

4. Results and Analysis

4.1 Overall Coverage

The first set of data to be presented and analyzed is the overall coverage of the vocabulary in terms of frequency ranges in Davies' (2006) frequency dictionary. Table 3 gives information of the number of lemmas that each textbook presented for each of the ten, 500 lemma ranges as determined by the frequency dictionary. The coverage, the percentage out of a total of 500 for each range, of each textbook is provided to the right of the amount of lemmas in that respective range.

Table 3.

Overall coverage of top 5,000 lemmas by range and textbook

Range	<i>Pido la palabra</i>		<i>¡Estoy listo!</i>	
	no.	%	no.	%
500	442	88.4	313	62.6
1000	340	68	181	36.2
1500	289	57.8	157	31.4
2000	217	43.4	109	21.8
2500	202	40.4	95	19
3000	155	31	78	15.6
3500	143	28.6	73	14.6
4000	118	23.6	37	7.4
4500	78	15.6	43	8.6
5000	74	14.8	33	6.6
TOTAL	2058	41.2	1119	22.4

The final line represents the total number of the 5,000 most frequent lemmas represented by the textbooks and the respective overall coverage of those 5,000 most frequent lemmas.

Table 3 shows that the first range, which represents the 500 most frequent lemmas, is well covered by both textbooks. *Pido la palabra* (1998), for example lacks only 58 of the words in this first frequency range. For a beginning level course, this high level of coverage makes sense. Without much, if any background in Spanish, the target learners need to have the basics of the language to successfully communicate in a second language environment. Also, as Lara (1993) claims, the first 1,451 words in Mexican Spanish represent a coverage of 75% of all cultural linguistic utterances (p. 11). Through the third range where this number lies, *Pido la palabra* presents more than half of the 1,500 lemmas. 1,071 of these lemmas are presented with a coverage of 71.4%. *¡Estoy listo!* (2003), on the other hand, presents 651, or 43.4%, of the same top three ranges. In terms of total numbers, *Pido la palabra* presents almost twice as many of the top 5,000 lemmas than *¡Estoy listo!*.

While *¡Estoy listo!* (2003) presents only slightly more than half of the total lemmas that *Pido la palabra* does, it shows similar priority towards the top ranges in relation to less frequent ranges. All three of the top three ranges of 500 cover a larger percentage of words than the overall coverage of 22.4% of the most frequent 5,000 lemmas. This is similar to the results of Davies and Face (2006), which found that the further down the scale of ranges, the less coverage there was. Besides this general trend, these data cannot be directly compared to the Davies and Face study because their article does not distinguish between textbooks at this level of coverage. Instead the researchers compiled all of the words from each level (first and second year), as if a learner were to use all three of the textbooks of one level at one time.

The current study also compared the total number of words that were presented in each textbook with how many of those words were in the list of the top 5,000 lemmas. Table 4 shows the total number of lemmas presented in the textbooks, the number of lemmas presented that are in the frequency dictionary, the number of lemmas presented that are not in the frequency dictionary, and the percentage of in-dictionary lemmas relative to the total number of lemmas presented.

Table 4.

Coverage by textbook: percentage of words in frequency dictionary

	Total no. lemmas	no. + dictionary	no. - dictionary	% + dictionary
<i>Pido la palabra</i>	2924	2058	866	70.38
<i>¡Estoy listo!</i>	1438	1119	319	77.82

The data here show that the majority of the vocabulary presented in both textbooks is frequent enough to be in the frequency dictionary. Although *¡Estoy listo!* (2003) does not present as many total lemmas as *Pido la palabra* (1998), the vocabulary that it does present is more likely to be encountered by the target learners in the “real world.” This could be due to the fact that the linguistic input for learners presented in *¡Estoy listo!* was written in a possibly more controlled way than the authentic readers of *Pido la palabra*. Authentic texts, like those in *Pido la palabra*, might be more likely to present more infrequent words than texts written with a beginning level Spanish learner in mind.

