

New Pathways to Terminological Minimum

Marek ŁUKASIK

Pomeranian University in Słupsk

E-mail: marek.lukasik@apsl.edu.pl

Abstract: It is widely accepted that terminological minimum is a practical means used for the selection of specialist vocabulary to be taught at various proficiency levels in LSP courses. This popular understanding of the notion renders it similar to the concept of lexical minimum; from the didactic perspective, the two notions seem to largely overlap, being in fact vocabulary lists based mainly on the frequency parameter. However, this picture has largely obscured the true ontological status of the two concepts, and it was not until only recently that the notion of lexical minimum was defined in a more comprehensive way, bringing to light some of its most salient characteristics. Likewise, it seems necessary to re-define the notion of the terminological minimum, if only on account of the specific nature of the lexical unit included in them, i.e. the term. The current paper attempts to elucidate the notion in question and to provide a methodological framework for the compilation of a modern terminological minimum.

Keywords: corpus methodology, lexical competence, lexical minimum, LSP, minimum vocabulary, metalexigraphy, terminology, terminography

1.

In professional settings where the transfer of specialist information depends on specialised texts, the knowledge of terminology is of great importance. This stems from the fact that in a natural-language communication act specialised texts are the fundamental exponents of specialised knowledge, while individual terms in such texts can be regarded as signs denoting individual concepts, i.e. chunks of specialised knowledge. In some exceptional contexts, even one term can be treated as a complete text. From this perspective, the entirety of terms in a specialist text constitute its conceptual backbone.

From the sociolinguistic perspective, it is important to emphasise that knowledge of lexis can be a threshold for becoming a member of a discourse community (P. Mamet 2002: 144–145). Accordingly, the significance of terminology tools cannot be underestimated in the contemporary world, with ever more specialised fields and subfields of human activity emerging and developing their own code of communication. International specialist communication as well as the didactic process require the establishment of a solid conceptual and terminological base in order to avoid distortion of specialist knowledge and ensure uninterrupted flow of information.

Moreover, most translation tasks commissioned worldwide concern specialised (i.e. non-literary) texts; this observation equally concerns big organisations, such as the UN, WTO, WHO, NATO or the EU (see e.g. D. Cao/ X. Zhao 2008). Also, new translation areas, such as those connected with software localisation, require the expansion of the translator's terminological competence – a state of affairs that directly affects translator education. According to D. Gouadec, translation training is nowadays far more specialised than it used to be, with graduates entering the translation market aiming straight away “for more high-end of the market (both in terms of domain and in terms of IT skills needed)”, not “just translation” (D. Gouadec 2010: 148). The increase in the use of computer-assisted translation software, which often includes terminology management tools, and development of automated (machine) translation increase the role of terminology resources even further.

On a local scale, recent years have seen an increased emphasis on vocational training put by the Polish Ministry of Education, with numerous vocational courses reopening after years of stagnation and new vocational exams announced.¹ The school year of 2014/2015 was even dubbed the “Year of Professionals” and the Ministry has since organised various conferences devoted to vocational education. As of the school year of 2017/2018, a new vocational schooling scheme has been introduced, with a new type of two-level vocational school established. Parallel to an already existing secondary technical school, such schools educate future professionals.

One important constituent of the vocational training curriculum in Poland, common to all professions, is foreign vocational language course². Practically speaking, the Ministry³ has named as many as 209 professions, which represent roughly 70 different specialised areas, viz. also specialised languages⁴. Yet, there has been glaring scarcity of educational language materials, such as LSP textbooks/coursebooks, workbooks, specialised learner's dictionaries, etc. (in most languages), while the market has been slow to meet the demand (M. Pławska 2016: 7). One of the underlying reasons, important from the perspective of the present paper, might be the lack of resources for such materials, including structured (level-graded) collections of authentic primary texts, i.e. specialised corpora, vocabulary sets, specialised dictionaries, encyclopaedia, etc.

It seems that appropriately compiled terminological minimums could close the existing gap in graded specialised lexical resources, which are necessary in most, if not all, areas and contexts mentioned above. On top of its function as a lexical educational resource or a benchmark for the assessment of lexical proficiency at a certain level, terminological minimum is also a cognitive control tool as well as an

¹ Regulation of the Minister of Education, dated 17th February 2012 on Core Curriculum in Vocational Training, Journal of Laws (Dz.U.) item 184.

