

# Marek Łukasik

---

## Lexical Minimum (Re)Defined

---

Lingwistyka Stosowana / Applied Linguistics / Angewandte Linguistik nr 23,  
47-63

---

2017

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach  
dozwolonego użytku.

**Marek ŁUKASIK**  
Pomeranian University in Słupsk

## **Lexical Minimum (Re)Defined**

### **Abstract:**

By popular definition, lexical minimum is a useful glottodidactic tool representing a selection of vocabulary to be taught at various language proficiency levels. Practically speaking, such minimums take the form of wordlists, extracted on frequency bases and drafted separately for each educational stage. However, the nature of language acquisition as well as the teaching/learning processes call for a much broader view of the lexical minimum, encompassing its primary, and possibly fundamental, meaning as a lexical competence of an individual and a conceptual measure for its expansion and evaluation. It is therefore claimed that the existing concept of lexical minimum as an exclusively didactic tool is outdated and needs redefining. Moreover, the methodology for a lexical minimum compilation should be revised, mainly on account of flawed compilation procedures applied by some authors. Accordingly, the aim of the present paper is to address some of the issues outlined by clarifying the notion of lexical minimum anew, discussing its multifarious characteristics and suggesting some methodological solutions to its construction.

### **Introduction**

Considerations presented in this paper have arisen as a result of my studies into the nature of terminological minimum. On the face of it, a terminological minimum is a lexical minimum, only presenting specialised vocabulary. This seems to be a view shared by some authors, who additionally do not make a clear terminological differentiation between the two concepts (cf. O.L. Yaroshenko 2014). Yet, the two notions, while related, require a clear demarcation of their scope, functions, compilation methodology and possibly presentation techniques. This stems from the features of the entry unit of each of the resources in question, i.e. a general word versus a specialised word/ term, as well as specific characteristics of prospective users and specific uses of the terminological minimum.

Analysing the necessary background, one can easily find out that neither notion has been convincingly defined nor elaborated on, with only a handful of their characteristics, mostly of glottodidactic relevance, discussed. Granted, the emergence of lexical minimum was linked to specific educational needs (assistance in developing reading skills), while terminological minimums have generally appeared as a response to the lack of LSP educational materials. Yet, the functions of the two minimums go well beyond the glottodidactic domain.

Given the unique nature of either concept, the article will present considerations related to the notion of lexical minimum, while a forthcoming paper will focus solely on the notion of terminological minimum. The succession is deliberate: under certain conditions lexical minimum can be viewed as a broader notion, and therefore some issues inherent in the former will be shared by the latter.

On a terminological note: both terms, i.e. ‘lexical minimum’ and ‘terminological minimum’ are considerably more popular in Slavic languages than elsewhere. English-writing authors prefer other expressions, such as ‘core vocabulary/terminology’, ‘elementary vocabulary/terminology’, ‘basic vocabulary/terminology’ or ‘minimum vocabulary/terminology<sup>1</sup>’, each possibly exhibiting some degree of divergence from another, which – for practical purposes – will not be discussed in this paper.

The following sections attempt to formulate the definition of lexical minimum, focussing on both the historical development of the notion and its ontological status (Section 1). Later, some suggestions are offered as regards the construction principles of lexical minimums as specific didactic tools (Section 2). This part sheds some light on erroneous, yet widely adopted, criteria and methods for lexical minimum compilation, and points to the most salient requirements for the creation of modern reference materials of this kind. In particular, the needs of prospective users as well as contemporary methods in digital humanities are discussed (Section 3 and 4). Finally, the paper suggests a new shape for lexical minimum as a specific multi-purpose didactic tool (Section 5).

## 1. Definition of lexical minimum

The term ‘lexical minimum’ has most frequently been used in the educational (glottodidactic) context, and refers to an appropriate choice of vocabulary to be acquired by a learner at a specific language proficiency level. In particular, such vocabulary should be ‘important’ and ‘useful’. In fact, word frequency has most often been the basic (or primary) criterion of vocabulary choice, while the minimum, understood as a specific type of a reference work, has become a frequency-driven word list. This prevalent conception of the notion in question can be exemplified by a definition included in the *Dictionary of Lexicography*:

**basic vocabulary** Those words selected by frequency counts and similar means, which are considered essential for communication in a language and thus also for purposes of learning that language. This selection (variously called ‘core’ or ‘elementary’ or ‘fundamental’ or ‘minimum’ vocabulary) tends to focus on words which are statistically representative, stylistically neutral and semantically powerful while at the same time

---

<sup>1</sup> The term ‘minimum vocabulary’ can cause some misunderstanding beyond linguistics, as the term is also used in the context of philosophy, and in particular logic (cf. J.G. Slater 1999: 16).

helping learners to cope in everyday communicative situations (R.R.K. Hartmann/ G. James 2002: 13, cross-references removed by M.L.)<sup>2</sup>.

The definition outlined above aptly encapsulates the major glottodidactic function of terminological minimum, and possibly indirectly addresses the fuzzy notion of ‘usefulness’ (the last line of the definition). It also mentions the vocabulary selection criteria, i.e. statistical, stylistic and semantic ones. In this respect, the definition represents a consensus, developed over the years when the notion crystallised (see below). Yet, neither the scope of application, nor the selection criteria for the lexical minimum represent a complete and final characteristics of the notion discussed in this paper. As will be demonstrated below, the glottodidactic function constitutes an important, albeit not the only role that a lexical minimum can assume. The prevailing educational role of lexical minimum can be better understood against an outline of its historical development.

