COMPARATIVE ANALYSIS OF IDIOM SELECTION AND SEQUENCING IN ESTONIAN BASIC SCHOOL EFL COURSEBOOKS

Rita Anita Forssten, Liljana Skopinskaja, Suliko Liiv

Abstract. The article investigates the selection and sequencing of the idioms encountered in two locally-produced and international coursebook series currently employed in Estonian basic schools. It is hypothesized that there exists a positive correlation between idioms' difficulty and coursebooks' language proficiency level. The hypothesis is tested through a statistical analysis of the idioms found which are categorized in terms of their analysability into three categories where category 1 includes analysable semi-literal idioms, category 2 comprises analysable semi-transparent idioms, and category 3 encompasses non-analysable opaque idioms, and then analysed through an online language corpus (British National Corpus). The results of the study reveal that the coursebook authors under discussion have disregarded idioms' frequency as a criterion for selection or sequencing, whereas the factor utilized to some extent is the degree of analysability.

Keywords: coursebook evaluation, corpus analysis, analysability, compositionality, English

1. Introduction

In Chitra Fernando's words, "No translator or language teacher can afford to ignore idioms or idiomaticity if a natural use of the target language is an aim" (1996: 234). Still, rather than being linguistic luxury, idioms are primarily a matter of comprehension; native speakers use idioms frequently and spontaneously. Language learners can without doubt express themselves by using literal language only, but to understand native speakers, who use idioms in everyday speech, they need at least a passive command of complex lexicalised word combinations, that is idioms. At first sight, idioms may appear burdensome: an abundance of cramming by heart. However, some idioms (e.g., *go off the rails*) are to some extent analysable, which makes them easier to understand and remember. What this article

aims to stress is that this particular aspect of idioms should be utilized in foreign language teaching. In his corpus analysis, Dilin Liu claims that most teaching materials of English idioms lack analytical reasoning and consist of infrequent phrases that are not particularly useful for learners, and therefore "one of the first issues to consider in idiom instruction is which idioms to teach and in what sequence" (2003: 671–672).

While idioms have been studied extensively from various linguistic perspectives, such as semantic, functional, socio-cultural, etc., the criteria for their inclusion in EFL coursebooks are still unclear and deserve comprehensive research. The study on EFL teaching materials has so far investigated design issues in idioms' teaching (Beloussova 2015) or their frequency analysis (Alavi, Rajabpoor 2015). None has actually addressed idiom introduction sequencing criteria which is the focus of the present study. Our aim is to establish whether the coursebook authors of locally-produced series (*I Love English*) and international series (*Upstream*) have adopted analysability as their criterion for selecting and sequencing idioms in the design of their student's books and workbooks to facilitate learning by starting with the most intelligible expressions, or whether they alternatively (or in coexistence with the sequencing criteria) have employed frequency of occurrence as their idiom selection criterion[1]. For this purpose, a taxonomy of idioms is constructed which is based on prior research (Grant, Bauer 2004, Nunberg et al. 1994) but modified to suit the present study goals.

The hypothesis is that there appears a positive correlation between the idioms' degree of difficulty (related to their analysability) and the language proficiency levels of the coursebooks. Furthermore, some consideration is given to the number of idioms: whether there are more idioms embedded in international coursebooks (the UP series) than in locally produced ones (the ILE series).

In conducting the research, the two coursebook series mentioned above have been chosen on the basis of their being registered in the Estonian Education Information System (EHIS), with the ILE series being a representative of locally produced coursebooks, and the UP series being that of international materials produced for international markets rather than a particular culture or country.

2. On the nature of idioms

Linguists have so far not reached a unified and strictly defined classification of idioms; the theories of phrasal idioms give different interpretations of the term from both semantic and grammatical perspectives, which in turn results in the complexity of idiom identification and classification (Cacciari, Tabossi 2014, Moon 1998, Liu 2008, Nunberg et al. 1994).

As stated in David Crystal's *A Dictionary of Linguistics and Phonetics*, idiom is "a term used in grammar and lexicology to refer to a sequence of words that is semantically and often syntactically restricted, so that they function as a single unit." He further extends the definition by explaining that the constituents of idioms are bound together even though "some idioms do permit a degree of internal change, and are somewhat more literal in meaning than others" (2008: 236).

[1]   Later in this article, *I Love English* and *Upstream* will be abbreviated as ILE and UP, and the expressions *student's book* and *workbook* as SB and WB respectively.

Institutionalization (or conventionality) is regarded as an integral part of an idiom. It is the process by which new material is incorporated into language; it involves the fact that any potential ambiguity of the original formation is ignored (Bauer 1983). However, the degree of conventionality varies from extremely frequent to rather infrequent (Moon 1998). For Fernando, institutionalization denotes that "idioms are conventionalized expressions, conventionalization being the end result of initially *ad hoc*, and in this sense, novel expressions" (1996: 3). Otherwise stated, a phrase that is originally created to suit a purpose in a specific situation becomes conventionalized when taken over into general knowledge and active use of a speech community.