² Maintained in the new curriculum for the new scheme of vocational training (see: <http://men.gov.pl/wp-content/uploads/2017/03/pelna-tresc-rozporzadzenia-1.pdf>, [Accessed: 8.10.2017]).

³ Regulation of the Minister of Education, dated 13th December 2016 on the Classification of professions in vocational education, Journal of Laws (Dz.U.) item 2094.

⁴ According to a preliminary study undertaken by the Author of this paper.

important instrument in terminology management, terminology standardisation and development of modern digital terminology solutions.

Therefore this paper attempts to capture the major characteristics of terminological minimum (TM) understood as a multilayered construct, representing certain cognitive capacity, conceptual construct, a terminological resource and a (glotto)didactic means. Since TM shares some characteristics with lexical minimum (LM), the discussion starts with some important considerations referring to the latter term. Then, an attempt is made to encapsulate the notion of TM, and to present some of its most general attributes, considered in light of two different approaches to terminology. These considerations constitute a point of departure for the elaboration of TM design criteria and construction methods. Finally, the paper expounds on some specific features of TM as a resource/ reference work, and offers an outlook of the role and shape of TM in the digital age.

2.

While researching the concept of TM, the author has naturally come across the notion of LM. A thorough analysis has revealed that a number of issues inherent to LM are shared by TM, mainly because the two notions traditionally have a common denominator of being graded (i.e. level-based) vocabulary lists selected on the basis of some criteria, usually quantitative ones. Therefore, the author first focussed on a much broader term of LM, with respective findings presented in his previous paper (M. Łukasik 2017b). Those results have provided some significant insights into the nature of lexical resources that can prove useful for the current considerations on TM. In order to maintain a continuity of reasoning and expand essential points of the general discussion on the notion of LM and TM, some of the most important conclusions from the previous study, supplemented with a few additional comments, are briefly presented below.

2.1.

It first needs to be noted that the two terms, i.e. ‘lexical minimum’ and ‘terminological minimum’, seem to be more prevalent in the Slavic language area⁵, compared to English-writing scholars, who prefer other expressions, such as ‘core vocabulary/terminology’, ‘elementary vocabulary/terminology’, ‘basic vocabulary/terminology’, ‘minimum vocabulary/terminology’ or ‘core vocabulary/terminology’, each possibly exhibiting some degree of divergence from another. Equivalent terms are also found across other languages e.g. in German (‘terminologische Mindestwortschatz’, ‘Basiswortschatz’, ‘Minimal-/Mindestwortschatz’, ‘Grundwortschatz’, ‘zentrale Wortschatz’, among other proposals) or in French (‘vocabulaire essentiel’, ‘vocabulaire fondamental’) (see e.g. A. Dörre 2010, P. Kühn 1979, 1991, U. Schnörch 2002). In the present paper the terms of ‘lexical minimum’ and ‘terminological minimum’ will be used throughout.

⁵ Cf. the respective Russian terms ‘Лексический минимум’ (*Leksičeskij minimum*) and ‘Терминологический минимум’ (*Terminologičeskij minimum*).

2.2.

For some scholars, LM is a broader concept, subsuming the notion of TM. This overarching nature of LM is reflected in Kühn's words: „Grundwortschatz, ist ein inflationärer Begriff [...]” (P. Kühn 1991: 1358). Granted, if the entire lexical richness of a natural language is taken into account (i.e. its horizontal and vertical variation, including all specialised languages and their variants) when compiling an LM, then, naturally, the notion can be seen as encompassing TM⁶ (LM>TM). This might be one of the reasons why some authors use the term ‘lexical minimum’ in works that in fact present specialised lexis (see e.g. the *Lexical minimum of general technical English for students of engineering* (O.L. Yaroshenko 2014: 104) or *Professional Lexical Minimum for Customs Students* (E.Ju. Gordeeva 2015))⁷.

