

A COMPARATIVE EVALUATION OF IDIOMS IN RUSSIAN-AUTHORED ENGLISH LANGUAGE TEXTBOOKS AND IDIOMS EXTRACTED FROM CURRENT AUTHENTIC SOURCES

A.G. Shilnov, J. Miller, P.J. Mitchell, V.M. Smokotin

Abstract. Many English language teaching (ELT) textbooks contain material that may not reflect native speaker usage. This is particularly true of less frequently used vocabulary such as idioms. The primary objective of this paper is to evaluate the selection of idioms in ELT textbooks published in Russia and used in Russian higher education institutions. Additionally, the paper examines a number of related issues surrounding ELT material development practices and language authenticity as well as idiom definition and usage. Regarding the question of idiom definition, the paper looks at the most commonly used terminology referring to idioms and related items (phrasemes, lexical bundles, etc.). A special focus is placed on such criteria of idiom definition as figurativeness, semantic transparency and opacity, and compositional invariance. Based on these criteria, the paper also explains why some categories of frequently recurrent and/or figurative language chunks remained outside the scope of the study. A major focal point of the study is the category of proverbs and sayings, which in the paper are placed under the umbrella of idioms. Due to their particularly low frequency of occurrence in authentic English language corpora revealed during the course of the study, the paper questions the viability of extensive coverage of proverbs and sayings in ELT materials. Driven mostly by pedagogical considerations, the paper provides a general overview of the idiom component in ELT textbooks. Examples of specific ELT contexts currently faced with the issue of textbook inauthenticity are given. In addition, the paper presents a brief historical account of ELT practices in Russia and links them to the issue of linguistic inauthenticity of some Russian-authored ELT materials. Highlighting the need to address concerns relating to textbook inauthenticity, the paper presents major advantages of relying on authentic sources in selecting idiomatic items to be taught to language learners. Specifically, the findings of the study empirically demonstrate that a bank of idioms compiled based on a random selection of contemporary authentic materials is likely to include more or less commonly used items that are present and explained in popular learner's dictionaries, feature adequate frequency in comprehensive authentic English language corpora and may therefore be safely included in syllabi. In the course of the study, idioms from three Russian-authored ELT textbooks (Corpus 1) were compared to idioms derived from contemporary authentic spoken texts featuring spontaneous, semi-spontaneous and prepared speech by native British and North American speakers (Corpus 2). The frequency of the idioms in the two corpora was then compared to the frequency of the same idioms in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). Idioms from Corpus 1 occurred with a considerably lower frequency than those in Corpus 2 in both the BNC and COCA. The results of the study indicate that many idioms in the Russian textbooks do not reflect authentic native speaker

use. Since this problem may be widespread wherever ELT materials are developed without native speaker input, textbook writers and teachers should therefore consider carefully what idioms they include in their material, and, when possible, make greater use of authentic data.

Keywords: idioms; English language teaching (ELT); language corpus; textbook; authentic materials; authenticity.

Introduction

The importance of idiom awareness and competence for language learners in regard to both their receptive and productive skills cannot be overemphasised. Idioms occur in a variety of different situations of authentic language discourse, including formal, informal, oral or written interactions, and almost any exposure to authentic language involves exposure to idioms. The figurativeness of idioms is often utilised to create various stylistic effects that contribute to the diversity of language, such as imagery or humour (e.g., based on word play) [1. P. 372]. In regard to encoding skills, the production of idiomatic language is considered crucial for advanced language learners, and figurative idiom competence is currently gaining significance in contexts where English is used institutionally as a second language ('the Outer Circle' countries, e.g. India, Malaysia [2]) or as a foreign language for international communication ('the Expanding Circle' countries, e.g., Japan, Russia [2; 3. P. 79]. Phraseological competence is a problem, however, even for advanced learners, as studies of learner corpora show [4].

Developing the idiomatic awareness of language learners is often a challenge, exacerbated by the fact that, like other lexical items, idioms can be restricted in their usage to a certain situation, register, type of discourse [5], geographical location, age group [1] or speech community. Mistranslating seemingly identical idiomatic counterparts (known as 'false friends') into or from their own language can also be a trap for learners [6. P. 107]. As a result, it is not uncommon to see learners struggling to recognise idioms in authentic speech, failing to use them properly in their own speech, or avoiding their use altogether. There is therefore a clear need for ELT courses and textbooks to have a specific focus on idioms and to equip learners with a sufficient number of idiomatic expressions for effective communication.

Defining idioms

Innumerable attempts to capture the nature of idioms and other related lexical units have resulted in multiple theories, terms and definitions. Nevertheless, the nature of idioms remains elusive, and the definition of this linguistic phenomenon can differ substantially from scholar to scholar, with the ensuing confusion tending to create hindrances for both researchers and ELT practitioners [7. P. 38]. It is not the purpose of this article to go into exhaus-

tive detail pertaining to idiom definition and classification. However, we shall attempt to cover the fundamental aspects of the issue in the light of idiom pedagogy and instruction.

In the last two decades, corpus-driven trends in linguistics and ELT have affected the way phraseology [1, 6, 8] and formulaic language [9, 10] are viewed by many academics. Many scholars, focusing on the notions of frequency, distribution and word co-occurrence, have adopted specific terminology such as *recurrent word combinations* [11], *clusters* [12], *n-grams* [13], *multiword units* (MWUs) [7. P. 38; 14 P. 549] and *phrasemes* (Dobrovolskij & Piirainen [6], contribution to the corpus-informed study of recurring language was made by Biber, Conrad and Cortes [15], who refer to 2005). In this regard, significant contribution to the corpus-informed study of recurring language was made by Biber, Conrad and Cortes [Ibid.], who refer to one of the most widely used notions in this type of research, lexical bundles (see also [16, 17]), defined as “the most frequently recurring sequences in a register” [15. P. 376].