An interest in the selection and grading of appropriate language material for classroom instruction arose on the grounds of second-language teaching. Realising the ineffectiveness of the so-called Direct Method, Edward L. Thorndike proposed in the twenties of the last century that for the teaching to be effective and economical, one needs to establish lexical items that are needed most. It gave rise to methods of vocabulary selection and control, based on frequency and range. It was presumed that “the frequency with which a word is used in normal reading matter is probably an objective standard by which to judge the importance of the word”. Accordingly, “frequency of occurrence should be the chief criterion for teaching vocabulary with a view to developing the four skills” (C.P. Verghese 2007: 58, 86). Researchers made several attempts to select a minimum of vocabulary, being in fact early word frequency lists (*ibid.*).

E.L. Thorndike’s *The Teacher’s Word Book*, published in 1921, was a list of 10,000 words from 41 different sources, making up a corpus<sup>3</sup> of approximately 4.5 million words. The reference was aimed to help novice as well as experienced teachers choose the most important words to be taught at a certain grade. In Thorndike’s words, *The Word Book* was “a real treasure-house of help [to teachers] in a form readily available” (E.L. Thorndike 1921: V). The scholar later authored or co-authored a few other works of the kind, including *The Teacher’s Word Book of 20000 Words*, published in 1931, and *The Teacher’s Word Book of 30000 Words*, released in 1941.

Another step was made by Michael West in 1953. The scholar published *A General Service List of English Words* – a frequency-based reference work that defined a limited list of 2,000 English words (cf. C.P. Verghese 2007: 87). The idea was to present English vocabulary of greatest coverage to learners and teachers of English as a second language to facilitate communication (knowledge of all items

---

<sup>2</sup> As has been mentioned in the introduction above, the term ‘basic vocabulary’, as used by R.R.K. Hartmann and G. James, is equivalent to the term ‘lexical minimum’ discussed here.

<sup>3</sup> Naturally, it did not constitute a ‘corpus’ in the modern meaning of the word (cf. L. Bowker/ J. Pearson 2002: 9).

was claimed to ensure the understanding of 80-95% of written texts and colloquial speech). Attempts to create ever more reliable core vocabulary lists still continue today. New lists created by present-day researchers are based on electronic corpora, extracted with the use of elaborate statistics and presenting ever greater coverage. Two such examples include the *New General Service List* (published alongside other similar lists, such as: *New Academic Word List*, *Business Service List* or *TOEIC Word List*) by Charlie Browne, Brent Culligan and Joseph Phillips<sup>4</sup>, and the *New General Service List* by Vaclav Brezina and Dana Gablasova (V. Brezina/D. Gablasova 2013).

As C.P. Verghese points out, frequency-based vocabulary lists are still useful in the process of textbook creation, providing adequate grading of the vocabulary. They are also a diagnostic tool, setting out the range of 1,500 to 2,000 words as “the minimum productive vocabulary for the student to write or speak fluently and comfortably on non-specialised subjects” (ibid.), providing a solid basis for a subsequent self-study stage.

H. Kaczmarek (2006) states that in view of the complexity of the teaching/learning process, the dominant role should be attributed not so much to the teacher, but rather to the appropriately prepared textbooks/ coursebooks and other didactic materials, developed on the basis of glottodidactic methods that incorporate relevant results of glottodidactic studies. The researcher adds that the need for the optimisation of the glottodidactic process requires an appropriately prepared lexical material, and the so called lexical minimums, or basic vocabulary, provide a specific base for the construction of textbooks (H. Kaczmarek 2006: 28). Therefore, a goal-oriented lexical minimum becomes a more universal reference tool whose aim is to directly (lexical minimum used as a self-study guide) and indirectly (lexical minimum as a source for textbooks, teaching materials, etc.) enhance the communicative competence of its users. Naturally, the variables necessitate appropriate grading of the material and possibly different content presentation, resulting in a series of lexical minimums (rather than one wordlist characteristic of the early works).

S. Grucza (2004) takes a broader approach to the term in question, stating that the lexical minimum<sup>5</sup> is in fact a linguistic term which refers to a vocabulary stock of a particular language selected (limited) according to some criteria. He adds that it can also be defined as a set of vocabulary that enables a particular group of speakers/listeners to communicate within their basic communication needs (S. Grucza 2004: 256). The definition presented by the scholar has far-reaching consequences for the understanding of the concept<sup>6</sup>. By agreeing that ‘lexical minimum’ is a linguistic term, we acknowledge that it belongs to the realm of linguistics, i.e. a study of language understood as an inherent property of a human

---

<sup>4</sup> Available from: <<http://www.newgeneralservicelist.org/>>, last accessed on 31.01.2017.

<sup>5</sup> As a matter of fact S. Grucza (2004) uses the Polish term ‘slownictwo minimum’ (en. ‘minimum vocabulary’) interchangeably with the term ‘minimum leksykalne’ (en. lexical minimum) in his work.

<sup>6</sup> Later in his paper S. Grucza uses the term in its narrow sense (cf. S. Grucza 2004: 257).