It needs be emphasised, however, that on consistency grounds, only in some cases should the term ‘lexical minimum’ stand for the term ‘terminological minimum’. Such substitution is principally possible in two cases: (1) where the choice of vocabulary is not restricted to terms, and also includes other lexical units found in specialist texts, (2) if the primary function of the reference work does not include standardisation or no terminological goal is attempted. However, if the lexical units included are terms proper⁸ or where normative nature (status) of a set of lexical units is to be established by way of the minimum, the term ‘terminological minimum’ is preferable (for more detailed considerations see below).

2.3.

Contemplating the notion of ‘minimum vocabulary’, which is often equated with the concept of LM⁹, it primarily needs to be considered in its most fundamental meaning, namely that of a lexical competence, or indeed the mental lexicon of an individual. This conclusion stems from the fact that vocabulary as such belongs to the realm of language, i.e. an inherent property of a human being (F. Grucza 1983: 296–297, 301–302). Depending on the language skills that interlocutors use, it is possible to differentiate between passive (for text reception) and active (for text production) minimum vocabulary. The qualifier ‘minimum’ refers here to a certain level of lexical competence allowing an uninterrupted transfer of information between speakers at a certain proficiency level. However, in spite of the fact that such lexical competence level is by convention gradable (i.e. elementary, intermediate, advanced), its evaluation will always be relative; it can only be measured against the competence of other individuals or groups of individuals. Practically speaking, the only viable

⁶ Regardless of the precision of the definition of specialised lexis (terminology), which, by convention, a TM is supposed to present.

⁷ The term ‘vocabulary’, and in particular ‘basic vocabulary’, is also frequently used for specialised vocabulary sets (e.g. *Grundwortschatz für Pflegeberufe*, R. Strack 2015), alongside the term ‘basic terminology’ or ‘basic concepts’ (e.g. *Glossary of basic terminology on disaster risk reduction*, UNESCO).

⁸ According to the traditional definition of the term (cf. H. Felber/ G. Budin 1994: 26).

⁹ At least in one of its meanings.

manner to undertake such assessment is through speakers' language exponents, i.e. texts. It might therefore be easier to evaluate active rather than passive vocabulary (competence) of an individual. Also, it is not (and will probably never be) possible to study/ evaluate an entire cognitive capacity, such as the lexical competence, of a person.

Despite the fact that the competence is hardly quantifiable, statistics has been used to correlate the number of internalised word stock and the amount of text understood by its readers (D. Hirsh/ P. Nation 1992, A. Dörre 2010: 20ff).

2.4.

One of the fundamental assumptions that is relevant to both LM and TM is that neither should be mistaken for the frequency dictionary. Although there is some degree of correspondence between the notions of LM/TM (understood as a specific type of reference work) and a frequency dictionary, that is the quantitative element behind the linguistic data, the major point of divergence lies in the fact that the latter is designed to inform about the frequency of words (or other linguistic units/ phenomena), while in the case of the former frequency data of lexical units are used at the initial stage of lemma selection procedure, at least according to the prevailing approach to LM/TM compilation. Another difference is that qualitative lemma selection procedure belongs more to LM/TM than the frequency dictionary design. Any frequency dictionary that presents other types of data beyond statistical ones should be considered a hybrid lexicographic work.

Another term worth delineating here is that of the 'minimum dictionary', which refers to a lexicographical work that represents LM¹⁰. As a matter of fact, the two terms have often been used interchangeably with regard to a reference work that includes a selection of vocabulary. From an ontological perspective, it could be said that LM is the content of a minimum dictionary, the latter featuring a lexicographical structure, sometimes equipped with supplementary materials, in particular useful indexes (e.g. *Słownik minimum języka polskiego* (see: H. Zgólkowa 2013). Yet, an extended lexicographic structure has also been found in lexical minimums (e.g. *Lexical minimum of English for Bulgarian Learners* (see: A.Ī. Danchev et al. 1982), and therefore no definite borderline can be drawn between the two terms.

Contrary to its primary function outlined above, the concept of 'minimum dictionary' has often been used as a badge of a frequency dictionary (see e.g. *Častotnyj anglo-russkij fizičeskij slovar'-minimum* (see: P. Alekseev et al. 1996), which supports my contention that the notions of LM/TM are predominantly seen either as frequency or frequency-based reference works (cf. definition of basic vocabulary in the *Dictionary of Lexicography*, R.R.K. Hartmann/ G. James 2002: 13). P.M. Alekseev (2005) seems to confirm this observation, at least in relation to the use of statistics in the choice of language material in glottodidactics and creation of minimum vocabularies. The scholar writes:

¹⁰ In yet a different approach, it is the didactic dictionary that is designed to present terminological minimum, or at least the most frequently used terms (S. Gajda 1990: 120).

