarked that the proportion of tokens and types found in *F2F* and *NEF* are not equally balanced: the STTR coefficient in *F2F* is 30.17 and 33.06 in *NEF*. Such a difference could be interpreted as the result of a heavier

pedagogical intervention in *NEF*, which partially distorts the normal lexical distribution in this textbook.

As to the behaviour of the types and tokens within ranges, we will first consider Range 1 as opposed to Ranges 2 and 3, on the one hand, and Range 1 as opposed to all the 14 ranges, on the other. As expected, Range 1 contains the highest number of words when compared against the number of words in each one of the other Ranges in isolation. Words in Ranges 2 and 3 as well as the rest of Ranges not measured by the processing software (off-ranges items) do not belong to this level, but all such ranges taken together include more than 50% of the total amount of types in the two course books (Fig. 2):

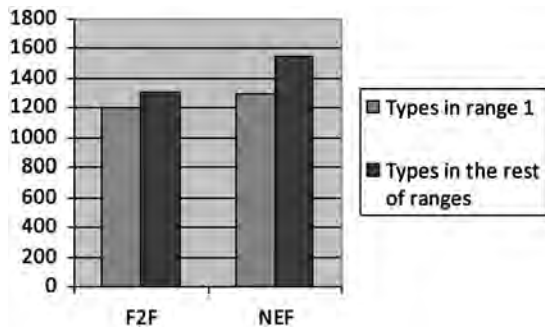


Figure 2. *Types in Range 1 and the rest of ranges taken together in F2F and NEF.*

The reason why less frequent words from ranges other than Range 1 are included in elementary course books is either rooted in the law of lexical distribution mentioned above or in “contextual requirements” (the necessary occurrence of words required by topic and context).

In terms of the distribution of words per range, Range 1 includes more types than Range 2, and Range 2 is higher in number than Range 3 (see Tables 1 and 2 for the exact figures and Fig. 3 for a visual representation of such figures):

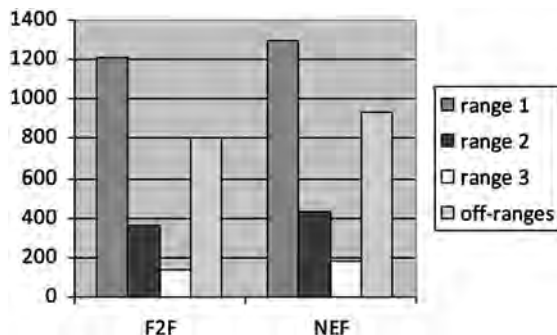


Figure 3. *Types per range in F2F and NEF (three first ranges + off-ranges).*

The number of words in Range 1 is high indeed in both textbooks (84.6% in *NEF* and 86.47% in *F2F*). But the proportion substantially changes regarding types (45.65% in *NEF* and 48.11% in *F2F*). In this respect, the proportional distribution of words per range is somehow more balanced in both textbooks regarding types than tokens, as can be more clearly perceived if we compare the descending curve referred to the amount of lexical items of the first 14 ranges (see Table 3 and Fig. 4 for *F2F* and Table 4 and Fig. 5 for *NEF*):

TABLE 3. *Tokens and Types in the first 14 first ranges in F2F.*

Ranges	TOKENS/%	TYPES/%
one	31,752/86.47	1,207/48.11
two	1,886/5.14	365/14.55
three	552/1.50	136/5.42
four	405/1.10	64/2.55
five	238/0.65	55/2.19
six	99/0.27	28/1.12
seven	86/0.23	19/0.76
eight	38/0.10	17/0.68
nine	37/0.10	16/0.64
ten	15/0.04	6/0.24
11	20/0.05	4/0.16
12	19/0.05	3/0.12
13	12/0.03	4/0.16
14	15/0.04	3/0.12
off-ranges:	1,547/4.21	582/23.20

