Lexical Profile of L2 Russian Textbooks

Ekaterina Talalakina

Tony Brown

Mikhail Kamrotov

Lexical Profile of L2 Russian Textbooks

Ekaterina Talalakina
Tony Brown
Mikhail Kamrotov

1. Introduction

Traditionally, the link between vocabulary mastery and reading comprehension has been examined through the prism of lexical thresholds and vocabulary coverage (Milton 2009). Lexical thresholds represent the most frequent words in a language (i.e., lemmas, or dictionary forms of a word) and usually come in increments of 1,000. In relation to the Russian National Corpus, knowledge of the 1,000 most frequent lemmas allows for comprehension of 60% of a text's vocabulary, 2,000 lemmas – 69%, and 10,000 – 85% (Lyashevskaya and Sharoff 2009, v). These figures support an earlier estimation by Brown (1996, 2), who claimed (without elaborating on what grounds) that a passive knowledge of the 8,000–10,000 most frequent lexemes allows for "reasonable confidence" in reading Russian for general purposes.

While vocabulary frequency and text coverage are one measure of the difficulties that a learner might have in reading authentic texts, the question of how many words learners need to know to demonstrate reading proficiency levels according to the ACTFL Guidelines (2012) was addressed by Hacking and Tshirner (2017). They investigated the relationship between second language (L2) Russian vocabulary size and ACTFL proficiency levels among US college students and postulated that the 1,000-word band correlated with a rating of Intermediate-Low, the 2,000-word band – Intermediate-Mid, the 3,000-word band – Advanced-Low, the 4,000-word band – Advanced-Mid, and the 5,000-word band – Advanced-High. These data on vocabulary knowledge and its link to reading proficiency levels raise interesting questions for our field; for example, do commonly available textbooks designed for the Intermediate and Advanced levels of instruction cover the 5,000 most frequent words in Russian, since, absent widely available graded or extensive reading programs, most classroom learners will likely

encounter new vocabulary through textbooks? Furthermore, the data raise the question of to what extent textbook authors should account for frequency data in constructing learning materials.

2. Literature review

Previous studies addressing vocabulary coverage in L2 textbooks have documented a significant deviation from frequency lists. Keller (1991) [1] compared five "core" textbooks (i.e., G. A. Bitextina and D. Davidson, *Russian: Stage One*; B. T. Clark, *Russian*, 3rd ed.; V. G. Kostomarov, *Russian for Everybody* [adaptation by R. L. Baker]; R. Leed and A. Nakhimovsky, *Beginning Russian* [2 vols.]; and G. and L. Stilman and W. E. Harkins, *Introductory Russian Grammar*) against a frequency list of 3,500 Russian nouns in order to determine both the lexical profile of the textbooks and the pedagogical value of the textbooks' most frequent words. The study revealed that vocabulary used in these textbooks reflected a significant departure from the frequency list. Keller concluded that textbook authors need to put greater emphasis on vocabulary recycling [2] in their materials.

Rifkin (1992) reiterated the importance of word recycling in connection with the influence of the communicative approach movement on Russian-language textbooks. He also noted that some introductory textbooks included vocabulary that had questionable usefulness for general purposes, such as *бетон* 'concrete' and *крановщица* 'female crane operator.' Based on a frequency dictionary produced by Lyashevskaya and Sharoff (2009), *бетон* falls within the most frequent 7,000 words, while *крановщица* does not even make the top 20,000-word cut and likely would be beneficial only to students specializing in construction. Although as of 2019 the textbooks Rifkin reviewed are no longer in use, the two aforementioned examples illustrate the inclusion of low-frequency words of a highly technical nature that potentially place an extra burden on students since these words distract their attention from learning more frequent words that could be used in a larger range of contexts. Even if they were included solely for the purpose of pronunciation or grammar training, high-frequency vocabulary items could have just

---

[1] Admittedly, Keller (1991) represents dated research, but a provides a useful point of departure, if for no other reason than to illustrate strides made in the field.
[2] In the present study, the term "recycling" refers to the repetition of a word in any form and, thus, indicates the number of tokens (i.e., running words in a text; Gardner 2008).

as easily fulfilled such a purpose, which raises the question of how textbook authors select vocabulary items. Rifkin likewise pointed out that, with few exceptions, Intermediate and Advanced textbooks generally fall short in terms of "selection, sequencing, and presentation of vocabulary" (p. 480), which observation further supports a call for a more careful consideration of lexis in textbook design.

As part of a case study of the beginning Russian textbook *Mezhdu nami*, Comer (2019) offered a comprehensive review of existing studies that examine vocabulary input in foreign language textbooks. Among the findings reported, the study shows that the textbooks introduced a relatively low proportion (32.1%) of the most frequent 5,000 words (based on the Russian National Corpus by Lyashevskaya and Sharoff (2009) and the minimum vocabulary expectations established by the Russian Federation for three major levels of the Test of Russian as a Foreign Language), which is consistent with data for other languages mentioned in the study. Comer attributed the above finding to the word composition of the corpus itself, which reflects language from written texts rather than spoken language.