It has been demonstrated that practically all selection criteria [for the choice of language material] are statistically based. Creation of basic vocabularies and dictionary minimums enters the sphere of class-room statistical lexicography, an important part of it being formed by bilingual, multilingual and semantic frequency dictionaries [...] (P.M. Alekseev 2005: 320).

Yet, despite an extended set of (ever more specific) statistical methods used in lexical studies, such as dispersion and distribution across corpus/ corpora (V. Brezina/ D. Gablasowa 2013: 3), it has been assumed that the qualitative element should in most instances be taken into account in the process of LM/TM construction, in so far as quantitative criteria alone cannot account for some qualitative variables, such as the communicative/ pragmatic aim¹¹.

Alternatively, it may well be that it is impossible to draw a definite demarcation line between the various terms/ concepts discussed above. This stance can be supported by a claim put forward by P.A. Fuertes-Olivera that regardless of their names, all reference works are ontologically the same; they all strive at satisfying user needs by way of utilising various technologies and innovations that can help the users access the lexicographic data they need, while the differences between individual works stem from the fact that there are different users in different situations (P.A. Fuertes-Olivera 2014: 28).

2.5.

Among the qualitative criteria for LM lemma selection one can list: the ease of learning, stylistic/ emotional neutrality, pragmatic relevance, usefulness, informative/ communicative value, word formation productivity, etc. (see F. Charalabopoulou et al. 2017, H. Kaczmarek 2006: 29, M. West 1953: IX-X). Appealing as the criteria might seem, they have not been clearly defined (if at all) and lack the necessary methodological principles¹²; there are hardly any guidelines as to how to apply them in specific projects. Accordingly, one needs to bear in mind that qualitative methods are inherently subjective and possibly intuitive.

2.6.

Construction of a reference work such as LM should follow a predefined chain of lexicographic activities, including design, user and usage study, corpus compilation, data extraction, database creation, writing of entries, proofreading, marketing, etc. One of the most fundamental variables that influences the design phase as well as any future corrections to the work are the users: their knowledge, skills, needs, and the tasks they will use LMs for. Practically speaking, it is necessary to undertake user studies and define a catalogue of universal user needs for specific user types. The principles to follow are derived from the Function Theory of Lexicography, as

¹¹ In fact, the mere fact of choosing specific statistical measures or setting a threshold on the lexical set to be presented may amount to a qualitative compilation element.

¹² One example relates to the vague notion of ‘usefulness’, which is either not defined at all or relates to a wide range of communicative or glottodidactic goals a TM is supposed to help achieve (see E. Jendrych 2012).

outlined by H. Bergenholtz and S. Tarp (H. Bergenholtz/ S. Tarp 2002, S. Tarp 2008, see also S. Tarp 2008: 168, P.A. Fuertes-Olivera/ S. Tarp 2014: 48-50).

2.7.

One of the major conclusions regarding the LM compilation criteria is that such reference tools should be based on authentic language samples, i.e. representative corpora. Representativeness is an all-important prerequisite here; the quality of any LM will be as good as the quality of the primary sources it is based on. However, despite a wide-ranging discussion on the issue of representativeness, few corpora can be considered representative. Moreover, representativeness is a relative feature, mainly because there are no objective ways to balance a corpus or measure the quality in question (T. McEnery/ R. Xiao/ Y. Tono 2006: 21). Also, practical implementation of the feature will depend on the horizontal and vertical characteristics of language units to be extracted from a corpus. Definitely, compilation of any reference tools, including LM or TM, on the basis of other reference materials, such as coursebooks/ textbooks, e-learning courses, etc., as suggested by some authors (e.g. E. Jendrych 2009b: 130, J. Tomaszczyk 2012: 47), is an erroneous method, unless some specific aim is attempted (e.g. comparison of the lexical scope in textbooks¹³) or the sources are used as secondary, i.e. supplementary¹⁴. Such secondary sources present an already selected material, often graded according to conventional proficiency levels, therefore considerably limiting the representativeness of the resources, and effectively LM/TM based on them. Often, the selection criteria in such sources remain unknown.