The Davies and Face (2006) study also showed these percentages of words in first year SFL language textbooks. The total numbers of all active vocabulary lemmas in the first year textbooks were 2,218, 1,616, and 3,217. The percentages of these words that were also in the frequency dictionary were 85%, 81%, and 78%, respectively. All three of these

figures are higher percentages than *¡Estoy listo!* and *Pido la palabra*. However, these comparisons across the SSL and SFL textbooks are not completely valid because Davies and Face only extracted active vocabulary, and the current study extracted all presented vocabulary. It might be the case, for instance, that had the passive vocabulary been included in the Davies and Face study, more infrequent lemmas would be included, lowering this percentage. On the other hand, had only active vocabulary been extracted in the two SSL textbooks studied, the amount of total lemmas from *¡Estoy listo!* would have been even less. Also, although *¡Estoy listo!* has a relatively high coverage rate, it only presents a total of 1,438 lemmas. This is significant because it shows that the textbook may not be providing enough input for the student and may need to be lexically supplemented by other materials. According to Renouf's (1984) study of EFL textbooks (as cited in Sinclair & Renouf, 1988), a textbook with 1,438 total lemmas would be at or near the bottom of the list of amount of lemmas presented. Such a comparison should not be seen as totally valid, however, because in this type of textbook analysis one can neither know how much is done with the vocabulary presented, how often it is used in the classroom, nor how well it is learned.

Another way to examine a textbook's vocabulary coverage was proposed by Davies and Face (2006). This measurement of coverage examines coverage in a different way. These researchers give the example of a quantitatively ideal textbook:

Suppose that a textbook has N number of words, e.g. 1,300 words. In the "best of all worlds" scenario, these 1,300 words would correspond to words #1-1,300 in the frequency dictionary. In other words, it would be as though the textbook vocabulary corresponded exactly to the listing in the dictionary. (p. 8)

The total numbers of vocabulary entries for both textbooks in this study are shown above in Table 4. *Pido la palabra* (1998) presents a total of 2,924 lemmas. Of those, 1,619 correspond to the words 1-2,924 in the frequency dictionary. This means that 55.37% of the 2,924 lemmas in *Pido la palabra* relate to words 1-2,924 in the frequency dictionary. *¡Estoy listo!* presents a total of 1,438 lemmas, and 634 of them correspond to the words 1-1,438 in the frequency dictionary. The coverage in this case is 44.09%. These numbers are significant because they mean that 44.63% and 55.81% of the lemmas in these respective textbooks do not relate to the first 2,924 and 1,438 words, respectively, in the frequency dictionary. This means that for the number of lemmas that each textbook presents, they both have relatively low coverage of the most frequent lemmas up to those respective numbers, which shows that infrequent entries may be taught at the expense of more frequent entries. It should be noted, however, that this is an artificial construct of the ‘ideal’ vocabulary in a textbook. This construct only addresses frequency and neglects aspects of semantic fields or themes. In terms of affective factors, teaching only frequent words, which include most function words, might be boring for both a teacher and his or her students.

Davies’ (2005) claims that the first 1,000 most frequent lemmas in Spanish constitute up to 80% of written and 88% of spoken Spanish. This number is slightly different than Lara’s (1993) assertion that the first 1,451 words represent 75% of Spanish, possibly due to differences in lemmatization, available corpus data, and frequency calculation. However, either way, it is clear from both of these sources that the first 2,000 most frequent lemmas are critical to producing and understanding Spanish. If nearly half or more of those critical entries are not presented to a first-year student, there could be critical problems dealing with everyday communication.

The previous tables, however, do not describe what kinds of lemmas are being presented and at what frequency. Similar to Table 4 one can also determine the same raw number coverage based on syntactic category. Table 5 relates the total amounts of lemmas with the amount of top 5,000 lemmas that each textbook presents in respect to syntactic categories.

Table 5.

Coverage by textbook and syntactic category: percentage of words in frequency dictionary

Pido la Palabra

	Total no. lemmas	no. + dictionary	no. - dictionary	% + dictionary	% - dictionary
Nouns	1570	1030	540	65.61	34.39
Verbs	521	435	86	83.49	16.51
Adjectives	624	404	220	64.74	35.26
Adverbs	101	90	11	89.11	10.89

¡Estoy Listo!