The terminology mentioned above commonly encompasses both non-figurative (having a literal / non-idiomatic meaning) and figurative (having an idiomatic meaning) items. As in our study the pedagogical component plays a significant part, we focus on those lexical items that present the greatest challenge to language learners in terms of semantics, structure and usage. Therefore, we rely on the term ‘idiom’ as one of the most commonly used terms, both in ELT and linguistics research [7, 18] and in ELT materials [19]. Based on the two key criteria which single idioms out from other units – opacity and invariance – we adopt Liu’s concept of idiom definition for L2 learners [18. P. 15–16]: ‘For L2 learners, idioms are perhaps best defined as multiword expressions that are invariant or variance-restricted in structure and often (not always) non- or semi-literal in meaning. In other words, all idioms are invariant or variance-restricted, but not all of them are non- or semi-literal, for there are quite a few fairly literal idioms that are fixed in form’. However, as will be discussed later in the article, only a limited number of idioms can be considered to possess full compositional invariability (examples from our corpora would include *on the other hand* or *from scratch*), with most idioms being subject to user variance in a specific context (e.g., *a blast from the musical past*).

Though opacity / figurativeness is deemed a primary criterion in idiom definition, it can vary significantly between idioms, and one and the same item may be more transparent to some speakers than to others, based on their knowledge and imagination as well as the context in which it occurs. Figurativeness makes idioms more challenging to learn than collocations, which, together with phrasal verbs, are outside the scope of our study. In this respect, it is worth mentioning that dictionaries (e.g., the *Oxford advanced learner’s dictionary* (henceforth OALD) [19] often include separate sections

for collocations and phrasal verbs. Some publishers also produce separate dictionaries of these language units, which indicates the specificity of these two categories.

We also decided not to include in our analysis those idiomatic and metaphorical phrases whose composition can be varied to such a degree that it does not fit the criterion of invariance. For example, in our analysis of authentic audio and video materials we came across the phrase *shift gears*, which in some dictionaries and studies is explicitly treated as an idiom [20. P. 435]. However, the significant number of variations expressing related meanings (*switch gears*, *change gears*, *go into another gear*, *step up a gear*, *take something up a gear*, etc.) leads us to believe that *gear* has developed a specific sense of 'pace, effort or mode of operation' which can therefore be presented as a separate meaning of the lexeme, as is the case in OALD. Similarly, idiomatic compound nouns (*bottom line*, *rock bottom*, etc.) presented in OALD as noun entries rather than idioms were not part of our analysis, though in some publications they are included in idiom lists [20. P. 526; 21. P. 435].

Finally, since a significant proportion of the items under investigation is represented by what is commonly referred to as proverbs – “short, generally known sentences of the folk that contain wisdom, truths, morals, and traditional views in a metaphorical, fixed, and memorizable form and that are handed down from generation to generation” [22. P. 597] – and sayings (such as *tastes differ*), we also include them under the broader heading of idioms, taking into account their opaque and relatively invariant nature as well as the fact that in some dictionaries (e.g. OALD) proverbs and sayings are labelled as idioms. In the findings below, these two types are conflated under the heading sayings / proverbs.

The Idiom Component in ELT Textbooks

Corpus research [5, 20] indicates that idioms are not frequently used in everyday speech, and proverbs even less so [23]. However, corpus frequency does not necessarily equate to usefulness, especially in terms of comprehension [5, 24]. In addition, the notion of 'frequency' in regard to idioms is a fairly arbitrary one [5]. Relative frequency, though, is important. If an idiom is rarely or never used in authentic native speaker English, then it is unlikely to be useful for language learners.

The concept of authenticity in ELT textbooks relates not just to the words themselves, but, among other things, to the frequency with which certain types of language (especially less commonly used expressions such as idioms and proverbs) appear in native speaker discourse [25]. Hence, as McCarthy and Carter [26. P. 338] warn, “importing real data into the classroom in the belief that authenticity is thereby guaranteed is no simple matter”. Vocabulary taught in a textbook should therefore take account not just

of form and meaning, but also use. Unfortunately, many ELT textbooks produced without native speaker input fail to address the issue of authenticity in any form.

Many ELT textbooks are idiosyncratic in their choice of English vocabulary. A corpus-based study of Swedish ELT school textbooks, for example, found that many words taught in the books were not used in English L1 communication, as demonstrated by a comparison of these Swedish textbooks and the *New General Service List* [27]. In research by Gouverneur [28] which relied on the TeMa (Textbook Material) corpus, three ELT best-sellers (*Cutting Edge*, *Inside Out* and *New Headway*) were strongly criticized for their presentation of collocational patterns of two high frequency verbs. In a study by Alavi & Rajaabpoor [29], idioms selected for materials used in a leading English teaching school in Iran did not reflect the ordinary frequency of these idioms in the BNC, Brown corpus or MICASE. There are, however, relatively few studies on the representation of idioms in ELT textbooks. It is therefore profitable to study materials produced in countries with large populations and in which teaching throughout the country often relies on the use of the same textbook. One such country is Russia.

Back in Soviet times (between the late 1940s and the late 1980s), learners and teachers alike were kept largely isolated from contemporary authentic English content by the Iron Curtain, with curricula dictated by the Ministry of Education [30. P. 317]. Textbook content was predominantly based on ideologically acceptable works by British writers such as Charles Dickens, and many idioms were derived directly from literary texts and therefore did not necessarily have a high frequency of usage in contemporary authentic speech. The impact of Soviet textbook creation methods and ELT practices in general can still be perceived today, as new and only slightly altered editions of textbooks originally published decades ago (e.g., *Practical English Course* by Arakin et al. [31–35] first published in 1972) continue to be used, and even relatively recently created materials often reveal traces of the Soviet patterns (e.g., a textbook by Merkulova, Filimonova, Kostygina, Ivanova and Papanova [36] is close to the textbook by Arakin et al. in terms of thematic content and activities).

Though corpus-based idiom research is a relatively recent area of study which started to emerge in the 1990s [37], it is now increasingly applied to textbook evaluation [29], with language corpora actively utilised to create a wide range of ELT materials [38. P. 82–83] and to inform ELT practices and research in general [39. P. 120]. Therefore, the research question addressed in this study can be considered timely and relevant: do the idioms and sayings taught in Russian ELT textbooks reflect the use of contemporary native speakers of English?

Although we focus here on Russia, the findings may indicate a problem common in ELT textbooks for other language groups, since there is al-

ready evidence [27, 29] that this problem exists elsewhere. The study is therefore important in terms of alerting textbook writers and teachers to the importance of using authentic, contemporary English for their students.