being (F. Grucza 1983: 296-297, 301-302). From this perspective, lexical minimum can be regarded as a part of language competence that encompasses vocabulary stock acquired by an individual (i.e. their lexical competence making up their mental lexicon), which enables him/her to communicate at a certain proficiency level. Therefore, similarly to language competence (F. Grucza 1989: 38), the notion of lexical minimum as a mental capacity should be regarded as a gradable and relative feature of human beings, definitely not limited to basic communicative needs only or understood as a universal feature across a language society. From these summary linguistic considerations it follows that distinguishing the various levels of lexical minimums (elementary, pre-intermediate, intermediate, etc.), as suggested by some authors<sup>7</sup>, ought to be seen as a convenient practical solution to compiling a progressive set of glottodidactic materials, rather than a true representation of the level of lexical competence acquired by students at a specific educational stage. In fact, such competence cannot be regarded as a discontinuous one-dimensional entity, and one must also take notice of both the fairly complicated network of interrelated notions and words it creates, and each individual's narrower (linguistic/communicative/cultural) and broader (world) knowledge, as well as skills. It further transpires that the only means of determining a level of lexical competence of a person, and, therefore, establishing whether they have attained a predefined level (certain-level lexical minimum)<sup>8</sup>, is through the exponents of their language, i.e. words and word combinations (texts). However, one needs to remember that it is necessary to account for both active and passive vocabulary, which translates into assessment of all traditional language skills (listening, reading, speaking, writing). Also, any evaluation of the lexical competence should be seen as relative in that it is undertaken against the competence of other listeners/speakers within a given language community. Besides, there is no objective threshold as any lexical minimum is a result of quantitative and qualitative analyses of a number of texts, representing language exponents of a number of individuals. Another variable here is the proper determination of the *tertium comparationis*, which might differ depending on the pragmatic aim of the communication act. With so many variables to be taken into account, some of which are hardly quantifiable, it needs to be recognised that establishment of any individual or polylectal lexical minimum should be regarded as an approximation. In fact, no competence can ever be fully evaluated by objective measures. S. Grucza rightly notices that lexical minimum is not a fixed set of vocabulary stock that one can establish in an absolute manner (S. Grucza 2004: 257).

The considerations presented above certainly call for a clear distinction between lexical minimum as a portion of someone's language competence, and a lexical minimum as a specific reference work (see below). This distinction seems to be of primary importance where lexical minimum is used as a benchmark for formal evaluation of language (vocabulary) proficiency at various levels. In such cases,

---

<sup>7</sup> None of whom overtly mentions, or refers to, the notion of the lexical minimum as a cognitive inherent property of an individual.

<sup>8</sup> Obviously, it cannot be done in absolute terms.

lexical minimums need to be understood twofold, both as a certain level of an individual's vocabulary competence (therefore sometimes passive and active vocabulary is referred to in the curricula outlining specific language competences), and as a specific reference tool setting the vocabulary standard (benchmark) at a given level – the latter also understood as a test tool that can be used to measure the former (such as, for example, the series *Lexical minimum of Russian as a foreign language*<sup>9</sup>). The inherent relative nature of lexical minimum may considerably limit its application as a testing tool to a specific language proficiency level, types of users, communicative situations, etc., for which the reference work has been designed. Still, for most educators vocabulary testing is believed to be an objective and quick measure of students' language competence (E. Jendrych 2013: 80).

As has already been mentioned, lexical minimum as a tool in glottodidactics can resemble a dictionary, in particular a mono- or a bilingual glossary. Such dictionaries are sometimes referred to as the 'minimum dictionary'<sup>10</sup>. S. Grucza uses the term with reference to a dictionary that presents a lexical minimum. Similarly to lexical minimums, minimum dictionaries are never definite and have to be viewed from a relative perspective, mainly on account of the fact that they are "functional dictionaries, whose structure is determined by language needs of their users" (S. Grucza 2004: 257, translation mine – M.Ł.).

At this point it is necessary to draw a distinction between the two notions/terms that are used interchangeably, namely "lexical minimum" (understood as a type of a reference work;  $\approx$  minimum dictionary<sup>11</sup>) and "frequency dictionary". The *Dictionary of Lexicography* offers the following definition of the latter:

**frequency dictionary** A type of reference work which provides information about the frequency of linguistic units, especially of the basic vocabulary. Work on frequency dictionaries has been motivated by, and beneficial to, various fields such as the design of shorthand systems [...], the teaching of reading to children [...], the selection of basic vocabularies for foreign-language teaching [...]. One particularly fruitful application of the frequency approach is the design and improvement of the learner's dictionary (R.R.K. Hartmann/ G. James 2002: 59; cross-references removed by M.Ł.).

Despite a clear indication that a frequency dictionary is strongly linked to the glottodidactic context, and likewise the basic vocabulary (lexical minimum), it needs to be emphasised that the two reference works differ to a great extent. First and foremost, the construction of frequency dictionary employs exclusively quantitative

---

<sup>9</sup> Available for various levels, i.e. basic, first and second certification level (see e.g. N.P. Andrjūsina/ T.V. Kozlova 2006).

<sup>10</sup> In lexicography, the term 'minimum dictionary' is sometimes used as a quantitative benchmark in dictionary typology to mean a reference work of certain (small) size (= number of entries), not necessarily compiled in accordance with lexical minimum criteria.

<sup>11</sup> Clearly, there may be some points of divergence between the lexical minimum as a reference work/tool and the minimum dictionary, the latter incorporating the lexicographic method in its construction, and possibly assuming a different form than the former. These aspects will be elaborated in greater detail in the upcoming paper. Here, the author assumes a definition of the minimum dictionary proposed by S. Grucza (2004: 257).

methods, with no qualitative (evaluative) compilation method envisaged. Second, its main role consists in delivering statistical data, and therefore any transformation of such lists into relevant glottodidactic means/tools/materials or its use in the construction of a lexical minimum should envisage appropriate (glottodidactic, pragmatic, etc.) evaluation and adaptation. Pavel M. Alekseev (2005) points out that

its [=frequency dictionary] main feature consists in its capacity to provide information on the frequency of words [...]. The dictionary entries may be organized either in the regressive order of their frequencies, starting with the most frequent one, or in the alphabetic order. Often the dictionary may consist of two lists: the alphabetic and frequency ones (P.M. Alekseev 2005: 312).

From this perspective, if a work is purported to be a frequency dictionary, but does contain a qualitative element, it should be regarded as a type of a hybrid dictionary. While it is clear that frequency lists/dictionaries may provide a valuable resource for lexical minimums (in fact to a greater extent than in the case of terminological minimums), they should be considered only a starting point in the lexical minimum compilation.