	Total no. lemmas	no. + dictionary	no. - dictionary	% + dictionary	% - dictionary
Nouns	797	584	213	73.27	26.73
Verbs	214	199	15	92.99	7.01
Adjectives	289	208	81	71.97	28.03
Adverbs	56	47	9	83.93	16.07

This table shows that in both textbooks, the verbs and adverbs that are presented are much more likely to be frequent than the nouns and adjectives that are presented. In the case of *Pido la palabra* (1998), nouns and adjectives are both more than twice as likely to be too

infrequent to be present in Davies' (2006) 5,000 lemma frequency list than verbs and adverbs. It is shown here that the number of nouns presented for both textbooks is more than the total of all the other syntactic categories combined. It is also of note the percentages of adverbs and verbs that *¡Estoy listo!* (2003) presents that are in the frequency dictionary. This shows that, although only a small number of adverbs and verbs were found in the textbooks, the ones that were presented were very likely to be frequent. Nouns and adjectives may have more likely to be concrete and fit better into a themed chapter than verbs or adverbs. This difference may possibly cause textbook authors to use more infrequent, theme-specific nouns and adjectives to fit existing conceptual ideas for a lesson.

The data from Table 5 may be potentially misleading, however, because they refer to a raw numbers and not relative coverages. In terms of raw numbers, *Pido la palabra* (1998) presents 940 more nouns than adverbs from list of the top 5,000 lemmas. This difference may not be relevant in the discussion of syntactic category coverage, because there are many more nouns in the frequency list than verbs. Table 6 represents the overall coverage of content-word lemmas (nouns, verbs, adjectives, and adverbs) across the ten different frequency ranges, relative to the total number of those respective items in the frequency list as categorized by syntactic category. These percentages compare the number of in-dictionary entries presented with the number of that respective syntactic category in the top 5,000 lemmas. For example, of 1,030 of the 2,511 nouns (85.98 %) in the top 5,000 lemmas were presented by *Pido la palabra* (1998).

Table 6.

Vocabulary coverage in percentage by frequency range and syntactic category

<i>Pido la palabra</i>					<i>¡Estoy listo!</i>				
Range	N	V	Adj	Adv	Range	N	V	Adj	Adv
500	85.98	90.44	84.88	93.33	500	57.93	63.97	62.79	57.78
1000	70.89	68.75	62.82	42.31	1000	43.04	29.86	28.21	19.23
1500	60.64	49.58	59.43	60.00	1500	34.14	19.33	37.74	25.00
2000	46.25	35.59	47.75	13.33	2000	25.30	15.25	20.72	13.33
2500	42.56	32.56	37.14	46.15	2500	21.45	6.98	17.14	23.08
3000	35.50	20.59	30.09	17.65	3000	19.08	7.84	13.27	11.76
3500	29.15	23.66	28.07	33.33	3500	18.82	5.38	11.40	0.00
4000	25.86	18.69	21.50	30.43	4000	8.37	4.67	7.48	8.70
4500	16.47	17.39	14.58	0.00	4500	12.45	3.26	6.25	0.00
5000	17.88	6.33	13.39	11.11	5000	8.03	1.27	4.72	11.11
TOTAL	41.02	40.43	37.03	43.69	TOTAL	23.26	18.49	19.07	22.82

According to the total coverage, both *Pido la palabra* (1998) and *¡Estoy listo!* (2003) are relatively consistent in representation across syntactic categories. The difference between the percentages of the most-covered and the least-covered categories (adverbs and adjectives, respectively) is only 6.66% and 4.77%, respectively.

4.2 Under-representation

This section discusses the lemmas that were under-represented in *Pido la palabra* (1998) and *¡Estoy listo!* (2003). This particular analysis is needed to better understand what

kinds of words are frequent in native Spanish speech and writing, but are not represented by the first year SSL textbooks. As shown in Table 3, *Pido la palabra* (1998) and *¡Estoy listo!* (2003) cover 442 and 313 of the first range of 500 in the frequency list, respectively. This means that the textbooks do not represent 58 and 187 of the very frequent lemmas (see Appendix A for complete list of these under-represented, highly frequent entries).

To better understand the under-representation of syntactic categories for each textbook, Table 6 can also be read to show the percentages of frequent lemmas that are not covered. For example, *¡Estoy listo!* (2003) only presents 18.49% of the total number of verbs in the most frequent 5,000 lemmas, meaning that 81.51% of the frequent verbs are not represented. Table 7 uses these data to show the total numbers and the percentage of under-represented frequent lemmas.