Method

In order to compare textbook content and authentic content, we created two corpora of idioms. Corpus 1 consists of 213 idioms (including proverbs and sayings) extracted from the following Russian-authored textbooks:

1. *Practical English Course (Years 1–5)* [31–35].
2. *A Course in English for Language Majors* [40].
3. *English for University Students. Reading, Writing and Conversation* [35].

These textbooks were designed for Russian university students and are among the most popular of the Russian-authored textbooks currently used in many ELT university courses, mostly in the course titled *Oral and written English practice*.

In Corpus 1, we only included those idioms that were presented in practice exercises and thus were singled out by the textbook creators as relevant for learners. In Merkulova's textbook, some items are presented as idioms but do not fit our definition and are not listed as idioms in OALD (e.g., *a stumbling block* is presented as a compound noun in OALD), and they were therefore not included in the corpus. The total number of idioms found in the textbooks was 213.

In order to provide a current authentic spoken corpus relevant to the needs of university age English language learners, we compiled Corpus 2 from idioms that we extracted from a range of freely available authentic sources of spoken British and North American English (BrE and NAmE) that students are likely to listen to. According to usage labels adopted in OALD, BrE refers to usage in the UK, whereas NAmE denotes usage in the US and Canada (e.g., *two / ten a penny (BrE)(NAmE a dime a dozen)*), unless a particular country or part of the UK is specified by other labels [41]. Primary criteria for the selection of authentic sources were as follows:

1. **Types of sources.** We relied on radio and television programmes (chiefly those available as podcasts) by popular British and American broadcasters: the BBC (British Broadcasting Corporation), NPR (National Public Radio) and CNN (Cable News Network). The second source was represented by YouTube videos by popular channels (with at least 10,000 subscribers) having no less than 10,000 views. Details are given in Appendix 1.

2. **Language content.** Idioms were extracted only from the speech of native English speakers from the British Isles and North America, as far as could be identified from the speaker's accent in a given audio or video. Materials were mostly represented by natural spontaneous or semi-spontaneous speech (interviews, panel discussions, conversations between presenters and

on-the-spot reporters) covering a variety of topics, from health and well-being to politics and economics, and to a lesser degree prepared speech, with the prevalence of semi-formal or informal register.

3. Time of production. For our analysis, we selected materials produced between January 2016 and August 2018.

The total duration of native English speech in the materials under investigation was approximately 21 hours. We stopped recording authentic materials when we had reached the target number of 213 discrete idioms (i.e. the same number of idioms as in the textbooks). The frequency of the idioms in our two corpora was compared to the frequency of the same idioms in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The BNC is described as “a 100-million-word collection of samples of a written and spoken language of British English from the later part of the 20th century”, consisting of ‘the bigger written part (90%, e.g. newspapers, academic books, letters, essays, etc.) and the smaller spoken part (remaining 10%, e.g. informal conversations, radio shows, etc.)’ [42]. COCA is referred to as “the only large and balanced corpus of American English” which is “equally divided among spoken, fiction, popular magazines, newspapers, and academic texts” [43]. On 1 November 2018 it contained more than 560,000,000 words. COCA is widely used in research into different language chunks such as, for example, situation-bound utterances [44] as well as in informing EFL material creation practices [45].

The measurement was performed with relation to both spoken and written subcorpora of the BNC and COCA, enabled by the Sketch Engine corpus manager [42] and BYU corpora access [43]. Idioms from Corpora 1 and 2 were compared and contrasted based on their corpus frequency. While running idioms through the BNC and COCA, we relied on general search guidelines and endeavoured to make our results as accurate and comprehensive as possible. We considered the fact that in authentic speech the composition of an idiom can vary in terms of word forms, punctuation and syntax, and that idioms are particularly challenging to deal with in corpus research due to the sometimes variable nature of their canonical forms [46. P. 13]. Following a related corpus-based idiom study by Simpson and Mendis [20], we sought to capture not only an idiom’s canonical form presented in dictionaries but also possible variations resulting from insertion (*a blast from the musical past*), truncation (*end of story* → *end of*), substitution of components by synonymous or related items including regional variations (*shut the door on something* → *slam the door on something*, *sweep something under the carpet* (British English) → *sweep something under the rug* (American English)) or syntactical variations (*to reinvent the wheel* → *the reinvention of the wheel*). Unlike Simpson and Mendis, however, who recorded what they call “performance [i.e. accidental] variations” – substitutions or rearrangements of idiom components resulting in semantic changes, e.g. *a thorn*

in your side → *side in your thorns* [20. P. 436], we only dealt with naturally occurring variants related to the meaning of the original idiom.

We should also mention that for some of the analysed items it was not feasible to arrive at the exact number of times they occur as actual idioms, since those items could be used both figuratively and literally, e.g., *under somebody's nose*. At this point, neither Sketch Engine nor BYU Corpora tools can enable the researcher to filter results based on literal or figurative usage. Therefore, for those items we used a sampling technique by setting a cut-off point of the first 20 tokens, among which we counted the proportion of idioms and then extrapolated the total estimated frequency based on that proportion.

Findings

Overall, we analysed the corpus occurrence frequency of the 213 idioms derived from the Russian-authored textbooks and the 213 idioms extracted from radio podcasts and YouTube videos featuring authentic English speech produced between September 2016 and August 2018. Nine idioms appeared in both corpora, so the total number of idioms in the study was 417. As of 1 November 2018, the BNC corpus included 96,134,547 words of written and spoken English. For the COCA corpus, the figure was approximately 560,000,000. The statistical analysis was carried out with the help of SPSS software.

In Tables 1 and 2 and Figures 1 to 4 we present the frequency of all our 417 idioms in the BNC and COCA corpora showing mean, minimum and maximum values as well as the number of times the total sum of all idioms combined appears in each of the corpora.