Idioms have an emotive aspect that is not typical of other lexical units (Nunberg et al. 1994): idioms conform to moods, feelings and attitudes of the objects they denote and they seldom portray emotionally neutral situations. A further distinguishing pragmatic feature of idioms is informality, which is a logical consequence of the emotive aspect. Informality denotes idioms' tendency to appear in speech rather than in writing and to represent colloquial rather than formal register. Another pragmatic property that is closely related to the aforementioned qualities is proverbiality which signifies the tendency of idioms to express resemblance between situations involving social interest and concrete or familiar phenomena, e.g., *chewing fat* or *spilling the beans* (Nunberg et al. 1994).

One of the most commonly used criteria of an idiom is non-compositionality which signifies that the meaning of an idiom cannot be inferred by adding together the meanings of its constituents (Fernando, Flavell 1981). Idioms are often, but not always, non-literal and Fernando's term semantic opacity (1996) refers to their non-compositionality. However, the degree of compositionality varies between different types of idioms. Sam Glucksberg (2014) gives *pop a question* as an example of a fully compositional idiom: both the verb and the noun phrase can be mapped directly onto their respective idiomatic referents. He suggests *kick the bucket* as an example of a less compositional idiom that does not allow constituent-to-constituent mapping since only the phrase as a whole can be mapped to the meaning of *die*.

A feature of idioms that is somewhat subject to dispute is inflexibility, or frozenness. Nunberg et al. (1994) maintain that syntactic inflexibility of idioms manifests itself in the limited extent of different structural constructions. Idioms are inflexible in a way that is not comparable to any other expressions, and rewording makes them, if not completely incomprehensible, then at least strained (Egan 2008). Fernando (1996) distinguishes idioms from general idiomaticity in stating that even though all idioms indicate idiomaticity, such idiomatic expressions that are unrestricted in their variants are not idioms. This property of idioms that Moon (1998) refers to as fixedness is lexicogrammatical, and it also includes syntactic frozenness.

Furthermore, Moon (1998) proposes that idioms are multiword expressions, simultaneously distinguished from phrasal verbs and compounds. In addition, intonation separates a non-compositional interpretation from a compositional one, thereby differentiating idioms from literal phrases. Similarly, Fernando (1996) perceives idioms as multiword expressions, using the term compositeness to describe that feature; yet, she admits that some linguists regard even one-word expressions as idioms. Although the majority of linguists appear to have found consensus on

that multiwordness is a criterion to be attached to idioms, even this is a matter of dispute (Makkai 1972).

3. Research materials and methods

The current study is based on descriptive statistics, which is used to describe and summarize the collected data. The measures of basic statistics included are the average values of variables and dispersion of variables. The aspect of the study is frequency distribution, and the knowledge function is primarily description. The first level analysis is the distribution analysis: counting the frequencies of idiom occurrences in local and international coursebook series. The second level analysis is variable-oriented: measuring the degree of correlation.

The hypothesis that there is a positive correlation between the degree of difficulty of the coursebook idioms and the language level of the coursebooks is accepted only if the null hypothesis that there exists no relationship between these two phenomena can be proven false. The possible positive correlation supports the view that either the ILE or UP series or both of them present idioms in a well-considered order starting from the most intelligible ones. Since frequency and comprehensibility cannot be integrated, they are measured and scrutinized separately.

The primary source of the analysis comprises fourteen ILE coursebooks, of which seven are SBs and seven WBs, and eight UP coursebooks, similarly consisting of both textbooks and workbooks. The first ILE set is intended for second or third graders, i.e. for students who are just starting their English studies at primary level, whereas the last part of the series is for ninth graders who are to reach B1 language level. Thus, the ILE series is a sequence of books stretching over the whole period of A-foreign language learning in Estonian basic schools. Correspondingly, the parts of the US series that cover levels from beginner (A1+) to pre-intermediate+ (B1+) are analysed. The four UP sets are divided into two parts in order to make the series comparable with the 7-piece ILE series.

3.1. Corpus-based study

The frequencies of the idioms manually extracted from the ILE and UP series are obtained from the spoken and written corpora of the British National Corpus (BNC). The reason for the exclusion of the Corpus of Contemporary American English (COCA) is that as a rule British English is the norm in Estonian schools, and most of the EFL teaching materials currently employed in secondary education have been developed in the UK.

To discover the correct frequencies of idiom occurrence, in some cases the entire lexeme needs to be searched for: when necessary, all verb inflections and changes in the form of nouns, i.e. the singular and plural forms and the nominative, objective, and possessive cases, are included in the search. Because a part of formal idioms is filled by any syntactically and semantically appropriate word or phrase, the character _ is being used in the corpus search. It symbolises any single word that can be used in the phrase being searched so that there is the required slot(s) in the phrase where an appropriate element can be filled in. Moreover, in some instances,

a phrase requires several word order variations. In addition to inflections, cases, and word order, variations in spelling need to be checked.