Table 7.

Under-representation based on syntactic category

Pido la Palabra

	Total no. lemmas in top 5,000	No. of top 5,000 lemmas NOT presented in textbook	Percentage NOT represented
Nouns	2511	1481	58.98 %
Verbs	1076	641	59.57 %
Adjectives	1091	687	62.97 %
Adverbs	206	119	56.31 %

¡Estoy Listo!

	Total no. lemmas in top 5,000	No. of top 5,000 lemmas NOT presented in textbook	Percentage NOT represented
Nouns	2511	1927	76.74 %
Verbs	1076	877	81.51 %
Adjectives	1091	883	80.94 %
Adverbs	206	159	77.18 %

Table 7 shows both how although there are large differences in the number of entries presented in terms of syntactic category and how the coverage of those syntactic categories is relatively consistent for each textbook.¹ In the Davies and Face (2006) study, adverbs were determined to have significantly less coverage than nouns, verbs, and adjectives. Interestingly, however, adverbs in the two textbooks in this study have the best and second-best coverage of the content-word syntactic categories. This may be unexpected not only

¹ Refer to Appendix E for a graphical representation of the segmentation of the top 5,000 lemmas

because of the results from the replicated study in which adverbs were found to be significantly under-represented (p. 9) but also because most adverbs in Spanish, like adjectival forms of verbs, can easily be formed using a simple suffixation rule: adjective + *mente*. According to Davies' (2006) frequency dictionary, of the 206 adverbs in the most frequent 5,000 lemmas, 116 are of this particular composition. If a part of speech had to be more under-represented than the others as active vocabulary, it thus would make sense for it to be adverbs. However, the presentation of adverbs might be common in passive vocabulary because a student may easily understand their meaning even if never presented out of context as long as the adjectival base was already known.

4.3 Over-representation

Another way to describe the lemmas that were presented by these Spanish language textbooks is to show what kinds of words were over-represented. This is an important measurement, as a textbook author should want to be efficient with the vocabulary presented. It would not benefit students to spend time and energy learning infrequent words at the expense of a significant amount of unrepresented frequent words. In this case, over-represented lemmas were operationalized as those not present in Davies' (2006) frequency dictionary. Table 4 shows how many of the total number of lemmas presented in the textbooks were also found in the frequency dictionary. Those lemmas that were in the textbooks but not in the frequency dictionary were considered over-represented. Also, the final column in Table 6 describes the percentages of words, based on their syntactic category, that were presented by the textbooks but are not found in the frequency dictionary. Of all the nouns, for example, that were presented in *Pido la palabra* (1998), 34.39% were not present in the frequency dictionary. In both textbooks, nouns and

adjectives represent the syntactic categories with the highest rate of not being covered in the dictionary.

Another way to better understand the over-represented lemmas in the textbooks is to separate the lemmas that are not in the dictionary and then divide them into their respective syntactic categories. This allows one to see the relative spread of the syntactic categories of the lemmas that were over-represented. It is of note that these data do not compare with any frequency assignment. It is unknown, for example, whether an entry in this section has a frequency assignment of 5,001 or 12,000. Table 8 shows the numbers of over-represented entries in terms of syntactic categories as well as their relative coverage compared to the other syntactic categories. Refer to Appendix F for a graphical representation of this table.

Table 8.

Over-represented lemmas

	<u>Total no. -dictionary</u>	<u>Nouns</u>		<u>Verbs</u>		<u>Adjectives</u>		<u>Adverbs</u>	
		no.	%	no.	%	no.	%	no.	%
<i>Pido la palabra</i>	857	540	63.01	86	10.04	220	25.67	11	1.28
<i>¡Estoy listo!</i>	318	213	66.98	15	4.72	81	25.47	9	2.83

This table shows that nouns are clearly more over-represented than the other syntactic categories. Davies and Face (2006) also found nouns to be much more over-represented than other syntactic categories in their study. One possible reason for nouns and adjectives being more likely to be over-represented is that they hold more obvious, teachable content and fit well into thematic chapters. A physical object, for example, can easily be seen as a picture or object, and its descriptions can be pointed to. This might lead textbook writers to use more infrequent nouns and adjectives because they can easily be taught visually. These

infrequent lemmas might also make for a more interesting lesson, in which students can learn different lemmas that are associated with semantic fields that are familiar to them like parties, food, clothing, and furniture. Verbs and adverbs, on the other hand, may be more difficult to teach because they are not as visually concrete or as easily exemplified physically.