Davies and Face (2006) looked at active vocabulary from Spanish textbooks and compared it to lemmas from the Corpus del Español and the new *Frequency Dictionary of Spanish: Core Vocabulary for Learners*. They discovered that "whatever N number of vocabulary words a textbook includes, only 10–50% of those are among the N most frequent lemma in the language" (p. 142). Thus, they found that some words numbered among the most frequent 1,000 lemmas were underrepresented, whereas other words that lie beyond the most frequent 5,000 lemmas were overrepresented. According to Davies and Face, such a finding speaks to the semantic fields chosen by the textbook developers and, hence, the need to align textbook vocabulary with real-word usage across a variety of genres, discourse types, and semantic fields.

Underrepresentation of high-frequency vocabulary in L2 textbooks was recorded in a number of other studies as well. Lipinski (2010) found that only 53–64% of vocabulary items from the 1,000 most frequent words [3] were found in the three introductory textbooks on L2 German examined in her study. Wagner (2015) compared textbooks of French to a Frequency Dictionary of French and found that first- and second-year L2 textbooks

---

[3] Leipzig/BYU Corpus of Contemporary German.

offer fewer than 1,464 words out of the most frequent 5,000, which finding led Wagner to conclude that students may be missing essential input for those levels. Similarly, according to research conducted by O'Loughlin (2012), three textbooks of English from Elementary to Intermediate level introduce a combined total of 1,500 out of 2,000 high-frequency words.[4] Findings from these studies highlight a lack of high-frequency vocabulary in lower-level textbooks.

Two studies posit explanations for the mismatch between vocabulary input and the stated level of a textbook. Catalán and Francisco (2008) analyzed vocabulary in four English as a Foreign Language textbooks and concluded that publishing houses do not agree on the number and type of words to which students should be exposed at a given proficiency level. This difference of opinion stems from a lack of explicit standards for vocabulary selection in materials design and likewise explains a gap in the stated level of learning materials and the vocabulary input associated with those levels. Criado and Sánchez (2009) illustrated this gap by way of an EFL textbook marketed for the Intermediate level, but which, judging by the use of high- versus low-frequency vocabulary, was more appropriate for higher levels of proficiency.

Although studies focused on textbook vocabulary input consistently argue in favor of a careful selection of lexis, no relevant studies to date have investigated the current state of affairs in the field of L2 Russian vocabulary coverage in textbooks pitched at Intermediate and Advanced levels. The current study addresses this gap by analyzing four L2 Russian textbooks in order to answer the following two research questions:

RQ1: To what extent does the lexis choice in the textbooks reflect proficiency levels targeted by these textbooks?

RQ2: To what extent is the lexical coverage in the Intermediate and Advanced textbooks under question representative of the vocabulary of contemporary works of fiction and texts in the mass media and, thus, contributive to students' ability to read a range of genres?

---

[4] Citing related research on this subject, O'Loughlin (2012) asserts that "high frequency vocabulary provides the most benefit to learners, as the most frequent 2,000 word families cover over 80% of text (Carroll et al. 1971) and account for nearly 95% of spoken language, thus providing learners with the lexical foundation to engage in everyday conversation (Adolphs and Schmitt 2003, 433)" (O'Loughlin 2012, 256).

3. Methodology

For the present study, the researchers compiled a corpus that comprised four textbooks published between 2007 and 2019 by Routledge and Yale University Press: Murray, J., *Intermediate Russian: A Grammar and Workbook* (T1); Kagan, O. E., A. S. Kudyma, and F. J. Miller, *Russian: from Intermediate to Advanced* (T2); Kagan, O. E., and A. S. Kudyma, *Russian Through Art for Intermediate to Advanced Students* (T3); and Rifkin, B. and O. Kagan, *Advanced Russian Through History* (T4).

Textbook choice partly paid homage to the late Olga Kagan—a champion of L2 curricular materials pitched at the Intermediate and Advanced levels and the person to whom this special issue of *Russian Language Journal* is dedicated. Recognizing the importance of analyzing alternative perspectives, the researchers selected Murray, J. and S. Smyth's *Intermediate Russian*, which, consistent with the Kagan et al. textbooks under consideration, targets the Intermediate and Advanced threshold. However, to a large extent, availability of eBooks governed the choice of whether to incorporate a textbook into the corpus, since comparative statistical analyses of tokens both within and across textbooks necessitated such a format as a point of departure. Three of the four textbooks used in this research were purchased from the publisher in eBook format, while the fourth was shared in electronic format by one of the authors.[5] Aside from availability, textbook choice reflected a deliberate attempt on the part of the researchers to represent equitably both traditional and content-driven textbooks.

T2 and T3 explicitly reference the ACTFL scale, whereas T1 and T4 do not, but these textbooks still state that the materials are intended for learners at both Intermediate and Advanced levels. Descriptions associated therewith contributed to the decision on the part of the researcher to include the textbooks in the corpus under consideration.