In compliance with their frequencies discerned by the corpus search, the spotted idioms are divided into six categories: class 1 includes idioms with 0–500 hits, class 2 consists of idioms with 501–1,000 instances, idioms with 1,001–1,500 matches comprise class 3, class 4 includes idioms with 1,501–2,000 solutions, class 5 contains idioms with 2,001–2,500 hits, and class 6 is composed of idioms with over 2,500 occurrences. The relative share of each class (1–6) in the total idiom quantity is calculated, likewise their mean frequency scores and standard deviations to indicate how much the data varies. Since the category system provides several values of discrete variables, cumulative percentages are presented to clarify the tendency. Subchapter 4.2 includes an overview of the results. The frequency distribution of the ILE and UP idioms can be found in Table 1 under subtitle 4.2.

3.2. Constructing typologies

The fundamental premise of our research is that an idiom is a conventionalized multiword unit. As a further framework for identification of an idiom, the study relies upon its following general properties: its meaning is not the sum of the meanings of its constituents, i.e. its meaning cannot be completely transparent; it is semantically opaque or semi-opaque, and/or metaphorical; it is both syntactically and lexically invariable – at least to a certain extent.

The analytical technique chosen for the purpose of classifying observations (i.e. the idioms found in the ILE and UP series) into meaningful categories is grouping. In order to categorize the idioms, the characteristics that distinguish different idiom groups from each other have to be identified. Therefore, what follows the previous procedure is the construction of typologies. The idioms are divided into categories in accordance with their degree of difficulty. To be able to discover their introduction sequence pattern or complete lack thereof, only the first occurrence of each idiom is registered in the teaching materials.

The idioms are categorized in compliance with their degree of analysability – which contributes to their degree of difficulty. The necessity of combining and modifying some of the existing taxonomies (Makkai 1972, Fernando, Flavell 1981, Nunberg et al. 1994, Grant, Bauer 2004, Liu 2008, Yoshikawa 2008, Cacciari, Glucksberg 1991) for the purposes of the present study, instead of using some of them as such, results from the reason that many of the existing categories are unnecessarily detailed and hierarchical for sequencing idioms for foreign language learning materials. For example, Makkai has dozens of idiom subcategories and structural sub-sub-categories (1972). Furthermore, Fernando and Flavell's (1981) categories of both semi-opaque and opaque idioms contain idioms most of which language learners would find opaque. Similarly, despite the fact that Cacciari and Glucksberg (1991) recognise a difference between categories of opaque but analysable and of non-analysable idioms, the phrases belonging to their category of opaque but analysable idioms are actually non-analysable for a beginner learner. For example, the phrase *spill the beans* is likely to appear as an opaque idiom for a beginner learner, although due to one of the meanings of the verb *spill*, it is logical that many researchers, e.g., Nurnberg et al. (1994), regard the phrase as analysable.

Yet, a beginner learner could hardly be expected to infer its meaning – presumably not even from context – because s/he, most likely, is not familiar with the meaning *reveal confidential information* of the verb *spill*.

The present categorization is predicated on the taxonomies of Grant and Bauer (2004) and Nunberg et al. (1994). Idioms in which some element preserves its literal meaning, i.e. semi-literal idioms, constitute category 1 (e.g., *surf the Internet*, *feel blue*, *rags to riches*, *fruits of _ labour* and *love _ to bits*). Category 2, i.e. the second most easily understood idiom group, is composed of idioms none of whose constituents preserves its literal meaning. Yet, these idioms are somewhat transparent due to some other factors: a conventionalized metaphor or allusion can function as a semi-transparent idiom, e.g., *keep the ball rolling*, *fit for a queen*, *put a smile on _ face*, *be in the wars*, *could eat a horse* and *Lot's wife*. The idioms in categories 1 and 2 are decomposable. Category 3 encompasses utterly opaque phrases commonly designated as core idioms, such as *red herring*, *of course*, *by the way*, *be up to*, *make one's way*. It is impossible for EFL learners to infer the meanings of core idioms because they are non-decomposable. In summary, category 1 consists of analysable idioms that are to some extent literal, category 2 is constituted by analysable idioms that are semi-transparent, while the non-analysable idioms of category 3 consist in opaqueness.

3.3. Correlation and dependence

At each language proficiency level of both coursebook series, the relative frequencies of different categories of idiom analysability are calculated. To explore whether there occurs any dependence between the two variables (i.e. the degree of idiom analysability and the coursebook language level), the possible correlation between them is measured: the relative frequencies obtained are compared to the rising language proficiency levels of the book series. The measure of correlation is calculated with the sample means and standard deviations of the two sets of the population data. To exclude the possibility that the differences could be explained by a sampling process error, one-tailed and paired-samples t-tests are executed. The interpretation of the correlation coefficients that is used is based on the guide suggested for the absolute value of r: correlation is trivial if $|r| \leq 0.1$, it is small if $0.1 < |r| \leq 0.3$, it is moderate if $0.3 < |r| \leq 0.5$, and it is strong if $|r| > 0.5$ (Cohen 1988).

4. Data analysis and discussion

4.1. Quantity

The quantities of idioms discovered in the ILE and UP series confirm the assumption that international coursebooks written within the target language culture utilize idioms more frequently than the local ones written in Estonia. The UP series contains twice the number of idioms compared to the ILE series: a total of 264 idioms that fulfil the criteria used to determine the idiom status in the present study can be found in the ILE series, while the corresponding number in the UP series is 550.

Regarding the ILE series, the quantity of new idioms introduced does not change between the first two parts, but beginning from the third part the number of idioms grows steadily until it increases substantially in the final part of the series (see Figure 1). In the UP series, the development of the idiom vocabulary size is not as linear as in ILE; there occurs even occasional decrease in their number between the different levels of the series (see Figure 2). Nonetheless, the number of idioms increases significantly at the pre-intermediate level. In either case, the evidence seems to strongly suggest that the coursebook authors operate from the premise that learners do not really develop the ability to handle idioms until the pre-intermediate language level.
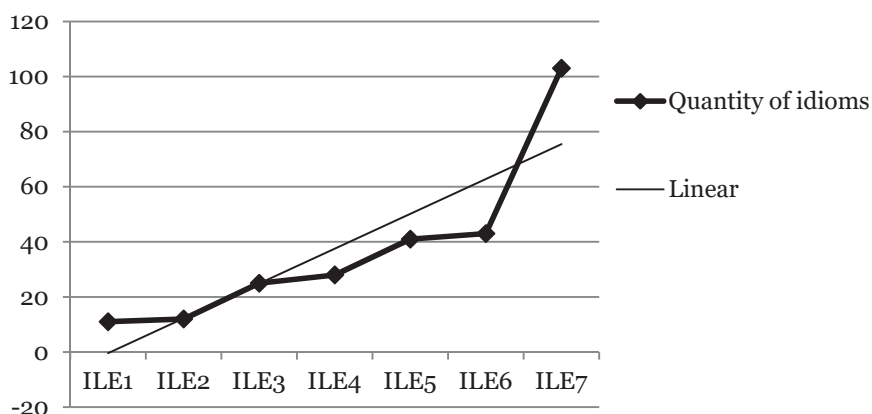

Figure 1. The quantity of idioms in the ILE series (x-axis: the ILE coursebook sets including both an SB and a WB, y-axis: the number of idioms)
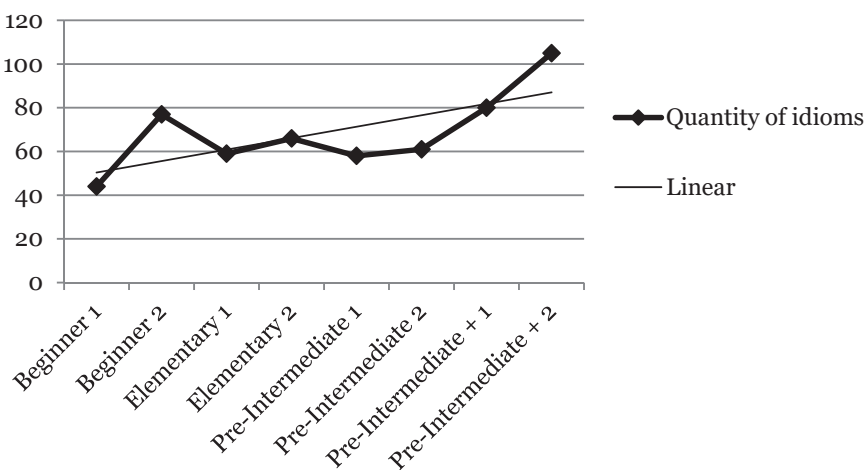

Figure 2. The quantity of idioms in the UP series (x-axis: the UP coursebook sets including both an SB and a WB, y-axis: the number of idioms). The variables on the horizontal axis are the parts of the UP series that cover the levels beginner (A1+), elementary (A2), pre-intermediate (B1) and pre-intermediate+ (B1+)

4.2. Frequency

In both coursebook series, the overwhelming majority of idioms (which constitutes 56.8% of the whole in ILE and 65.0% in UP) are significantly infrequent (with occurrences ranging from null to 500) in the BNC. Class 6, which is the second largest idiom group both in the ILE and UP series, comprises phrases with multitudinous matches in the corpus, ranging from 2,500 to over 40,000. However, the relative shares of ILE and UP idioms in this class are comparatively small: in ILE 17.8 and in UP 12.7%. Idioms belonging to class 2 (501–1,000 hits) form the third largest proportion of idioms both in the ILE (13.3%) and in the UP (9.5%) series.

Furthermore, in both ILE and UP, class 3 (1,001–1,500 occurrences per idiom), class 4 (1,501–2,000 occurrences per idiom), and class 5 (2,001–2,500 occurrences per idiom) represent only a marginal proportion of the total quantity of ILE/UP idioms: the proportions of these three categories vary between 3.0 and 5.3%.

The frequency distribution of the ILE and UP idioms described above can be found in Table 1. When usefulness is considered to be determined by the frequency of occurrence predicted by the BNC, some observations on the usefulness of the idioms selected by the authors and generalizations about the idiom frequency in the BNC can be made on the basis of the present study.