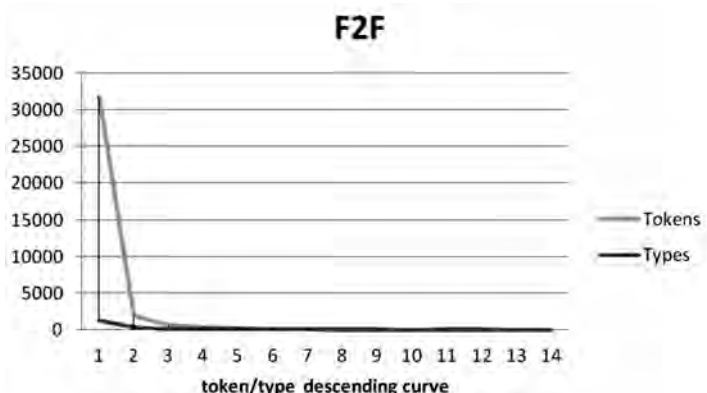


Figure 4. *Token/type descending curve in F2F.*

TABLE 4. *Tokens and Types in the 14 first in the first 14 ranges of NEF.*

Ranges	TOKENS/%	TYPES/%
one	26,102/84.60	1,297/45.65
two	1,876/6.08	429/15.10
three	677/2.19	185/6.51
four	350/1.13	99/3.48
five	224/0.73	60/2.11
six	190/0.62	49/1.72
seven	66/0.21	22/0.77
eight	28/0.09	14/0.49
nine	33/0.11	14/0.49
ten	24/0.08	13/0.46
11	10/0.03	4/0.14
12	20/0.06	6/0.21
13	20/0.06	9/0.32
14	15/0.05	8/0.28
off-ranges:	1,220/3.95	632/22.25

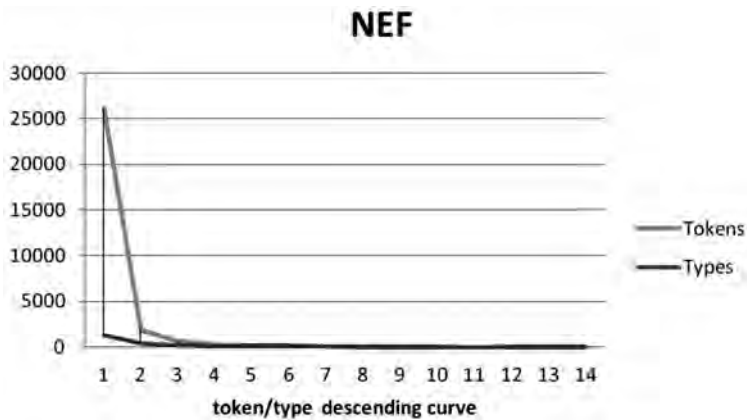


Figure 5. *Token/type descending curve in the first 14 ranges of NEF.*

6.3. Repetitive practice activities

Let us turn our attention now to the second pedagogical tool indicated in the *Introduction* to foster lexical repetition: repetitive practice activities. Our aim here is to identify repetitive practice activities in both textbooks in the sections devoted to lexis. In this respect, we have only considered the activities in the sections explicitly

reserved for vocabulary learning and which are labelled with the heading “Vocabulary”. Needless to say, we are aware of the fact that practice with vocabulary underlies any activity where language is involved, be it focused on grammar, pronunciation or a receptive or productive skill.

For the sake of operationalizing repetitive practice activities in this study, we have regarded as repetitive practice those activities a) in which the same or similar lexical elements have to be re-used several times, either receptively or productively, in a rather mechanical way, or b) whenever repetition is overtly required in more communicatively based activities or meaningful drills. This is the case, for example, of the following invented series of activities for practising the numbers in English:

1. *Can you say these numbers?*
2. *Write the hotel room numbers.*
3. *Take turns to say five numbers.*
4. *How do you say these phone numbers?*
5. *Listen and write the phone numbers.*
6. *Ask three students their phone number.*

Activities such as “Tick the words you know” or “Tell other students about the clothes and colours you wear”, if appearing only once in the very same unit, disconnected from other related activities which require re-using the same or similar words, are not regarded as repetitive practice activities. Also, it should be borne in mind that we do not attempt to offer conclusions about the actual efficacy of the activities found in any of the analysed textbooks, but to identify repetitive practice activities out of all the vocabulary activities in an attempt to infer up to what point enough opportunities for lexical repetition are offered to students.

In the case of *F2F*, most of the sections focused on vocabulary learning involve two or more activities in which the use of the same or semantically related words are required more than once to perform such activities (see following example, p. 30):

- a) Tick the phrases you know. Then do the exercise in Language Summary 4 (p. 128).