The data from the textbooks were analyzed using R statistical software (ver. 3.5.2[6]), while the words were lemmatized using Yandex (MyStem 3.1). The corpus contained a total of 180,695 tokens (i.e., running words in the text in Cyrillic, including headers and footers, appendices, and

---

[5] Special thanks to Benjamin Rifkin for graciously sharing electronic files of essays in Rifkin, B., O. Kagan, and A. Yatsenko, *Advanced Russian Through History*, for which Yale University Press does not offer an eBook.

[6] The following packages were used: readtext 0.74, future.apply 1.2.0, data.table 1.12.2, stringr 1.4.0, readr 1.3.1, tidyverse 1.3.1, and matrixStats 0.54.0.

glossaries). Whenever a word was composed of multiple parts without a hyphen, each part was handled as a separate token (e.g., *потому что* was handled as two separate words), and all hyphenated words were addressed as single tokens. In solving contextual disambiguation issues, we relied on a built-in MyStem 3.1 algorithm that proved to be highly efficient in handling cases like *есть* ('there is' vs 'to eat'), as our experiments showed.

Lemmatization was needed in order to compare the textbook tokens to the 5,000-word general vocabulary frequency lists by Sharoff, Umanskaya, and Wilson (2013) and fiction and mass media lists by Lyashevskaya and Sharoff (2009), both of which appear in the form of lemmas. In their study, Hacking and Tschirner (2017) explicitly cited the aforementioned lists in regards to aligning the most frequent Russian words with ACTFL reading proficiency levels. The use of the same frequency list by Sharoff, Umanskaya, and Wilson (2013) guaranteed the validity of inferences regarding the levels of proficiency in the present study.[7]

In order to analyze the lexis of L2 Russian textbooks, the researchers used the framework of the Lexical Frequency Profile (LFP) by Laufer and Nation (1995). Within this framework, the words in the text are classified according to their belonging to the first, second, third, fourth, or fifth thousand frequency band. The LFP shows the lexical richness and sophistication of a text by providing the percentage of the text covered by the words from each of those bands. Findings by Hacking and Tschirner (2017) enabled the researchers to match coverage levels with reading proficiency levels and served as the methodology by which the researchers answered the first research question regarding the frequency portrait of the words in the textbook.

The researchers subsequently checked lemma frequency from the textbooks against fiction and mass media frequency lists compiled by Lyashevskaya and Sharoff (2009), which were based on the Russian

---

[7] The decision to use the list by Sharoff, Umanskaya, and Wilson represents the outcome of a careful selection process from a number of important lexical lists, including one by Lyashevskaya and Sharoff (2009), based on the Russian National Corpus and the lexical minimums for each level of the Test of Russian as a Foreign Language (TORFL)/ Тест по русскому языку как иностранному (ТРКИ) (Andriushina and Kozlova 2006; Andriushina 2008; Andriushina 2009). The former represents a slightly earlier version of Sharoff, Umanskaya, and Wilson (2013), so the deviation in results should not necessarily be significant. The latter is pegged to the TORFL levels, which have not been empirically validated in relation to the ACTFL levels. Thus, for the sake of validity, the researchers chose the same source as the one used in the Hacking and Tschirner (2017) study.

National Corpus. Such an approach gave primacy to genres—in particular, (1) fiction, (2) mass media, (3) fiction and mass media, and (4) neither fiction nor mass media—and helped to determine the percentage of words from the textbooks in each category. The data that emerged enabled the researchers to make inferences about the lexis types in each of the textbooks and thus answer the second research question regarding the contributive effect of lexical coverage on students' ability to read a range of genres.

4. Results

Table 1 presents frequency data for the total number of tokens (including proper nouns and abbreviations) and the number of types (i.e., the number of unique lemmas). In addition, Table 1 sets forth data regarding lemmas grouped (G) in 1,000 word increments (G1=1,000; G2=2,000; G3=3,000; G4=4,000; G5=5,000; G6=6,000; and beyond) and the corresponding reading proficiency level from the Hacking and Tschirner (2017) study.

Table 1. Frequency profile of textbooks by coverage

	T1	T2	T3	T4
Stated level(s)	Intermediate	Intermediate/ Advanced	Intermediate/ Advanced	Advanced
Number of tokens	27,214	79,151	37,504	36,826
Number of types	4,313	6,209	3,301	5,877
G1 coverage corresponding to IL Reading	64.5%	60.2%	51.8%	52.8%
G2 coverage corresponding to IM Reading	8.7%	9.8%	10.8%	11.2%
G3 coverage corresponding to AL Reading	4.9%	4.8%	6.6%	5.6%
G4 coverage corresponding to AM Reading	2.7%	3.3%	2.9%	3.6%
G5 coverage corresponding to AH Reading	2.1%	2.8%	3.2%	3.3%
G6+ [8] coverage	17.1%	19.1%	24.7%	23.5%

---

[8] Words beyond the 5,000 frequency band.