Table 1. The BNC Frequency Distribution of the ILE and UP Idioms

	Hits in the BNC	Qty in ILE	ILE %	ILE cum. %	Qty in UP	UP %	UP cum. %
Class 1	0–500	150	56.8	56.8	358	65.0	65.0
Class 2	501–1,000	35	13.3	70.1	52	9.5	74.5
Class 3	1,001–1,500	14	5.3	75.4	29	5.3	79.8
Class 4	1,501–2,000	8	3.0	78.4	22	4.0	83.8
Class 5	2,001–2,500	10	3.8	82.2	19	3.5	87.3
Class 6	2,500–	47	17.8	100.0	70	12.7	100.0
Total		264	100.0		550	100.0	

Based on the above statistics, it would appear that neither the ILE nor UP authors have paid sufficient attention to the usefulness of idioms entailing frequency rates as high as possible. If the above category is transformed into a six-point scale in which the most frequent idioms get six points, the mean frequency score of the ILE idioms is 2.4 points out of six, and the corresponding result of the UP idioms is 2.1. The standard deviation in the ILE set is 1.95 and 1.79 in the UP set; since the range is five, the estimate of the standard deviation is 1.25.

The comparison between the corpus study results of the ILE and UP idioms produces very similar outcomes. One reason for this is that over 54% of the idioms that appear in the ILE series can also be found in the UP series. Another reason for the resemblances could be that the established frequency distribution structure reflects the volumes of idiom frequency not only in the analysed coursebooks but in general: viz. that the quantity of very infrequent idioms is considerably more extensive than that of frequently used idioms. However, it should be taken into consideration that different idiom frequency findings tend to be somewhat distorted by the inconsistency in defining idioms. The following remarks may give some explanations to the abovementioned idiom frequency distribution.

First, certain types of idioms seem to be frequent in several other corpora too. Some instances of the coursebook idioms that have a high frequency score in several corpora are *of course* with 29,651 BNC hits and *such as* with 31,465 BNC hits. What is common to these phrases is that they are notably short expressions; they lack both the subject and predicate; and their intended use, which is versatile, differs significantly from that of rather infrequent idioms both at sentence and register levels. All of the ILE and UP idioms that are frequent in several corpora (excluding *such as*) appear on top of Liu's list of the most frequent idioms in all three corpora he searched – Corpus of Spoken, Professional American English; Michigan Corpus of Academic Spoken English; and Spoken American Media (Liu 2003: 695–700).

Second, another frequent group of idioms consists of idiomatic phrasal verbs, although their frequency rates are considerably lower than those of the previous type. Some instances of frequent idiomatic phrasal verbs are *carry on* with 3,858 hits and *turn up* with 2,856 hits. However, a challenging problem in measuring the frequency of phrasal verbs remains unsolved: the corpus does not sort multi-sense phrasal verbs, e.g., *to work out,* by their meanings. Therefore, the phrasal verb statistics in the present study are for illustrative purposes only and do not in every case represent the frequency of occurrence per a figurative meaning.

Third, the wide-ranging affordances demonstrate the relevance of idioms of class 4, class 5, and class 6 for language learners. Yet, the quantities of these idioms in both series are remarkably smaller than those of rare idioms; even together these categories constitute less than 25% of the idioms in the ILE and UP series. The disappointing result is somewhat expected; according to Liu, even many English idiom teaching and reference materials for ESOL learners consist of idioms that demonstrate "primarily the authors' intuition rather than any empirical data, and a substantial number of them are rarely used" (Liu 2003: 672).

On balance, frequency is a somewhat complex indicator of usefulness. It appears that some idioms (like short and predicate-less idioms and idiomatic phrasal verbs) are less register-sensitive and not limited to use in certain pragmatic contexts, which results in versatility; hence their frequency. Still, spoken and informal language may include a considerable number of idioms that are classified as rare. Liu (2003) points out that when selecting idioms for language teaching purposes the relevant register(s) should be kept in mind.

4.3. Introduction sequence

The results indicate that there exists a statistical relationship between the degree of idiom analysability and the proficiency levels of the given coursebooks. The relationships are presented visually in the form of line charts (see Figure 3 and Figure 4). The calculated probability supports the opposite of the null hypothesis: all the p-values vary between 0.0008 and 0.0031. One of the p-values (p = 0.0008) is statistically even highly significant.

In the ILE series, the correlation coefficient of category 1 – which contains the idioms with a literal component like *laugh your head off* and *speak one's mind* (ILE 6 SB 2004) – is −0.78 (p = 0.0031). It indicates variables that are considered highly correlated: the relative share of category 1 significantly decreases when the language proficiency level rises. In the UP series, the corresponding figure is

−0.97 (p = 0.0012). The negative correlation between the most easily analysable idioms and the rise in the language proficiency level in both series is noteworthy: the share of easily analysable idioms with a literal element diminishes towards the pre-intermediate level.
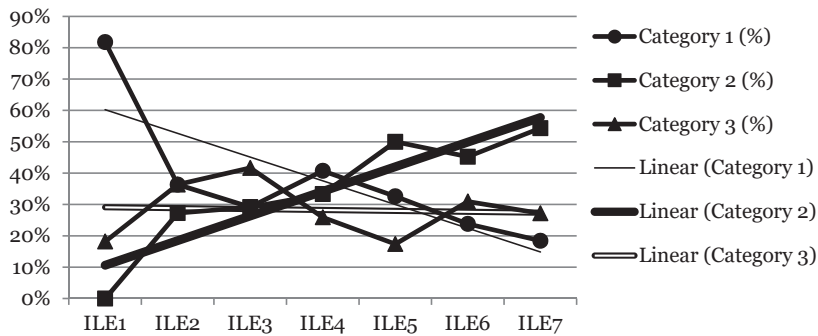

