

Vocabulary Selection and Word Repetitions in Beginner L2 Spanish Textbooks

Selección de vocabulario y número de repeticiones en libros de texto de ELE

LANI LÓPEZ BASTIDAS
CLAUDIA SÁNCHEZ-GUTIÉRREZ
UNIVERSITY OF CALIFORNIA, DAVIS

This article examines vocabulary selection and treatment in four beginner Spanish textbooks from major publishers in the US, specifically in the chapters that focused on: 1) regular verbs; and 2) food vocabulary. The corresponding chapters from each book were analyzed using AntConc and AntWordProfiler in order to get information about: 1) the number of times each different word is repeated; and 2) how frequent these words are in Spanish. Results revealed that lexical frequency is not the guiding principle in textbook vocabulary selection and that even highly frequent and relevant words are generally repeated only once per chapter, which indicates a deficient approach to vocabulary learning and teaching in mainstream L2 Spanish textbooks. Nonetheless, differences arise between the two types of chapters and pedagogical proposals should take such differences into account when making general suggestions about what vocabulary should be introduced in the textbooks and how it needs to be treated.

Keywords: *lexical frequency; vocabulary selection; vocabulary repetition; Spanish L2 textbooks; Spanish textbooks*

Este artículo examina la selección de vocabulario de cuatro libros de texto en español publicados por importantes editoriales en los Estados Unidos. El análisis se centra principalmente en los capítulos que presentan: 1) los verbos regulares; y 2) el vocabulario de la comida. Los capítulos correspondientes a los temas mencionados se analizaron en cada libro de texto usando AntConc y AntWordProfiler con el fin de obtener información sobre: 1) el número de repeticiones de cada palabra; y 2) la frecuencia de estas palabras. Los resultados revelan que la frecuencia léxica de las palabras no se toma en cuenta a la hora de seleccionar el vocabulario, ya que palabras de alta frecuencia se repiten una sola vez por capítulo. Esto muestra una deficiencia en el aprendizaje y enseñanza de vocabulario en los libros de texto de ELE. Sin embargo, hay diferencias notables entre los dos tipos de capítulos analizados, las cuales se deben tomar en cuenta al momento de decidir el tipo de vocabulario apropiado para cada capítulo.

Palabras clave: *frecuencia léxica; selección de vocabulario; repetición del vocabulario; libros de texto de ELE*

1. INTRODUCTION

1.1 *Lexical frequency*

Research in vocabulary acquisition and teaching has demonstrated the usefulness of lexical frequency lists in selecting the words that need to be prioritized in second language (L2) classrooms. Indeed, lexical frequency counts are extracted from corpora of authentic native speakers' conversations and written texts, which provide rich information about what words are more and less used and, thus, more and less useful for day-to-day real-life interactions. In this context, corpus data (and the resulting frequency counts) offer a sound selection criterion when deciding which words should be taught, and in what order, to facilitate the student's learning experience.

In terms of the exact number of words that should be prioritized in the L2 classroom, researchers propose that the first 3,000 to 4,000 words would be a reasonable goal. This conclusion comes from analyses of lexical coverage in different types of oral and written text. For instance, in his study of 2006, Nation found that with knowledge of the first 4,000 word-families in English the reader will be familiar with 95.06% of the words in a novel of over 120,000 tokens, such as *Lady Chatterley's Lover*. This means that readers will encounter one unknown word in approximately every 20 words, which does not significantly hinder overall comprehension but is not enough for fluent reading without the support of a dictionary (Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010). In a shorter literary text, such as *The Picture of Dorian Gray*, the first 3,000 word-families plus proper nouns reached 98.86% of text coverage, even though this type of high coverage, which does grant a more fluent reading experience, is generally achieved with more than 4,000 words in most texts. For instance, in his analysis of the movie *Shrek*, a vocabulary size of 7,000 word-families plus proper nouns was needed to reach 98.08% coverage, while 8,000 to 9,000 words would be necessary to reach 98% of coverage in newspapers or most novels.

Schmitt and Schmitt (2014), partly replicate these findings but they argue that the first 3,000 word-families, and not 4,000, in English should be considered as high frequency vocabulary whereas any word beyond that threshold would be classified as mid- or low- frequency. The authors offer several arguments for this proposal: 1) the amount of coverage accounted for by each rank of 1,000 words drops considerably after the first 3,000, such that learning the next 1,000 words (words 3,001 to 4,000) would contribute less than 1% of additional coverage for most texts. It is thus reasonable to propose that, after the first 3,000 words, learners should focus on the vocabulary that most directly responds to their specific communicative needs and interests; 2) The authors also noticed that graded readers for language learners rarely present more than 3,000 different words; and 3) the 3,000 most frequent words tend to correspond to the ones included in L2 learners dictionaries, which indicates that lexicographers' intuitions about what words should be taught in L2 classrooms generally match objective frequency criteria.

Although lexical frequency by itself is not enough to ensure learners' success, and it needs to be combined with other factors, this section aimed to demonstrate how the consideration of the 3,000 most frequent words as central to language learning and teaching should be common practice. However, frequency is not only important in terms of vocabulary selection but also in how it affects vocabulary learning, especially when it comes to incidental learning. The next

section will outline arguments to pay closer attention to how many times important words are repeated in the materials that are used in the classroom.