Table 1

**Frequency statistics of idioms extracted from Russian ELT textbooks
(Corpus 1, 213 idioms)**

	BNC Frequency	COCA Frequency
Mean frequency value	30.26	151.31
Minimum frequency value	0	0
Maximum frequency value	448	2,827
Sum total of recurrence	6,445	32,230

Table 2

Frequency statistics of idioms extracted from authentic sources (Corpus 2, 213 idioms)

	BNC Frequency
Mean frequency value	148.42
Minimum frequency value	0
Maximum frequency value	5,311
Sum total of recurrence	31,613

As can be seen, the cumulative occurrence of the textbook extracted idioms in either of the corpora is much lower than that of the authentic

sources idioms (6,445 vs. 31,613 for the BNC frequency and 32,230 vs. 181,305 for the COCA frequency). The same applies to the mean values (30 vs. 148 and 151 vs. 851) and the maximum values (448 vs. 5,311 and 2,827 vs. 23,007).

As far as the minimum values are concerned, the textbook idiom corpus, Corpus 1, includes a significant number of items with zero frequency in both of the corpora (the vast majority of those items belong to the subcategory of sayings / proverbs). For instance, 42 of the textbook idioms do not appear in the BNC at all, while 41 items are not found in COCA, as shown in Figures 1 and 2.

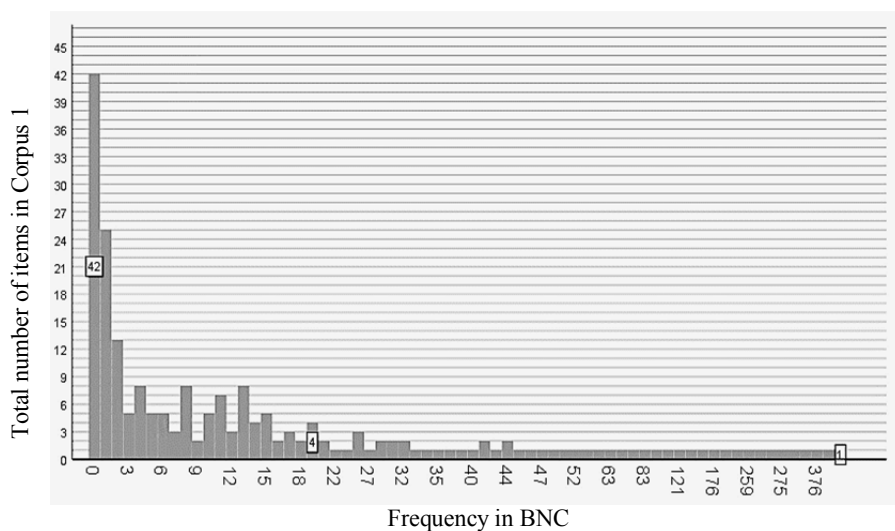


Fig. 1. Frequency values in the BNC of idioms extracted from Russian ELT textbooks (Corpus 1)

By contrast, only six of the idioms from authentic sources fail to appear in the BNC, while the COCA corpus search results reveal that the most infrequently used items derived from the authentic sources (*You're telling me!* and *a shot across / over the bows*) recur at least three times in this corpus, as illustrated in Figures 3 and 4.

Generally, as illustrated by Figures 1 and 2, the vast majority of the textbook extracted items feature a rather low frequency of occurrence in both corpora (below 22, i.e. less than 0.22 per million words in the BNC and less than 0,041 per million words in COCA). It should also be noted that only nine of the items from both corpora overlap, viz.: *from scratch*, *face to face*, *out of hand*, *the other way around / round*, *go hand in hand*, *get / come to grips with somebody / something*, *out of sight out of mind*, *can't make head or / nor tail of something*, *a castle in the air*. In the textbook corpus, most of

the overlapping idioms were derived from the more recent textbook by Y.B. Yastrebova et al. [40], in which they are grouped thematically (‘head’ idioms, ‘face’ idioms, etc.).

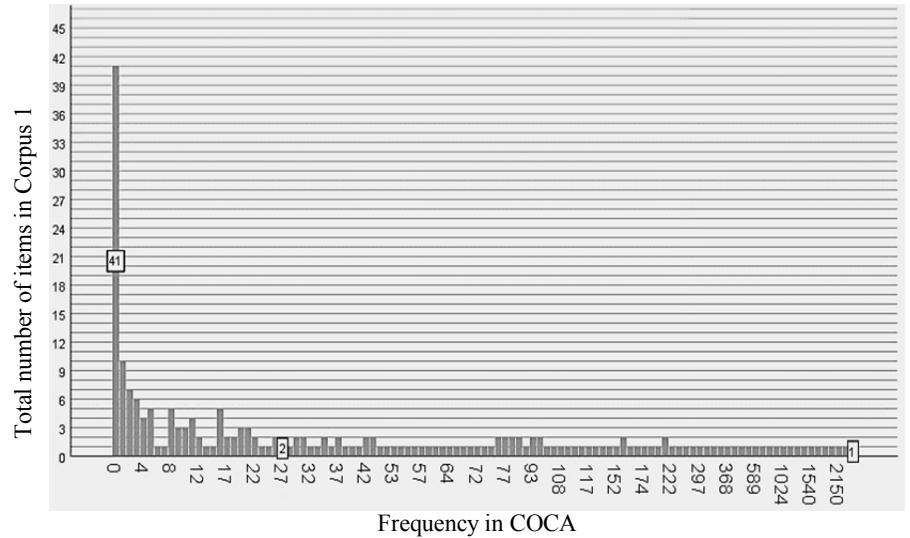


Fig. 2. Frequency values in COCA of idioms extracted from Russian ELT textbooks (Corpus 1)