Figure 3. The occurrence of analysable idioms with a literal element (category 1), analysable idioms without a literal element (category 2), and unanalysable idioms (category 3) in the ILE series
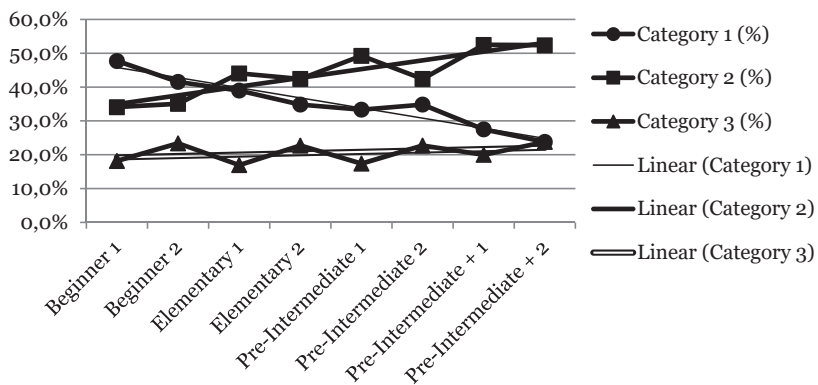

Figure 4. The occurrence of analysable idioms with a literal element (category 1), analysable idioms without a literal element (category 2), and unanalysable idioms (category 3) in the UP series. The variables on the horizontal axis are the parts of the UP series that cover the levels beginner (A1+), elementary (A2), pre-intermediate (B1) and pre-intermediate+ (B1+)

The r-value of 0.92 (p = 0.0014) indicates a strong positive correlation between category 2 – which covers the analysable idioms without a literal element like *ups and downs* and *hand in hand* (ILE 6 SB 2004) – and the rise of the proficiency level in the ILE series. The corresponding calculated correlation coefficient 0.89 (p = 0.001) in the UP series indicates a trend similar to the ILE series. To compensate for the drastic reduction of category 1 idioms, the relative proportion of category 2 idioms grows substantially in the second part of the ILE series. This increase is, however, not as sharp as the decrease of the proportion of idioms with a literal element because there also appears steep short-term growth in the share of the idioms belonging to category 3.

The correlation coefficients reflecting the relationship between the relative proportion of category 3 – which encompasses the unanalysable idioms like *of course* and *by the way* (ILE 4 SB 2006) – and the language proficiency levels are in strong contrast to those of category 1 and category 2. The r-value of −0.07 (p = 0.002) in the ILE series indicates only a negligible negative relationship, while the r-value of 0.35 (p = 0.0008) in the UP series reveals a moderate uphill linear relationship: the statistical data exhibit a barely detectable decrease in the proportion of unanalysable idioms towards B1 language level in the ILE series and a rather moderate increase in the UP series.

Although there appears a relationship between the degree of analysability and the coursebook language level in both series, the results concerning the correlation are not unequivocal. It would seem that while the language proficiency level goes up, analysable idioms with a literal constituent (category 1) are gradually replaced by analysable idioms without a literal element (category 2); this trend supports the hypothesis that there is a positive correlation between the idioms' degree of difficulty and the coursebook language levels. However, the data concerning unanalysable idioms are not compatible with the hypothesis; the language proficiency level does not affect the sequencing of core idioms. In the ILE series, the variation in the quantity of unanalysable idioms appears to be random, whereas in the UP series the quantity of unanalysable idioms remains rather invariable throughout all coursebook levels.

Even though the present study does not investigate why some idioms are more frequently used than others, it is worth noting that, according to Moon's research (1998), core idioms are very rare. Therefore, it may be that even if the degree of difficulty of idioms were their introduction sequence criterion, there might not be enough core idioms to compensate for the reduction in the quantity of analysable idioms in such a way that it would count as authentic language usage. Alternatively, the coursebook authors may have postponed the extensive introduction of difficult core idioms to more advanced levels.

5. Conclusion

The theoretical literature on idioms highlights them as expressions characteristic of language, while simultaneously objecting to their insufficient and unorganized treatment in language teaching materials. Therefore, the present study was undertaken to explore the treatment of idioms in the two EFL coursebook series (*I Love English* as contrasted to *Upstream*) utilised in Estonian schools, more precisely, the application of the criteria for idiom selection and introduction sequencing. The study also sought to ascertain whether there is any noticeable difference in the quantity of idioms being taught between international and local EFL coursebooks.