*Read books/magazines watch DVD's/videos play tennis take photos
go skiing go swimming go running go dancing listen to music
listen to the radio watch sport on TV*

- b) Work in pairs. Take turns to ask and answer questions about the free time activities.

A. Do you watch sport on TV?

B. *No, never / Yes, every weekend.*

Vocabulary activities in *NEF* are appreciably lower in number and do not require such an intensive repetition of the same forms (see following example, p. 60):

Vocabulary. Shopping

a) Match the words and pictures.

- Postcards
- Batteries
- A (camera) film
- T-shirts
- A mug

b) In pairs, cover the words and test your partner.

Fig. 6 below offers the raw frequencies of repetitive practice activities for both textbooks:

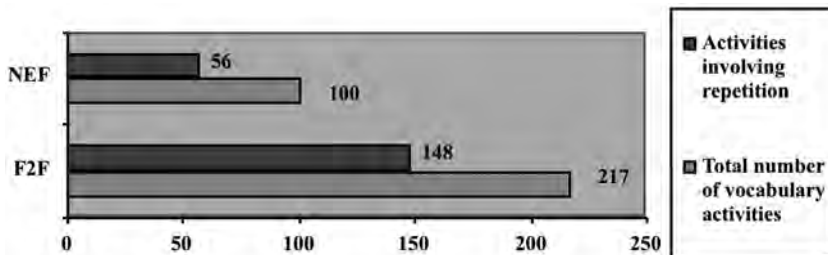


Figure 6. Total number of vocabulary activities and activities involving lexical repetition.

From the data above it can be concluded that lexical repetitive practice is abundant in both textbooks, although *F2F* puts more emphasis on it than *NEF* (68.2% vs. 56% respectively). Overall the results are quite similar in that more than half of the overall vocabulary activities are of a repetitive nature. We consider this proportion to be adequate so as to develop students' consolidation of lexis, although we are aware that this should be tested in experimental studies.

7. CONCLUSION

Frequency and repetitive practice seem to be at the very heart of general knowledge acquisition. This applies to the specific sphere of vocabulary learning as well. Studies such as the present one should make language textbook designers aware of the tools they have

at their disposal in order (i) to elaborate and structure the lexical component presented, and (ii) check the results of their work.

Our analysis shows that both textbooks are similar in their lexical profile: (i) they offer a noticeably high amount of types in Range 1, well over the learning potential of learners (according to the experimental research indicated in sections 3 and 6.1), (ii) they do not deviate much from the expected pattern of normal distribution of words in texts, (iii) they seem to exert a certain pedagogical control on the materials presented, and hence on the lexis included.

On the other hand, differences are less salient in two aspects: (i) both textbooks are not fully equivalent in the amount of tokens and types and in the STTR; in the case of *NEF*, the larger amount of types and lower amount of tokens could be a sign of a more conscious pedagogical manipulation of lexis, which somehow distorts the normal lexical distribution; (ii) although both textbooks include more than half of vocabulary repetitive practice activities out of the overall lexically focused activities, *F2F* seems to place more emphasis on repetitive practice activities.

The pedagogical practical implications of the analysis are straightforward for teachers and learners, as well as for authors of teaching materials. The results of the quantitative lexical analysis make it clear that each one of the textbooks is very similar in most of the parameters investigated. We could hypothesise that the use of any of them in the classroom will not probably imply significant effects on the teaching/learning of vocabulary. This should be a most valuable source of information for textbook users.

Also, the quantitative analysis shows additional and convincing evidence on the constraints that textbook authors must face: the laws governing the normal distribution of words in texts limit their freedom of lexical choice, even if they readjust or even somehow “distort” the language presented to the students. CLT, the current approach in FLT, advocates the presentation of authentic language. If this criterion prevails, we must also face its consequences: this implies less control over the lexical items the student will be exposed to. Textbook authors should find a balance between those two perspectives.

As to researchers, careful descriptions of textbooks such as that presented in this study may constitute a preliminary step in the design of quasi-experimental studies targeted at testing the efficacy of language teaching materials which follow specific patterns in lexical treatment. The results may enrich our state of knowledge regarding the efficacy of certain pedagogical practices as well as generating valuable information about SLA processes, which could in turn inform future research.