2.8.

Summarising the above considerations, it is worth emphasising that LM should be seen as a multidimensional concept. Its primary definition encompasses the following meanings: a) a language competence of an individual, i.e. their lexical competence (LM as a cognitive capacity), b) a relative measure tool of an individual's lexical competence (LM as a cognitive benchmark and – when externalised – a diagnostic tool, i.e. a testing device), c) a graded lexical resource for development of learning and/or teaching materials (LM as a general educational resource), d) an educational reference in its own right (LM as a specific educational resource, e.g. a self study material).

LM as a specific source or reference work is not a homogeneous entity. Its content will vary both horizontally and vertically, reflecting the conceptual and lexical complexity as well as pragmatic/ communicative value of lexical units presented. Richness of the structure and content of LM as a specific reference work can be a basis for their extended typology (P. Kühn 1991: 1358–1359). One important design prerequisite is that the structure of LM should be kept simple and the resource itself should not develop into an all-in-one dictionary. This simplicity calls for even more

¹³ Admittedly, it would be a far-fetched goal for any LM/TM.

¹⁴ Some exception from this rule might be employment of frequency dictionaries, provided that (a) the coverage is compatible with the scope of projected LM/TM, (b) robust quantitative methods have been applied for its compilation.

rigid compilation procedure. It is proposed that in most cases LM could take a form of a modern database, accessible online or becoming part of other lingware, whose content could be displayed at user's request and according to their current needs. The design of the database should also be subjected to rigid compilation procedures. As P. Hanks points out, the absence of space constraints (in modern electronic environment) calls for more, not less intellectual discipline in the selection and arrangement of information (P. Hanks 2003: 164).

3.

Following on the discussion on LM, it first needs to be emphasised that TM should primarily be seen as a specific lexical competence of a person. However, from the ontological perspective, it might be difficult to differentiate between LM and TM, mainly due to the difficulties in defining the operational unit in each case. If LM is considered a wider concept, then all internalised lexical units comprising the lexical competence of an individual are taken into account. If it is necessary to differentiate between general language lexical competence and specialised language lexical competence, then one should first define specialised language lexical units. Such differentiation is of utmost importance where is considered TM as a subset of the entire lexical competence.

If linguists cannot fully agree on a nature of word as such, then it is even more difficult to find a universal definition of a specialised-language vocabulary unit. Similarly to other linguistic units, specialised lexis does not form a homogeneous category, and scholars traditionally differentiate between various types of specialised vocabulary. The notion central to this discussion is that of the term.

3.1.

Traditionally, term is seen as a sign that is permanently linked to a distinctive concept or a group of concepts (H. Felber/ G. Budin 1994: 26). Among the most salient characteristics of terms is standardisation (terms should be conventional units)¹⁵ and univocity (terminology should follow the one term—one concept rule). The traditional (Vienna school) approach to terminology assumes that a clear delineation between concepts is possible on the basis of comparison of concept characteristics, which are derived from introspection. The identification of terms is therefore connected with precisely defining the concept. The obvious problem (and a basis for criticism) is seen in the accessibility of the mental (conceptual) layer, the reliability of its observation and presentation as well as the high degree of subjectivity, even in cases where a field specialist is involved (cf. R. Temmerman 2000: 45).

In the traditional approach to terminology, lexical units that fail to follow the prerequisites outlined above, cannot be considered terms proper. In particular, any non-standardised units or synonymous expressions will be labelled non-preferred (or even forbidden). Figurative language and connotative language are disregarded in traditional approach to terminology (R. Temmerman 1997: 54–55).

¹⁵ As a matter of fact, it is seen as the fundamental aim of terminology.

Adoption of the above presented view of terminology narrows the catalogue of lexical units considered in TM to terms. An issue requiring further discussion is the design of an evaluation tool to measure lexical competence in such a limited scope.

3.2.