1.1.1 *Number of repetitions for incidental learning*

It is not completely clear how many repetitions of a word are needed to gain long term knowledge of it, yet there is no denying that frequency does play a central role in the lexical acquisition of L2 learners. Some authors suggest that 5 repetitions are sufficient while others propose that at least 10 repetitions of a word are needed in order to ensure proper future recognition and recall (Webb 2007; Pellicer Sanchez & Schmitt 2010; Sanchez-Gutierrez, Perez Serrano & Robles García, 2019). These differences may be due to what has constituted the definition of what knowing a word is in the different studies cited above. Indeed, lexical knowledge is a multifaceted construct that includes, among other aspects, knowledge about what a word means, how it relates to other words from the same morphological family or which words appear more frequently next to them.

In his 2007 study, Webb investigated the effects of encountering a series of target words 1, 3, 7, and 10 times in a text on the attainment of lexical knowledge and confirmed that knowledge of vocabulary increases as words are increasingly repeated in the text. However, while this may hold true for most aspects of vocabulary knowledge, some of them may require more or less encounters in the text to start developing. Concretely, Webb measured ten different aspects of productive and receptive knowledge, orthographic form, meaning and form, grammatical functions, syntax, and associations. The results indicated that with three encounters of a word, there was a significant gain of receptive knowledge for orthography, grammatical functions, and syntax to start developing. The productive knowledge of associations also presented a significant gain with three encounters, but seven encounters were necessary to observe significant gains in most other aspects of productive knowledge, such as orthography, or meaning and form. Overall, however, the author found that knowledge gains of every tested knowledge aspect were greater as the encounters kept increasing.

Similarly, Pellicer Sanchez and Schmitt (2010) tested different aspects of word knowledge, namely spelling recognition, word class recall, meaning recall and meaning recognition, after different amounts of word encounters: 1, 2-4, 5-8, 10-17, and 28+. One of the differences between this study and Webb's (2007) is that Webb used adapted readings which ensured that learners knew all the words in the reading except the 10 made up words used to measure word knowledge, while Pellicer Sanchez and Schmitt used a raw and authentic novel where the words expected to be unknown came from a different language. Thus, their participants were native Spanish speakers, and advanced learners of English with over 10 years of experience with the language. The participants read a book in English, with only a few words in a language that they were not familiar with. Their experience is, therefore, quite different from that of a beginner or intermediate L2 learners reading a text in their L2. Although, their results were similar to those of Webb (2007), the authors found that incidental learning occurred in all of the aspects of word knowledge tested but a substantial knowledge gain was overall noticed after 10-17 encounters. Measurable learning was gained in 28% of the cases or in 9.39 out of the 34 words tested. Meaning recognition was the aspect of knowledge with the most gain with 43% of the cases or 14.45 out of the 34 target words learned on average.

Furthermore, Sánchez Gutiérrez et al. (2019) tested whether the minimum number of occurrences could be limited to less than 10 encounters if the target words were presented in bold,

thus promoting more saliency. Their results confirmed that ten encounters with a word incremented the results in two vocabulary knowledge tests respectively focused on 1) meaning recognition and 2) spelling recognition. Interestingly, in an immediate post-test, only five repetitions were necessary to enhance student outcomes in the tests, but these results faded after two weeks in the delayed posttest. This indicates that, while five repetitions seem to offer sufficient encounters to notice a word in the short term, such noticing is not strong enough to maintain knowledge in the longer term. In terms of the relationship between the number of repetitions and typographical enhancement through bold-typing, words seen five times in bold were better recognized in the spelling recognition test than words seen five times without bolding. However, again, this difference did not hold in the long term and disappeared in the delayed post-test. These results demonstrate that frequency of exposure cannot be “tweaked” by simply increasing the degree of noticeability of the items. While this may “do the trick” in the short term, it does not promote longer term learning goals. This confirms the central relevance of frequency of exposure in the context of incidental vocabulary learning.

In this context, it seems necessary to better understand what current L2 students are exposed to in terms of the words they are expected to learn and how those are presented and repeated. One way to approach this is by studying to what extent beginner L2 textbooks 1) present vocabulary selected among the 3,000 most frequent words and 2) rely on providing several exposures and opportunities for practice of those most relevant words. Indeed, textbooks are at the origin of most of the classroom interactions and L2 courses’ curriculum (Allen, 2008; McDonough, Shaw & Masuhara, 2013; McGrath, 2013), thus vocabulary selection and presentation in those books should offer a good proxy of what is ultimately happening in the classroom.

2. TEXTBOOK VOCABULARY PRESENTATION

2.1 Vocabulary selection in textbooks

In 2006, Davies and Face conducted a study to analyze the type of vocabulary included in beginner and intermediate Spanish textbooks. The authors gathered six textbooks in total, three for beginner and three for intermediate Spanish. They analyzed the active vocabulary in the textbooks, namely the vocabulary lists included at the end of each chapter, and each word was subsequently lemmatized. Then, the frequency of each word was assessed with data from Davies’ (2005) *Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Results showed that about 93% of the first most frequent 500 words are included in at least one of the textbooks analyzed, 82% of the second 500, 67% of the third range of 500 words, and the number decreased drastically every 500 words. Furthermore, first year textbooks offered a better coverage of most frequent vocabulary, although, second year textbooks did not offer coverage of the less common words as it would be expected. A qualitative investigation into why textbooks might not rely on lexical frequency as much as would be expected for vocabulary selection revealed that textbooks are organized by semantic fields (e.g. food, health, travel, etc.), which allows for the inclusion of words that are not frequent among native speakers but that fit into these fields because they refer to concrete objects that are relevant for the field. For instance, when teaching the vocabulary of the clothes, words in that semantic field may not be extremely frequent, but they are relevant when one wants to cover the lexical items related to that specific field. On the other hand, this reliance on vocabulary lists based on semantic relatedness allows for the exclusion of words that are

frequent among native speakers because they do not fit into any of the semantic fields included in the chapters.

Godev (2009) replicated the study conducted by Davies and Face's (2006), however, unlike Davies and Face, who analyzed beginner and intermediate Spanish textbooks, Godev focused on five beginner Spanish textbooks published in the United States. She also used Davies (2006) to assess the frequency of active vocabulary as well as the words included in one reading per textbook to analyze reading coverage and opportunities for vocabulary acquisition. The results revealed that only between 23% and 30% of the words included at the end of the chapters pertained to the first 1,000 most frequent words in the frequency dictionary. The reading coverage analysis indicated that those most frequent words offered between 73% and 89% of coverage, which, according to the author, is not sufficient to promote vocabulary acquisition and fluent reading. Finally, the author pointed out that the most frequent 1,000 words in the textbooks do not agree with the 1,000 most frequent words in the frequency dictionary, and stated that the elementary Spanish textbooks analyzed in her study did not seem to be designed by using frequency dictionaries or any other systematic and explicit "guiding principle".

Finally, in a recent study by Sanchez-Gutierrez, Marcos Miguel and Olsen (2019) 16 textbooks were analyzed to further investigate the lexical frequency and lexical characteristics of the words included in the glossaries of eight beginner and eight intermediate textbooks. All textbooks were published in the US and broadly used in American universities. They proposed that a logical goal for a two-year language program, such as the ones usually in place at US universities, would be for students to be exposed to the first 1,500 during the first year of Spanish courses in college, while the second year could focus on the second 1,500, resulting in a knowledge (at least at the receptive level) of the first 3,000 most frequent words in Spanish. These numbers would promote an easy transition for students who would, then, enroll in culture, literature or linguistics courses in Spanish, as part of a major or minor in that language. Their results showed that none of the beginner textbooks analyzed in their study included the 1,500 most frequent words of the Spanish language. Similarly, the intermediate textbooks analyzed did not include most of the 1,501 to 3,000 most frequent words in the Spanish language. According to the authors, students were exposed to only 59% of the vocabulary from the first 1,000 most frequent words in the Spanish language, and about 33% of the second 1,000 most frequent words.

Overall, the results presented in this section show that vocabulary selection in L2 Spanish textbooks designed for first- and second- year college students is not based on the criteria proposed by vocabulary learning researchers, namely that students should be exposed to the 3,000 most frequent words of the language in their first years of L2 learning (Schmitt & Schmitt, 2014). While these results are compelling in themselves, an interesting second step is to know how many repetitions those most frequent words present in the textbooks. Concretely, are the high-frequency words included in the textbooks repeated enough times that students get sufficient exposure to learn them? The next section offers a review of studies that have looked at the number of repetitions that target words present in L2 textbooks.

2.1.1 Number of repetitions

Matsuoka and Hirsh (2010) analyzed all the words included in a higher-intermediate L2 English textbook to investigate how big a vocabulary is needed to read the whole textbook, and to analyze the number of repetitions offered in the text and the vocabulary learning opportunities it offers.

Their results indicated that knowledge of the first 2,000 most frequent words is needed for adequate comprehension, along with recognition of proper nouns, technical, textual, technology words and knowledge of academic or specialized words. Furthermore, the text did not offer sufficient opportunities for deepening word knowledge, since 33.3% of the first 2,000 words only appear once in the text, followed by 31% that appear 5 times, 21.2% that appear 7 times, and 12.1% that appear 10 or more times.

Martini (2012) analyzed a corpus comprised of three L2 English textbooks along with their workbooks, with the purpose of analyzing vocabulary frequency and recycling, high-frequency underrepresented words in the textbooks and representation of mid-frequency words. The frequency distribution results indicated that the textbooks included most of the high-frequency words, specifically 98% of the first 1k, 95% of the 2k and 85% of the 3k words. On the other hand, the recycling of the words did not show similarly positive results, with 94% of the words from the first 1k repeated 10 or more times, but only 65% of the 2k, and 27% of the 3k reaching that same criterion. According to the results, most mid-frequency 4k-8k are missing, and even high frequency words in the corpus are not recycled enough times to ensure learning, since only few of them occur 10 times or more. Finally, the author created a list containing the underrepresented high-frequency words in the corpus with the purpose of teaching them explicitly.

Interestingly, no study of this sort has been carried out with L2 Spanish textbooks to date, and those in English have mostly focused on intermediate proficiency levels. The current study aims to fill this gap in the literature by providing the first data about high-frequency word selection and repetition in beginner Spanish textbooks.

3. RESEARCH QUESTIONS

In this study we aim to answer the following two questions:

- 1) Is vocabulary selection based on frequency criteria in two different textbook chapters (i.e., the food chapter and the regular verbs chapter)?
- 2) How many of the words included in the textbooks which are among the first 3,000 most frequent words are repeated 10 times or more, thus providing more learning opportunities to students?