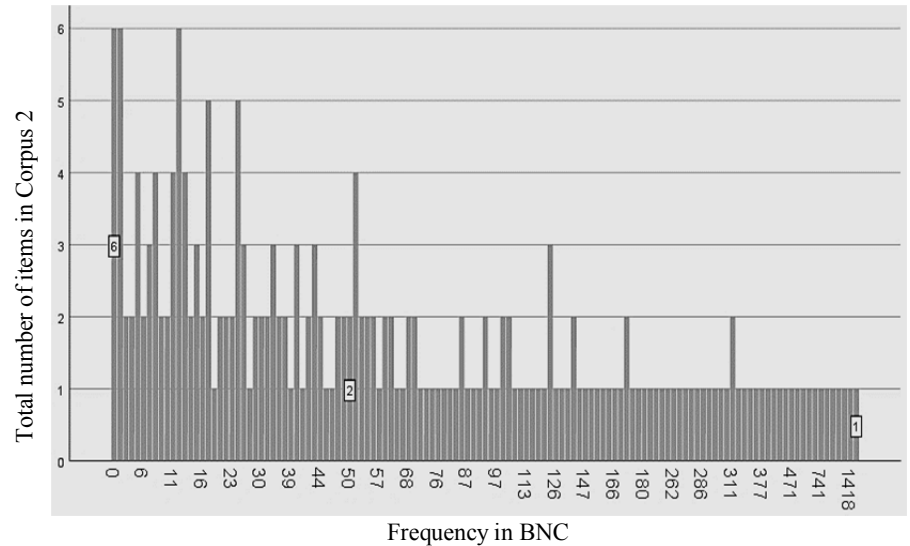


Fig. 3. Frequency values in the BNC of idioms extracted from authentic sources (Corpus 2)

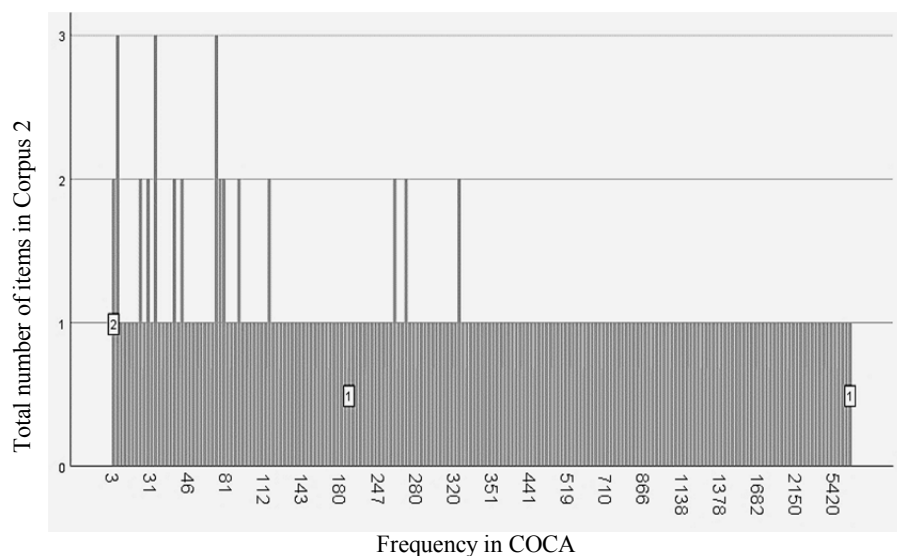


Fig. 4. Frequency values in COCA of idioms extracted from authentic sources (Corpus 2)

As mentioned earlier, we paid particular attention to sayings / proverbs, assuming that this type of expression can differ substantially in terms of its corpus occurrence, which can have important implications for language learners. Figure 5 demonstrate the proportion of sayings / proverbs and other types of idioms (labelled ‘other idioms’) in the two corpora.

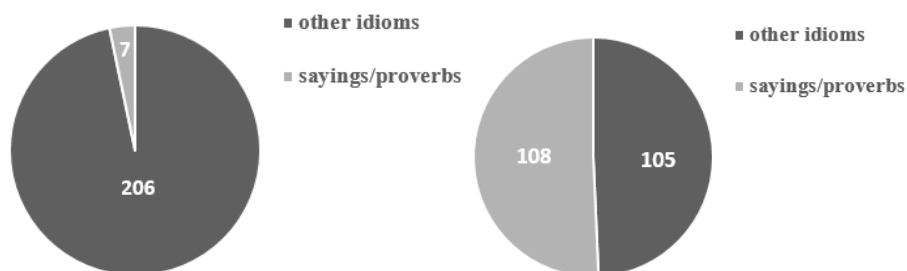


Fig. 5. Proportion of items extracted from textbooks (left) and from authentic sources (right)

As can be seen, out of the 213 idioms from current authentic English videos and audios, only seven were classified as sayings / proverbs. The analysed textbooks, on the other hand, contain a much greater number of sayings / proverbs, which in fact even outnumber other types of idioms included in the textbooks (108 vs. 105 respectively). Table 3 compares the occurrence of sayings / proverbs with that of other idioms in the two corpora. As can be seen, regardless of the corpus, sayings/proverbs feature substantially lower mean frequency values than other idioms. Additionally, despite outnumber-

ing other types of idioms in the textbook corpus, sayings / proverbs still have significantly smaller sum frequency values (488 vs. 5 958 in the BNC and 1 686 vs. 30 544 in COCA).

Table 3

Frequency statistics of sayings / proverbs versus other idioms in the BNC and COCA

Idioms Extracted from Russian ELT Textbooks			
BNC Frequency of Authentic Sources Idioms			COCA Frequency of Authentic Sources Idioms
Other idioms	Mean	56.73	290.90
	N	105	105
	Sum	5,957	30,544
	Minimum	0	0
	Maximum	448	2,827
Sayings / proverbs	Mean	4.52	15.61
	N	108	108
	Sum	488	1686
	Minimum	0	0
	Maximum	53	113
Total	Mean	30.26	151.31
	N	213	213
	Sum	6445	32,230
	Minimum	0	0
	Maximum	448	2,827

Idioms Extracted from Authentic Sources			
BNC Frequency of Authentic Sources Idioms			COCA Frequency of Authentic Sources Idioms
Other idioms	Mean	152.19	867.13
	N	206	206
	Sum	31,351	178,628
	Minimum	0	3
	Maximum	5311	23,007
Sayings / proverbs	Mean	37.43	382.43
	N	7	7
	Sum	262	2,677
	Minimum	0	4
	Maximum	164	1,442
Total	Mean	148.42	851.20
	N	213	213
	Sum	31,613	181,305
	Minimum	0	3
	Maximum	5,311	23,007

Overall, the findings of this study indicate that all but nine of the 213 idioms in the Russian textbook corpus appear much less frequently in the corpus of authentic sources. Moreover, the high frequency of say-

ings/proverbs in the Russian textbook corpus is disproportionate to the number of sayings/proverbs used in current-day authentic discourse.

Discussion and Recommendations

The findings have led us to consider ELT practices in general and material creation in particular. First of all, based on a frequency comparison, it could be argued that the selection of idioms in the Russian-authored textbooks under investigation can be called into question. The analysis of idioms from these textbooks has revealed that a considerable number of them have very low corpus frequencies regardless of corpus creation time and English language variety. This is in line with previous studies on English idioms in a variety of corpora [5, 20]. The concern, however, is not that idioms are used with relatively lower frequency than some other vocabulary items, but that most of the idioms in the Russian textbook corpus are used even less frequently than those in the corpus of authentic materials. This means that students are being taught idioms that will not help them comprehend current materials or communicate effectively with their English-speaking peers.

Secondly, sayings / proverbs as a specific type of idiom tend to have a low frequency of occurrence in contemporary authentic English discourse as compared to other idiomatic items [23. P. 325]. Therefore, careful consideration should be exercised when deciding on the proportion of proverbs and sayings to be included in ELT materials. From the viewpoint of frequency in everyday language, the number of sayings / proverbs contained in the Russian-authored textbooks under investigation is disproportionately high. This may reflect Hinkel's [47] comment on students' general fondness for learning such colorful expressions, but it could also be detrimental to communication with native speakers, since proverbs seem to be used less frequently by English L1 speakers than by Russian L1 speakers. For example, in the Russian Web Corpus of 147,930,261 words [42], the frequency of the saying *О вкусах не спорят* is 53 (0.28 per million), whereas its English counterpart *Tastes differ*, found in our textbook corpus, occurs only 3 times in the BNC (0.03 per million) and 8 times in COCA (0.01 per million). Research also indicates that university-aged native speakers of English use far fewer idioms and proverbs than older speakers [48].

Finally, the study's findings contribute to the broader debate concerning approaches towards ELT materials, specifically the potential dichotomy between textbooks and authentic materials [25]. The idioms in the Russian textbooks in this study are indeed far from reflecting the English that learners would find useful in communicating with native English speakers. Authenticity needs to apply not just to the existence of an expression in English, but to its use in different situations. Therefore, it is recommended that ELT materials creators carefully consider the sources that they draw upon while

selecting vocabulary content for textbook activities. Teachers also need to consider the needs of their students and involve them more in the search for useful materials. Creating a corpus of authentic texts such as the one in this study, for example, is an activity that university students could be encouraged to undertake as part of their English language studies. This would enable them to examine the language of specific groups so that they can build their own bank of useful expressions. If lexical inauthenticity of materials is a major concern for ELT in Russia, this may well reflect the situation in many other countries. New materials creation practices worldwide could bring ELT closer to contemporary authentic English.

Conclusion

The results of the research indicate that the Russian-authored textbooks under investigation include low-frequency idioms, frequently represented by sayings and proverbs, many of which are not to be found in modern English corpora. Conversely, the extraction and analysis of the same number of idioms from modern authentic discourse reveal their high rate of recurrence in both British and American corpora that include multiple varieties of texts. Therefore, the study confirms the findings of similar studies for other countries [28, 29], further reinforcing the notion that textbook content can be subject to lexical inauthenticity in terms of vocabulary selection generally and idiom selection in particular – an issue that is highly relevant in modern ELT practices. Further studies on the teaching of English idioms in other countries would confirm if this is a worldwide problem. In the meantime, however, the current growth of balanced million-word corpora and the proliferation and accessibility of authentic language content online constitute precious resources that language teaching practitioners and textbook writers should not ignore.

References

1. Miller J. (2013) Phraseology across continents. In J. Szerszunowicz (Ed.), *Research on phraseology across continents* (Vol. 2) (pp. 370–386). Bialystok, Poland: University of Bialystok Publishing House.
2. Kachru B.B. (1992) Teaching world Englishes. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (2nd ed., pp. 355–365). Urbana, IL: University of Illinois Press.
3. Tran H.Q. (2017) Figurative idiomatic competence: An analysis of EFL learners in Vietnam. In K. Kimura & J. Middlecamp (Eds.), *Asian-focused ELT research and practice: Voices from the Far Edge* (pp. 66–86). Phnom Penh, Cambodia: IDP Education (Cambodia) Ltd.
4. Meunier F., Gouverneur C. (2007) The treatment of phraseology in ELT textbooks. Paper presented at the *Corpora in the foreign language classroom: Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6)*. University of Granada, Spain.

5. Liu D. (2003) The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, 37(4), 671–700.
6. Dobrovolskij D., Piirainen E. (2005) *Figurative language: Cross-cultural and cross-linguistic perspectives*. Oxford, England: Elsevier.
7. Grant L., Bauer L. (2004) Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25, 38–61. <https://dx.doi.org/10.1093/applin/25.1.38>
8. Granger S., Meunier F. (2008) Phraseology in language learning and teaching: Where to from here? In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247–252). Amsterdam/Philadelphia: John Benjamins Publishing Company.
9. Wray A. (1999) Formulaic language in learners and native speakers. *Language Teaching*, 32(4), pp. 213–231.
10. Demetriou T. (2016) *Predicting IELTS ratings using vocabulary measures* (Doctoral dissertation, University of the West of England, Bristol, UK). Retrieved from <http://eprints.uwe.ac.uk/28665>
11. Altenberg B. (1998) On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 101–122). Oxford: Oxford University Press.
12. Hyland K. (2008) Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), pp. 41–62.
13. Stubbs M. (2007) An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 89–105). Amsterdam, Netherlands: Radopi.
14. Siyanova-Chanturia A., Martinez R. (2014) The idiom principle revisited. *Applied Linguistics*, 36(5), pp. 549–569. <https://dx.doi.org/10.1093/applin/amt054>
15. Biber D., Conrad S., Cortes V. (2004) If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), pp. 371–405.
16. Liu D. (2012) The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), pp. 25–35.
17. Chen Y.H., Baker P. (2010) Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), pp. 30–49.
18. Liu D. (2008) *Idioms: Description, comprehension, acquisition, and pedagogy*. New York, NY: Routledge.
19. Oxford University Press. (2019a) *Oxford advanced learner's dictionary*. Retrieved from <https://www.oxfordlearnersdictionaries.com/>
20. Simpson R., Mendis D. (2003) A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37(3), pp. 419–441.
21. Nunberg G., Sag I. A., Wasow T. (1994) Idioms. *Language*, 70(3), pp. 491–538.
22. Mieder W. (1996) Proverbs. In J. Harold (Ed.), *American folklore: An encyclopedia*, (pp. 597–601), New York, NY: Garland Publishing.
23. Arnaud P., Moon R.E. (1993) Frequency and use of English and French proverbs. *Familiar phrases: Tropes, stereotyped expressions, and clichés*, pp. 323–341. Paris, France: Kimé.
24. Martinez R. (2013) A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67(2), pp. 184–198. <https://dx.doi.org/10.1093/elt/ccs100>
25. Gilmore A. (2007) Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40, pp. 97–118.
26. McCarthy M., Carter R. (2001) Size isn't everything: Spoken English, corpus, and the classroom. *TESOL Quarterly*, 35(2), pp. 337–340. <https://dx.doi.org/10.2307/3587654>
27. Norberg C., Nordlund M. (2018) A Corpus-based study of lexis in L2 English textbooks. *Journal of language teaching and research*, 9(3), pp. 463–473. doi: <http://dx.doi.org/10.17507/jltr.0903.03>