The traditional approach to terminology has seen a growing criticism over the years, mainly on account of its prescriptive nature, including the imposed link between concept and term and the overemphasis on standardisation. Standardisation and univocity have been deemed unrealistic in natural language professional communication (cf. V. Tsakona 2007: 122). It has been stressed that terms should be analysed before concepts in their natural environment, i.e. in texts. J.C. Sager stresses that terms are not context-independent entities, and the meaning of terms can change, depending on the communicative intention of the text (or in fact that of the sender's), which can be transmitted at various textual levels (J.C. Sager 1990: 101). Accordingly, the link between concept and term can in fact only be reconstructed as an approximation, even if standardisation procedures have rendered some concept-term pairs relatively stable. Admittedly, the link in the minds of individuals is not only highly idiosyncratic, but also hardly identifiable.

Clearly, the new approach allows for a wider catalogue of units to be considered potential headwords in TM. In fact, specialised vocabulary in its entirety should be taken into account, provided that such selection meets the needs of users. Accordingly, along with standardised terms and nomenclature names, representing the conventional part the lexicon, also other lexical units that have yet to be standardised (such as pre-terms), general scientific/ technical words, hypoterms (general-language units occurring in specialised texts) or units emerging from natural professional communication, such as professionalisms or slang¹⁶ – are all seen equally part of the specialised lexicon and are recognised on a par with standardised terms. Moreover, regarded as natural language communication phenomena, lexical units resulting from polysemy and homonymy as well as abbreviated forms are also taken into account.

From such a wide lexical perspective, it is more difficult to set clear boundaries as to what should count as a TM operational unit. Definitely, according to the latest views on terminology, the expression 'term' subsumes most of the types of specialised vocabulary outlined in the preceding paragraph. This in turn brings it closer to a common (=non-specialised) explication of the notion, presented e.g. in general language dictionaries. For example, according to the *Oxford English Dictionary* term is "A word or phrase used to describe a thing or to express a concept, especially in a particular kind of language or branch of study." (Oxford English Dictionary Online). For such a broad set of lexical units, an overarching synonymous expression 'specialised vocabulary/ lexical unit' can be introduced.

¹⁶ It has been argued that such units (i.e. professionalisms and slang units) are highly valuable in professional communication and their information-carrying and communicative potential is equal to that of terms, despite their limited scope of use (see M. Łukasik 2017a: 311).

3.3

The lexical considerations presented above call for at least two different approaches to, and two different types of, terminological minimum: a prescriptive and a descriptive one (Fig. 1).

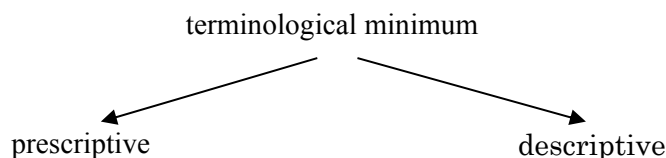


Fig. 1 Fundamental types of terminological minimum.

(i) From the perspective of the prescriptive (traditional) approach to terminology, the base for considering units of TM is the terminological lexicon¹⁷ of a domain/ field, and more specifically formally standardised terms. Accordingly, TM can be defined as a specific lexical competence, an evaluation measure, a terminological tool, a resource or a reference work covering base terms (and hence core concepts) of a domain/ field. As a lexical competence of an individual, TM may help distinguish language speakers with an appropriate conceptual-linguistic core knowledge characteristic of the domain/ field and its LSP, while TM as an evaluation measure can help assess such competence. However, due to a highly individual nature of knowledge, it must be emphasised that any thresholds set cannot be seen as absolute or objective. Presenting core concepts, TM is specifically useful as a tool in preventing undesired and unsubstantiated transgressions beyond the cognitive basis of a domain/ field¹⁸, simultaneously constituting an authoritative and prescriptive resource for various didactic uses, terminological applications, and beyond. For example, S.E. Wright and G. Budin (1997) suggest that standardised terminology ensures that legal regulations are enforced according to the “letter of the law”, while “adherence to the sense of specified terms in technical and legal contexts is designed to guard public health and safety, provide liability protection, and even guarantee human rights under the law”. In a similar vein, chemical nomenclatures provide systems for naming compounds, and are designed to “eliminate proliferation of conflicting names and to enable scientists to understand and benefit from each other’s work” (S.E. