

# 1 Technical Vocabulary in Languages for Special Purposes:

## 2 The Corpus-Based Russian Economics Word List

3

### 4 Highlights

- 5 • This research develops a corpus-based economics word list in Russian for L2 learners.
- 6 • This study adopts a data-driven method to maximize the list's out-of-sample reliability.
- 7 • The developed list provides a shortcut to increased coverage of economics texts.
- 8 • This methodology can be used to develop corpus-based lists in other languages.

9

### 10 Abstract

11 In this article, we address the topic of language for specific purposes (LSP) through the  
12 creation and validation of a technical, corpus-based word list in Russian. We build on an  
13 existing body of research on developing lists of specialized vocabulary in English as a second  
14 language (ESL). We introduce some methodological innovations in compiling a list of Russian  
15 vocabulary for students with a professional interest in economics. This study illustrates that a  
16 corpus-based list of specialized vocabulary, which includes 507 lemmas united in 324 word  
17 families, can provide a hypothetical shortcut to the increased coverage of discipline-specific  
18 texts compared with the coverage achieved using the next 1,000-word frequency increment of  
19 general vocabulary. Therefore, such a list could facilitate LSP mastery at lower levels of

20 proficiency, which yields important pedagogical implications for the foreign language  
21 classroom. The methodology could potentially be used to compile similar lists in other  
22 languages.

23 Keywords: frequency, word lists, corpora, technical vocabulary, language for special purposes,  
24 Russian

25

## 26 1. Introduction

27 Although the learning outcomes of language for specific purposes (LSP) are usually  
28 measured through performance-based procedures with less emphasis on grammar and  
29 vocabulary (Basturkmen and Elder 2004), it is undeniable that vocabulary, as a core language  
30 component, is vital for learners' functional second language (L2) use, both in receptive and  
31 productive tasks. Thus, it is essential to investigate the conditions under which L2 learners  
32 can obtain the maximum return on their time invested in the mastery of language features in  
33 general and vocabulary in particular.

34 Corpus-based frequency word lists have been proven to be useful in facilitating the  
35 measurement of the lexis needed for comprehension, judging by text coverage. Regarding  
36 English, Nation (2006) found that 98% coverage is needed for adequate comprehension, which  
37 learners can reach by recognizing the most frequent 8,000-9,000 word families in English.  
38 However, the 98% threshold is by no means a necessary and sufficient condition for  
39 comprehension. Laufer and Ravenhorst-Kalovski (2010) pointed out that even a small increase  
40 in lexical coverage may be just as beneficial to comprehension as a larger increase in  
41 coverage. The authors provided a 95% minimal coverage threshold for adequate  
42 comprehension in addition to an optimal one. In Russian, similar data exist for lemmas but not  
43 for word families: the 10,000 most frequent lemmas from the Russian National Corpus cover  
44 only 85% of texts (Lyashevskaya and Sharoff 2009). The optimal 98% threshold is difficult to

45 achieve since the values of simple statistics are heavily influenced by the sample size, and  
46 corpus data exhibit natural variation in the analyses of textual data (Baayen 2002).  
47 Nevertheless, these numbers raise the question of what goals can realistically be set for  
48 vocabulary learning within an L2 curriculum.

49 The abovementioned data build on the measurement of the vocabulary size and depth,  
50 which serve as cornerstones in vocabulary acquisition. However, another essential dimension  
51 in this domain is the type of vocabulary under discussion. Nation and Kyongho (1995)  
52 examined frequency, text coverage, and range to distinguish between general service  
53 vocabulary and special purpose vocabulary. They argued that a general service  
54 vocabulary in English is beneficial for students up to the 2,000-word level. After this threshold,  
55 a special purpose vocabulary is more advantageous for learners with professional interests.

56 For the present research, we posited that specialized vocabulary acquisition in Russian  
57 is suitable for language learning starting from the intermediate proficiency level and underlies  
58 specialized vocabulary lists. We designed this study to provide a tool in the form of a  
59 specialized list that in future research can be tested for being a shortcut for L2 Russian  
60 learners in increasing discipline-specific reading comprehension compared with mastering the  
61 next 1,000-word band of general purpose vocabulary.

62 2. The construct of technical vocabulary

63 In L2 curriculum design, determining a special purpose vocabulary is part of a needs  
64 analysis (Nation and Macalister 2010). In higher education, a needs analysis can be viewed  
65 through the prism of social theory, that is, applying a social, practice-based (text-based)  
66 language curriculum built on academic literacy and disciplinary knowledge (Mickan 2012).  
67 Thus, in this context, the notion of a special purpose vocabulary embraces both vocabulary for  
68 academic purposes and vocabulary for professional purposes, which are extensively studied in  
69 regard to English as a second language (ESL).

70 The construct of vocabulary for academic purposes in ESL manifests in several lists,  
71 and the two most recent are the Academic Word List (AWL) by Coxhead (2000) and the  
72 Academic Vocabulary List (AVL) by Gardner and Davis (2014). Although each was generated  
73 through different methodologies, both lists present lexis common to various academic  
74 disciplines and not typical of other genres, such as fiction. Several studies have investigated  
75 the coverage of the AWL in field-specific texts in applied linguistics, finance and economics,  
76 which show quite a substantial presence of academic words in these professional domains (Li  
77 and Qian 2010; Sun and Lei 2013; Vongpumivitch et al. 2009). Nevertheless, both the AWL and  
78 AVL focus exclusively on "core academic" vocabulary in contrast to "technical academic  
79 vocabulary," which is characteristic of a particular discipline. This cross-disciplinary, generic  
80 nature of academic vocabulary is viewed as a disadvantage by some researchers (Hyland and  
81 Tse 2007).

82 A separate strand of ESL literature explores technical academic vocabulary in different  
83 disciplines to create field-specific AVLs. Some of these lists show a significant overlap with  
84 the AWL: the Chemistry AWL covers 57% of the AWL, the Nursing AWL covers 56%, and the  
85 Linguistics AWL covers 67% (Islamizadeh 2016; Valipouri and Nassaji 2013; Yang 2015). Thus,  
86 these lists combine the “core academic” and “technical academic” lexis, which proves their  
87 value in studying academic disciplines in ESL. At the same time, several studies have focused  
88 on purely technical vocabulary, which is typical not only of academic texts but also of other  
89 genres. In one study, specialized words that appear for students in trades such as plumbing  
90 and fabrication were assembled from a variety of materials by using both written and spoken  
91 corpora (Coxhead et al. 2018; Coxhead and Demecheleer 2018). Likewise, Tongpoon-  
92 Patanasorn (2018) went beyond academic texts in creating a technical word list for finance  
93 and included mass media items in the corpus. In all cases, the lists meet the professional  
94 needs of L2 learners to help them master disciplinary content through relevant technical  
95 vocabulary.

96 Given the variety of approaches to technical vocabulary, it is essential to provide a clear  
97 definition of the actual term. Chung and Nation (2004, p. 252) suggested that technical  
98 vocabulary possesses three features, namely, it is “subject related, occurs in a specialist  
99 domain, and is part of a system of subject knowledge.” The researchers compared the  
100 following four approaches to identify technical vocabulary: an expert rating scale; context

101 clues; dictionaries; and a computational approach. They concluded that the most valid is a  
102 rating scale that involves a specialist in the field who manually assigns a technicality rating to  
103 words in a text. The rating takes into account the distinction of vocabulary levels by Nation  
104 (2001) into high-frequency, academic, technical, and low-frequency vocabulary. However, the  
105 authors of this study acknowledged that technical vocabulary, although representing a  
106 separate level, can also come from other categories (Chung and Nation 2003).

107         Similarly, a number of studies have indicated that the term “technical vocabulary”  
108 includes both technical terms from specialized dictionaries and polysemic general service  
109 words with technical meanings (Ha and Hyland 2017; Lei and Liu 2016; Muñoz 2015; Todd  
110 2017). Thus, it is common to apply a mixed-methods approach to identify technical vocabulary  
111 that involves a corpus-based computer analysis, such as a vocabulary classification and  
112 keyword identification, along with expert opinions, such as expert ratings or dictionary  
113 consultations (Kwary 2011; Tongpoon-Patanasorn 2018).

114         Historically, studies on special purpose vocabulary in Russian have been focused on  
115 technical terms that increase reading comprehension in scientific texts. Spurred by the need  
116 to keep pace with Russian advances in the physical sciences, a boom in learning materials  
117 occurred in the middle of the 20th century. As an illustration, over ten textbooks concerning  
118 scientific Russian were in use in the US by the 1950s (Yakobson and Apanasewicz 1959). In  
119 the late 1960s, an Indian scholar published a cross-disciplinary technical vocabulary list

120 assembled manually for individuals studying physics, chemistry, mathematics, mechanics, and  
121 engineering (Bakaya 1967). Afterward, however, there was a prolonged gap in the scholarly  
122 interest in Russian frequency lists until the 2009 frequency dictionary produced by  
123 Lyashevskaya and Sharoff (2009). Most recently, Author et al. (2020) proposed a Russian AVL  
124 that includes the most frequently used words across various disciplines, although it excludes  
125 technical vocabulary.

126       Clearly, the abundance of studies on technical vocabulary in ESL stands in sharp  
127 contrast with the lack of such vocabulary lists and studies in L2 Russian. At the same time, an  
128 extensive body of research on tailoring language materials to the professional needs of L2  
129 learners in ESL could benefit curriculum designers in other languages for specific purposes.  
130 For the present study, we set out to illustrate this by using the case of a specialized  
131 vocabulary list in Russian. We built on major ESL trends in developing special purpose  
132 vocabulary lists and added some methodological innovations. In constructing a list of L2  
133 Russian vocabulary for learners with a professional interest in economics, we sought to  
134 answer the following questions:

- 135       1.     What constitutes the construct of Russian specialized vocabulary in economics?
- 136       2.     How does this list relate to text coverage, an important consideration for L2  
137 readers?
- 138       3.     What are the implications for implementing such a list in L2 Russian curricula?



### 139 3. Methodology of technical list generation

#### 140 3.1. Corpus composition

141 Since our ultimate goal was to obtain a comprehensive list with comparable coverage  
142 across genres, we rejected the idea of a homogenous corpus that consists solely of Russian  
143 academic texts (Wang et al. 2008; Yang 2015). Thus, we compiled a customized corpus to  
144 include two subcorpora, specifically, mass media texts and academic texts, with both covering  
145 the same timespan (2016–2018). Mass media texts from the ten most reputable business  
146 press outlets comprise 3,218,737 running words. The Russian academic section includes  
147 7,270,140 running words from eight high-ranking economics journals that cover all economics  
148 subfields (see Appendix A for the outlets and journals list). Although these two subcorpora are  
149 not balanced in size, the methodology that we employed (see below) addresses this issue, and  
150 the final list shows almost an equal coverage of both subcorpora. Additionally, since  
151 increasing the overall corpus size usually leads to more robust results (Miller and Biber 2015),  
152 we combined the two genres so that the final sample contained 10,488,877 running words,  
153 with 30.7% coming from mass media texts and 69.3% being provided by academic texts.