28. Gouverneur C. (2008) The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis. In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 223–243). Amsterdam/Philadelphia: John Benjamins Publishing Company
29. Alavi S., Rajabpoor A. (2015) Analyzing idioms and their frequency in three advanced ILI textbooks: A corpus-based study. *English Language Teaching*, 8(1), pp. 170–179.
30. Ivanova V., Tivyaeva I. (2015) Teaching foreign languages in Soviet and present-day Russia: A comparison of two systems. *Zbornik Instituta za pedagogika istrazivanja*, 47(2), pp. 305–324. <https://dx.doi.org/10.2298/ZIP115023051>
31. Arakin V.D., Antrushina G.B., Kirillova E.P., Levina E.L., Petrushin S.I., Samokhina T.S. (2003) *Prakticheskij kurs anglijskogo jazyka. 5 kurs. 4-e izdaniye* [Practical English course. Year 5. 4th ed.]. M., Russia: Vlos.
32. Arakin V.D., Novikova I.A., Afanasyeva O.V., Bronnikova S.N., Davidenko T.G., Saakyan, A.S. ... Shabadash G.A. (2006) *Prakticheskij kurs anglijskogo jazyka. 3 kurs. 4-e izdaniye* [Practical English course. Year 3. 4th ed.]. M., Russia: Vlos.
33. Arakin V.D., Novikova I.A., Aksenova-Pashkovskaya G.V., Bronnikova S.N., Guryeva Y.F., Dianova E.M. ... Petrushin S.I. (2007) *Prakticheskij kurs anglijskogo jazyka. 4 kurs. 5-e izdaniye* [Practical English course. Year 4. 5th ed.]. M., Russia: Vlos.
34. Arakin, V.D., Selyanina, L.I., Gintovt, K. P., Sokolova, M. A., Shabadash, G.A., Krylova, N.I. ... Denisova, V.S. (2012). *Prakticheskij kurs anglijskogo jazyka. 1 kurs. 6-e izdaniye*. [Practical English course. Year 1. 6th ed.]. Moscow, Russia: Vlos.
35. Arakin V.D., Selyanina L.I., Kutsenko A.V., Lyubimova L.G., Mikheyeva I.V., Popova L.G. ... Denisova V.S. (2008) *Prakticheskij kurs anglijskogo jazyka. 2 kurs. 7-e izdaniye* [Practical English course. Year 2. 7th ed.]. M., Russia: Vlos.
36. Merkulova Y.M., Filimonova O.E., Kostygina S.I., Ivanova Y.A. Papanova L.V. (2002) *Anglijskij jazyk dlya studentov universitetov. Chteniye, ustnaya i pismennaya praktika* [English for university students. Reading, writing and conversation]. SPb., Russia: Soyuz.
37. Moon R.E. (1998) *Fixed expressions and idioms in English: A corpus-based approach*. Oxford, England: Clarendon Press.
38. McEnery A.M., Hardie A. (2012) *Corpus linguistics: Methods, theory and practice*. Cambridge, England: Cambridge University Press.
39. Mitchell L.A., Gural' S.K. (2015) *Ispol'zovaniye internet-resursov v obuchenii inozazychnoy grammatike* [The use of Internet resources in teaching foreign grammar] // *Jazyk i kul'tura – Language and Culture*. 2 (30). pp. 117–122.
40. Yastrebova Y.B., Vladykina Y.G., Ermakova M.V. (2014) *Kurs anglijskogo jazyka dlya studentov jazykovykh vuzov. Coursebook for upper-intermediate students. [A course in English for language majors. Coursebook for upper-intermediate students]*. M., Russia: MGIMO University.
41. Oxford University Press. (2019b) *Oxford learner's dictionaries. Guide to symbols and labels*. Retrieved from <https://www.oxfordlearnersdictionaries.com/about/english/labels>
42. Lexical Computing CZ (2018) *Sketch Engine. British National Corpus (BNC)*. Retrieved from <https://www.sketchengine.eu/>
43. BYU Corpora (2018) *The Corpus of Contemporary American English (COCA)*. Retrieved from <https://corpus.byu.edu/coca/>
44. Mitchell L.A., Mitchell P.J., Vozdvizhenskiy V.V. (2015) The role of actual situational contexts in the interpretation of situation-bound utterances. *Procedia-Social and Behavioral Sciences*, 200, pp. 313–317.
45. Huo Y.J. (2014) Computer aided design of Chinese college English teaching materials based on COCA corpus. In *Applied Mechanics and Materials* (Vol. 590, pp. 916–919). Trans Tech Publications Ltd.
46. Vrbinc A., Vrbinc M. (2016) Canonical forms of idioms in online dictionaries. *GEMA Online Journal of Language Studies*, 16(2), pp. 1–15. <https://dx.doi.org/10.17576/gema-2016-1602-01>

47. Hinkel E. (2017) Teaching idiomatic expressions and phrases: Insights and techniques. *Iranian Journal of Language Teaching Research*, 5(3), pp. 45–59.
48. Miller J. (2012) Dictionaries without borders: Expanding the limits of the academy. *Journal of Learning Design*, [S.l.], 5(1). pp. 43–51.