Responding to these questions will hopefully enhance instructors' and textbook writers' awareness about the importance of not only selecting relevant words but also providing enough recycling opportunities for students to develop a strong long-term knowledge of the most useful words of their L2.

4. METHOD

Four elementary Spanish textbooks from different American publishers were selected, which are presented in Table 1. The textbooks were directed to adults or university students. Two chapters were selected to further analyze in each textbook: 1) the chapter where all regular verbs in the present tense are introduced; and 2) the chapter where food vocabulary is presented for the first time. Both are subjects typically treated in all textbooks, and present different types of words, which may influence how vocabulary for those chapters is selected and treated. Concretely, while food vocabulary includes mostly nouns, the chapter about regular verbs will include mostly verbs. The food chapter will also be expected to be more centered on explicit vocabulary teaching

whereas the focus of the chapter on regular verbs is expected to be more on the conjugation of those verbs, rather than the learning of the meaning of the verbs themselves. By choosing chapters that are different in their scope and goals, we expect to observe differences both in selection criteria and number of repetitions. For this study we used textbooks that were sent by publishers. As we did not have access to the online platforms we limited the analysis to the written texts and exercises included in the books.

Table 1: Textbooks included in the study

Textbook	Publisher	Regular Verbs	Food
<i>Arriba</i>	Pearson	Chapter 2	Chapter 6
<i>Conectados</i>	Cengage	Chapter 2	Chapter 6
<i>Pura Vida</i>	Wiley	Chapter 2	Chapter 7
<i>Tu Mundo</i>	Mc Graw Hill	Chapter 3	Chapter 9

The two chapters selected from all textbooks were scanned and transformed into *.txt* format. Then, all the English words, the instructors' notes, and the numbers from the texts were removed from the text. After, an AntConc lemmatization list was used to lemmatize the words. In order to ensure accuracy, the automatic lemmatization was then manually revised and corrected, when necessary. AntConc 3.5.8.0 was also used to extract the number of repetitions of the words in every selected chapter in each textbook. Subsequently, we used Sublime Text to divide the words in four frequency groups, which were listed in separate text files:

- words with 1 repetition
- words with 2-4 repetitions,
- words with 5-9 repetitions
- words with 10 or more repetitions.

The frequency of the words in each group was assessed using the frequency lists from the *Corpus del Español* (Davies, 2002-...). The words from the corpus were then divided into frequency ranks of 1,000 words, so that rank 1 would include the first 1,000 most frequent words in the corpus, rank 2 would correspond to the second 1,000 most frequent words in the corpus, etc. These ranks were separated in different lists, so that we would obtain a list of word that was in rank 1, with 5-9 repetitions, rank 2 with 5-9 repetitions, rank 1 with 1 repetition, etc. This allowed us to organize words 1) by number of repetitions and 2) by lexical frequency, answering questions such as: how many times are words in the first rank repeated in each chapter? how frequent are the words that are repeated the most in each chapter?

Finally, the words that were not repeated enough times in the chapters were selected. The first step taken for this analysis was to compare the words from the textbooks to the words in the Spanish corpus using AntWordProfiler. Then, all the regular verbs that pertained to the first 3,000 most frequent words in the *Corpus del Español* which were not repeated 10 or more times in the verb chapters were selected. A similar approach was used for nouns denoting food, utensils, and meals in the chapters about food. This resulted in a list of words for each chapter that required more repetitions than the ones currently provided in the textbooks.

5. RESULTS

Up to this point, several aspects of vocabulary have been discussed, including aspects of lexical knowledge, repetition, frequency, and the steps taken to extract the vocabulary from the analyzed textbooks. Now, the results obtained from the comparison of the four textbooks analyzed here and the *Corpus del español* (Davies 2002) will be presented. This analysis was conducted in order to find how well these textbooks are preparing students to speak Spanish in real-world situations by using high frequent vocabulary, and if the textbooks promote long term vocabulary learning.

5.1 Vocabulary selection and repetition

Table 2 shows the total number of words included in both the regular verbs chapter and the food chapter, organized by number of repetitions. One specific detail to point at is the vast difference between the food and the regular verbs chapters, even from the same textbooks. Let's consider the following numbers: *Conectados* included 519 more words in the food chapter than in the verbs chapter, followed by *Pura vida* which included 515 more words, *Arriba* included 316 more words, and *Tu mundo* included 108 more words in the food chapter than in the verb chapter. This may be due to the verb chapter being focused more on conjugation and less on vocabulary lists, therefore limiting the number of items to learn but using those to create sentences or practice other skills. The food chapter, alternatively, may rely more on lists of words, such as those that can appear in real life, like menus or grocery lists. Therefore, naturally, those two chapters present different patterns in the number of items included and their treatment.

Another data point to consider about this table is the high number of words with a single occurrence. Almost half of the vocabulary included in the chapters is repeated only once, which certainly does not promote long term retention. On the other hand, the words with the highest potential for long term retention, as evidenced by their high number of repetitions, cumulate a very small percentage of the total words included in the chapters. This suggests that the inclusion of a word in a chapter does not necessarily mean that the word is given the necessary attention. However, given that this first general analysis focuses on all words and not specifically on the ones that are among the 3,000 most frequent in Spanish, it may well be that low-frequency words are the ones repeated the least and more frequent words are actually repeated enough times to promote their learning. This hypothesis will be explored in the next paragraphs, which specifically address the repetition of words among the first 3,000 in the two chapters that were included in the study.