## 2. Lexical minimum construction criteria

Researchers generally agree that the construction of a lexical minimum (= selection of the vocabulary), similarly to lexicographic projects, should embrace both quantitative and qualitative methods. One of the most compelling statements in support of the view that the significance of word frequency should not be overestimated comes from Robert de Beaugrande and Wolfgang Dressler (1981), who write:

The correlation between information value and any absolute frequency of the occurrence is certainly not straightforward. The actual effects of an occurrence in its context can always be upgraded or downgraded via appropriately planned settings. Hence, frequency is useful, especially if computed for a very large set of texts, but neither sufficient nor reliable (R. de Beaugrande/ W. Dressler 1981: 150)<sup>12</sup>.

Compilers of frequency lists realised the deficiencies of the frequency measures quite early. The first word frequency lists, created for second-language teaching purposes, such as the one by E.L. Thorndike, soon proved to be inadequate, and it became clear that also other measures, such as the “usefulness of a word at a particular stage in the learning of a language also has to be taken into account” (C.P. Verghese 2007: 58). Michael West (1953: IX-X) in his *General Service List* provides such additional ‘qualitative’ criteria as:

- the ease of learning,
- necessity,
- cover,
- stylistic and emotional neutrality.

---

<sup>12</sup> It is worth mentioning, however, that some scholars have convincingly proven that statistical methods are alone highly reliable in a generation of core vocabulary lists (cf. V. Brezina/ D. Gablasowa 2013) – see the end of the current section.

Questionable as the criteria may appear to quantitative linguists (see below), West's proposal is important as it goes beyond statistics and takes into consideration the glottodidactic element (ease of learning), the communicative element (cover) and the semantic value (necessity)<sup>13</sup>. An interesting example of a total departure from frequency parameter comes from Charles Kay Ogden, who authored the 1930 *Basic English* – a work of 850 words compiled on the basis of logical and philosophical principles (the aim was to create a simplified international auxiliary language). The dictionary is divided into the following sections: names of things (600 words, of which 400 are common names, or 'general things', and 200 are names of particular objects, or 'picturable things'), names of qualities (150 words – qualifiers, i.e. adjectives and adverbs, as well as words denoting 'general quality' and 'opposite qualities'), and words denoting operations, grouped into verbs, pronouns, prepositions, determiners, question words, conjunctions, numbers (150 words)<sup>14</sup>.

Currently, scholars generally advocate a combination of quantitative and qualitative proposals. Hanna Kaczmarek lists the following criteria for lexical minimum vocabulary selection:

- frequency;
- stylistic neutrality;
- communicative value;
- word-formation productivity (H. Kaczmarek 2006: 29).

The researcher points out that from the perspective of the frequency parameter alone, lexical minimums can be constructed on the basis of appropriately compiled corpora or as a compilation of already-existing frequency-based lexical lists<sup>15</sup>. Yet, H. Kaczmarek rightly emphasises that the use of the frequency parameter alone must be considered controversial, because each extracted lexeme is representative of the corpus it has been obtained from<sup>16</sup>, and therefore other criteria, such as the communicative and pragmatic relevance, as well as informative and semantic value of the lexemes should also be taken into account. Other scholars seem to share the view that the frequency criterion should not be used in isolation from other variables. Mateusz Warchał (2013), for example, mentions the value of association tests<sup>17</sup> in the construction of a lexical minimum, the aim of which is to limit language interference in the process of learning a language (M. Warchał 2013: 294).

Also S. Grucza defines three methodological steps in the determination of the lexical minimum. These include:

---

<sup>13</sup> Words that were of low frequency, but were 'necessary' from the semantic point of view to define other, higher-ranking words on the list – had been included in the list.

<sup>14</sup> <<http://ogden.basic-english.org/>>, last accessed on 31.01.2017.

<sup>15</sup> Such lists may take the form of frequency dictionaries.

<sup>16</sup> The issue of representativeness represents a long-standing problem in corpus linguistics (see, e.g. D. Biber 1993).

<sup>17</sup> The association test is one of the methods for teaching vocabulary, which aims to encourage learners to "draw connections between what they do know and words they encounter they do not know" (J. Kruidenier 2002: 67).

1. Compilation of a list of the most frequent lexemes in a language. The procedure relies on statistical studies of lexical frequencies, of both written and spoken texts. In the next step the list is arranged and organised in a desired manner.

2. Extension of the lexical material by adding so-called functional lexemes (such as conjunctions and prepositions).

3. Pragmatic extension and supplementation, at which stage lexical gaps are identified and filled on the basis of such criteria as: lexical relationship, communicative usefulness, communicative necessity, communicative effectiveness and glottodidactic relevance (S. Gruzca 2004: 256).

As has already been mentioned, the ‘qualitative’ approach, for example that of Michael West’s, is criticised by corpus/quantitative linguists, who denounce it as fraught with subjectivism when compared to modern linguistic research (V. Brezina/D. Gablasowa 2013: 3). The scholars suggest a purely quantitative approach to the extraction of basic vocabulary, claiming that while they generally agree with West’s position that “frequency alone is not a reliable measure for selecting words important for learners”, instead of applying any qualitative (subjective) criteria, as West did, they use a combination of three quantitative measures: frequency, dispersion and distribution across language corpora to successfully create a comprehensive wordlist (*ibid.*). Interestingly enough, scholars applying lexicostatistical methods aim to quantify the entire process of selection and compilation as well as undertake quantitative assessment of a minimum ‘usefulness’ (coverage) by defining the percentage range of the degree to which a learner should be able to understand or produce texts. Undeniably, there are some lexical minimum compilation parameters, such as those related to users themselves, that are difficult to measure by means of objective (quantitative) means, and yet are necessary for adequate construction of the final product. Having said that, the quantitative methods seem to be more applicable in the case of the compilation of glottodidactic-oriented general lexical minimums than in the construction of a terminological minimum of similar characteristics.