The above findings indicate that all four textbooks under consideration generally fall within the respective level, as evidenced by target lexis at the Intermediate level that ranges from 70.0–73.2% (G1+G2 combined coverage, T2 and T1 respectively) and at the Advanced level from 75.3–76.5% (G1–G5 combined coverage, T3 and T4 respectively). All four textbooks present a considerable number of high-frequency vocabulary from the first 2,000 most frequent words, in contrast to the number of words at the 3,000-5,000 range (9.7% in T1, 10.9% in T2, 12.7% in T3, and 12.5% in T4).

Figure 1. Frequency profile of textbooks by lemma types

■ G1 □ G2 ▨ G3 ▨ G4 □ G5 ■ G6+ □ G6+ (excl. proper nouns and acronyms)
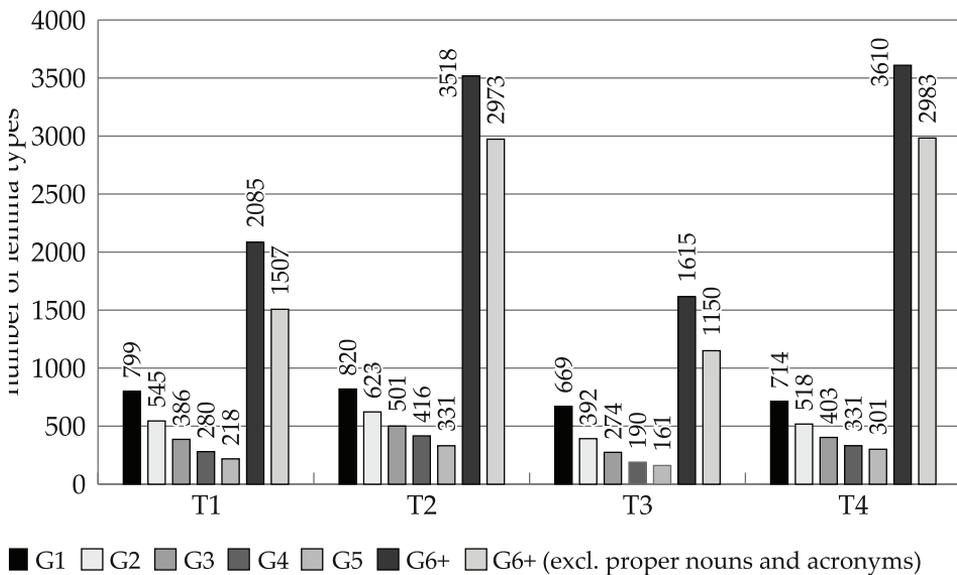
Figure 1 illustrates the profile of each textbook by number of word types within each increment of the most frequent 5,000 words. Count by word types reveals a sizeable imbalance in favor of low-frequency vocabulary: G6+ word types, even with proper nouns and acronyms excluded, account for a considerable portion of the types and play an important role in targeting a culture-specific component of language studies.

Although G6+ vocabulary used in the textbooks accounts for a large percentage of lemma types, it occupies a smaller percentage of tokens, as illustrated in Figure 2 below.

202

Figure 2. Percentage of G6+ words in types and tokens
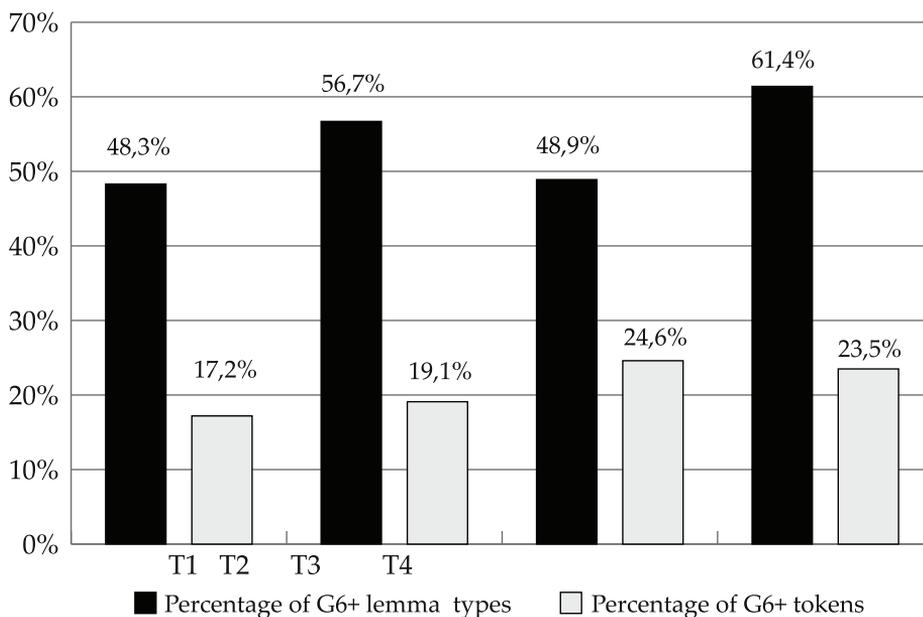
Given the large number of G6+ words, determining the degree to which textbooks recycle them becomes essential in order to make inferences about incidental vocabulary learning (i.e., vocabulary learning without an intent to learn; Laufer and Hulstijn 2001). In a comparison of reading exposure to incidental learning of lexis, Schmitt (2010) asserts that exposure must occur 8–10 times in order for learners to develop a passive knowledge of words. Table 2 breaks down G6+ frequency information by textbook.