The findings of the research, based upon 814 idioms found in 22 student's books and workbooks, support the hypothesis that the idioms' degree of analysability correlates with the language levels of the coursebooks, contradicting at the same time the presumption of frequency as being the criterion for the selection and sequencing of idioms in the given coursebooks. The usefulness of idioms, which is measured by their frequency in the BNC, appears not to have been taken into account. The BNC occurrences demonstrated in the present study indicate that in the majority of cases the idioms employed in the analysed coursebooks are notably

infrequent both in spoken and written British English. The BNC search of over half of the idioms gave only 0–500 instances. The idioms that are significantly frequent (with over 2,500 matches in the corpus) constitute only eighteen per cent of the ILE idioms and thirteen per cent of the UP idioms. Furthermore, the results support the presumption that the authors of international coursebooks are more prone to use idioms than those of locally produced ones.

The study has also identified a considerable positive correlation between the quantity of analysable idioms without a literal element and the rise in the coursebook language proficiency levels, and a corresponding negative correlation regarding the quantity of analysable idioms with a literal element and the coursebook level, which suggests that the degree of idioms' analysability has been employed as an introduction sequence criterion. However, the results of the comparison between the usage of opaque idioms and coursebook levels appear to conflict with the hypothesized relationship: the quantity of unanalysable idioms is rather low and invariable in every part of both coursebook series. On the other hand, the moderate quantity of core idioms throughout both series could be incorporated in the hypothesis as regards the relationship between idiom analysability and coursebook proficiency level if it were assumed that the volume of opaque idioms has been restricted on purpose due to their challenging nature and that they have been sequenced to be introduced to students at more advanced language proficiency levels.

A further survey on the idioms of the *Upstream* sequels would be required to either prove or disprove the reasoning that a large-scale introduction of opaque core idioms has been postponed to more advanced language levels. Nevertheless, this study provides some aspects that could be taken into consideration in the design of coursebooks and typologies for categorising and sequencing idioms particularly for coursebook evaluation purposes.

References

Alavi, Sepideh; Rajabpoor, Aboozar 2015. Analyzing idioms and their frequency in three advanced ILI textbooks: A corpus-based study. – English Language Teaching, 8 (1), 170–179. http://dx.doi.org/10.5539/elt.v8n1p170

Bauer, Laurie 1983. English Word-Formation. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165846

Beloussova, Violetta 2015. Idiom Learning Materials for Estonian Secondary School Students. MA thesis. Tartu: University of Tartu. http://hdl.handle.net/10062/47069

Cacciari, Christina; Glucksberg, Sam 1991. Understanding idiomatic expressions: The contribution of word meanings. – Greg P. Simpson (Ed.), Understanding Word and Sentence. Amsterdam: Elsevier, 217–240.

Cacciari, Christina; Tabossi, Patrizia (Eds.) 2014. Idioms: Processing, Structure, and Interpretation. 2nd ed. New York: Psychology–Taylor.

Cohen, Jacob 1988. Statistical Power Analysis for the Behavioural Sciences. 2nd ed. New Jersey: Erlbaum

Crystal, David 2008. A Dictionary of Linguistics and Phonetics. 6th ed. Malden: Blackwell.

Davies, Mark 2004. BYU-BNC. (Based on the British National Corpus from Oxford University Press). http://corpus.byu.edu/bnc/ (21.8.2015).

Egan, Andy 2008. Pretense for the complete idiom. – Noûs, 43 (3), 381–409. http://dx.doi.org/10.1111/j.1468-0068.2008.00686.x

Fernando, Chitra 1996. Idioms and Idiomaticity. Oxford: Oxford University Press.

Fernando, Chitra; Flavell, Roger H. 1981. On Idiom: Critical Views and Perspectives. Exeter: University of Exeter.

Glucksberg, Sam 2014. Idiom meanings and allusional content. – Christina Cacciari, Patrizia Tabossi (Eds.), Idioms: Processing, Structure, and Interpretation. 2nd ed. New York: Psychology–Taylor, 3–26.

Grant, Lynn; Bauer, Laurie 2004. Criteria for re-defining idioms: Are we barking up the wrong tree? – Applied Linguistics, 25 (1), 38–61. http://dx.doi.org/10.1093/applin/25.1.38

Liu, Dilin 2003. The most frequently used spoken American English idioms: A corpus analysis and its implications. – TESOL Quarterly, 37 (4), 671–700. http://dx.doi.org/10.2307/3588217

Liu, Dilin 2008. Idioms: Description, Comprehension, Acquisition, and Pedagogy. London–New York: Routledge.

Makkai, Adam 1972. Idiom Structure in English. The Hague: Mouton. https://doi.org/10.1515/9783110812671

Moon, Rosamund 1998. Fixed Expressions and Idioms in English: A Corpus-Based Study. Oxford: Clarendon.

Nunberg, Geoffrey; Sag, Ivan A.; Wasow, Thomas 1994. Idioms.– Language, 70 (3), 491–538. https://doi.org/10.1353/lan.1994.0007

Yoshikawa, Hiroshi 2008. International intelligibility in World Englishes: Focussing on idiomatic expressions. – Intercultural Communication Studies, 17 (4), 219–226. http://web.uri.edu/iaics/2008-vol-17-no-4/ (16.8.2015).

Coursebook references

Evans, Virginia; Dooley, Jenny 2005. Upstream Beginner A1+: Student's Book. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2013. Upstream Beginner A1+: Workbook. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2008. Upstream Elementary A2: Student's Book. 5th ed. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2007. Upstream Elementary A2: Workbook. 3rd ed. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2007. Upstream Level B1+: Student's Book. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2008. Upstream Level B1+: Workbook. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2006. Upstream Pre-Intermediate B1: Student's Book. 3rd ed. Newbury: Express.

Evans, Virginia; Dooley, Jenny 2007. Upstream Pre-Intermediate B1: Workbook. 4th ed. Newbury: Express.

Jõul, Mare; Kurm Ülle 2008. I Love English: Student's Book 5. Tartu: Studium.

Jõul, Mare; Kurm Ülle 2007. I Love English: Student's Book 4. Tartu: Studium.

Jõul, Mare; Kurm Ülle 2005. I Love English: Student's Book 3. Tartu: Studium.

Jõul, Mare; Kurm Ülle 2008. I Love English: Workbook 5. Tartu: Studium.

Jõul, Mare; Kurm Ülle 2006. I Love English: Workbook 4. Tartu: Studium.

Jõul, Mare; Kurm Ülle 2005. I Love English: Workbook 3. Tartu: Studium.

Kurm, Ülle 2004. I Love English: Student's Book 2. Tartu: Studium.

Kurm, Ülle 2011. I Love English: Student's Book 1. Tartu: Studium.

Kurm, Ülle 2004. I Love English: Workbook 2. Tartu: Studium.

Kurm, Ülle 2011. I Love English: Workbook 1. Tartu: Studium.

Kurm, Ülle; Jõul, Mare 2004. I Love English: Student's Book 6. 2nd ed. Tartu: Studium.

Kurm, Ülle; Jõul, Mare 2004. I Love English: Workbook 6. 3rd ed. Tartu: Studium.

Kurm, Ülle; Soolepp, Ene; Fyfe, Nicola 2010. I Love English: Student's Book 7. Tartu: Studium.

Kurm, Ülle; Soolepp, Ene; Fyfe, Nicola 2014. I Love English: Workbook 7. 3rd ed. Tartu: Studium.

IDIOOMIDE ESITAMISJÄRJESTUSE JA VALIKUKRITEERIUMIDE VÕRDLEV ANALÜÜS EESTI PÕHIKOOLIDES KASUTATAVATES INGLISE KEELE ÕPPEKOMPLEKTIDES

Rita Anita Forssten, Liljana Skopinskaja, Suliko Liiv
Tallinna Ülikool

Artikkel analüüsib võrdlevalt idioomide valikut ja esitamisjärjestust Eesti põhikoolides kasutatavates inglise keele õppekomplektides "I Love English" ja "Upstream". Uurimuse käigus püstitatakse hüpotees, et idioomide keerukusastme tõusu ja õppekomplektide keeleoskustaseme vahel eksisteerib statistiliselt oluline seos. Lähtudes eelnevatest idioomide tüpoloogiatest (Grant, Bauer 2004, Nunberg jt 1994) jaotatakse idioomid antud uurimuses kolme kategooriasse oma keerukusastme poolest: läbipaistvad otsetähenduses keelenditega idioomid, läbipaistvad otsetähenduses keelenditeta idioomid ja läbipaistmatud idioomid. Idioomide osakaalu muutumist õpikuseeriates uuritakse korrelatsioonikoefitsiendi abil BNC segakorpuse põhjal. Õpikutes kasutatud idioomide esinemissagedust vaadeldakse eraldi.

Uurimusest selgub, et läbipaistvad otsetähenduses keelenditega idioomid ja läbipaistvad otsetähenduses keelenditeta idioomid korreleeruvad tugevasti õpikute keeleoskustasemega, kuigi läbipaistmatute idioomide korrelatsioon on minimaalne. Võib täheldada, et ehkki mõlema õpikuseeria autorid on arvestanud idioomide keerukusastet ja esitamisjärjestust õpikute koostamisel, on idioomide valik neis puudulik. Samuti pole antud õpikutes kasutatud idioomide esinemissagedust kõnes ja kirjas nende valiku- ega esitamisjärjestuse kriteeriumina.

Võtmesõnad: õpikuanalüüs**,** korpusanalüüs, läbipaistvus, kompositsionaalsus, inglise keel

Rita Anita Forssten (Tallinn University) was awarded a Master of Arts in Education by Tallinn University in 2016 and her research interests are vocabulary acquisition, coursebook evaluation and intercultural communication.
Narva mnt 29, 10120 Tallinn, Estonia
rita.forssten@gmail.com

Liljana Skopinskaja's (Tallinn University) primary research interests are intercultural communication and its assessment, coursebook evaluation, foreign language teaching methodology and language teacher education.
Narva mnt 29, 10120 Tallinn, Estonia
liljana.skopinskaja@tlu.ee

Suliko Liiv's (Tallinn University) research interests are contrastive linguistics, intercultural communication, language policy, testing, methods of foreign language teaching.
Narva mnt 29, 10120 Tallinn, Estonia
liiv@tlu.ee