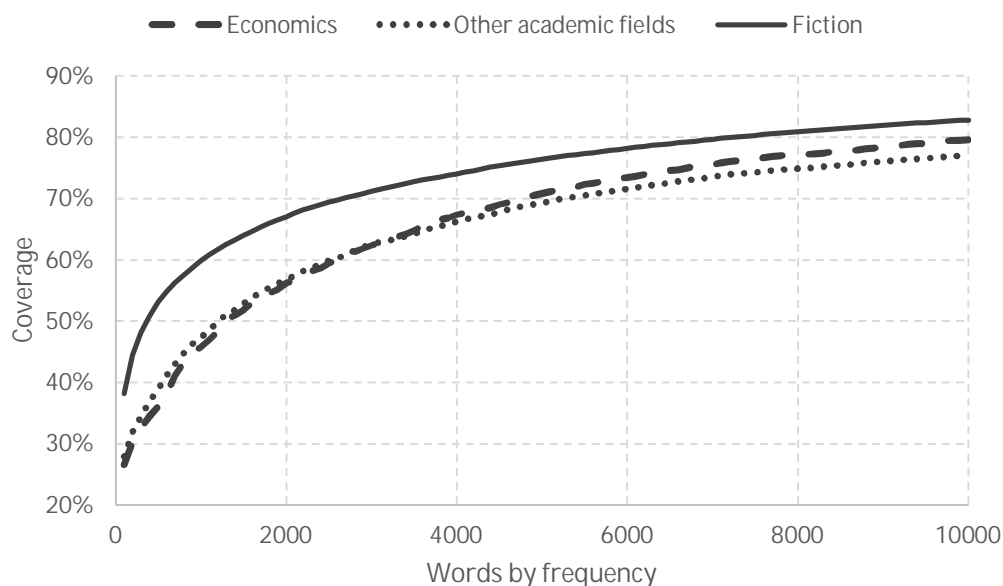
154 Moreover, we used two comparison corpora for validation purposes. The first is a fiction  
155 corpus based on texts from 20 contemporary Russian authors, aligned with the guidelines by  
156 Gardner and Davis (2014), who emphasized that a corpus needs to represent contemporary  
157 language. The fiction corpus contains 3,058,875 running words, which is comparable to the

158 one used in the study by Coxhead (2000). We used this fiction corpus to determine the various  
159 patterns that different types of vocabulary exhibit in terms of text coverage. The second  
160 comparison corpus is an academic corpus of 7,916,522 running words; it encompasses  
161 research articles published from 2016 to 2018 and is balanced regarding the number of words  
162 across the humanities, science and technology, and the social sciences (excluding  
163 economics).

164 We first studied the lexical profiles of the corpora in terms of the most frequent general  
165 vocabulary by using the frequency lists of Sharoff et al. (2013), who presented their word lists  
166 in the form of lemmas. We lemmatized the corpus by employing the lemmatization algorithm  
167 developed by Yandex (MyStem 3.1), which also can remove homonymy from context. After  
168 lemmatization, we made additional manual corrections to address some well-known  
169 difficulties when the algorithm was unable to distinguish between two homonyms (e.g., *bank*  
170 'bank' vs. *banka* 'jar').

171 We checked the lemmatized corpora against the general vocabulary list to discover that  
172 5,000 lemmas cover 71.4% of the economics corpus, 76.3% of the fiction corpus, and 69.3% of  
173 the academic corpus. A more granular distribution is shown in Fig. 1, which demonstrates that  
174 Russian word frequencies are roughly proportional to word ranks in the frequency table, which  
175 is in line with Zipf's law. Fig. 1 also reveals that when the frequency rank is lower, the  
176 additional coverage is lower. This lexical profile of the corpus indicates 95% coverage of the

177 text calls for significantly more than 10,000 words previously indicated by Brown (1996) in  
178 relation to Russian. Such figures raise the question of whether there is a shortcut to achieving  
179 full comprehension, rather than consistently learning thousands of words.



180  
181  
182  
183

Fig. 1 Distribution of the most frequent general vocabulary increments in different corpora.

184 To create the Russian Economics Word List (REWL), we used the criteria of the ratio,  
185 degree of dispersion, and minimum occurrence threshold in connection to the economics  
186 corpus, which we collected specifically for this study from academic and mass media texts of  
187 over 10 million tokens.

188 First, we employed a frequency-based approach to identify the words that are more  
189 specific to the economics corpus than other corpora. A popular alternative would be to use the  
190 likelihood-based approach (Rayson & Garside 2000). Overall, the likelihood and frequency

191 ratio statistics are competing keyness statistics, and their relative performance is highly case-  
192 specific (Pojanapunya & Todd 2018). However, previous research on the use of these two  
193 approaches for compiling word lists in Russian has shown that likelihood-based statistics to  
194 construct vocabulary lists can produce highly corpus-specific results (Author et al. 2020). For  
195 this reason, we adopted the frequency-based approach. In particular, we measured the ratio  
196 criterion as a relation between a word's frequency in the economics corpus and its frequency  
197 in the other corpora. The ratio is used to remove general high-frequency words and words that  
198 are inherent to other academic fields (Gardner and Davis 2014; Lei and Liu 2016). We selected  
199 words that were at least 1.5 times more frequent in the economics texts than in the other  
200 texts. As there is no generally accepted value for the ratio measure and scholars usually select  
201 a value at their discretion, we proposed a more systematic approach based on cross-  
202 validation (see below).

203         Second, dispersion makes it possible to exclude words that are specific to a small part  
204 of the corpus and cannot be removed based on the ratio criterion. As noted by Biber et al.  
205 (2016), Julliland's D measure (Juilland et al. 1970), although widely used, suffers from a  
206 significant dependence on the size and partition of the corpus, which makes it difficult to find  
207 an appropriate threshold value. The degree of dispersion (DP) is a measure developed by  
208 Gries (2008) that can be viewed as the most robust approach to date for computing word  
209 dispersion across parts of a corpus. For consistency purposes, we applied the cross-validation

210 approach to select a threshold for DP. This resulted in removing all words with DP values  
211 exceeding 0.25. The main advantage of our approach is that it is data-driven for the most part,  
212 with as little discretion used as possible. The threshold for DP is a result of the proposed  
213 cross-validation procedure, which is explained in detail below.

214 Finally, a mild minimum occurrence criterion helped us to only select words that  
215 occurred at least 10 times in the corpus. This criterion filtered out words that are extremely  
216 rare but spread evenly across the corpus. We set this threshold at our discretion. Given the  
217 size of the corpus data, the chosen value corresponds to a normalized frequency (per million  
218 words) of approximately 1, which is significantly below the lowest frequency of 2.29 observed  
219 in the Sharoff et al. list (2013). Therefore, we believe that the applied threshold has not led to  
220 the exclusion of influential lemmas from the REWL. Although one can find additional  
221 measures suggested in the literature for word-selection purposes, incorporating a large  
222 number of these measures might undermine the robustness and generalizability of the  
223 findings because of overfitting, as each additional measure requires choosing another  
224 threshold value. As such, too many thresholds might lead to a list of words that has good in-  
225 sample properties but does not provide enough out-of-sample coverage.

226 The fine-tuning of the model parameters is crucial to fitting any model to data. In the  
227 context of the present research, this process involved selecting a minimum ratio and  
228 maximum DP values. One of the well-known approaches to calibrating parameters is cross-

229 validation. Unfortunately, this procedure inevitably leads to a smaller sample size as the  
230 corpus is split into training and test parts. According to Miller and Biber (2015), even large  
231 corpora cannot guarantee the reliability of the word list, which means that it is highly desirable  
232 to keep the sample size as large as possible. In this case, combinatorically symmetric cross-  
233 validation (CSCV), as suggested by Bailey et al. (2016), appears to be the most appropriate  
234 method.

235 We chose the out-of-sample coverage as the target performance statistics for the word  
236 list and optimized it with respect to the ratio and DP. The CSCV procedure involved the following  
237 steps:

- 238 1) splitting the corpus into eight disjointed parts of equal size;
- 239 2) creating all possible combinations of these parts in groups of four;
- 240 3) making a training set by joining four parts that constitute one combination;
- 241 4) making a test set from the remaining parts of the corpus;
- 242 5) using a training set to produce a word list; and
- 243 6) evaluating the coverage of the word list by using training (in-sample coverage) and test  
244 (out-of-sample coverage) sets.

245 Splitting the corpus into eight parts resulted in 70 training and test set combinations.

246 The main objective of this procedure was to uncover the relationship between the in-sample  
247 (IS) and out-of-sample (OOS) coverage of a word list obtained with a particular ratio and DP

248 values. Splitting the corpus into smaller subsamples can help to evaluate its internal  
249 representativeness, which is a prerequisite to any quantitative linguistic analysis but is often  
250 overlooked by researchers (Miller and Biber 2015). The CSCV procedure leverages multiple  
251 partitions of the corpus, which makes it possible to collect statistics on internal  
252 representativeness by using a single corpus.

253       Following Bailey et al. (2015), we summarized the relationship between IS and OOS  
254 coverage with a linear regression that describes a word list performance degradation:

$$255 \qquad \text{Coverage}_{\text{OOS}} = \alpha + \beta \cdot \text{Coverage}_{\text{IS}} + \varepsilon \qquad (1)$$

256       If the corpus is homogenous and possesses high internal representativeness for  
257 compiling a word list, then we can expect a positive  $\beta$  and a high  $R^2$  of regression (1).

258       Each combination of ratio and DP values produced a set of  $\beta$  and  $R^2$  values. The latter  
259 can be used to select the optimal ratio and DP values that constitute the most internally  
260 representative word selection model. The set of ratio values that we considered includes  
261 values from 1.25 to 2 with 0.25 increments, whereas the DP values range between 0.15 and  
262 0.3 with 0.05 increments. Furthermore, since the goal was to compile a list of no more than  
263 1,000 words, we also used the number of top words to be selected (which ranged from 300 to  
264 1,000 with increments of 100) as another constraint. In total, these constraint values produced  
265 128 sets of parameters, for which we computed  $\beta$  and  $R^2$ .

266 The procedure employed for compiling the word list is clearly extendable to collocations  
267 and multiword expressions. However, we did not include these in the developed list for two main  
268 reasons. First, moving beyond single words would dramatically increase the size of the list,  
269 thereby creating challenges for its practical use for L2 learners, particularly for those without a  
270 basic knowledge of economics. Second, multiword expressions can exhibit a considerably larger  
271 degree of variability across genres of economic literature than single-word expressions. The  
272 development of genre-specific lists of economic collocations might be useful for L2 learners;  
273 nevertheless, this research task not only goes far beyond the scope of this article but also  
274 requires serious pedagogical and methodological considerations.

#### 275 4. Resulting output

##### 276 4.1. The words in the list

277 Fig. 2 presents the CSCV results. Each combination of ratio, DP, and size constraints is  
278 characterized by the average OOS coverage across all 70 permutations of the training and test  
279 sets and the  $R^2$  statistic. We omitted the  $\beta$  regression coefficient from Fig. 2 since its values  
280 were always nonnegative and could therefore not be used to distinguish between high and low  
281 levels of internal representativeness.



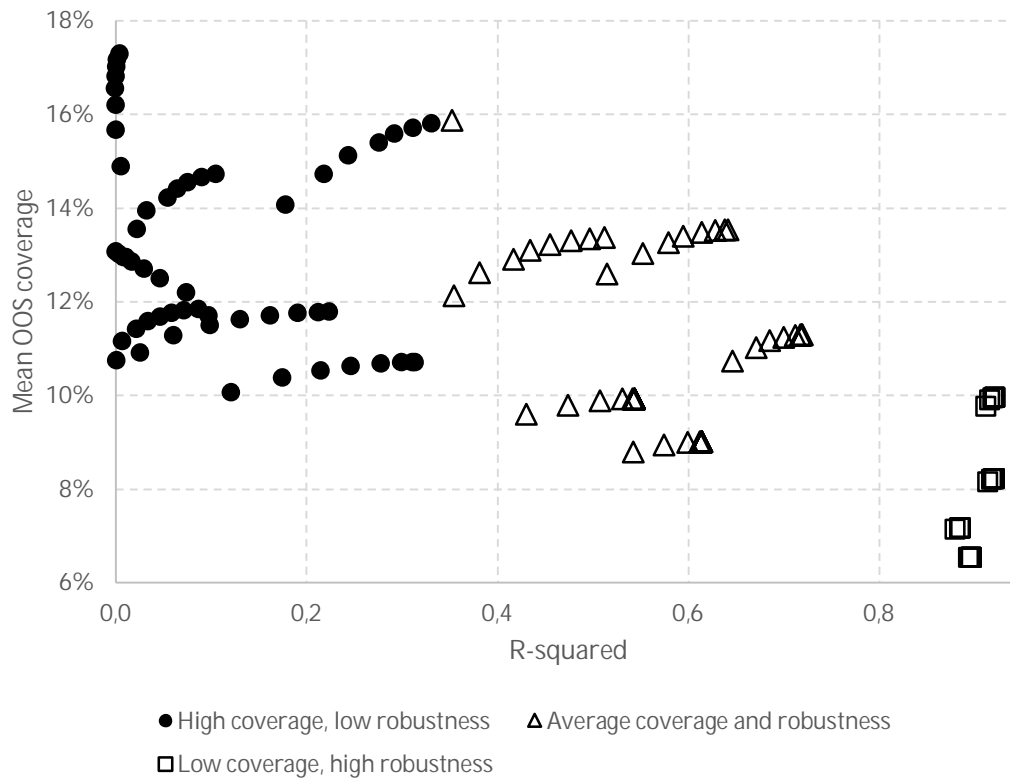


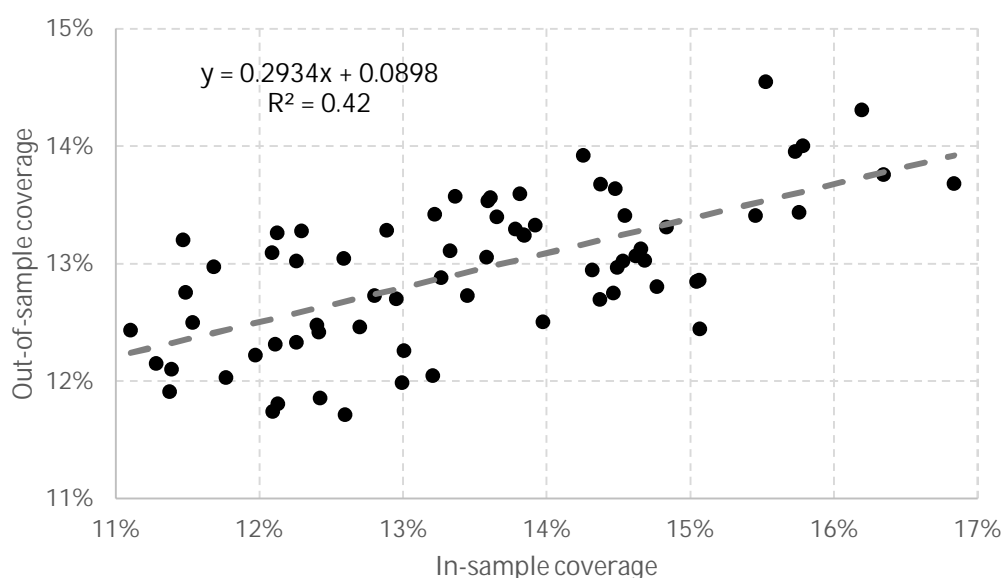
Fig. 2 CSCV results.

The variation in parameters induces changes in  $R^2$  in the range of 0 to 0.92 and changes in the mean OOS coverage in the range of 6.5% to 17.3%. Due to the reverse dependence between the  $R^2$  statistic and mean OOS coverage, parameter optimization entails a trade-off between these two measures. To find meaningful groups within the cross-validation results, we applied a widely used k-means clustering approach (Kaufman and Rousseeuw 2009) with the optimal number of clusters determined by the average silhouette width (see Appendix C for the cluster selection results). The model identified three clusters. The left cluster contains points with a nonsignificant dependence between IS and OOS coverage. Although high OOS coverage is the key feature of the word list, the association between IS and OOS is weak for this cluster, and the selected word list will not have satisfactory text coverage beyond the given corpus. The

294 right-hand side cluster, in contrast, includes points with high  $R^2$  and significant dependence,  
295 but its low OOS coverage downgrades the word list's practical utility. Thus, we chose the central  
296 cluster with an optimal range of parameters that ensure both reliability and coverage.

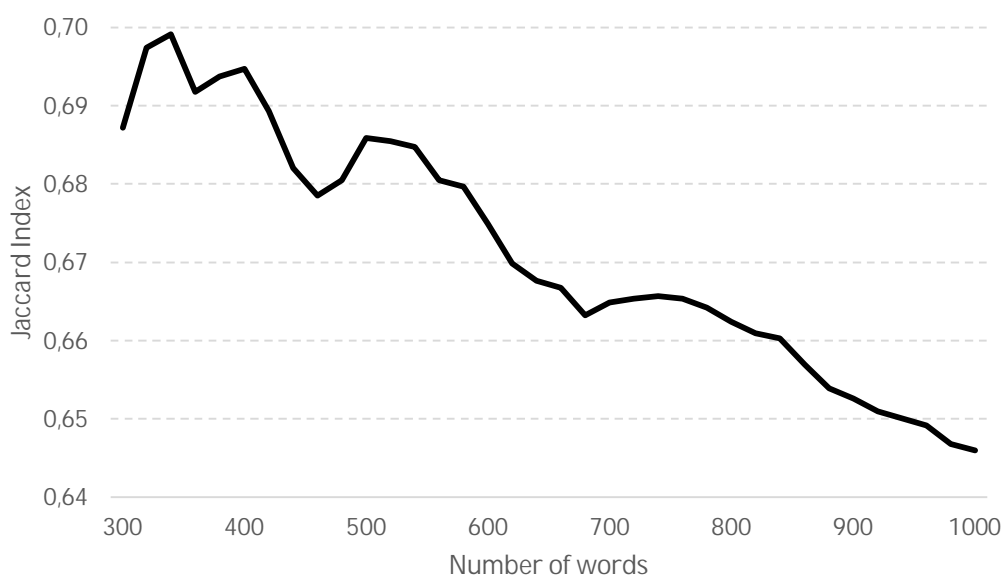
297 Points from this cluster correspond to all tested sizes of the word list (300–1,000) but  
298 are limited to only two combinations of the ratio and DP values: (1.25, 0.2) and (1.5, 0.25). The  
299 latter is preferred, as a higher ratio presumably better distinguishes technical economics  
300 words from general vocabulary and the vocabulary of other academic fields.

301 A typical performance degradation pattern for the selected ratio and DP values is  
302 presented in Fig. 3. Generally, OOS coverage is lower than IS coverage, but the relationship  
303 between the two is positive and significant. The model shows that maximizing IS coverage in  
304 the given corpus also allows maximizing OOS coverage.



305 Fig. 3 Typical performance degradation pattern for the selected ratio and DP.  
306 Each point corresponds to a single partition of the corpus as defined by the CSCV.

307 Since the points in the central cluster covered all sizes of the word list, it was  
308 impossible to select an optimal size with a performance degradation analysis. The goal was to  
309 compile a reliable word list that would be as stable as possible out of sample. The size of a  
310 word list may be optimized with respect to its robustness across different samples. We  
311 employed a simplified interpretation of stability as a set of words that remained unchanged  
312 across the subsamples, without taking the word ranking into account, and we used the  
313 Jaccard index as a stability measure, with lower values corresponding to lower stability.



314 Fig. 4 Robustness of the word list as measured by the Jaccard index.  
315

316 We computed the Jaccard index for different CSCV permutations by compiling two word  
317 lists for the training and test sets and averaging the Jaccard index values for all permutations.  
318 Fig. 4 presents a typical pattern of the association between an average Jaccard index and the  
319 size of the list. After multiple trials, we selected 507 as the optimal size, as it corresponds to a

320 local maximum of average Jaccard index values and consequently defines a relatively stable  
321 word set across corpus partitions. The REWL accounts for 12.7% of tokens in the economics  
322 corpus. Based on the equation of dependence between IS and OOS coverage, we expected an  
323 OOS coverage of 12.5%, which is slightly below the IS value.

#### 324 4.2. Validation of the REWL

325 We validated the REWL from six perspectives. First, we computed the coverage in  
326 relation to the three academic subcorpora used to compile the REWL, namely, academic texts  
327 from 1) the humanities, 2) science and technology, and 3) other social science fields  
328 (excluding economics). The REWL coverage of the other social sciences corpus is 3.5% and,  
329 as one would expect, is the highest of the three. Nevertheless, this value is more than 3.5  
330 times lower than the REWL coverage of economics texts. The science and technology and  
331 humanities corpora have coverages of 2% and 1.2%, respectively. These results clearly  
332 indicate that the REWL contains words that are highly specific to the field of economics.

333 Second, to ensure that the REWL does not include general high-frequency words, we  
334 validated the list against a corpus of fiction texts that we did not utilize in the word selection  
335 procedure. The coverage was only 0.6%, which supports the specificity of the REWL.

336 Third, the economics vocabulary list for international students proposed by Andryushina  
337 et al. (2015) can serve as a direct point of comparison with the REWL. The economics list by  
338 Andryushina et al. (2015) contains 490 lemmas, is roughly the same size as the REWL, and

339 has the same coverage of the economics corpus (12.7%). However, the content similarity of  
 340 both lists is rather limited. The lists share only 102 lemmas (20% of the REWL), and more  
 341 importantly, the list by Andryushina et al. (2015) has a significantly higher coverage of our  
 342 corpus in other academic fields (Table 1). This implies that rather than containing a pure  
 343 economics vocabulary, the list by Andryushina et al. (2015) is a combination of general  
 344 academic and economics lexis. This inference is further confirmed by the fact that the list by  
 345 Andryushina et al. (2015), unlike the REWL, has 92 words in common with the Russian AVL  
 346 (Author et al. 2020). Therefore, the REWL is more specific to economics than the list by  
 347 Andryushina et al. (2015).

348

Field	The humanities	Science and technology	Social sciences (excluding economics)
Coverage by the REWL	1.2%	2%	3.5%
Coverage by the economics vocabulary list by Andryushina et al. (2015)	3.9%	8.3%	8.4%

349 Table 1 Comparison of the noneconomic academic corpus coverage between the REWL and  
 350 the list by Andryushina et al. (2015).

351 Fourth, building on the idea suggested by Lei and Liu (2016), we compared the REWL to  
352 two economics dictionaries. The comparison revealed that 315 words from the REWL (62.1%)  
353 were mentioned in at least one of the two dictionaries. This suggests that the REWL provides  
354 learners with high-frequency economics words beyond terminology. For instance, words not  
355 included in the dictionaries are certain verbs such as *akkumulirovat'* "to accumulate" and  
356 *vyplačivat'* "to make payments," nouns such as *neft'* "oil" and *benzin* "gasoline," adjectives  
357 such as *vzaimovыgodnyj* "mutually beneficial" and *vysokokvalificirovannyj* "highly qualified," all  
358 of which are obviously related to the field of economics.

359 Fifth, we compared the coverage of the REWL to two economics subcorpora that make  
360 up the corpus used to develop the REWL, specifically, academic and mass media texts. The  
361 REWL covers 13.1% of the academic economics texts, 11.7% of mass media economics texts,  
362 and 12.7% of the combined corpus.

363 Sixth, we estimated OOS coverage by the REWL with a new corpus. The literature  
364 review on the discourse of economics texts allowed us to identify the following four main  
365 genres in this discipline: research articles; academic textbooks; magazine articles; and reports  
366 issued by financial institutions such as central banks (Bondi-Paganelli, 1996; Resche, 2000;  
367 Parodi, 2015; Simanungkalit, 2016). We used two of the four main genres to compile the  
368 corpus for developing the REWL (research articles and magazines), and we utilized the  
369 remaining two to validate the REWL.

370 The first part of an OOS chunk of data consisted of textbooks for future economists  
371 with L2 Russian. This part of the validation corpus included textbooks by Fedotova and  
372 Khoronenko (2012), Čekina and Kapasova (2020), Popova et al. (2014), and Dubinskaja and  
373 Paj (2007), which totaled 73,077 running words. The second part of the validation corpus  
374 included Bank of Russia annual reports from 2016 to 2020, which totaled 600,793 running  
375 words.

376 The REWL showed 12.6% and 16.6% coverage of academic textbooks and reports,  
377 respectively (see Table 2 for the summarized statistics). Although the coverage exhibits  
378 variation across the genres of economics discourse, both OOS results match the coverage of  
379 the corpus used to derive the REWL. This supports our argument for the robustness of the  
380 proposed algorithm. To answer the first research question, it is possible to suggest that these  
381 validation results prove that the REWL represents the construct of specialized vocabulary that  
382 goes beyond terminology and is equally applicable to both academic and mass media texts.

Genre	IS/OOS	Coverage by the REWL
Research articles	IS	13.1%
Magazine articles	IS	11.7%
Academic textbooks	OOS	12.6%

Reports issued by financial institutions	OOS	16.6%
--	-----	-------

383 Table 2 REWL coverage by genre in economics discourse.

384           Regarding how the REWL compares to the general service vocabulary in the coverage of  
385 economics texts, it is important to examine the additional coverage that the list provides in  
386 contrast to the additional coverage by groups of words from the list of Sharoff et al. (2013) (see  
387 Table 3). To compute the additional coverage, we considered ranges of ranks in increments of  
388 1,000 from Sharoff et al. (2013). For each range, we excluded all words from the REWL that are  
389 present in the list by Sharoff et al. (2013) and have ranks within this range or higher. We referred  
390 to the coverage by the remaining words in the REWL as “additional,” which is presented in Table  
391 3. The results show that the 507 lemmas from the REWL provide a larger additional coverage of  
392 economics texts than all 1,000 increments of the 5,000 word list of general service vocabulary,  
393 except for the first one. Therefore, the list could hypothetically act as a shortcut in increasing  
394 the text coverage with fewer words.

Rank in Sharoff et al. list	1-1000	1001-2000	2001-3000	3001-4000	4001-5000	Outside Sharoff et al.'s
-----------------------------	--------	-----------	-----------	-----------	-----------	--------------------------



						5,000 words
Coverage by words from Sharoff et al.'s list with given ranks	46.2%	10.6%	6.3%	5.3%	3.5%	NA
Additional coverage by the REWL	12.7%	10.8%	8.2%	7.2%	5.2%	4.4%

395 Table 3 A comparison of the economics text coverage by the REWL and Sharoff et al.'s (2013)  
396 list based on Sharoff et al.'s (2013) frequencies.

397 Another comparison revealed that the REWL differs substantially from the general  
398 vocabulary lists proposed by Andryushina (2006, 2008, 2009). Only 17 words of the REWL are  
399 found in the three combined lists, which indicates that the REWL has considerable value for  
400 readers of economics texts by addressing economics-specific vocabulary.

#### 401 4.3. Main characteristics of the REWL

402 The REWL has several important characteristics. First, it can increase the coverage of  
403 texts from the two main genres in economics (academic and mass media articles) in contrast  
404 with the 1,000 increments of general vocabulary outside the first 1,000 words associated with

405 the ACTFL<sup>1</sup> Intermediate Low level of L2 Russian proficiency in the study by Hacking and  
406 Tschirner (2017). The study by Hacking and Tschirner (2017) suggests that so-called  
407 vocabulary thresholds tailored to the ACTFL levels of proficiency exist. In their study, each  
408 level of proficiency was associated with a certain frequency-based vocabulary volume linked  
409 to reading comprehension. Although vocabulary knowledge is not the only factor that  
410 contributes to reading comprehension alongside lexical inferencing and grammatical  
411 knowledge (Comer 2012a; Comer 2012b), vocabulary quality and size can make a difference in  
412 L2 acquisition. As such, it is critical to emphasize the idea of the threshold beyond which  
413 specialized vocabulary might give more returns on the effort invested in vocabulary acquisition  
414 by learners of L2 Russian. Based on the present study, we hypothesized that this threshold for  
415 economics-related texts might be 1,000 lemmas.

416         The REWL has a manageable length, which is in line with the argument by Nation  
417 (2016) who mentioned that foreign language vocabulary lists of over 2,000 words could be  
418 seen as overwhelming to instructors and unsuitable for a single course. The 507 words of the  
419 REWL can be studied within a one-semester Russian language course and given to students  
420 for independent study.

#### 421 4.4. Potential application of the REWL

---

<sup>1</sup> ACTFL stands for the American Council on the Teaching of Foreign Languages; it publishes proficiency guidelines that specify the major levels of language mastery.

422 The issue of the target audience for the REWL is open to debate, as it is difficult to  
423 estimate the exact number of its potential users. Nevertheless, statistics published by the  
424 Ministry of Science and Higher Education of the Russian Federation (2018) allow for a rough  
425 estimation of the size of this audience. These statistics indicate that 39% of international  
426 students major in economics and management. At the same time, according to the Federal  
427 Service of State Statistics (2021), there were 278,000 international students in Russia in 2018,  
428 with 79,300 coming from countries other than former Soviet republics. Although there are no  
429 further data on the major preferences for the latter subgroup, several assumptions still  
430 emerge from this evidence. Namely, tens of thousands of students with L2 Russian choose  
431 economics-related fields as their majors in Russia. This constitutes a significant number of  
432 REWL users, not to mention those who study Russian and have an interest in economics  
433 outside Russia.

434 The REWL presents a validated source of specialized vocabulary for instructional  
435 materials. However, it is important to exercise caution in its implementation. Toth and  
436 Moranski (2018) highlighted the gap between researchers and teaching practitioners and  
437 emphasized the tailored implementation of any research findings in a foreign language  
438 classroom. One concern over the difference between research and learning material is that  
439 the words from the vocabulary lists should not appear in L2 classrooms in a decontextualized

440 manner (Brezina and Gablasova 2017). Instead, engaging students in context exploration to  
441 avoid rote learning can add value to vocabulary acquisition.

442 The list also needs careful consideration from curriculum designers and Russian  
443 language instructors in terms of vocabulary acquisition strategies. First, in a foreign language  
444 classroom, the list can be put into practice by taking into account the basic principles and  
445 recommendations discussed by Laufer (1990) and Barcroft (2004). In practice, the list should  
446 be augmented with strategies for independent study and exercises such as mapping, as  
447 examined by Morin and Goebel Jr. (2001). Thus, the list should appear in the learner's foreign  
448 language landscape in conjunction with strategies to work with these words (Chung and  
449 Nation 2003; Yamamoto 2014).

450 Research on vocabulary acquisition suggests that vocabulary lists could be useful  
451 primarily for reading comprehension before the mastery of such vocabulary transfers into  
452 productive use in speaking and writing (Laufer 1998; Schmitt 2010). This concern over the gap  
453 between the receptive and productive utility of vocabulary lists has received considerable  
454 scholarly attention, with the general agreement that students' writing deviates significantly in  
455 vocabulary coverage from expert writing (Durrant 2016; Gardner and Davis 2016; Malmström  
456 et al. 2018). However, regarding Russian for special purposes, the degree to which students  
457 might use the list in speaking or writing depends on their personal needs and can be left to  
458 their discretion.

459 Furthermore, the REWL presents valuable opportunities to promote learner autonomy,  
460 which has increased with the advent of new technologies in foreign language learning (Little  
461 et al. 2017; Liu et al. 2018; Luke 2006; Rivera–Mills and Plonsky 2007; De Saint Léger 2009).  
462 Learner autonomy can be partially achieved through data-driven learning (Kennedy and Miceli  
463 2001) in which students can pose a question, pick a search strategy, observe examples, and  
464 make inferences. In such self-directed learning, students can be encouraged to discover the  
465 collocational patterns of REWL words. Such activity can increase both the depth and range of  
466 learners' vocabulary acquisition. Thus, the REWL yields new learning opportunities both in  
467 group settings, such as in an L2 Russian classroom, and in study abroad programs, which  
468 emphasize learner autonomy.

469 Additionally, for students who are reasonably proficient in Russian derivational  
470 morphology (i.e., who understand the basic affixes), the REWL can be clustered into a set of  
471 324 word families, with each family gathering words with the same root under one umbrella.  
472 This expands the vocabulary beyond 507 words and allows learners to explore new affixes of  
473 existing word-family members. For instance, five lemmas from the word family of *import*  
474 "import" are included in the REWL—*import* "import, noun," *importer* "importer," *importirovat'*  
475 "to import," *importnyj* "import, adj.," and *importozameščenie* "import substitution"—but the  
476 word family of *upravljaemyj* "run by, managed by" does not have any other family members in  
477 the REWL. Knowledge of Russian word formation can help add associated derivatives such as

478 *upravljat* "to run, to manage," *upravljenje* "management," and *upravljenec* "manager." Because  
479 family members in the REWL are represented unevenly (abundant in some word families and  
480 scarce in others), the REWL can encourage L2 Russian learners to acquire additional  
481 vocabulary from their professional field, which increases text coverage and comprehension.

482 Finally, the REWL contains a number of words that can serve as cognates with other  
483 languages. Specifically, Russian language learners who already speak English can easily  
484 identify approximately 100 word families out of 324 since they are cognates with English (e.g.,  
485 *import* "import," *bank* "bank," *trend* "trend," and *embargo* "embargo"). Research on cognates  
486 shows that their presence can significantly ease the vocabulary acquisition process, although  
487 "it may be more challenging for second language learners to use cognates than noncognates,  
488 at least shortly after learning" (Rogers, Webb and Nakata 2015, p. 9). Therefore, the presence  
489 of cognates might increase the learnability of one-third of the REWL.

490 Our study has a number of limitations that suggest opportunities for further research.  
491 First, the study is limited by the corpora used to create the word list. Although the corpus  
492 compilation is in line with previous research in the field, word lists are still corpus-sensitive  
493 (Miller and Biber 2015). This means that a larger corpus might yield somewhat different  
494 results. However, the cross-validation methodology that we used considerably decreases  
495 possible deviations from the obtained outcomes. Second, some of the REWL characteristics  
496 are tested against one specific general service list (Sharoff et al. 2013). A comparison to yet

497 another general vocabulary list might bring different results, but this should not affect the  
498 composition of the REWL, since the list itself was not generated by using a general service  
499 list. Finally, we focused only on the field of economics; therefore, research in other  
500 professional areas is needed to investigate whether such a hypothetical shortcut is possible in  
501 other disciplines. Regardless of these limitations, the methodology for the creation of the  
502 REWL has robust pedagogical implications and can improve LSP curriculum design for other  
503 professional areas of L2 Russian application.

## 504 5. Conclusion

505 Our results offer a useful tool for LSP students and aid instructors who share the vision  
506 of vocabulary tools as essential components of language mastery. Potentially, the REWL could  
507 benefit L2 Russian students with a professional interest in economics and provide a shortcut  
508 to a better comprehension of specialized texts before learners reach general advanced  
509 proficiency. The methodology can also be used by instructors of other languages interested in  
510 developing lists of specialized vocabulary to meet the professional needs of their students.

## 513 Appendix A

### 514 Sources of Corpus Data

515 Academic sources:

- 516 1. Economy of Region  
517 2. Foresight and STI Governance  
518 3. Russian Management Journal  
519 4. Studies on Russian Economic Development  
520 5. The HSE Economics Journal  
521 6. The Journal of the New Economics Association  
522 7. Vestnik of Saint Petersburg University. Management  
523 8. Voprosy Ekonomiki  
524

525 Mass media sources:

- 526 1. Forbes  
527 2. Biznes zhurnal  
528 3. Biudzheth  
529 4. Harvard Business Review  
530 5. Kommersant  
531 6. Neft' i kapital  
532 7. RBC  
533 8. Finansovaja gazeta  
534 9. Ekonomika i zhizn'  
535 10. Expert  
536

537 Academic textbook sources:

- 538 1. Čekina, E.B., and D.A. Kapasova. 2020. Russkij jazyk. Učebnoe posobie dlja  
539 ekonomistov. Almaty: Kazak Universitet.  
540 2. Dubinskaja, E.V., and L.P. Paj. 2007. Russkij jazyk buduščemu ekonomistu. Moscow:  
541 MADI  
542 3. Fedotova, I.E, and S.S. Khoronenko. 2012. Russkij jazyk dlja ekonomistov. Izučaem  
543 professional'nuju leksiku. Minsk: BGEU  
544 4. Popova, I.M, E.B. Patrakeeva, and M.M. Glazkova. 2014. Russkij jazyk dlja ekonomistov.  
545 Tambov: FGBOU VPO "TGTU".  
546

547 Reports sources:

- 548 1. Bank of Russia Annual Report for 2016  
549 2. Bank of Russia Annual Report for 2017  
550 3. Bank of Russia Annual Report for 2018  
551 4. Bank of Russia Annual Report for 2019  
552 5. Bank of Russia Annual Report for 2020



## 555 The REWL in lemmas grouped by word families in Cyrillic

556 *А в т о м о б и л ь н ы й; а г р а р н ы й, а г р о п р о д о в о л ь с т в е н н ы й;*  
557 *а д м и н и с т р и р о в а н и е; а к к у м у л и р о в а т ь; а к т и в; а к ц и я;*  
558 *а л ь т е р н а т и в а; а м б и ц и о з н ы й; а с с о р т и м е н т; а у к ц и о н;*  
559 *а у т с а й д е р; б а н к, б а н к о в с к и й; б а н к р о т с т в о;*  
560 *б е д н о с т ь; б е з н а л и ч н ы й; б е н з и н; б и з н е с; б и р ж е в о й;*  
561 *б о н у с; б у м; б у х г а л т е р с к и й; б ю д ж е т, б ю д ж е т н ы й;*  
562 *в а к а н с и я; в а л ю т а; в в о з; в з н о с; в к л а д ы в а т ь,*  
563 *в л о ж е н и е, о т к л а д ы в а т ь; в л а д е л е ц; в н е д р я т ь;*  
564 *в н е ш н и й; в о д о с н а б ж е н и е; в о з в р а т; в о з д е й с т в о в а т ь;*  
565 *в о з о б н о в л я т ь; в о л а т и л ь н о с т ь; в с п л е с к;*  
566 *в ы б о р о ч н ы й, п о д ы б о р к а; в ы г о д н о, в ы г о д н ы й,*  
567 *н е в ы г о д н ы й, в з а и м о в ы г о д н ы й; в ы и г р ы в а т ь,*  
568 *п р о и г р ы в а т ь; в ы п у с к, в ы п у с к а т ь; в ы р у ч к а;*  
569 *в ы с о к о к в а л и ф и ц и р о в а н н ы й; в ы с о к о т е х н о л о г и ч н ы й,*  
570 *т е х н о л о г и я; г а з, г а з о п р о в о д; г и б к о с т ь; д е - ф а к т о;*  
571 *д е в а л ь в а ц и я; д е л о в о й; д е м о г р а ф и я; д е н е ж н ы й;*  
572 *д е п о з и т; д и в е р с и ф и к а ц и я, д и в е р с и ф и ц и р о в а т ь;*  
573 *д и з е л ь н ы й; д и н а м и ч н ы й; д и с б а л а н с; д о б ы в а т ь; д о л г,*  
574 *д о л г о в о й, з а д о л ж е н н о с т ь; д о л г о с р о ч н ы й,*  
575 *к р а т к о с р о ч н ы й, с р е д н е с р о ч н ы й, п р о с р о ч и в а т ь;*  
576 *д о л е в о й, д о л я; д о м а ш н и й; д о п у щ е н и е; д о р о ж н ы й;*  
577 *д о с т а т о ч н о с т ь; д о с т и г н у т ы й; д о х о д, д о х о д н о с т ь,*  
578 *д о х о д н ы й; е ж е г о д н о, е ж е г о д н ы й, с р е д н е г о д о в о й;*  
579 *е ж е м е с я ч н ы й; ж е с т к и й, у ж е с т о ч е н и е; ж и л ь е;*  
580 *з а в ы ш а т ь, з а в ы ш е н и е, п о в ы ш а т ь с я, п о в ы ш е н и е,*  
581 *п р е в ы ш е н и е; з а е м, з а е м н ы й, з а е м щ и к; з а к а з,*  
582 *з а к а з ч и к; з а к р ы т и е; з а к у п а т ь, з а к у п к а, з а к у п о ч н ы й;*  
583 *з а м е д л е н и е, з а м е д л я т ь; з а м е щ е н и е,*  
584 *и м п о р т о з а м е щ е н и е; з а н я т о й, з а н я т о с т ь; з а п а с;*  
585 *з а т р а т а, з а т р а т н ы й; з а я в к а; з д р а в о о х р а н е н и е;*  
586 *и г р о к; и з д е р ж к а; и з н о с; и м п о р т, и м п о р т е р,*

587 импортировать, импортный; инвестирование,  
588 инвестировать, инвестиционный, инвестиция,  
589 инвестор; индекс; индикатор; индустриальный;  
590 инфраструктурный; ископаемое; исход; исчерпание;  
591 капитал, капитализация, капиталовложение,  
592 капиталый; кварталый; коммунальный;  
593 компания; компенсировать, компенсироваться;  
594 комфортный; конкурент, конкурентный,  
595 конкурентоспособный, конкуренция,  
596 конкурировать; консолидация, консолидированный,  
597 консолидировать; консультационный;  
598 контейнерный; контракт; конъюнктура;  
599 координация, координировать; корзина;  
600 корпоративный, корпорация; корректировать,  
601 корректировка, коррекция, скорректировать;  
602 косвенный; кредит, кредитный, кредитование;  
603 кризис, кризисный, предкризисный; крупный,  
604 укрупненный; либерализация; лидер, лидировать;  
605 ликвидация; ликвидность, ликвидный; логарифм;  
606 логистика, логистический; логично; льгота,  
607 льготный; маловероятный; материнский;  
608 металлургический; мировой, общемировой,  
609 среднемировой, всемирный; модернизировать;  
610 монополист; надбавка; накапливать; налаживать;  
611 налог, налоговый; намечаться; намного; нанимать;  
612 население; начислять; невысокий; негативно;  
613 неизменный; нелегальный; неопределенность;  
614 непрофильный; нетто; нефтегазовый, нефтепродукт,  
615 нефть, нефтяной; нивелировать; низкий, занижать,  
616 снижение; ниша; новация; номинальный; норматив;  
617 обеспеченность; облигация; обозримый;  
618 оборудование; обострение; обслуживание,  
619 обслуживать; объем; ожидание, ожидать, ожидаться;  
620 окупаемость; опережать, опережающий; оплата,  
621 оплачивать; оптовый; ориентир, переориентация;  
622 ослабление; отраслевой, отрасль, подотрасль;  
623 отставание, отставать; отток, приток; отчетность,

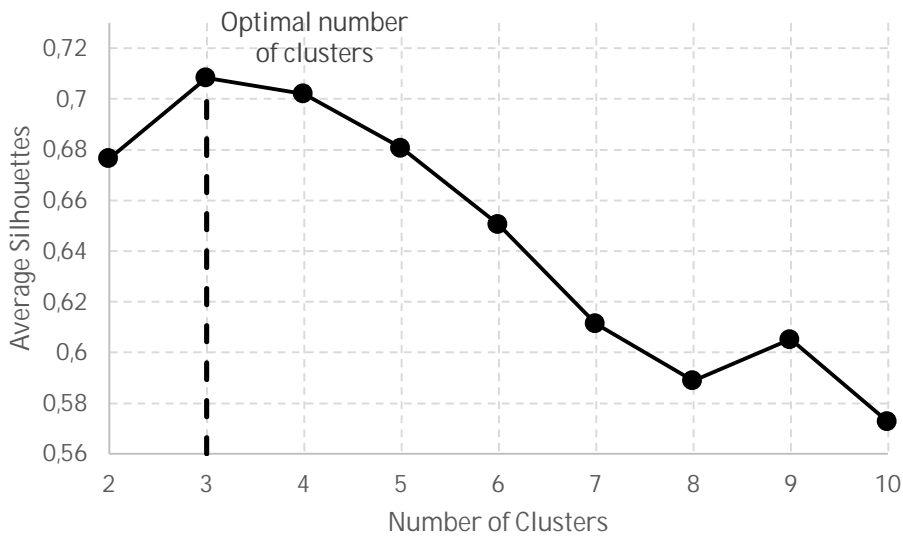
624 *отчетный; пай; паритет; парниковый; партнерский;*  
625 *пенсия, пенсионер, пенсионный, пенсия; перевозка;*  
626 *передовой; переработка; пересматривать; период;*  
627 *перспектива; пилотный; пищевой; плановый,*  
628 *запланировать; плата, платеж,*  
629 *платежеспособность, платежеспособный,*  
630 *выплачивать; погашение; поддержка; подключение;*  
631 *подразделение; показатель; покупатель,*  
632 *покупательный, покупка; политика,*  
633 *геополитический; получатель; портфель; поставка,*  
634 *поставщик, доставка; постепенный; потребитель,*  
635 *потребительский, потреблять; пошлина;*  
636 *правдоподобие; предприниматель, предприятие;*  
637 *предсказуемый, предсказывать; премия;*  
638 *преференция; прибыль, прибыльность, прибыльный;*  
639 *приватизация; привлекательность,*  
640 *привлекательный; примерно; приоритетный;*  
641 *приостановка; приходить; провал; прогноз,*  
642 *прогнозировать, прогнозироваться; продавец;*  
643 *продукт, продуктовый, продукция; проект;*  
644 *прожиточный; прозрачность; производитель,*  
645 *производить, производство; промышленность;*  
646 *пропорция, диспропорция; прорывной;*  
647 *протекционизм; процентный; рабочий, заработок;*  
648 *развивающийся, развитый; размещать;*  
649 *разнонаправленный; распределительный,*  
650 *перераспределение; рассчитывать,*  
651 *рассчитываться; расход, расходный; реагировать;*  
652 *реальный; регион, региональный; регрессия; резерв;*  
653 *резидент, нерезидент; резкий; рейтинг;*  
654 *рентабельность; реструктуризация; ресурс; риск,*  
655 *рискованный; розничный; рост, прирост,*  
656 *наращивание, быстрорастущий, расти; рублевый;*  
657 *рынок, рыночный, нерыночный; санкция;*  
658 *сбалансированность, сбалансированный,*  
659 *сбалансировать; сбережение; сбыт; сводный; сделка;*  
660 *сдерживание, сдерживать; сектор, секторальный;*

661 сельский; сланцевый; смягчение; совокупный;  
662 создаваемый; сокращать; сопоставимый,  
663 составлять; спад; специализироваться; спрос;  
664 средний, средневзвешенный; ссуда; стабилизация,  
665 стабилизироваться, стабильно, нестабильный,  
666 нестабильность; ставка; стагнация; статистика;  
667 стимулирование; стоимость, себестоимость;  
668 страна, страновой, межстрановой; стратегически;  
669 строительный, строительство,  
670 машиностроительный; субсидирование,  
671 субсидировать, субсидия; суверенный; счет; сырье,  
672 сырьевой, несырьевой; тариф, тарифный; темп;  
673 технически; товар, товарный; топливный, топливо;  
674 торги, торговля, торговый, внешнеторговый;  
675 тормозить; транзакция; трансграничный;  
676 транспортировка, транспортный; тренд;  
677 туристический; убыточный; увязывать;  
678 удешевление; удорожание; укрепление; улучшаться;  
679 управляемый; ускоренный; услуга; устойчиво;  
680 усугублять; ухудшать, ухудшаться, ухудшение; учет,  
681 учетный; фармацевтический; филиал;  
682 финансирование, софинансирование,  
683 финансировать, финансироваться, финансовый,  
684 финансы; фонд, фондовый; хозяйство; целесообразно;  
685 целлюлозно-бумажный; цена, ценовой,  
686 ценообразование, недооценивать, переоценка;  
687 централизация, централизованный; чреватый;  
688 чрезмерный; экзогенный; экология;  
689 эконометрический, экономика, экономически,  
690 экономический, экономия, макроэкономический;  
691 экспансия; экспортер, экспортировать,  
692 экспортный; экстерналии; электроника; эмбарго;  
693 энергетика, энергетический, энергоноситель,  
694 энергоресурсы, электроэнергетика,  
695 электроэнергия; эффект, эффективность,  
696 неэффективность, неэффективный.  
697