Table 2. Text frequency of G6+ words in textbooks

Textbook	Number of words that appear 8+ times	Number of words that appear only once
T1	70	1167
T2	471	1491
T3	308	559
T4	156	2201

An examination of words counted in the second column of Table 2 reveals that, in most cases, word choice fits the topics discussed in the textbooks (e.g., T3=308 and T4=156 represent specialized vocabulary and proper nouns relevant to art and history respectively). However, in the case of T1, one finds words with dubious general-purpose usefulness such as калоша 'galosh' and ди-джей 'DJ,' which appear 23 and 10 times respectively, whereas words of higher usefulness, such as продуктовый 'grocery,' угощение 'treat,' задерживаться 'to be late,' побеспокоить 'to disturb,' скандалить 'to make a scandal,' гибкий 'flexible,' замужем 'married,' and рейс 'flight' appear only once. Data in column 3 reveal significant skewness towards incidental vocabulary in all cases except T3. Overall, the choice of vocabulary beyond the 5,000 most frequent words aligns well with textbook topics.

*Table 3. Types from fiction and mass media lists by percentage of words and coverage*

	T1	T2	T3	T4
% of lemma types (and number) shared by textbook and fiction list	50.6% (2184)	40.5% (2516)	49.1% (1622)	34.2% (2011)
% of lemma types (and number) shared by textbook and mass media list	54.1% (2335)	48.5% (3012)	56.2% (1856)	45.6% (2681)
Coverage by shared lemmas from fiction list	83.1%	78.9%	73.6%	71.6%
Coverage by shared lemmas from mass media list	84.5%	84.9%	79.7%	80.2%

Genre characteristics of vocabulary likewise represent a point of interest in terms of the vocabulary profile of the textbooks. Textbook data were checked against fiction and mass media frequency lists. These lists were compiled using corpora of texts dated from 1950–2007 (45 million tokens and 49 million tokens in fiction and mass media corpora respectively). Lists included all functional parts of speech. As in the present research, each service word containing multiple parts was treated as multiple separate words (e.g., потому что was handled as two separate words). Words in lists were characterized by statistics on frequency, range,

and dispersion. The latter was measured by Juilland's D coefficient which shows how uniformly a word is distributed across different parts of the corpus. Table 3 lays out the percentage of types found in fiction and mass media frequency lists, as well as their coverage of the text. Given that fiction and mass media lists overlap, Figure 3 illustrates the total number of types distributed across four categories, in particular, (1) words outside mass media and fiction lists, (2) words present in both lists, (3) words from mass media list, and (4) words from fiction list.

The above data indicate that, although lexis from fiction and mass media accounts for 34.2–56.2% of all word types in the textbooks, the coverage of those words ranges from 71.6–84.9%, thus suggesting that the majority of material spans both genres. Figure 3, which illustrates significant overlap between fiction and media words in the texts, reinforces the claim of a universal nature of vocabulary genre presented in all four textbooks, despite the fact that two of the four pertain to specialized fields, namely art and history.

Figure 3. Number of lemmas per textbook lexis across four categories involving text genre
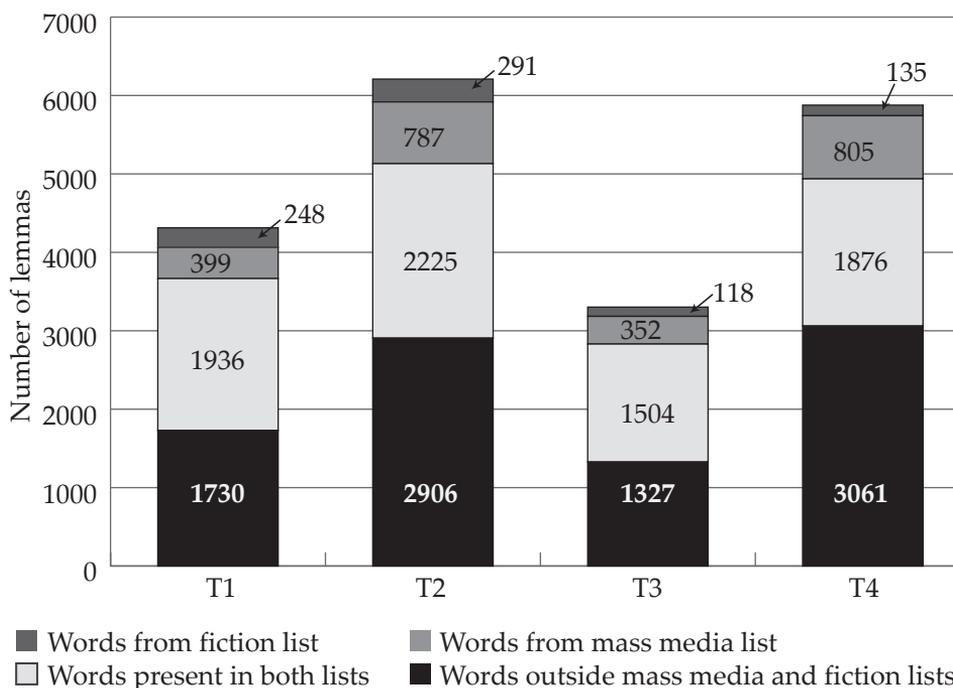
5. Discussion