Also similar to the results of Davies and Face (2006), the nouns that were over-represented in these textbooks tend to refer to concrete concepts. These researchers operationalized infrequent entries as not occurring more than 100 times in Davies' (2002) 20-million word, *Corpus del Español*. This definition will be used to describe the extent of infrequency of example entries. Examples of over-represented concrete concepts in the textbooks in this study include *chuparrosa* [hummingbird] (0 occurrences in Davies' (2002) corpus), *hyena* [hyena] (13), *tornasol* [sunflower] (5), and *ventanal* [large window] (79).

Davies and Face (2006) only found four over-represented nouns and two verbs that could be considered abstract. While most of the over-represented lemmas found in *Pido la palabra* (1998) and *¡Estoy listo!* (2003) are similarly concrete, there are also many more abstract concepts represented than in the Davies and Face study. For example, *adverbio* [adverb] (27 occurrences in Davies' 2002 corpus), *agrado* [charm, affability] (97), *atribución* [attribution] (41), and *astucia* [astuteness] (94) represent four such abstract words from the A's alone. The presence of so many more abstract concepts in this study might be due to the fact that all presented lemmas were extracted from the textbooks instead of only active vocabulary. Similarly over-represented abstract lemmas might also be present but not active in the textbooks studied by Davies and Face.

4.4 Mexican Vocabulary

Entries that were not present in Davies' (2006) frequency dictionary were investigated further in more extensive, bilingual dictionaries (see section 3.3.3). If an entry in these bilingual dictionaries mentioned that a word was particularly used in Mexican or Latin American variations, it was coded with "(Mex)" preceding its definition. This allowed the researcher to determine to what extent these texts were dialect specific as well as to help balance the fact that a multi-dialectal, Spain-dominated corpus was used to create the frequency dictionary. Also, because the target learners for the textbooks in this study are second language learners in Mexico, one might think that much of the vocabulary presented would be Mexico-specific. Table 9 represents the number of these Mexican or Latin American vocabulary entries relative to the total number of lemmas extracted from the textbooks.

Table 9.

Percentage of Mexican/Latin American lemmas

	<u>Total no. of lemmas</u>	<u>no. of (Mex) lemmas</u>	<u>% of total entries</u>
<i>Pido la palabra</i>	2924	97	3.32
<i>¡Estoy listo!</i>	1438	49	3.41

At 3.32% and 3.41% of their total vocabulary being specifically of Mexican or Latin American varieties of Spanish, neither textbook strays too far away from vocabulary that is frequent across all Spanish dialects. This is significant in terms of the question regarding the differences between SFL and SSL textbooks. In the case of these second language textbooks, there is only a small percentage of vocabulary particularly exclusive to Mexico or Latin America.

Because the entries were tagged for general dialect during the frequency assignment process, they could be compared to Lara's (1993) generally frequency-based *Diccionario fundamental del español de México* [Fundamental dictionary of Mexican Spanish]. Table 10 represents how many of the words originally shown to be Mexican or Latin American in the textbooks were represented in Lara's dictionary.

Table 10.

Coverage of (Mex) lemmas in Mexican Spanish dictionary

	Total no. of (Mex)	no. in Mexican Spanish Dictionary	% of (Mex) lemmas
<i>Pido la palabra</i>	97	15	15.46
<i>¡Estoy listo!</i>	49	10	20.41

The data here show that of the total Mexican- and Latin American-specific entries, a relatively small portion of them were frequent in Mexican Spanish. This gives some support to Moreno de Alba's (2005) claim that there would not be many different, variation-specific lemmas in a frequency list of a given variation of Spanish that would not be present in a frequency list based on Spanish as a whole. That is to say, of the Mexican-specific entries that were presented, only a small percentage of them (15.46% and 20.41%) were determined to be frequent in the variation as a whole. However, this dictionary used to determine frequency in Mexican Spanish may not be an appropriate or reliable instrument for such comparisons to Mexican Spanish as a whole (see sections 3.2.3, 3.4.2).