Table 2: Total number of words per band of repetitions and type of chapter

Repetitions	<i>Arriba</i>		<i>Conectados</i>		<i>Pura Vida</i>		<i>Tu Mundo</i>	
	Verbs	Food	Verbs	Food	Verbs	Food	Verbs	Food
1	381	535	300	514	354	538	411	388
2-4	275	375	178	404	261	423	343	414
5-9	107	145	110	159	106	208	144	178
10+	94	118	112	142	111	178	127	153
Total	857	1,173	700	1,219	832	1,347	1,025	1,133

Table 3 presents the total number and percentage of words among the first 3,000 most frequent words (including proper nouns) of the Spanish language for the different number of repetitions listed in this study (i.e., 1, 2-4, 5-9, and 10+) in the regular verbs chapter. For instance, in *Arriba*, 260 words, or 68.25% of the words that are repeated only once, are among the first 3,000 most frequent words in Spanish. In that same book, 94 words, or 100% of the words that occur 10 or more times in the textbook, are among the 3,000 most frequent words in Spanish. While this result indicates that, in this book, words that are repeated many times are among the most frequent in Spanish, it also demonstrates that only a few of those highly frequent words actually are included in the textbook and repeated enough times. This trend is observable in most books, as the number of words in each band diminishes with the number of repetitions, namely: there are many more words that are repeated only once than words that are repeated 10+ times. This distribution may seem natural in a text for native speakers where no emphasis is made on vocabulary learning but might fall short when it comes to ensuring that learners get sufficient exposure to new words in their L2.

Similar patterns can be observed in all the books, with only little differences among them. In *Conectados*, only 104 out of 700 words included in the verb chapter are repeated ten or more times and are within the 3,000 most common words of the Spanish language. *Pura Vida* includes 211 or (59.6%) words with one repetition in the first 3,000 most frequent words, and less than half that number are repeated ten or more times in this chapter. This means that only 103 words out of the total 832 words included in the chapter promote long term vocabulary retention and are within the most frequent words among native Spanish speakers.

In sum, none of the textbooks seems to 1) base vocabulary selection for the regular verbs chapter on frequency criteria or 2) provide sufficient repetitions of relevant, frequent, words for learners to increase their lexicon through exposure to frequent words. Additionally, what could at first-sight be interpreted as apparent differences between books results from the higher numbers of total words in some of the textbooks, rather than genuine distinctions in their vocabulary selection and repetition patterns. For instance, *Arriba* includes over 13% more words within the first 3,000 most frequent Spanish words than *Tu Mundo*. However, if the total number of words included in those percentages is compared, *Arriba's* 260 words in the 3K range and *Tu Mundo's* 224 does not seem that significant and the observed difference in percentages was only due to the higher number of total words included in the regular verbs chapter in *Tu Mundo* than in *Arriba*. Furthermore, *Tu Mundo* is the textbook with the highest amount of words that are not among the first 3,000 most frequent words in Spanish, as it includes 320 of those lower frequency words, followed by *Pura Vida* with 232 words of these characteristics, and finally *Arriba* and *Conectados*, with 197 and 193 respectively. Nonetheless, those differences can be, once again, interpreted as a byproduct of the significantly higher number of total words in *Tu Mundo* and not as revealing any underlying pedagogical design decision in any of the textbooks.

Table 3: Distribution of high frequency words per number of repetitions in the regular verbs chapters

	Rep.	<i>Arriba</i>	# of words	<i>Conectados</i>	# of words	<i>Pura Vida</i>	# of words	<i>Tu Mundo</i>	# of words
P. nouns	1	16.54%	63	16.33%	49	10.17%	36	11.68%	48
1-3K		51.71%	197	46.33%	139	49.43%	175	42.83%	176
Total		68.25%	260	62.66%	188	59.6%	211	54.51%	224
P. nouns	2-4	16.36%	45	13.48%	24	6.13%	16	11.08%	38
1-3K		60.73%	167	59.56%	106	67.82%	177	58.6%	201
Total		77.09%	212	73.04%	130	73.95%	193	69.68%	239
P. nouns	5-9	16.82%	18	5.45%	6	3.77%	4	7.64%	11
1-3K		71.03%	76	71.81%	79	83.96%	89	79.17%	114
Total		87.85%	94	77.26%	85	87.73%	93	86.81%	125
P. nouns	10+	5.32%	5	0.89%	1	4.50%	5	4.72%	6
1-3K		94.68%	89	91.96%	103	88.29%	98	87.41%	111
Total		100%	94	92.85%	104	92.79%	103	92.13%	117
Total 3K			660		507		600		705
Total chapter			857		700		832		1025
Total not in 3K			197		193		232		320

Table 4 presents the total number and percentage of words included in the first 3,000 most frequent words (including proper nouns) of the Spanish language for the different bands of repetitions listed in this study (i.e., 1, 2-4, 5-9, and 10+) in the food chapter. While the total number of words in the food chapters is overall significantly higher than in the regular verb chapters, the distribution of those words across repetition bands is similar in both types of chapters. In *Arriba*, only 102 words are both among the 3,000 most frequent in Spanish and are repeated 10+ times. *Conectados* includes 132 words with these characteristics, *Pura Vida* and *Tu Mundo* include 153 and 112 of those words respectively. Given that the chapters contain between 1,133 and 1,347, these numbers indicate that only around 10 percent of the words in the food chapters were relevant from the perspective of their high lexical frequency and presented enough repetitions to favor long term retention.