Clearly, word lists are corpus-dependent and, accordingly, their utility varies from one corpus to another. As such, exclusive focus on frequency information from one list can prove pedagogically unsound owing to a certain degree of deviation between lists, which likewise supports findings by Keller (1991). Based on an overview of textbook studies, Milton (2009) asserts that language textbooks often display a balanced number of high- and low-frequency vocabulary, and the textbooks analyzed in this study reinforce that claim.

In terms of level-appropriate vocabulary by textbook, findings from this research indicate that G1 vocabulary gives way to lower-frequency vocabulary in proportion to increases in proficiency level; however, one should also take into consideration the number of mid-frequency words in a given language. Mid-frequency vocabulary in English ranges from 4,000 to 8,000 words, which, according to Schmitt and Schmitt (2014), represents an essential lexicon for proficient language use in spite of a number of pedagogical challenges associated with the acquisition of such vocabulary items. In regards to Russian, findings from this research indicate that textbooks underrepresent vocabulary in the 3,000–5,000 range. Consulting frequency lists may help authors of future upper-level textbooks to include more words from these ranges.

With regard to the amount of lexis presented in each of the textbooks, a lemma-to-token ratio offers important insights into word recycling; in particular, the lower the ratio, the higher the chance of word recycling. For T1 and T4, the ratio equals 0.159 (each lemma appears 6 times on average in the corresponding corpus); for T2 and T3—0.079 (13 times on average) and 0.088 (11 times on average), respectively. Accordingly, words in T2 and T3 are recycled more frequently than words in T1 and T4—a finding supported by data in Table 2 in relation to G6+ words that appear over eight times. That being said, all of the textbooks have a large number of words that appear only once (from 1,167 to 2,201), with T3 standing in sharp contrast to the other three (559). This raises the question of long-term benefits associated with exposing L2 learners to large numbers of "truly incidental" words (i.e., words that appear only once), since it is doubtful that students will learn those words through absent recycling. Alternatively, both Schmitt (2008) and Milton (2009) assert that maximizing exposure does not lead to vocabulary overload if

the students are properly engaged with the material.

In addition, research carried out by Laufer and Rozovski-Roitblat (2015) shows that one or two carefully designed exercises involving a word can benefit students more in the way of retention than 18–21 exposures to a word. In light of such findings, material developers should consider tracking so-called "one-timers" in their texts and evaluating their usefulness. In the case of words of low usefulness, material designers might consider omitting or substituting them with mid-frequency synonyms; conversely, in the case of words of high usefulness, incorporating them in exercises may increase the chances of L2 learners acquiring them.

6. Conclusions

A frequency measure of vocabulary use can assist textbook designers in making data-driven decisions regarding the content of foreign language textbooks with a communicative emphasis. Past research addressing frequency data in textbooks shows that significant deviations from frequency lists occur, and this finding has particular relevance for curriculum designers. This study presents a frequency analysis of four textbooks and reveals that although vocabulary frequency in the textbooks reflects the word knowledge needed at specific levels of reading proficiency, a gap exists in the type and treatment of vocabulary introduced from the 3,000–5,000-word range.

And yet the data also reveal that all the textbooks include a large number of vocabulary items representative of the 6,000 and beyond threshold (G6+). By implementing both intentional and incidental learning techniques, L2 learners stand to benefit from exposure to vocabulary in these textbooks owing to a balanced selection of vocabulary across genres. In order to ensure mastery of vocabulary at the Intermediate and Advanced levels, material designers should make a conscious effort to limit the amount of lexis beyond the 6,000 threshold in favor of lexis in the 3,000–5,000 range.

In addition, this research demonstrates how frequent vocabulary recycling can provide learning opportunities within each textbook and illustrates the potential for uninterrupted learning and maximal retention across levels. All in all, the present study shows that vocabulary measures can inform data-driven decision making in material and curriculum design.

7. Limitations and directions for future research

Given the similar composition of author teams for the materials under consideration, one may reasonably expect to find a certain lexical similarity across textbooks. Future research examining frequency data in Russian L2 textbooks likely will benefit from increased diversity with regard to source material authorship.

Although Schmitt (2010, 63) characterizes frequency as "arguably the single most important characteristic of lexis that researchers must address," others criticize the approach for promoting ambiguity, since the most frequent words often have multiple meanings (polysemous), which eventually leads to processing difficulty (Crossley, Cobb, and McNamara 2013). While the lemmatization algorithm used in the present study claims to distinguish between words that have the same spelling but different meanings (context homonyms), it does not distinguish between polysemous words, so words may be used in two textbooks but with different meanings.