698

699 Appendix C

700 Selecting optimal number of clusters for k-means clustering



701

702 Fig. C.1 Average silhouette values vs number of clusters.

703

704

705

706 Appendix D

707 The REWL with frequency ratios and log-likelihood scores (the economics corpus against

the other corpora)

Lemma	Frequency	
	Ratio	LL-score
ликвидность	528.7	1548.8
несырьевой	215.9	310.0
заемщик	173.9	1735.7
портфель	134.4	2464.7
агропродовольственный	123.0	172.8

Lemma	Frequency	
	Ratio	LL-score
волатильность	100.7	1399.2
кредитование	97.5	2569.1
рублевый	91.6	759.2
макроэкономический	87.3	2765.7
доходность	83.7	3335.6
ценовой	80.1	877.6
импортозамещение	75.3	1642.5
целлюлозно-бумажный	70.6	191.5
облигация	70.4	1718.1
секторальный	67.9	183.8
капитализация	66.9	903.5
эмбарго	66.0	356.6
выручка	62.9	2030.3
предкризисный	56.2	149.7
сбережение	54.1	1292.2
валюта	53.3	3809.8
окупаемость	49.6	261.0
фондовый	49.6	1302.7
рентабельность	49.4	1363.9
импорт	46.3	5011.7
газопровод	44.3	230.4
депозит	43.9	1311.4
банк	43.3	18563.8
нерезидент	42.8	332.0
кредит	41.1	6876.9
контейнерный	40.4	207.5
ликвидный	39.4	252.3
машиностроительный	37.4	380.3
экспортер	36.9	1029.5
удорожание	33.6	337.0
субсидирование	33.5	546.0
просрочивать	33.1	413.7
девальвация	31.8	790.6
тарифный	30.5	414.3
экспортный	30.1	2490.7

Lemma	Frequency	
	Ratio	LL-score
а у к ц и о н	29.8	1029.6
а к т и в	29.4	6652.6
э к о н о м е т р и ч е с к и й	28.9	816.4
ц е н а	28.1	20920.2
т а р и ф	28.1	1716.5
п р о ц е н т н ы й	26.9	3387.2
с т р а н о в о й	26.6	353.2
э л е к т р о н и к а	26.1	346.8
и м п о р т н ы й	26.0	1376.3
т о п л и в н ы й	25.9	218.0
с т а в к а	25.7	7921.6
к о м п а н и я	25.2	35189.6
с ы р ь е в о й	24.1	1173.0
н е ф т е п р о д у к т	23.8	846.5
б е н з и н	23.8	422.2
п р о д у к т о в ы й	23.3	440.3
и н в е с т о р	22.8	4524.7
к р е д и т н ы й	22.5	3342.2
п р и в а т и з а ц и я	22.5	1053.1
с р е д н е с р о ч н ы й	22.3	809.5
в з а и м о в ы г о д н ы й	22.2	129.8
и н в е с т и ц и я	22.0	8975.9
э к с п о р т и р о в а т ь	22.0	512.8
с п р о с	21.8	7663.6
и н в е с т и ц и о н н ы й	21.5	6126.9
с е б е с т о и м о с т ь	21.3	470.4
к о р з и н а	21.2	491.7
б е з н а л и ч н ы й	20.9	144.8
б а н к о в с к и й	20.8	3874.5
с с у д а	20.5	188.5
т е м п	20.4	7146.4
р о з н и ч н ы й	20.1	2092.7
в н е ш н е т о р г о в ы й	19.7	470.9
в ы с о к о т е х н о л о г и ч н ы й	19.6	1075.0
б о н у с	19.5	155.6

Lemma	Frequency	
	Ratio	LL-score
прибыльность	18.9	278.3
резерв	18.9	1733.7
протекционизм	18.6	167.5
задолженность	18.3	1336.1
номинальный	18.1	1158.2
ценообразование	18.1	567.9
совокупный	18.0	2170.6
инвестировать	17.9	856.5
долговой	17.8	1152.4
платежеспособность	17.2	151.9
монополист	16.6	127.4
заемный	16.5	343.8
логистика	16.4	286.1
долгосрочный	16.4	4002.9
рынок	16.2	28484.3
покупательный	15.0	241.1
инвестирование	15.0	830.0
прирост	14.8	3587.5
биржевой	14.8	377.1
потребительский	14.7	2755.4
игрок	14.7	1971.0
ассортимент	14.6	278.2
экономика	14.5	23226.2
убыточный	14.5	198.5
конкурент	14.5	1114.7
покупатель	14.3	1671.7
межстрановой	14.2	404.0
завышение	14.2	118.9
закупать	14.1	369.7
прибыль	13.9	3531.1
погашение	13.7	527.7
покупка	13.7	2092.4
реструктуризация	13.6	423.9
спад	13.6	1254.5
стагнация	13.5	560.8



Lemma	Frequency	
	Ratio	LL-score
к р а т к о с р о ч н ы й	13.3	1499.9
п о ш л и н а	12.9	698.4
н е ф т е г а з о в ы й	12.9	853.4
у д е ш е в л е н и е	12.8	117.7
н е р ы н о ч н ы й	12.7	168.1
л ь г о т н ы й	12.6	410.6
и з д е р ж к а	12.6	791.9
с р е д н е в з в е ш е н н ы й	12.6	305.9
э л е к т р о э н е р г и я	12.5	1166.2
б ы с т р о р а с т у щ и й	12.3	148.9
п р о м ы ш л е н н о с т ь	12.3	4794.7
п р о и з в о д и т е л ь	12.2	2904.0
п а р н и к о в ы й	12.1	194.2
п о с т а в к а	12.1	2435.2
з а е м	12.1	798.0
н а ч и с л я т ь	12.1	157.0
с е к т о р	12.1	8882.5
с у б с и д и я	12.1	963.0
у ж е с т о ч е н и е	11.9	462.9
р а с п р е д е л и т е л ь н ы й	11.8	153.0
с о ф и н а н с и р о в а н и е	11.8	153.0
и м п о р т е р	11.8	327.1
м е т а л л у р г и ч е с к и й	11.8	502.2
б и з н е с	11.7	10535.6
к о н к у р е н т о с п о с о б н ы й	11.7	426.8
к о н ь ю н к т у р а	11.7	702.6
р ы н о ч н ы й	11.5	4145.4
т о п л и в о	11.5	1037.9
к р и з и с н ы й	11.5	1013.8
к в а р т а л ь н ы й	11.3	253.8
а к ц и я	11.3	3769.0
п л а н о в ы й	11.2	575.1
н а р а щ и в а н и е	11.0	616.6
с у б с и д и р о в а т ь	11.0	116.9
п р о д у к ц и я	10.9	7151.5

Lemma	Frequency	
	Ratio	LL-score
энергоситель	10.8	333.6
экстерналии	10.8	125.0
доходный	10.8	332.6
диверсификация	10.7	727.9
выпуск	10.6	3866.7
сбалансированность	10.6	323.7
паритет	10.5	170.5
бюджет	10.4	5376.1
платеж	10.3	1501.8
капитал	10.3	9233.9
поставщик	10.3	1861.4
рост	10.2	23321.6
потребитель	10.2	3708.1
предприятие	10.2	9374.9
стоимость	10.1	6317.5
прорывной	10.1	113.2
доход	10.0	10285.6
товар	10.0	7212.7
транзакция	9.9	451.9
прогнозироваться	9.9	174.9
прожиточный	9.8	264.0
занятость	9.8	2922.0
платежеспособный	9.6	141.8
банкротство	9.6	805.0
затрата	9.5	3888.2
заявка	9.5	858.3
развивающийся	9.4	1348.7
оборудование	9.3	2252.1
отчетность	9.3	1023.6
прибыльный	9.2	157.4
отрасль	9.1	7978.0
улучшаться	9.0	324.3
диверсифицировать	9.0	202.3
пенсионный	9.0	2627.0
прогноз	9.0	3709.0

Lemma	Frequency	
	Ratio	LL-score
санкция	8.7	1936.4
отток	8.7	637.5
товарный	8.7	1096.1
ввоз	8.6	265.4
среднегодовой	8.6	650.4
бюджетный	8.6	4267.2
замедление	8.6	758.9
логистический	8.6	366.8
торговля	8.5	4017.7
бухгалтерский	8.5	265.0
нефть	8.4	4575.7
опережающий	8.3	259.2
демография	8.3	93.3
стимулирование	8.3	1204.9
пенсионер	8.3	533.3
расход	8.2	5980.1
отраслевой	8.2	1617.3
финансовый	8.2	10786.9
дизельный	8.1	97.6
нефтяной	8.1	990.3
логарифм	8.1	364.1
капитальный	8.0	593.9
долг	8.0	2356.0
либерализация	7.9	365.3
неэффективность	7.7	231.8
партнерский	7.7	166.7
консолидированный	7.7	293.2
диспропорция	7.6	152.4
продавец	7.6	519.1
конкурентный	7.6	1516.2
укрупненный	7.6	93.5
занятой	7.5	161.1
нетто	7.5	91.6
обеспеченность	7.5	500.8
импортировать	7.4	218.8

Lemma	Frequency	
	Ratio	LL-score
в з н о с	7.4	595.5
п р о и з в о д с т в о	7.4	11785.8
к о н с у л ь т а ц и о н н ы й	7.3	95.1
ф и н а н с ы	7.3	991.1
и н ф р а с т р у к т у р н ы й	7.2	541.4
ф о н д	7.1	4774.3
н е п р о ф и л ь н ы й	7.0	89.5
о т ч е т н ы й	7.0	244.5
к а п и т а л о в л о ж е н и е	7.0	226.8
з а к а з	7.0	735.2
з а к у п к а	7.0	1819.1
и с к о п а е м о е	7.0	462.3
п р о г н о з и р о в а т ь	7.0	649.2
о б щ е м и р о в о й	6.9	120.4
т о р г и	6.9	722.6
с м я г ч е н и е	6.9	323.9
с л а н ц е в ы й	6.9	172.6
с н и ж е н и е	6.8	6276.8
т о р г о в ы й	6.8	2175.4
п е н с и я	6.7	1334.3
д о б ы в а т ь	6.7	526.8
л ь г о т а	6.7	672.5
ж и л ь е	6.6	1078.0
а в т о м о б и л ь н ы й	6.6	440.8
э к з о г е н н ы й	6.5	352.3
с р е д н е м и р о в о й	6.4	96.2
п е р е в о з к а	6.4	734.9
а у т с а й д е р	6.4	148.2
о ж и д а т ь с я	6.3	499.1
э к о н о м и ч е с к и й	6.2	16449.1
с б ы т	6.2	317.4
з а к а з ч и к	6.1	848.7
и з н о с	6.1	197.4
п р е д п р и н и м а т е л ь	6.1	1947.7
к р и з и с	6.1	3633.4

Lemma	Frequency	
	Ratio	LL-score
запланировать	6.1	275.8
сдерживать	6.0	365.7
экономия	6.0	634.4
стабилизироваться	6.0	126.9
ежемесячный	6.0	169.2
энергетика	6.0	615.8
водоснабжение	5.9	92.5
денежный	5.8	2633.9
оптовый	5.8	357.6
скорректировать	5.8	358.5
финансировать	5.8	251.1
подвыборка	5.8	110.1
доля	5.7	9015.0
индикатор	5.7	1774.3
здравоохранение	5.7	1020.2
лидировать	5.6	314.6
премия	5.6	682.1
отставание	5.6	330.5
корректировка	5.5	434.6
экономически	5.5	525.4
выплачивать	5.5	308.6
сдерживание	5.4	104.1
тренд	5.4	1177.5
приостановка	5.4	76.7
контракт	5.4	1643.6
консолидировать	5.3	113.0
пай	5.3	98.9
налог	5.3	2205.3
объем	5.2	8485.5
надбавка	5.2	178.5
ослабление	5.2	469.3
финансирование	5.2	2570.6
сырье	5.2	905.5
нанимать	5.1	136.3
достаточность	5.1	142.5