Table 4: Distribution of high frequency words per number of repetitions in the food chapters

	Rep.	Arriba	# of words	Conectados	# of words	Pura Vida	# of words	Tu Mundo	# of words
P. nouns	1	4.86%	26	7.98%	41	4.28%	23	6.68%	26
1-3K		57.56%	308	46.89%	241	50.93%	274	49.87%	194
Total		62.42%	334	54.87%	282	55.21%	297	56.55%	220
P. nouns	2-4	4.53%	17	6.93%	28	1.89%	8	3.86%	16
1-3K		66.13%	248	59.66%	247	58.63%	248	48.07%	199
Total		70.66%	265	66.59%	269	60.52%	256	51.93%	215
P. nouns	5-9	2.07%	3	2.52%	4	3.85%	8	3.93%	7
1-3K		60.7%	88	72.96%	116	63.94%	133	57.3%	102
Total		62.77%	91	75.48%	120	67.79%	141	61.23%	109
P. nouns	10+	1.69%	2	1.41%	2	2.25%	4	1.96%	3
1-3K		84.74%	100	91.54%	130	83.71%	149	71.24%	109
Total		86.43%	102	92.95%	132	85.96%	153	73.2%	112
Total 3K			792		803		847		656
Total chapter			1173		1219		1347		1133
Total not 3K			381		416		500		477

However, some differences are observable across textbooks and, contrary to what happened in the regular verb chapters, these cannot simply be interpreted as a result of the higher number of total words in one chapter over the others. For example, 70.66% of the words that are repeated 2-4 times in *Arriba* are within the 3,000 most frequent Spanish words but in *Tu Mundo* this number descends to 51.93%. Thus, these two textbooks exhibit a difference in this repetition band of around 19% with a somewhat similar number of total words in both chapters: 1,173 in *Arriba* and 1,133 in *Tu Mundo*. A similar pattern is observed between *Tu Mundo* and *Conectados*, in the 10+ repetitions band, where a difference of 20% is also observed. Although results at the extremes (one repetition and 10+ repetitions) are similar across textbooks, some distinct tendencies are observable at the intermediate bands. This indicates that some books, do not reach the 10+ repetitions for some of the most frequent words in the food chapter, but they do present more repetitions than other books, which tend to present frequent words only once.

These differences have pedagogical implications, in the sense that instructors using one textbook or the other will need to compensate for the insufficient repetitions to a different extent and focus on specific words. These words, which would require more attention in the classroom because they are 1) relevant for the students due to their high frequency and 2) not repeated enough to promote long-term retention, will be coined as *underrepresented words* for the purposes of this study. The next section will present the amount and distribution of these words in the two types of chapters analyzed here.

5.2 Underrepresented vocabulary

In order to carry out this analysis, we included only the list of regular verbs in the chapter about the verbs and only the nouns related to food names or utensils for the food chapter. This was a way to approximate the active vocabulary that the textbook writers considered as being the central focus of those particular chapters. Because the previous analyses were performed including all the words from the chapters, including function words, or words not directly related to the theme at hand in the chapter, we now want to turn more specifically to the words that are supposed to be highlighted in each chapter. This analysis aims to shed light on the actual number of active vocabulary words that instructors should put more emphasis on during their classes when working on these two chapters.

As can be seen in Table 5, the difference in the total number of underrepresented words between the verbs chapter and the food chapter is significant across all books. For instance, in *Pura Vida* 27 highly frequent verbs were repeated only once, 25 were repeated 2-4 times and 10 were repeated 5-9 times, making it a total of 62 underrepresented verbs. This number drastically drops in the food chapter, with only 16 underrepresented words in that same book. This difference is due to the fact that fewer active vocabulary words in the food chapters were among the 3,000 most frequent words in Spanish. Indeed, fruit or vegetable names tend to present a lower frequency than regular verbs of action, such as *comer* [to eat] or *beber* [to drink].

Table 5: Underrepresented Words by Chapter and Frequency Bands.

	<i>Arriba</i>		<i>Conectados</i>		<i>Pura Vida</i>		<i>Tu Mundo</i>	
	Verbs	Food	Verbs	Food	Verbs	Food	Verbs	Food
1 repetition								
1K	13	0	3	0	17	0	8	1
2K	9	1	3	0	6	1	10	0
3K	1	3	3	0	4	2	2	4
Total	23	4	9	0	27	3	20	5
2-4 repetitions								
1K	13	0	6	0	17	1	20	0
2K	0	3	3	3	7	1	5	3
3K	0	3	1	7	1	4	6	2
Total	13	6	10	10	25	6	31	5
5-9 repetitions								
1K	11	2	15	1	5	0	10	1
2K	3	3	3	5	3	3	3	4
3K	0	3	1	0	2	4	1	4
Total	14	8	19	6	10	7	14	9
Total all chapters	50	18	38	16	62	16	65	19

6. DISCUSSION

The aim of this study was to answer the following questions: 1) How many words in the regular verbs and food chapters are among the 3,000 most frequent words in Spanish? and 2) How many

of these words are repeated 10 times or more? In the next sections, we will address these questions, based on the results of our analyses, compare them to those in previous literature and propose avenues for improvement in the area of vocabulary selection and teaching.