In addition to frequency, usefulness and difficulty play an important role in lexis selection (Laufer and Nation 2012; He and Godfroid 2019). "Usefulness" refers to the capacity of lexis to help satisfy general needs that learners have in regards to a second language. It also includes lexis for special needs, as in the case of T3 and T4—textbooks with specialized vocabulary that cater to the needs of students in art and history, respectively. "Usefulness" also refers to words in multi-word units, which present-day computer algorithms can only partially tackle, so the judgment of the material designer in some instances may prove more reliable than a computer algorithm. Studies in ESL have shown that native speakers make reasonable judgments on word usefulness up to the 7000-word level in terms of frequency (Okamoto 2015). As for difficulty, one can measure it in two ways: (1) through the set of criteria outlined in the study by Laufer (1990) that involves morphology, synonymy, connotations, etc.; and (2) through readability formulas (Chen 2016), which are not widely available for Russian. Thus, research that looks beyond frequency measures as a way of building the lexical profile of textbooks stands to benefit the field at large.

As for the type of lexis presented in the textbooks, a comparison with fiction and mass media lists reveals an overall balance. The field of ESL benefits from the aforementioned LFP, which enables one to compare words with an academic vocabulary list (Hsu 2009); however, to date, Russian lacks such an academic vocabulary list and stands to benefit greatly from

future research carried out in this regard. In addition, as Flowerdew (2009) notes, research on corpus linguistics reveals that frequency patterns may vary across registers. Thus, future comparisons with word lists reflecting the most frequently used words in spoken language corpora could shed needed light on the peculiarities of a lexical profile.

In the case of T2—one of the longer Advanced textbooks in the field—one would expect to find a more exhaustive coverage of vocabulary relative to the first 1,000 most frequent words, and yet the data suggest otherwise. Discerning whether this incongruity stems from the lexicon under consideration simply not appearing frequently in written texts or from some other reason will require additional investigation.

With regard to materials design, this research points out that all four textbooks reveal insufficient coverage of G2–G5 vocabulary, which represents core vocabulary items across the levels of reading proficiency. Accordingly, the researchers recommend that materials designers target words from the most frequent 2,000–5,000 words more intentionally and incorporate them into scaffolding exercises. In this regard, the study also raises the question of vocabulary-building exercises in Intermediate- and Advanced-level textbooks and the degree to which such exercises actually target words from levels appropriate to the book's intended audience. While such a question exceeds the scope of the current study, the authors view it as a potential point of departure for future research.

Ultimately, findings from this research suggest that the textbooks under consideration expose learners to around 3,000–6,000 lemma types, the mastery of which positions a learner for achieving Advanced-level reading proficiency. This research further confirms an observation made by O'Keeffe, McCarthy, and Carter (2007), namely, that concentrating on words beyond the 6,000-word band yields a limited return on investment in terms of vocabulary acquisition. Ultimately, material designers promote language uptake by integrating scaffolding exercises in the form of intentional lexical work that promotes long-term learning of lexis.

References

ACTFL. (2012). ACTFL Performance Descriptors for Language Learners [Electronic version]. Accessed September 14, 2019. http://www.actfl.org/sites/default/files/pdfs/ACTFLProficiency Guidelines2012_FINAL.pdf.

Adolphs, Svenja, and Norbert Schmitt. 2003. "Lexical coverage of spoken discourse." *Applied linguistics* 24, 4: 425-438.

Andriushina, N. P., and T. V. Kozlova. 2006. *Leksicheskii minimum po russkomu iazyku kak inostrannomu. Bazovyi uroven'. Obshchee vladenie*. 3rd ed. Moscow: Zlatoust.

Andriushina, N. P., ed. 2008. *Leksicheskiiminimum po russkomu iazyku kak inostrannomu. Pervyi sertifikacionnyi uroven'. Obshchee vladenie*. 4th ed. Moscow: Zlatoust.

Andriushina, N. P., ed. 2009. *Leksicheskii minimum po russkomu iazyku kak inostrannomu. Vtoroi sertifikacionnyi uroven'. Obshchee vladenie*. 2nd ed. St. Petersburg: Zlatoust.

Brown, Nicholas J. 1996. *Russian Learners' Dictionary: 10,000 Words in Frequency Order.* Routledge, London.

Carroll, John Bissell, Peter Davies, Barry Richman, and Peter Davies. 1971. *The American Heritage word frequency book*. Boston: Houghton Mifflin.

Catalán, Rosa María Jiménez, and Rocío Mancebo Francisco. 2008. "Vocabulary Input in EFL Textbooks." *Revista española de lingüística aplicada* 21: 147–166.

Chen, Alvin Cheng-Hsien. 2016. "A critical evaluation of text difficulty development in ELT textbook series: A corpus-based approach using variability neighbor clustering." *System* 58: 64-81.

Comer, William. 2019. "Measured Words: Quantifying Vocabulary Exposure in Beginning Russian." *Slavic & East European Journal* 63 (1).

Criado, Raquel, and Aquilino Sánchez. 2009. "Vocabulary in EFL Textbooks. A Contrastive Analysis against Three Corpus-Based Word Ranges." In *A Survey on Corpus-based Research / Panorama de investigaciones basadas en corpus*, 862–875.

Crossley, Scott A., Tom Cobb, and Danielle S. McNamara. 2013. "Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications." *System* 41 (4): 965–981.

Davies, Mark, and Timothy L. Face. 2006. "Vocabulary Coverage in Spanish Textbooks: How Representative is It." *Selected proceedings of the 9th Hispanic Linguistics Symposium*, 132–143.

Flowerdew, John. 2009. "Corpora in Language Teaching." In *The Handbook of Language Teaching*, edited by Michael H. Long and Catherine J. Doughty, 327–350.

Gardner, Dee. 2008. "Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading." *Reading in a Foreign Language* 20 (1): 92–122.

Hacking, Jane F., and Erwin Tschirner. 2017. "The contribution of vocabulary knowledge to reading proficiency: The case of college Russian." *Foreign Language Annals* 50 (3): 500–518.

He, Xuehong, and Aline Godfroid. 2019. "Choosing Words to Teach: A Novel Method for Vocabulary Selection and Its Practical Application." *TESOL Quarterly* 53 (2): 348-371.

Hsu, Wenhua. 2009. "College English Textbooks for General Purposes: A Corpus-based Analysis of Lexical Coverage." *Electronic Journal of Foreign Language Teaching* 6 (1): 42–62.

Keller, Howard H. 1991. "Word Frequency and Pedagogical Value: Contrasting Textbook Vocabulary Lists with Word Frequency Counts." *Slavic and East European Journal*, 228–244.

Laufer, Batia. 1990. "Ease and Difficulty in Vocabulary Learning: Some Teaching Implications." *Foreign Language Annals* 23 (2): 147–155.

Laufer, Batia, and Jan Hulstijn. 2001. "Incidental vocabulary acquisition in a second language: The construct of task-induced involvement." *Applied linguistics* 22 (1): 1–26.

Laufer, Batia, and Paul Nation. 1995. "Vocabulary size and use: Lexical richness in L2 written production." *Applied linguistics* 16 (3): 307–322.

Laufer, Batia, and I. S. P. Nation. 2012. "Vocabulary." In *The Routledge Handbook of Second Language Acquisition*, edited by Susan M. Gass and Alison Mackey, 163–176. Routledge.

Laufer, Batia, and Bella Rozovski-Roitblat. 2015. "Retention of new words: Quantity of encounters, quality of task, and degree of knowledge." *Language Teaching Research* 19 (6): 687–711.

Lipinski, Silke. 2010. "A Frequency Analysis of Vocabulary in Three First-Year Textbooks of German." *Die Unterrichtspraxis / Teaching German* 43 (2): 167–174.

Lyashevskaya, Olga, and Serge Sharoff. 2009. "Frequency Dictionary of the Modern Russian Language (on the base of the National Corpus of the Russian Language)." Azbukovnik.

Milton, James. 2009. *Measuring Second Language Vocabulary Acquisition*. Vol. 45. Multilingual Matters.

O'Loughlin, Richard. 2012. "Tuning in To Vocabulary Frequency in Coursebooks." *RELC Journal* 43 (2): 255–269.

Okamoto, Mayumi. 2015. "Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection." *System* 51: 1–10.

O'Keeffe, Anne, Michael McCarthy, and Ronald Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press.

Rifkin, Benjamin. 1992. "The Communicative Orientation of Russian-language Textbooks." *Slavic and East European Journal*, 463–488.

Rifkin, Benjamin. 2005. "A ceiling effect in traditional classroom foreign language instruction: Data from Russian." *The Modern Language Journal* 89 (1): 3–18.

Schmitt, Norbert. 2008. "Instructed second language vocabulary learning." *Language Teaching Research* 12 (3): 329–363.

Schmitt, Norbert. 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Springer.

Schmitt, Norbert, and Diane Schmitt. 2014. "A reassessment of frequency and vocabulary size in L2 vocabulary teaching." *Language Teaching* 47 (4): 484–503.

Sharoff, Serge, Elena Umanskaya, and James Wilson. 2013. *A Frequency Dictionary of Russian: Core Vocabulary for Learners*. Routledge.

Wagner, Jennifer L. 2015. "Vocabulary, Variation, and Culture in American University Textbooks of French." PhD diss., University of South Australia.

Textbooks:

Kagan, Olga E., Kudyma S. Anna, and Frank J. Miller. 2015. *Russian: From Intermediate to Advanced*. Routledge.

Kudyma, Anna S., and Olga E. Kagan. 2019. *Russian Through Art: For Intermediate to Advanced Students*. Routledge.

Murray, John, and Sarah Smyth. 2013. *Intermediate Russian: a grammar and workbook*. Routledge.

Rifkin, Benjamin, Olga Kagan, and Anna Yatsenko. 2007. *Advanced Russian Through History*. Yale University Press.