### **3. Resources for lexical minimum**

The quantitative-qualitative approach adopted for the construction of a lexical minimum is in fact derived from modern corpus-based lexicographic methodology. It goes without saying that corpora have become an important source in linguistics, language teaching/learning, lexicography, psychology, translation studies and translation practice, terminology and terminography, among other areas of research and practical activity. In particular, (meta)lexicographers emphasise the need to focus on corpus quality in the first place, and – during the construction phase – on word distribution (P. Hanks 2003: 58), syntax patterns, collocational value, context, etc. One of the major advantages of the data obtained from corpora is their higher degree of objectivity. However, corpus methods are not without its cons, one of the major flaws being that existing corpora are largely non-representative, which stems from their imbalance in terms of the inclusion of only selected text types or genres (in unnatural proportions), or the bias resulting from overrepresentation of written

texts in corpora (cf. D. Biber 1993: 247-248). The quality of the resource depends on the quality of texts that constitute the corpus. Accordingly, appropriate procedure of corpus compilation may ensure the required parameters of the corpus.

One note needs to be made here in reference to an observed inclination towards choosing textbooks/ coursebooks as primary sources for a lexical minimum compilation. From methodological perspective this is an erroneous solution (procedure), insofar as (a) the lexical items present in textbooks had already been selected during their construction. Accordingly, the authors of lexical minimums in fact present a second-hand selection. Besides gathering the entire vocabulary stock of a textbook /textbooks, no new vocabulary is introduced. There is every likelihood that the statistics will be flawed anyway, as there is usually thematic progression in textbooks/coursebooks and content words, including terminology, might not be repeated frequently across the textbook; in the same vein, (b) statistical analyses of textbook-based 'corpus' aiming to establish the importance of lexical units and the sequence of their introduction in the didactic process are overtly flawed, on account of their non-representativeness against naturally occurring texts/utterances (they are representative of the corpus compiled, for example a corpus of textbooks in question). Moreover, (c) the quality of the original (textbook) vocabulary selection procedure, including the original sources, often remains unknown to the author of the lexical minimum.

Accordingly, the minimums created on the basis of textbooks: (i) should not be used for the creation of new textbooks (as is often suggested by various authors), since they do not provide reliable data or new material; (ii) must not be used for the evaluation of other textbooks: such analyses may amount to a mere comparison of vocabulary content of the textbooks considered<sup>18</sup>; (iii) can be used for competence evaluation only to the extent to which the assessment concerns the textbooks used for the construction of the minimum at a particular language level. A separate case here concerns the up-to-dateness of the resources themselves; unless there is another research purpose, such as an analysis of the lexical change in textbooks, evaluation of vocabulary content of present day textbooks against lexical minimums published some time ago, even as recently as 10-15 years ago, amounts to a methodological oversight (cf. H. Kaczmarek 2006: 30-31)<sup>19</sup>.

Notwithstanding some reservations related to the objectivity of corpus data analyses in general, corpora are still the preferred primary resource for the construction of lexical minimums, providing easy access to quantitative data and reliable authentic material. It seems that Adam Kilgariff's words continue to be

---

<sup>18</sup> Surprisingly, such procedures are undertaken (H. Kaczmarek 2006: 31-33), despite clear reservations of their authors as to the representativeness of corpora for the compilation of a lexical minimum (ibid: 29).

<sup>19</sup> H. Kaczmarek, for example, evaluates modern textbooks for teaching German as a foreign language (published in 2002, 2003 and 2004) against a lexical minimum dating back to 1982. Needless to say, the minimum had been compiled on the basis of similar coursebooks, published in Poland and prepared by Polish authors (H. Kaczmarek 2006: 31; P. Chmiel 1982).

highly pertinent here: “Although there are limitations on corpus design, and although we can never entirely escape subjective interpretations, corpora allow a “degree of objectivity” about some central questions, “where before we could only speculate” (A. Kilgarriff 1997: 137, as cited in M. Stubbs 2004: 108). Possibly, a do-it-yourself corpus constructed with the use of cyclic compilation procedure might meet the quality requirements, in particular a satisfactory degree of representativeness (M.W. Bauer/ B. Aarts 2003: 29; cf. M. Łukasik 2014: 79-80).

The data (and content) that can be obtained from a corpus, relevant from the perspective of a lexical minimum, include:

- word frequency (raw, standardised, etc.);
- word/phrase distribution across the texts of the corpus;
- n-gram statistics;
- collocations statistics;
- keywords, including terminology (against a reference corpus);
- contexts (concordances);

and, in the case of annotated corpora, searches or extraction of:

- words according to their part-of-speech parameter;
- words according to their semantic value;
- syntactic patterns and their distribution;
- pragmatic patterns and their distribution;
- stylistic characteristics;
- learners’ error distribution, etc. (see: G. Leech 2005; T. McEnery 2003: 453q459; P. Rayson et al. 2004; S. Granger 2003).