The results gathered from both types of chapters reveal that the vocabulary included in the textbooks is not selected based on frequency criteria. This conclusion is reached based on the great amount of infrequent vocabulary included in the chapters, which coheres with the findings in previous studies (Davies & Face, 2006; Matsuoka & Hirsh, 2006). Furthermore, even when a chapter included many high frequent words, these were not all treated equally, and the majority of them did not receive enough attention to promote long term retention. This seems to indicate that 1) no clear vocabulary selection criterion was used that took lexical frequency into account and 2) no specific guidelines seem to be established about the number of times that the most relevant active vocabulary should appear in a chapter.

While this trend was observable in both chapters, some significant differences did emerge between them. Specifically, in terms of the number of underrepresented words, the food chapters included less of those than the regular verb chapters. This is partly due to the lower frequency of food names as compared to highly frequent action verbs, as well as the highly context-dependent use of the former. Indeed, most regular verbs in the chapters describe actions that may occur in several different situations of life, and several times a day while food-related events only happen a limited amount of times per day. And even when a verb refers to those food-related events, it can be repeated more times than the names of the foods themselves. Let's consider the example of the verb *comer* [to eat] and the food names *tomate* [tomato], *carne* [meat] or *pescado* [fish]. Every time you eat any of those edibles, you will use the same verb to describe the action of eating each three of them, but the name of the specific food that you will be eating will be repeated only once for every time you actually eat it. To make matters worse, words related to food tend to present high dialectal variety. Thus, for example, all chapters will include a word that refers to the English *banana* but some of them will include the Spanish *plátano* and others the also Spanish *banana*. Some books will even include both options, which will, again, reduce the overall frequency of the words and their number of repetitions in the textbooks.

The differences between both chapters also arise from their respective learning goals. The food chapter requires students to learn longer lists of words that they can then use for real-life tasks, such as ordering food in a restaurant. In these chapters, learning new words is a goal per se, that will then offer students the tools they need to complete real-life tasks. The regular verbs chapters, alternatively, aim to teach students to conjugate regular verbs in the present tense. Thus, the emphasis is not on learning a long list of verbs but rather on conjugating those in fill-in-the-blank exercises first and, eventually, in communicative activities. The distinct goals of each chapter type may be key to how many times each word is repeated. Indeed, in the regular verbs chapter, around 90% of the words that were repeated 10+ times were among the first 3,000 most frequent ones, whereas this percentage only reached 90% in one of the books in the food chapters and is at 73% in another one. This indicates that fewer of the highly frequent words present in the chapters were repeated enough times in the food chapter than in the verbs chapter, which makes sense given that there were less frequent words in the first place and the activities did not necessarily promote the repetition of the same words. The regular verbs chapter, on the other hand, included more high-frequency words that were repeated more often because the goal of the chapter was for the students to conjugate those in the complete inflectional paradigm of the present tense.

In this context, given the different inherent characteristics and goals of these chapters, it is possible that adopting a single vocabulary selection criterion for both could be somehow flawed. If a purely frequency-based approach is used to choose the words that one should be teaching and paying more attention to in the classroom, the recommendation for both chapters would be to stick to the highly frequent words that are relevant to the subject and repeat them more. However, while this is reasonable for a chapter where the active vocabulary does include several highly frequent words, it would definitely limit learners to a very poor lexical environment when it comes to food names vocabulary, which are generally less frequent. For instance, they would not be able to learn words that would appear in most menus across the Spanish-speaking world, such as *filete [steak]* or *pescado [fish]*, because they are not technically among the first 3,000 most frequent words in Spanish. The logical question that arises here is whether frequency alone can be used as the guiding principle for all the subjects or tasks that are encountered in the language classroom. If the final aim is for students to communicate in meaningful contexts, using mostly high-frequency words makes sense in most contexts but can be extremely limiting in others, where the tasks at hand (e.g., ordering food in a restaurant, making a groceries list, following a recipe) require the knowledge, and use, of lower frequency words.

In cases, such as the food chapter, where frequency cannot be the unique selection criterion, effort could still be made to include all relevant words that are among the first 3,000, even though they do not represent the totality of the active vocabulary that students are expected to learn for that particular chapter. Other words may be chosen from realia, such as menus or pictures from actual grocery stores, thus using the relevance of the words for real-life situations as a second criterion that complements frequency in a useful way. Those words should then all be repeated thoroughly and series of exercises should ensure that students get multiple chances to practice each of the words. Finally, a better integration of grammatical and lexical contents may also offer more natural opportunities for repeated practice. For instance, the fact that verbs are not learned for the sake of learning their meaning but rather to be used in sentences with an adequate conjugation makes it possible for each verb to be used in multiple sentences, in different inflectional forms. While nouns do not present such inflectional richness, they could be used several times in various sentences with different verbs, thus promoting additional exposures. Essentially, this would mean that words with a simpler flectional paradigm could be taught in the same chapter with other words which do present such varied endings. In this way, words could be repeated instead of just being used once or twice in a single activity, which is what happens when a word only appears in a menu on the first page of the chapter and is never mentioned again.