Lemma	Frequency	
	Ratio	LL-score
пищевой	5.1	418.7
преференция	5.1	125.7
долевой	5.1	114.2
закупочный	5.1	137.3
конкуренция	5.0	1950.1
финансироваться	5.0	110.7
хозяйство	5.0	2532.0
сделка	5.0	1971.6
плата	5.0	2466.9
ухудшение	4.9	528.8
показатель	4.9	9993.3
правдоподобие	4.9	89.7
расти	4.9	2367.3
сбалансированный	4.8	160.9
модернизировать	4.8	69.4
учетный	4.8	122.0
вложение	4.8	808.9
сокращать	4.8	824.0
ухудшаться	4.8	186.2
строительный	4.7	437.0
рейтинг	4.7	1612.8
ускоренный	4.7	208.7
трансграничный	4.7	162.6
норматив	4.7	318.1
обозримый	4.7	82.2
газ	4.7	1805.9
компенсировать	4.6	372.4
коммунальный	4.6	216.1
расходный	4.6	112.1
регион	4.6	8657.4
риск	4.6	4112.0
исчерпание	4.6	100.5
услуга	4.6	4484.1
заработок	4.5	262.1
деловой	4.5	940.9

Lemma	Frequency	
	Ratio	LL-score
выгодный	4.5	321.6
выгодно	4.5	168.3
рассчитываться	4.5	462.1
статистика	4.5	1356.8
корпоративный	4.4	1767.9
оплата	4.4	1107.4
стратегически	4.4	77.6
развитый	4.4	1649.4
дисбаланс	4.4	162.7
завышать	4.4	184.5
подотрасль	4.4	67.9
занижать	4.3	151.4
неэффективный	4.3	282.6
привлекательность	4.3	336.7
стабильно	4.2	145.0
прозрачность	4.2	262.2
стабилизация	4.2	347.1
отстаивать	4.2	272.8
резидент	4.2	247.2
де-факто	4.2	76.0
пилотный	4.1	157.0
индустриальный	4.1	343.9
налоговый	4.1	2105.0
страна	4.1	14700.4
бум	4.1	97.7
ожидать	4.1	1689.9
рискованный	4.1	144.2
ежегодно	4.0	346.7
индекс	4.0	2716.6
доставка	4.0	224.2
координация	4.0	285.6
лидер	4.0	1173.9
устойчиво	4.0	151.3
передовой	4.0	248.9
энергоресурсы	4.0	174.6

Lemma	Frequency	
	Ratio	LL-score
к о р п о р а ц и я	4.0	824.2
п е р е о р и е н т а ц и я	3.9	61.7
п о в ы ш е н и е	3.9	3267.3
п р и о р и т е т н ы й	3.9	462.9
о п е р е ж а т ь	3.9	256.7
п е р е о ц е н к а	3.9	118.5
э ф ф е к т	3.9	3536.8
г и б к о с т ь	3.9	251.7
о б с л у ж и в а т ь	3.9	154.6
в ы и г р ы в а т ь	3.8	193.8
е ж е г о д н ы й	3.8	310.0
з а п а с	3.8	747.9
с т р о и т е л ь с т в о	3.8	1506.4
э л е к т р о э н е р г е т и к а	3.8	158.4
ф и л и а л	3.7	225.8
п о л у ч а т е л ь	3.7	228.1
п о л и т и к а	3.7	5340.0
с п е ц и а л и з и р о в а т ь с я	3.7	138.3
п о в ы ш а т ь с я	3.7	357.7
т р а н с п о р т и р о в к а	3.7	161.5
о п л а ч и в а т ь	3.6	187.5
т е х н о л о г и я	3.6	3184.8
в н е д р я т ь	3.6	254.1
п о д к л ю ч е н и е	3.6	77.8
в а к а н с и я	3.5	106.9
р е г р е с с и я	3.5	458.6
т о р м о з и т ь	3.5	88.6
о б с л у ж и в а н и е	3.5	498.5
п е р е р а с п р е д е л е н и е	3.5	326.7
п р и т о к	3.5	423.7
н е о п р е д е л е н н о с т ь	3.5	664.7
у х у д ш а т ь	3.5	77.7
н е в ы г о д н ы й	3.4	56.6
р е г и о н а л ь н ы й	3.4	2640.2
п р е в ы ш е н и е	3.4	177.4



Lemma	Frequency	
	Ratio	LL-score
фармацевтический	3.4	92.7
ликвидация	3.4	293.3
сопоставимый	3.4	386.3
технически	3.4	58.0
энергетический	3.4	505.6
бедность	3.3	255.7
выборочный	3.3	87.7
сбалансировать	3.3	77.3
транспортный	3.3	846.0
население	3.3	4624.3
перспектива	3.3	1560.6
усугублять	3.3	69.3
владелец	3.3	528.4
счет	3.3	3185.6
замещение	3.3	229.7
продукт	3.3	2229.6
затратный	3.2	56.1
консолидация	3.2	136.2
крупный	3.2	3175.6
нестабильность	3.2	154.1
нивелировать	3.2	55.4
подразделение	3.2	448.3
геополитический	3.2	126.8
ресурс	3.2	3193.4
мировой	3.2	2381.4
проект	3.2	4271.9
эффективность	3.2	2363.1
динамичный	3.2	126.3
ожидание	3.2	803.7
переработка	3.2	293.9
сводный	3.1	99.0
альтернатива	3.1	347.2
всемирный	3.1	273.0
компенсироваться	3.1	58.3
конкурировать	3.1	192.6

Lemma	Frequency	
	Ratio	LL-score
с у в е р е н н ы й	3.1	116.4
п р о в а л	3.1	148.8
п о т р е б л я т ь	3.0	134.4
а д м и н и с т р и р о в а н и е	3.0	80.2
в о з о б н о в л я т ь	3.0	122.6
п р о п о р ц и я	3.0	120.2
п р и м е р н о	2.9	951.9
а г р а р н ы й	2.9	217.2
в ы с о к о к в а л и ф и ц и р о в а н н ы й	2.9	81.3
н е л е г а л ь н ы й	2.9	78.9
р а з н о н а п р а в л е н н ы й	2.9	52.0
к о м ф о р т н ы й	2.9	61.9
н е г а т и в н о	2.8	186.4
к о с в е н н ы й	2.8	265.5
з а к р ы т и е	2.8	141.7
н и ш а	2.8	94.3
к о р р е к т и р о в а т ь	2.8	74.5
п р о и г р ы в а т ь	2.8	81.3
ц е н т р а л и з а ц и я	2.8	54.2
р а с с ч и т ы в а т ь	2.7	336.4
а м б и ц и о з н ы й	2.7	48.2
о р и е н т и р	2.7	170.9
к о р р е к ц и я	2.7	117.5
р е а л ь н ы й	2.7	1800.2
п р и в л е к а т е л ь н ы й	2.7	177.0
к о о р д и н и р о в а т ь	2.7	49.8
н е с т а б и л ь н ы й	2.7	52.0
з а м е д л я т ь	2.6	53.1
м а т е р и н с к и й	2.6	72.5
э к с п а н с и я	2.6	101.6
о т к л а д ы в а т ь	2.6	106.9
ц е н т р а л и з о в а н н ы й	2.6	81.3
а к к у м у л и р о в а т ь	2.6	38.6
э к о л о г и я	2.6	61.4
д о с т и г н у т ы й	2.6	41.5

Lemma	Frequency	
	Ratio	LL-score
целесообразно	2.6	147.6
приходиться	2.5	1028.7
туристический	2.5	76.9
увязывать	2.5	38.4
недооценивать	2.5	43.5
период	2.5	3628.2
возврат	2.5	163.1
всплеск	2.5	46.2
чреватый	2.4	41.2
чрезмерный	2.4	102.6
укрепление	2.4	248.4
выпускать	2.4	220.1
управляемый	2.4	42.2
низкий	2.4	1867.2
вкладывать	2.4	160.3
поддержка	2.4	1356.6
дорожный	2.4	138.3
рабочий	2.4	1165.0
реагировать	2.3	133.2
учет	2.3	1402.0
исход	2.3	111.9
резкий	2.3	333.5
составлять	2.3	3473.6
размещать	2.3	172.1
накапливать	2.3	57.4
невысокий	2.3	122.1
новация	2.2	45.6
намечаться	2.2	55.8
предсказывать	2.2	97.5
домашний	2.2	221.1
обострение	2.2	39.5
маловероятный	2.2	39.0
предсказуемый	2.1	39.8
сельский	2.1	550.9
допущение	2.1	61.0

Lemma	Frequency	
	Ratio	LL-score
жесткий	2.1	234.4
пересматривать	2.1	70.4
средний	2.1	1938.6
создаваемый	2.1	29.4
логично	2.0	26.3
налаживать	2.0	34.0
внешний	1.9	813.7
постепенный	1.9	93.0
намного	1.9	70.9
воздействовать	1.8	43.2
неизменный	1.8	66.6
производить	1.8	257.5

708

709

710 References

- 711 Andryushina, N.P., and T.V. Kozlova. 2006. Leksicheskij minimum po russkomu jazyku kak  
712 inostrannomu. Bazovyj uroven'. Obshchee vladenie. 3rd ed. Moscow: Zlatoust
- 713 Andryushina, N.P., and T.V. Kozlova. 2008. Leksicheskij minimum po russkomu jazyku kak  
714 inostrannomu. Pervyj sertifikacionnyi uroven'. Obshchee vladenie. 4th ed. Moscow:  
715 Zlatoust
- 716 Andryushina, N.P., and T.V. Kozlova. 2009. Leksicheskij minimum po russkomu jazyku kak  
717 inostrannomu. Vtoroj sertifikacionnyi uroven'. Obshchee vladenie. 2nd ed. St. Petersburg:  
718 Zlatoust
- 719 Andryushina, N.P et al. 2015. Requirements for First level of Russian as a foreign language.  
720 Common language. Professional Language. 3rd ed. St. Petersburg : Zlatoust.
- 721 Author et al. 2020.
- 722 Baayen, R. H. 2002. Word frequency distributions. Vol. 18. Springer Science & Business Media.
- 723 Bailey, D. H., J. Borwein, M. Lopez de Prado, and Q.J. Zhu. 2016. "The probability of backtest  
724 overfitting." *Journal of Computational Finance* 20: 39–69
- 725 Bakaya, R. M. 1967. "An Experiment in Compiling a Minimal Vocabulary for Reading Scientific-  
726 Technical Literature in Russian." *Babel* 13 (3): 163–168.

- 727 Barcroft, J. 2004. "Second language vocabulary acquisition: A lexical input processing  
728 approach." *Foreign Language Annals* 37 (2): 200–208.
- 729 Basturkmen, H., and C. Elder. 2004. "The practice of LSP." In *The handbook of applied*  
730 *linguistics*, edited by A. Davies and C. Elder, 672–94. Oxford, England: Blackwell.
- 731 Biber, D., R. Reppen, E. Schnur, and R. Ghanem. 2016. "On the (non) utility of Juilland's D to  
732 measure lexical dispersion in large corpora." *International Journal of Corpus Linguistics* 21  
733 (4): 439–464.
- 734 Bondi-Paganelli, M. 1996. "Language variations across genres: Quantifiers and worlds of  
735 reference in (and around) economics textbooks." *ASp. la revue du GERAS* (11-14): 33-53.
- 736 Brezina, V., and D. Gablasova. 2017. "How to produce vocabulary lists? Issues of definition,  
737 selection and pedagogical aims. A response to Gabriele Stein." *Applied Linguistics* 38 (5):  
738 764–767.
- 739 Brown, N. J. 1996. *Russian Learners' Dictionary: 10,000 Words in Frequency Order*. London:  
740 Routledge.
- 741 Chung, T. M., and P. Nation. 2003. "Technical vocabulary in specialised texts." *Reading in a*  
742 *foreign language* 15 (2): 103.
- 743 Chung, T. M., and Paul Nation. 2004. "Identifying technical vocabulary." *System* 32 (2): 251–  
744 263.