4.5 Summary of results

This section summarizes the results and addresses the initial questions posited above (see section 3.1). The first of these questions asked about how well did these two

SSL textbooks represent frequent lemmas. This question was answered in two ways; the first described how both textbooks represented the first 500 most frequent lemmas relatively well (see section 4.1). However, using another method in which the total number of lemmas that were presented was used as a baseline for the cut-off number in the frequency list, only 55.37% and 44.09% of the N number of lemmas presented are amongst the 2,924 and 1,438 most frequent lemmas in *¡Estoy listo!* (2003) and *Pido la palabra* (1998), respectively. This shows significant room for possible coverage improvement in these two Spanish textbooks. Also, as shown in Tables 4 and 5, both textbooks present significant amounts of vocabulary that is not present in the frequency dictionary.

The second question posited asks about the under-representation and over-representation of the vocabulary presented. As shown in Table 6, there were not large differences between syntactic categories amongst the textbooks relative to the coverage of those categories in the frequency list. However, *Pido la palabra* (1998) presented about twice as many frequent items as *¡Estoy listo!* (2003). In other words, the under-representation could also be described by comparing the volume of unique lemmas presented from one textbook compared to the other. Also, as seen in Table 7, nouns and adjectives are significantly more likely to be under-represented than verbs and adverbs. This is interesting because even though many more frequent nouns and adjectives are presented, of the vocabulary in these textbooks, a noun or adjective is much more likely to be infrequent than a verb or an adverb that is presented (see Table 5). Because nouns represent an overall much greater number of entries in the frequency list (2,511 out of 5,000), a variable that might obstruct better coverage is the space. There might not be enough room in the materials or time in a single course to incorporate many more concepts.

In terms of over-representation, it was determined (see Table 7) that nouns and adjectives are much more likely to be over-represented in the textbooks than verbs and adverbs. One possible explanation for this is that the semantic fields associated with the chapters' themes might involve more concrete concepts that are easily taught and make for more a more interesting lesson. This is interesting because nouns were also found to be more likely to be under-represented than verbs and adverbs (see Table 6). This means that of the 5,000 most frequent nouns, a smaller percentage is represented in the textbooks. However, nouns were also found to be the most over-represented part of speech, accounting for the majority of infrequent lemmas in both textbooks studied (see Table 7). This is how nouns are able to be both under- and over-represented.

The third question posited asked about potential differences between these SSL textbooks and the SFL textbooks studied by Davies and Face (2006). As discussed in the previous results sections, there were several similarities between Davies and Face's results that were based on SFL textbooks. Like in the Davies and Face study, for example, the textbooks in the current study gave better coverage to higher frequency ranges than lower ones. However, both studies found that the programs that use these textbooks could be significantly improved by including neglected very high frequent lemmas (see section 4.1). Finally, as a small inquiry, it was interestingly found that the second language textbooks examined in the current study were not very variation specific (see Table 8), supporting Moreno de Alba's (2005) claims of the commonality of frequent lemmas across different varieties of a given language.

Overall, the results of this study show that both of the textbooks examined represent the extremely frequent vocabulary well (see Table 7 and Appendix A). However, both also present significant amounts of vocabulary that is not highly frequent, possibly at the

expense of moderately frequent lemmas. Textbooks, including those studied in this investigation, often use themes or situations as a way to create a lesson or chapter. It may not be feasible for a single textbook to cover thousands of frequent words as they appear in a frequency list because of this type of organization. Instead, in a chapter involving foods, several infrequent vocabulary words might be presented because they are conceptually relevant to the lesson. Also, there were some dialectal specific vocabulary entries in both textbooks. However, although these textbooks were designed for students studying in or wanting to study in Mexico, there were much fewer Mexican-specific entries than lemmas common across regional dialects. This might be because there may be little difference between frequent lemmas across different variation; such *mexicanismos* may be seen as more appropriate for more advanced learners; or the textbooks may have been modeled after other, pre-existing resources.