7. CONCLUSION AND LIMITATIONS

The goal of the present study was to explore 1) whether the vocabulary included in elementary Spanish textbooks is selected based on frequency criteria, and 2) if highly-frequent words in these textbooks are repeated enough times to promote long term retention. Results indicate that the vocabulary included in all eight of the chapters analyzed here is not selected based on frequency and that only a very small number of high-frequency words is repeated enough times to ensure long term retention. Our data also revealed that not all chapters have the potential of benefiting from frequency dictionaries alone when it comes to vocabulary selection. This was evidenced in the food chapter, where reliance on only frequency would have limited students to such a short list

of words that it would have prevented them from actually being able to properly communicate in situations that require the knowledge of food vocabulary.

In sum, we propose that textbooks should ideally be revised to substitute low-frequency vocabulary by more relevant vocabulary for the beginner levels, such as high frequent vocabulary. The fact that most words present only one repetition should also be revised, as it indicates that no attention is currently being paid to actually favoring the learning of most of the words that are included in the textbook chapters. It is shocking to notice, for example, that in the food chapters three of the four textbooks analyzed included over 500 words with only one repetition, but the textbook with the highest number of words with ten+ repetitions only included 153 words. Words with one repetition are not intended to (or will most probably not) be learned by students, then why not try to find a better way to reduce the words with only one repetition and increase the words with 10 or more repetitions, especially among high-frequency and relevant vocabulary?

While the present study offers relevant insights on the treatment of highly-frequent vocabulary, we believe that follow-up studies would be necessary that take two new aspects into account. On the one hand, the study of these highly frequent words in the audiovisual materials included in the online platforms of the books would offer interesting additional information. However, this entails complicated discussions with publishers concerning the types of access that researchers can have to those materials. On the other hand, the study focused on single words and no attention was paid here to collocations. This is due to the fact that no information is yet available about the frequency of collocations. While frequency lists exist for words, no such list exists, to the best of our knowledge, for collocations. Nonetheless, future studies would greatly benefit from this additional source of data and more studies are definitely necessary that include this type of lexical chunks in their analyses.

REFERENCES

- Allen, H. W. (2008). Textbooks, materials and foreign language teaching: Perspectives from the classroom. *NECTFL Review*, 62, 5-28. Retrieved from <https://www.nectfl.org/wp-content/uploads/2016/06/review62.pdf#page=11>
- Davies, M. (2002-...). Corpus del español. Retrieved from <https://www.corpusdelespanol.org/resources.asp>
- Davies, M. & Face, T. L. (2006). Vocabulary coverage in Spanish textbooks: How representative is it? In *Selected proceedings of the 9th Hispanic Linguistics Symposium* (pp. 132-143). Retrieved from <http://www.lingref.com/cpp/hls/9/paper1373.pdf>
- Godev, C. B. (2009). Word-frequency and vocabulary acquisition: an analysis of elementary Spanish college textbooks in the USA. *Revista de Lingüística Teórica y Aplicada*, 47(2), 51-68. doi: [10.4067/S0718-48832009000200004](https://doi.org/10.4067/S0718-48832009000200004)
- Keck, C. (2012). Corpus linguistics in language teaching. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*, Malden, MA: Blackwell.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From Humans Thinking to Thinking Machines* (pp. 316-323). Clevedon: Multilingual Matters.

- Laufer, B. & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15-30. Retrieved from <https://files.eric.ed.gov/fulltext/EJ887873.pdf>
- Martini, J. (2012). *High frequency vocabulary in a secondary Quebec ESL textbook corpus*. PhD dissertation, Concordia University. Retrieved from https://spectrum.library.concordia.ca/974698/1/Martini_MA_F2012.pdf
- Matsuoka, W. & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22(1), 56-70. Retrieved from <https://files.eric.ed.gov/fulltext/EJ887877.pdf>
- McGrath, I. (2013). *Teaching Materials and the Roles of EFL/ESL Teachers*. London: Bloomsbury.
- McDonough, J., Shaw, C. & Masuhara, H. (2013). *Materials and Methods in ELT. A Teacher's Guide*. Oxford: Wiley-Blackwell.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. doi: [10.3138/cmlr.63.1.59](https://doi.org/10.3138/cmlr.63.1.59)
- Pellicer-Sánchez, A. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do "Things Fall Apart"? *Reading in a Foreign Language*, 22(1), 31-55. doi: [10.1017/S0272263115000224](https://doi.org/10.1017/S0272263115000224)
- Sánchez Gutiérrez, C., Pérez Serrano, M. & Robles García, P. (2019). The effects of word frequency and typographical enhancement on incidental vocabulary learning in reading. *Journal of Spanish Language Teaching*, 1-18. doi: [10.1080/23247797.2019.1590000](https://doi.org/10.1080/23247797.2019.1590000)
- Sánchez-Gutiérrez, C. H., Marcos Miguel, N. & Olsen, M. (2019). An analysis of vocabulary coverage and lexical characteristics in L2 Spanish textbooks. In P. Ecke & S. Rott (Eds.), *Understanding Vocabulary Learning and Teaching: Implications for Language Program Development. American Association of University Supervisors, Coordinators, and Directors of Language Programs (AAUSC) Volume 2018*. Cengage, 78-98. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/69783/2018_05.pdf
- Schmitt, N. & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503. doi: [10.1017/S0261444812000018](https://doi.org/10.1017/S0261444812000018)
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, 28(1), 46-65. doi: [10.1093/applin/aml048](https://doi.org/10.1093/applin/aml048)