Surprisingly, the list of possible analyses with the use of corpora, most of which are of quantitative nature, is relatively long and much more extensive than a presentation of the language material in a frequency dictionary. However, in view of the corpus quality issues, including corpus annotation accuracy rates<sup>20</sup>, parameters of lexical units other than their statistical measure in a corpus are called upon. The characteristics already mentioned above include the pragmatic relevance, informative and semantic value, communicative value/usefulness or stylistic markedness. The dilemma here lies in the fact that no author has provided any clear methodological guide as to how to apply the parameters in question (for example, how to measure usefulness or relevance, or to comprehensively account for communicative situations). As is often the case, qualitative analyses rely on subjective way of reasoning and experience of the researcher/compiler, and possibly on intuitive non-quantifiable and sometimes indefinable methods, which nevertheless lead to some objective conclusions (cf. C. Ratner 2002). This might be the reality of genuine language studies, in which a high degree of individualism and

---

<sup>20</sup> In fact, the automatic annotation accuracy rates are fairly high, yet might still be insufficient for some qualitative studies. For example, the USAS (UCREL Semantic Analysis System) achieves a 91% precision rate when compared to manually tagged test corpus (P. Rayson et al. 2004: 7), while Part-of-Speech taggers go as high as 97.3% (Ch.D. Manning 2011: 171).

relativism may (and most probably will) influence results obtained by quantitative means. Still, qualitative discussions and lexical minimum supplementations can be supported by:

- (i) corpus-based evidence (e.g. through context analysis);
- (ii) other electronic linguistic repositories, such as Princeton's WordNet, i.e. a semantics-based reference system resembling ontologies (P. Hanks 2003: 59);
- (iii) genuine ontologies;
- (iv) other reference works, such as dictionaries, encyclopaedia, grammars, provided their quality can be attested;
- (v) results of relevant linguistic studies.

#### 4. Users' needs

Undoubtedly, any reference materials should be tailored to specific needs of users, providing for specific compilation criteria. This equally concerns lexical minimums, regardless of whether these will be used by coursebook authors or end users themselves as a ready-to-use self-study reference. In practice, the authors of lexical minimums, before commencing any compilation work, should define and characterise the user group, as well as determine the thematic scope suitable for the group (H. Kaczmarek 2006: 29). Mateusz Warchał points out that each lexical minimum should be addressed to specific users, taking into account their age and language/communicative competence as well as the conditions in which the teaching/learning processes take place (M. Warchał 2013: 492).

If lexical minimums, and likewise minimum dictionaries, are user-centred functional tools, as suggested by S. Grucza (2004: 257), then the most appropriate approach to user studies is offered by the Function Theory of Lexicography, or the Theory of Dictionary Functions, developed by Henning Bergenholtz and Sven Tarp (cf. H. Bergenholtz/ S. Tarp 2002; S. Tarp 2008). In their view, there are four areas that the lexicographer needs to take interest in: (1) potential users, (2) user situation, (3) user needs, and (4) dictionary assistance (S. Tarp 2008: 140).

Sven Tarp and Pedro A. Fuertes-Olivera identify a set of lexicographically-relevant user characteristics, including these connected with their dictionary skills. By way of example, the following compilation of questions might prove useful in the design of a user-friendly lexical minimum for didactic purposes.

1. What is the mother tongue of the user/learner?
2. What is the learner's proficiency level in their mother tongue?
3. What is the learner's proficiency level in the foreign language in question?
4. What is the learner's level of cultural knowledge?
5. What is the learner's level of encyclopaedic knowledge?
6. What is the learner's knowledge of a specific discipline (layman, semi-expert, expert)<sup>21</sup>?
7. What is the learner's motivation to learn the foreign language in question?

---

<sup>21</sup> A parameter specifically important for the construction of the terminological minimum.

8. What does the foreign-language learning process look like (self-study/classroom setting/ e-learning)?
9. What is the broader context in which the foreign language instruction/learning takes place (e.g. Does the learning take place in an L2 community?)?
10. Is the learner exposed to their mother tongue during the learning process?
11. Does the learner use a specific textbook or is exposed to a specific didactic system?
12. Does the learner use (is exposed to) a specific didactic method?
13. Is the learning process related to a specific subject?
14. What is the user's general experience of lexicographical consultations?
15. What specific experience do the users have with a specific dictionary (e.g. a minimum dictionary)?
16. Does the user possess a device with access to the Internet? etc. (based on S. Tarp 2008: 168; P.A. Fuertes-Olivera/ S. Tarp 2014: 48-50).

User situations are connected with extralexigraphic language tasks that the user might encounter and in which they might need to consult a reference work, for example a lexical minimum. Such situations evoke the four traditional reception and production skills, namely: reading, listening, speaking and writing, as well as translation. In a fine-grained analysis of specific situations, detailed tasks might be considered, for example proofreading or text correction. Alternatively, the user might need to acquire encyclopaedic knowledge to solve a specific problem (S. Tarp 2013: 464). In our case, this type of information influences the decision of whether to compile a passive (for text reception) or an active (for text production) lexical minimum.

User needs arise as a result of specific users finding themselves in specific situations. As S. Tarp puts it “[n]eeds must be determined based on the user characteristics that are relevant in connection with each type of user situation” (ibid.). An example of a specific user positioned in a specific situation might be as follows: an intermediate adult learner of English as a foreign language, a specialist in his/her field, who may need to read a paper in a medical journal. The primary need may be connected with obtaining particular information from the article, while the lexicographic reference work might be indispensable for decoding mainly general words in L2, and probably only some of the L2 terms – through their equivalents in L1 (not necessarily through definitions). This calls for a specific set of parameters of the lexical (terminological) minimum to be designed specifically for the user (or a group of similar users).

From the perspective of our current discussion, the user research proposal outlined above seems to satisfactorily fill an important methodological gap in the lexical minimum compilation procedure. User research, coupled with other multilayer studies relevant to the functioning of a specific glottodidactic system, as well as quantitative and qualitative analyses of the source materials, provide the necessary data that influence the final shape of the lexical minimum.