In conclusion, analysing textbooks with the methodology proposed in this study deserves our attention, since researchers, learners, teachers and textbook designers may benefit from their conclusions. Indeed, we believe that both “hard sciences” (Psycholinguistics, Neurolinguistics and SLA) and the FLT research field should go hand by hand as much as possible so as to improve language pedagogy and foster more efficient instructional practices grounded on cognitive parameters.

NOTES

- * Correspondence to: Raquel Criado Sánchez. Dpto. Filología Inglesa. Facultad de Letras. Universidad de Murcia. Campus de La Merced. C/Sto. Cristo, 1. 30071 Murcia. E-mail: rcriado@um.es
 This paper has been financed by the Spanish Ministry of Science and Innovation, research project Ref.: FFI2009-07722 (*Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica*).
1. The BNC Base Lists in RANGE are described by the author as follows (see Instructions in file "Range program with British National Corpus list" at <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>): *The fourteen word family lists were made by Paul Nation from the British National Corpus (BNC). These word lists were made from the file at Lists 1.1 complete lists (1_1_all_fullalpha.txt) available at <http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html>. Only the headwords of the lemmas were kept in the list. [...] The procedure for making each list was as follows. Lemma members were removed for the list as were proper nouns, foreign words, numbers and letters. The words in the source list were sorted by the range figure in column 5 and a suitable range cut off point was chosen which would eventually result in about 1000 word families. This section of the list was then sorted by frequency (column 4) and words below a certain frequency point were put back in the source list. The remaining words in the section were then sorted by dispersion (column 6) and words below a certain dispersion figure were put back in the source list. So, for example, the headwords of all of the word families in the first 1000 list have a range of 98 or higher, a frequency of 68 or higher, and a dispersion of 93 or higher.*
 2. We include information on word families because their number correlates well to that of lemmas, which are slightly higher in amount.
 3. Lemmas have been estimated according to the proportion of real types and lemmas in several textbooks analysed. As an average, lemmas amount to ca. 70% of the types found in textbooks.
 4. RANGE does not calculate word families of lexical items outside the three ranges. The figure annotated here is the result of a probabilistic projection based on the proportion of word families vs. types in ranges 1, 2 and 3.
 5. Given that this graph has not been performed with the English version of Word but with the Spanish one, there appears a comma for decimals instead of a dot.

REFERENCES

- Alcaraz, G. 2011. *Quantity and Rate of Vocabulary Acquisition in the Context of Non-systematic input: Elementary Education Students of English as a Foreign Language*. PhD Dissertation. University of Murcia, Spain.
- Anderson, J. R. 2010. (1980). *Cognitive Psychology and its Implications*. (7th ed.). New York: Worth Publishers.
- Atkinson, R. C. and R. M. Shiffrin. 1968. "Human memory: a proposed system and its control processes". *The Psychology and Learning of Motivation* 2. Eds. K. Spende and J. Spence. New York: Academic Press. 89-195.
- Behrens, H. 2009. "Usage-based and emergentist approaches to language acquisition". *Linguistics* 47 (2): 383-411.
- Council of Europe. 2001. *Common European Framework of Reference for Language Learning, Teaching and Assessment*. Cambridge: Cambridge University Press.
- Criado, R. and A. Sánchez. 2012. "Vocabulary learning: the word frequency factor in textbooks and in the teacher's implementation of activities in the classroom". Paper