Information about authors:

Shilnov A.G., M.Ed., Senior Lecturer, National Research Tomsk State University (Tomsk, Russia). E-mail: shilnoff_andrey@mail.ru

Miller J., Ph.D., Senior Lecturer, School of Education, University of Adelaide (Adelaide, Australia). E-mail: julia.miller@adelaide.edu.au

Mitchell P.J., P.Sc. (Edukation), Head of Department of Translation and Language Communication, National Research Tomsk State University (Tomsk, Russia); Senior Research Fellow, Ural Federal University (Ekaterinburg, Russia). E-mail: peter_mitchell@mail.ru

Smokotin V.M., D.Sc. (Philosophy), Associate Professor, National Research Tomsk State University (Tomsk, Russia). E-mail: vladimirmokotin@yandex.ru

Appendix 1

Data sources for the corpus of authentic spoken texts

The idioms contained in Corpus 2 were extracted from the following sources:

1. Randomly selected podcasts from the following series produced by BBC World Service between September 2016 and August 2018: *Global News*, *Newshour Extra*, *The Bid Idea*, *The Documentary*, *Outlook* and *The Real Story*.

2. Randomly selected podcasts from the following series produced by the USA's *National Public Radio* (NPR) in March-August 2018: *Fresh Air*, *Morning Edition* and *All Things Considered*

3. Randomly selected *CNN 10* news shows produced by CNN in May – June 2018.

4. Randomly selected videos produced by the following YouTube channels between November 2016 and August 2018: *FORMULA 1* (2,472,993 subscribers on 1 July 2019), *The White House* (1,172,254 subscribers on 1 July 2019), *Channel 4 News* (743,367 subscribers on 1 July 2019) and *The New York Times Conferences* (17,304 subscribers on 1 July 2019).

Language produced by non-native speakers of English (identified from the accent and the content of a given clip) was not included in the analysis, since the study focusses on the occurrence of idioms solely in native speaker discourse.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИДИОМ, ПРЕДСТАВЛЕННЫХ В РОССИЙСКИХ УЧЕБНИКАХ АНГЛИЙСКОГО ЯЗЫКА, И ИДИОМ, ПРЕДСТАВЛЕННЫХ В СОВРЕМЕННЫХ АУТЕНТИЧНЫХ ИСТОЧНИКАХ

Шильнов Андрей Геннадьевич – магистр педагогики, ст. преподаватель, Национальный исследовательский Томский государственный университет (Томск, Россия). E-mail: shilnoff_andrey@mail.ru

Миллер Джулия – доктор философии, старший преподаватель, Школа педагогики, Университет Аделаиды (Аделаида, Австралия). E-mail: julia.miller@adelaide.edu.au

Митчелл Пётр Джонович – доктор педагогики, заведующий кафедрой перевода и языковых коммуникаций, Национальный исследовательский Томский государственный университет (Томск, Россия); старший научный сотрудник, Уральский федеральный университет им. первого Президента России Б.Н. Ельцина (Екатеринбург, Россия). E-mail: peter_mitchell@mail.ru

Смокотин Владимир Михайлович – доцент, доктор философских наук, профессор, Национальный исследовательский Томский государственный университет (Томск, Россия). E-mail: vladimirsmokotin@yandex.ru

DOI: 10.17223/19996195/49/10

Аннотация. Во многих учебниках английского языка содержится материал, не отражающий в полной мере употребление языка его носителями. В частности, это касается употребления низкочастотной лексики, которая включает в себя идиомы. Основная цель данной статьи – провести анализ идиом в учебниках английского языка, опубликованных в России и используемых в российских вузах. Авторы проводят обзор наиболее употребительных терминов в данной сфере (фраземы, лексические связки и т.д.). Особое внимание уделяется таким критериям в определении термина «идиома», как образность, семантическая прозрачность / непрозрачность и структурная инвариантность. Исходя из данных критериев, авторы аргументируют свое решение оставить определенные категории часто воспроизводимых и / или образных языковых отрезков за рамками исследования. Важное место в работе занимает категория пословиц и поговорок, которые рассматриваются как частный случай идиом в английском языке. Ввиду особо низкой встречаемости пословиц и поговорок в корпусах аутентичного английского языка, выявленной в ходе исследования, авторы ставят под сомнение целесообразность отведения данным единицам значительной роли в учебниках английского языка. В статье представлен обзор фразеологических компонентов учебных пособий и рассмотрены конкретные примеры проявления языковой неаутентичности учебников. Заявлены преимущества использования корпусов аутентичного английского языка с целью подбора идиом, которые могут использоваться в языковом обучении. В частности, результаты исследования эмпирически демонстрируют, что список идиом, составленный на основе произвольно выбранных современных аутентичных материалов, с большой долей вероятности будет состоять из часто употребляемых идиом, объяснения которых можно найти в англоязычных учебных словарях. Такие идиомы имеют высокую частотность употребления согласно данным поиска в корпусах и, следовательно, могут быть с уверенностью включены в учебный план курса. В ходе данного исследования было проведено сравнение идиом из трех отечественных учебников (корпус 1) с идиомами, выявленными в аутентичных видео- и аудиоматериалах, содержащих спонтанную, полуспонтанную и подготовленную речь носителей британского и американского вариантов английского языка (корпус 2). Авторами была проанализирована частотность данных идиом в Британском национальном корпусе (British National Corpus) и корпусе современного американского английского (Corpus of Contemporary American English). В ходе анализа выявлено, что представленные в учебниках идиомы имеют гораздо более низкую частотность в обоих корпусах. Результаты исследования показывают, что представленные в учебниках идиомы не отражают реальное употребление английского языка его носителями. Следовательно, как разработчикам методических материалов, так и преподавателям необходимо уделить должное внимание вопросу подбора идиом и, по возможности, активно использовать аутентичные источники.

Ключевые слова: идиомы; преподавание английского языка; языковой корпус; учебное пособие; аутентичные материалы; аутентичность.

Received 4 February 2020