## 5. The final product

The final set of parameters of the lexical minimum as a specific reference work will depend on a number of variables. First and foremost, the fundamental function of the minimum will greatly influence the final design criteria (e.g. a minimum for testing lexical competence vs. a minimum for self-study). Secondly, the specific groups of minimum users and their characteristics, such as their language proficiency, experience, general knowledge, methodological background, degree of expertise, etc., will have to be reflected in the final composition of a minimum (e.g. textbook authors vs. teachers vs. students). Those specific types of users may have different expectations (needs) with regard to the lexical minimum (e.g. textbook authors requiring source materials vs. students at a certain language proficiency level).

Another issue is the choice of source materials for the minimum. Any decisions at this stage may put a burden of subjectivity on the analysis stage and likewise the final product. Moreover, the quality of primary texts will influence quantitative and qualitative analyses. A problem lexical minimum compilers can encounter is the lack of resources of required value, and therefore one precaution must be observed: in view of lack of primary sources of specific characteristics, no shortcuts are allowed. The quality that should not be compromised at this stage is the representativeness of the material. Only a quality corpus can ensure that the lexical minimum exhibit features and content of relevant and acceptable value in a particular situation.

Undeniably, corpus analysis procedures will be of utmost importance here, including inclusion/exclusion criteria for the lexical items (lexemes). If later additions are necessary (function words, related vocabulary, culture-specific words/terms/names, etc.), the procedure should follow a methodology that guarantees high quality and coherence of the end product. It may be necessary to use the existing large corpora, such as the BNC or COCA, or to compile an auxiliary corpus, and analyse such resources with the use of modern lexical profiling software, such as the Sketch Engine (A. Kilgarriff/ M. Rundell 2002).

S. Grucza is probably right when he writes that it is impossible to empirically and definitely establish a lexical minimum, as it is not a fixed set of words (S. Grucza 2004: 257). Therefore, it is impossible to establish a fixed number of lexemes to be included in a lexical minimum, given its changeability depending on the users and user needs as well as its dynamic nature as a reference work reflecting authentic language use (= vocabulary range). However, discussions of the first half of the previous century still reverberate amongst scholars nowadays, and the numbers ranging from 800 to 2,000 or more words (to be included in a minimum) are frequently cited (cf. H. Kaczmarek 2006: 29).

Last but not least, a lexical minimum as a specific reference source does not have to assume the shape of a glossary or dictionary. On the contrary, modern digital methods envisage modern solutions to lexicographic products, which are more and more often incorporated in other lingware. Lexicographic products do not have to be run separately as standalone programs, but can exist as plug-ins or parallel background applications. A lexical minimum could in fact primarily become a

dynamic (= extended) universal database, developed and updated in a cyclic procedure aiming at enhancing its quality (cf. the cyclic corpus methods: M.W. Bauer/ B. Aarts 2003: 29). Such database would enable generation of user-defined lexical minimums on demand. Alternatively, an interactive algorithm working in an interconnected environment could gather data and deliver the content as required in real-time with no database needed. As a matter of fact, modern tools in e-lexicography aim at just such automatic or semi-automatic systems (P.A. Fuertes-Olivera/ S. Tarp 2014: 101).

In any case, the structure of the lexical minimum as a final product should be kept simple, so as to deliver the kind and amount of information that is exactly required by the users in a certain situation. In no way should such work turn into an all-in-one dictionary. However, this apparent simplicity must be supported by sound methodology and solid compilation procedures.

## 6. Conclusions

The paper has attempted to demonstrate that lexical minimum should not be considered as a one-dimensional notion associated with a specific type of a reference work, but ought to be viewed as a multi-dimensional concept, exhibiting multifaceted characteristics. Taking its ontological status into account, it is primarily a quantum of a person's language competence, related to his/her internalised vocabulary, passive and/or active, i.e. their lexical competence. It is also a relative measure of such lexical competence, usually implemented with the use of a lexical minimum as a specific testing tool. Based on a selection of a number of texts, which *de facto* comprise externalised strings of vocabulary items and make up a corpus, lexical minimum can fulfil a role of a representative language resource, important from the point of view of glottodidactics; it can be used as both a source for textbooks and other didactic materials, and a self-study reference. Most importantly, however, lexical minimum as a reference is a dynamic object, whose content and structure ought to be adjusted to the needs of users.

Accordingly, lexical minimum and the minimum dictionary representing the lexical minimum do not form a homogeneous category; in fact, lexical minimums will vary horizontally, depending on the scope of the vocabulary presented (including terminology, which – provided specific compilation criteria are met – will give rise to terminological minimum), and vertically, exhibiting variance as regards conceptual complexity, pragmatic/communicative value, etc. Other selection criteria can also be used, such as, for example, the prototypicality of notions. Definitely, the compilation of the lexical minimum should no longer be associated with lexicostatistical methods only, but also with the qualitative approach that would involve, *inter alia*, the communicative as well as extralinguistic variables. As a matter of fact, the qualitative element makes lexical minimum – traditional or electronic – different from a text corpus, the latter requiring further analyses and data processing, and these in turn – a certain degree of language competence. There are also new technologies on the horizon that will most certainly part from traditional approach to lexicographic work compilation.