- presented at *EuroSLA22 - 22nd Annual Conference of the European Second Language Association*. Adam Mickiewicz University. Poznan, Poland. 5-8 September.
- Davies, M. and T. L. Face. 2006. "Vocabulary Coverage in Spanish Textbooks: How Representative is It?" *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Eds. N. Sagarra & A. J. Toribio. Somerville, MA: Cascadilla Proceedings Project. 132-143.
- DeKeyser, R. M. 2001. "Automaticity and automatization". *Cognition and second language instruction*. Ed. P. Robinson. New York: Cambridge University Press. 125-151.
- Donzelli, G. 2007. "Foreign language learners: words they hear and words they learn: a case study". *ELIA* 7: 103-125.
- Ellis, N. C. 2002. "Frequency effects in language processing". *Studies in Second Language Acquisition* 24: 143-188.
- Erman, B. 2007. "Cognitive processes as evidence of the idiom principle". *International Journal of Corpus Linguistics* 12 (1): 25-53.
- Ito, T. & J. Bauman. 1995. "The Acquisition of Basic Vocabulary by College-Age Japanese Students in an Intensive EAP Program". Paper given at the 34th Annual Meeting of the Japan Association of College English Teachers, Tokyo.
- Jiménez, R. M. and R. Mancebo. 2008. "Vocabulary Input in EFL Textbooks". *RESLA* 21: 147-165.
- Kapur, S., F. I. M. Craik, E. Tulving, A. A. Wilson & G. M. Brown. 1994. "Neuro-anatomical correlates of encoding in episodic memory: Levels of processing effect". *Proceedings of the National Academy of Sciences of the United States of America* 91 (6): 2008-2011.
- Laufer, B. & Rozovski-Roitblat, B. 2011. "Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination". *Language Teaching Research* 15 (4): 391-411.
- Mandelbrot, B. 1965. "Information Theory and Psycholinguistics". *Language: Selected Readings*. Eds. R. C. Oldfield & J. C. Marchall. London: Penguin Books. 550-562.
- Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. *Range* (computer software). [Available at <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>].
- Nation, I. S. P. Instructions BNC [Available at <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>].
- Newell, A., & P. S. Rosenbloom. 1981. "Mechanisms of skill acquisition and the law of practice". *Cognitive Skills and Their Acquisition*. Ed. J. R. Anderson. Hillsdale, NJ: Erlbaum. 1-55.
- Oxenden, C., C. Latham-Koenig & P. Seligson. 2004. *New English File Elementary Student's Book*. Oxford: Oxford University Press.

- Paradis, M. 1994. "Neurolinguistic aspects of explicit and implicit memory: implications for bilingualism in SLA". *Implicit and Explicit Learning of Languages*. Ed. N. C. Ellis. London: Academic Press. 393-418.
- Paradis, M. 2009. *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins.
- Redston, C. & G. Cunningham. 2005. *Face2Face Elementary Student's Book with CD-ROM/Audio CD*. Cambridge: Cambridge University Press.
- Roeper, T. 2007. "What frequency can do and cannot do". *Frequency Effects in Language Acquisition*. Eds. I. Gülzow & N. Gagarina. Berlin: Mouton de Gruyter. 23-50.
- Rosa, E. & Leow, R. P. 2004. "Computerized task-based instruction in the L2 classroom: The effects of explicitness and type of feedback on L2 development". *Modern Language Journal* 88: 192-216.
- Rundus, D. J. 1971. "Analysis of rehearsal processes in free recall". *Experimental Psychology* 88: 63-77.
- Sánchez A. & P. Cantos. 1997. "Predictability of word forms (types) and lemmas in linguistic corpora". *International Journal of Corpus Linguistics* 2 (2): 251-272.
- Schmitt, N., S. Grandage & S. Adolphs. 2004. "Are corpus-derived recurrent clusters psycholinguistically valid?" *Formulaic Sequences*. Ed. N. Schmitt. Amsterdam: John Benjamins. 127-152.
- Scholfield, P. 1991. "Vocabulary in course books. Living with an unstable lexical economy". *Proceedings of the 5th Symposium on the Description and/or Comparison of English and Greek*. Tassaloniki: Aristotle University. 12-32.
- Scott, M. 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Tomasello, M. 2003. *Constructing language: A usage-based theory of language acquisition*. Cambridge: Cambridge University Press.
- Tomasello, M., & P. Brooks. 1999. "Early syntactic development: A Construction Grammar approach". *The Development of Language*. Ed. M. Barrett. London: University College London Press. 116-190
- Ullman, M. T. 2004. "Contributions of Memory Circuits to Language: the Declarative / Procedural Model". *Cognition* 92: 231-270.
- Waring, R. & M. Takaki. 2003. "At what rate do learners learn and retain new vocabulary from reading a graded reader?" *Reading in a Foreign Language* 15 (2): 130-163.
- Zipf, G. K. 1935. *The Psychobiology of Language*. Boston: Houghton-Mifflin.
- Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, Massachusetts: Addison-Wesley.