- 745 Comer, W. 2012a. "Lexical inferencing in reading in L2 Russian." *Reading in a Foreign Language*  
746 24 (2): 209-230.
- 747 Comer, W. 2012b. "The role of grammatical knowledge in reading for meaning in Russian." *The*  
748 *Slavic and East European Journal* 56 (2): 232-255.
- 749 Coxhead, A. 2000. "A new academic word list." *TESOL quarterly* 34 (2): 213–238.
- 750 Coxhead, A., and M. Demecheleer. 2018. "Investigating the technical vocabulary of  
751 plumbing." *English for Specific Purposes* 51: 84–97.
- 752 Coxhead, A., E. McLaughlin, and A. Reid. 2019. "The development and application of a  
753 specialised word list: the case of Fabrication." *Journal of Vocational Education &*  
754 *Training* 71 (2): 175–200.
- 755 Čekina, E.B., and D.A. Kapasova. 2020. *Russkij jazik. Učebnoe posobie dlja ekonomistov.*  
756 *Almaty: Kazak Universitet.*
- 757 De Saint Léger, D. 2009. "Self - assessment of speaking skills and participation in a foreign  
758 language class." *Foreign Language Annals* 42 (1): 158–178.
- 759 Dubinskaja, E.V., and L.P. Paj. 2007. *Russkij jazyk buduščemu ekonomistu.* Moscow: MADI
- 760 Durrant, P. 2016. "To what extent is the Academic Vocabulary List relevant to university  
761 student writing?" *English for Specific Purposes* 43: 49–61.
- 762 Federal Service of State Statistics. 2021. *Russian statistical yearbook.* 2018 (in Russian and  
763 English). Accessed on September 27, 2021: [https://gks.ru/bgd/regl/b18\\_13/Main.htm](https://gks.ru/bgd/regl/b18_13/Main.htm)

- 764 Fedotova, I.E, and S.S. Khoronenko. 2012. Russkij jazyk dlja ekonomistov. Izučaem  
765 professional'nuju leksiku. Minsk: BGEU
- 766 Gardner, D., and M. Davies. 2014. "A new academic vocabulary list." *Applied linguistics* 35 (3):  
767 305–327.
- 768 Gardner, D., and M. Davies. 2016. "A response to "To what extent is the Academic Vocabulary  
769 List relevant to university student writing?"." *English for Specific Purposes* 43: 62–68.
- 770 Gries, S. T. 2008. "Dispersions and adjusted frequencies in corpora." *International journal of*  
771 *corpus linguistics* 13 (4): 403–437.
- 772 Ha, A. Y. H., and K. Hyland. 2017. "What is technicality? A Technicality Analysis Model for EAP  
773 vocabulary." *Journal of English for Academic Purposes* 28: 35–49.
- 774 Hacking, J. F., and E. Tschirner. 2017. "The contribution of vocabulary knowledge to reading  
775 proficiency: The case of college Russian." *Foreign Language Annals* 50 (3): 500–518.
- 776 Hyland, K., and P. Tse. 2007. "Is there an "academic vocabulary"?" *TESOL quarterly* 41 (2):  
777 235–253.
- 778 Moini, R., and Z. Islamizadeh. 2016. "Do we need discipline-specific academic word lists?  
779 Linguistics Academic Word List (LAWL)." *Journal of Teaching Language Skills* 35 (3): 65–  
780 90.
- 781 Juilland, A. G., D. R. Brodin, and C. Davidovitch. 1971. *Frequency dictionary of French words*.  
782 The Hague: Mouton de Gruyter.



- 783 Kaufman, L., and P.J. Rousseeuw. 2009. Finding groups in data: an introduction to cluster  
784 analysis. Vol. 344. John Wiley & Sons.
- 785 Kennedy, C., and T. Miceli. 2001. "An evaluation of intermediate students' approaches to  
786 corpus investigation." *Language learning & technology* 5 (3): 77–90.
- 787 Kwary, D. A. 2011. "A hybrid method for determining technical vocabulary." *System* 39 (2):  
788 175–185.
- 789 Laufer, B. 1989. "What percentage of text-lexis is essential for comprehension." In *Special*  
790 *language: From humans thinking to thinking machines*, edited by C. Lauren and M.  
791 Nordman, 316–323. Clevedon: Multilingual Matters.
- 792 Laufer, B. 1990. "Ease and difficulty in vocabulary learning: Some teaching  
793 implications." *Foreign Language Annals* 23 (2): 147–155.
- 794 Laufer, B. 1998. "The development of passive and active vocabulary in a second language:  
795 same or different?." *Applied linguistics* 19 (2): 255–271.
- 796 Laufer, B., and G.C. Ravenhorst-Kalovski. 2010. "Lexical threshold revisited: Lexical text  
797 coverage, learners' vocabulary size and reading comprehension." *Reading in a Foreign*  
798 *Language* 22 (1): 15–30.
- 799 Lei, L., and D. Liu. 2016. "A new medical academic word list: A corpus-based study with  
800 enhanced methodology." *Journal of English for academic purposes* 22: 42–53.

- 801 Li, Y., and D. D. Qian. 2010. "Profiling the Academic Word List (AWL) in a financial  
802 corpus." *System* 38 (3): 402–411.
- 803 Little, D., L. Dam, and L. Legenhausen. 2017. *Language learner autonomy: Theory, practice  
804 and research*. Bristol: Multilingual Matters.
- 805 Liu, M.-C., Y.-M. Huang, and Y.-H. Xu. 2018. "Effects of individual versus group work on  
806 learner autonomy and emotion in digital storytelling." *Educational Technology Research  
807 and Development* 66 (4): 1009–1028.
- 808 Luke, C. L. 2006. "Fostering learner autonomy in a technology - enhanced, inquiry - based  
809 foreign language classroom." *Foreign Language Annals* 39 (1): 71–86.
- 810 Malmström, H., D. Pecorari, and P. Shaw. 2018. "Words for what? Contrasting university  
811 students' receptive and productive academic vocabulary needs." *English for Specific  
812 Purposes* 50: 28–39.
- 813 Mickan, P. 2012. *Language curriculum design and socialisation*. Bristol: Multilingual Matters.
- 814 Ministry of Science and Higher Education. 2018. Accessed on September 27, 2021:  
815 <https://www.5top100.ru/news/93717/>
- 816 Miller, D., and D. Biber. 2015. "Evaluating reliability in quantitative vocabulary studies: The  
817 influence of corpus design and composition." *International Journal of Corpus  
818 Linguistics* 20 (1): 30–53.

- 819 Milton, J. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual  
820 Matters.
- 821 Morin, R., and J. Goebel Jr. 2001. "Basic vocabulary instruction: Teaching strategies or  
822 teaching words?." *Foreign Language Annals* 34 (1): 8–17.
- 823 Muñoz, V. L. 2015. "The vocabulary of agriculture semi-popularization articles in English: A  
824 corpus-based study." *English for Specific Purposes* 39: 26–44.
- 825 Nation, I. 2006. "How large a vocabulary is needed for reading and listening?." *Canadian  
826 modern language review* 63 (1): 59–82.
- 827 Nation, I. S. 2016. *Making and using word lists for language learning and testing*. Amsterdam,  
828 Netherlands: John Benjamins Publishing Company.
- 829 Nation, I. S. P. 2001. *Learning vocabulary in another language*. Cambridge, UK: Cambridge  
830 University Press.
- 831 Nation, I. S. P., and J. Macalister. 2010. *Language curriculum design*. New York, NY:  
832 Routledge.
- 833 Nation, P., and H. Kyongho. 1995. "Where would general service vocabulary stop and special  
834 purposes vocabulary begin?." *System* 23 (1): 35–41.
- 835 Parodi, G. 2015. "Variation across university genres in seven disciplines: A corpus-based study  
836 on academic written Spanish." *International Journal of Corpus Linguistics* 20(4): 469–499.

- 837 Popova, I.M, E.B. Patrakeeva, and M.M. Glazkova. 2014. *Russkij jazyk dlja ekonomistov*.  
838 Tambov: FGBOU VPO "TGTU".
- 839 Rayson, P. and R. Garside. (2000). "Comparing corpora using frequency profiling." In  
840 Proceedings of the Workshop on Comparing Corpora, ACL 2000. 1-8 October 2000, Hong  
841 Kong, pp. 1 - 6.
- 842 Resche, C. 2000. "Hedging across genres: An approach for non-native students of English for  
843 Economics." *ASp. la revue du GERAS* (27-30): 289-308.
- 844 Rivera - Mills, S. V., and L. Plonsky. 2007. "Empowering students with language learning  
845 strategies: A critical review of current issues." *Foreign Language Annals* 40 (3): 535-548.
- 846 Rogers, J., S. Webb, and T. Nakata. 2015. "Do the cognacy characteristics of loanwords make  
847 them more easily learned than noncognates?" *Language Teaching Research* 19 (1): 9-27.
- 848 Schmitt, N. 2010. *Researching vocabulary: A vocabulary research manual*. Springer.
- 849 Sharoff, S., E. Umanskaya, and J. Wilson. 2013. *A frequency dictionary of Russian: Core  
850 vocabulary for learners*. New York, NY: Routledge.
- 851 Simanungkalit, V. G. 2016. "Metadiscourse Of Economics Texts with Different  
852 Genres" (Doctoral dissertation, UNIMED).
- 853 Sun, Y., and L. Lei. 2013. "Academic Lexical Items in Economics Journal Papers." *Chinese  
854 Journal of Applied Linguistics* 36 (3): 354-367.

- 855 Sutarsyah, C., P. Nation, and G. Kennedy. 1994. "How useful is EAP vocabulary for ESP? A  
856 corpus based case study." *RELC Journal* 25 (2): 34–50.
- 857 Todd, R. W. 2017. "An opaque engineering word list: Which words should a teacher focus  
858 on?." *English for Specific Purposes* 45: 31–39.
- 859 Tongpoon-Patanasorn, A. 2018. "Developing a frequent technical words list for finance: A  
860 hybrid approach." *English for Specific Purposes* 51: 45–54.
- 861 Toth, P. D., and K. Moranski. 2018. "Why haven't we solved instructed SLA? A sociocognitive  
862 account." *Foreign Language Annals* 51 (1): 73–89.
- 863 Valipouri, L., and H. Nassaji. 2013. "A corpus-based study of academic vocabulary in chemistry  
864 research articles." *Journal of English for Academic Purposes* 12 (4): 248–263.
- 865 Vongpumivitch, V., J. Huang, and Y.-C. Chang. 2009. "Frequency analysis of the words in the  
866 Academic Word List (AWL) and non-AWL content words in applied linguistics research  
867 papers." *English for Specific Purposes* 28 (1): 33–41.
- 868 Wang, J., S.-I. Liang, and G. Ge. 2008. "Establishment of a medical academic word list." *English  
869 for Specific Purposes* 27 (4): 442–458.
- 870 Yakobson, H. B., and N. Apanasewicz. 1959. "Selected Bibliography of Teaching Materials for  
871 Russian." *The bulletin of the National Association of Secondary School Principals* 43 (245):  
872 211–215.

873 Yamamoto, Y. 2014. "Multidimensional vocabulary acquisition through deliberate vocabulary  
874 list learning." *System* 42: 232–243.

875 Yang, M.-N. 2015. "A nursing academic word list." *English for specific purposes* 37: 27–38.

876

877

878

879