## References

- Alekseev, P.M. (2005), *Frequency dictionaries*. In: R. Köhler/ G. Altmann/ R.G. Piotrowski (eds.), *Quantitative Linguistics. An International Handbook* (HSK 27). Berlin/ New York, 312–324.
- Andrjūsina N.P./ T.V. Kozlova (2006), *Leksičeskij minimum po ruskomu jazyku kak inostrannomu. Bazovyj uroven'*. Sankt-Peterburg/ Moskva/ Zlatoust.
- Bauer M.W./ B. Aarts (2003), *Corpus Construction: a Principle for Qualitative Data Collection*, In: M.W. Bauer/ G. Gaskell (eds.), *Qualitative Researching with Text, Image and Sound. A Practical Handbook*. London / New Oaks/ New Delhi.
- de Beaugrande, R./ W. Dressler (1981), *Introduction to Text Linguistics*. London/ New York.
- Bergenholtz, H./ S. Tarp (2002), *Die modernelexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen*. In: "Lexicographica" 18, 253–263.
- Biber, D. (1993), *Representativeness in Corpus Design*. In: "Literary and Linguistic Computing" 8 (4), 243–257.
- Bowker L./ J. Pearson (2002), *Working with Specialized Language. A Practical Guide to Using Corpora*. London.
- Brezina, V./ D. Gablasova (2013), *Is There a Core General Vocabulary? Introducing the New General Service List*. In: "Applied Linguistics" 2013, 1–23.
- Chmiel, P. (1982), *Minimum leksykalne w nauczaniu języka niemieckiego*. Zielona Góra.
- Fuertes-Olivera, P.A./ S. Tarp (2014), *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin/ Boston.
- Granger, S. (2003), *Error-Tagged Learner Corpora and CALL: A Promising Synergy*. In: "CALICO Journal" 20 (3), 465–480.
- Grucza, F. (1983), *Zagadnienia metalingwistyki. Lingwistyka – jej przedmiot, lingwistyka stosowana*. Warszawa.
- Grucza, F. (1989), *Język a kultura, bilingwizm a bikulturyzm: lingwistyczne i glottodydaktyczne aspekty interlingwalnych i interkulturowych różnic oraz zbieżności*. In: F. Grucza (ed.), *Bilingwizm, bikulturyzm, implikacje glottodydaktyczne*. Materiały z XII Sympozjum zorganizowanego przez Instytut Lingwistyki Stosowanej UW, Zaborów, 18-20 września 1986 r. Warszawa, 9–49.
- Grucza, S. (2004), *Dydaktyka translacji. Terminologiczna preparacja dydaktycznych tekstów specjalistycznych*. In: J. Lewandowski (ed.), *Języki specjalistyczne 4. Leksykografia terminologiczna – teoria i praktyka*. Warszawa, 243–267.
- Hanks, P. (2003), *Lexicography*. In: R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford, 48–69.
- Hartmann, R.R.K./ G. James (2002), *Dictionary of Lexicography*. London/ New York.
- Jendrych, E. (2013), *Przydatność minimum terminologicznego angielskiego języka biznesu w ocenie jego przyszłych użytkowników*. In: "Linguodidactica" XVI, 73–93.
- Kaczmarek, H. (2006), *Minimum leksykalne a orientacja komunikacyjna w nauczaniu języka obcego. Kilka uwag na przykładzie niemieckiego jako języka*

- obcego*. In: "Prace Naukowe Akademii im. Jana Długosza w Częstochowie: Studia Neofilologiczne" T. 5, 27–35.
- Kilgarriff, A. (1997), Putting frequencies in the dictionary. In: "International Journal of Lexicography", 10(2), 135–155.
- Kilgarriff, A./ M. Rundell (2002), *Lexical profiling software and its lexicographic applications: A case study*. In: "Proceedings from EURALEX 2002". Copenhagen, 807–811.
- Kruidenier, J. (2002), *Research-Based Principles for Adult Basic Education. Reading Instruction*. Washington.
- Leech, G. (2005), *Adding Linguistic Annotation*. In: M. Wynne (ed.), *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford, 17–29.
- Łukasik, M. (2014), *Compiling a Corpus for Terminographic Purposes*, In: "Komunikacja Specjalistyczna" 7, 71–83.
- Manning, Ch.D. (2011), *Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?* In: A.F. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*. Berlin/ Heidelberg, 171–189.
- McEnery, T. (2003), *Corpus Linguistics*. In: R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford, 448–463.
- Rayson, P./ D. Archer/ S. Piao/ T. McEnery, (2004), *The UCREL semantic analysis system*. In: "Proceedings of the Beyond Named Entity Recognition Semantic Labeling for NLP Tasks Workshop". Lisbon, 7–12.
- Ratner, C. (2002), *Subjectivity and Objectivity in Qualitative Methodology*. In: "Forum: Qualitative Social Research" 3 (3). (URL <http://www.qualitative-research.net/index.php/fqs/article/view/829>). [Accessed 31.01.2017].
- Slater, J.G. (1999), *Russell's Conception of Philosophy*. In: A.D. Irvine (ed.), *Bertrand Russell. Critical Assessments*. Vol. 3: Language, Knowledge and the World. London/ New York, 1–20.
- Stubbs, M. (2004), *Language Corpora*. In: A. Davies/ C. Elder (eds.), *The Handbook of Applied Linguistics*. Oxford, 106–132.
- Tarp, S. (2008), *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen.
- Tarp, S. (2013), *Lexicographic functions*. In: R.H. Gouws/ U. Heid / W. Schweickard / H.E. Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume (HSK 5.4)*. Berlin/ Boston, 460–468.
- Thorndike, E.L. (1921), *The Teacher's Word Book*. New York City.
- Vergheese, C.P. (2007), *Teaching English As a Second Language*. New Dehli.
- Warchał, M. (2013), *Minimum leksykalne w edukacji dzieci - użycie korpusu językowego w nauczaniu interkulturowym*. In: A. Žele (ed.), *Družbena funkcijskost jezika (vidiki, merila, opredelitve) / The social functionality of language: (aspects, criteria, definitions)*, 32. Ljubljana, 491–496.
- West, M. (1953), *A General Service List of English Words*. London/ New York.
- Yaroshenko, O.L. (2014), *Principles of lexical minimum selection for teaching general technical English to engineering students*. In: "Advanced Education" 2, 104–110.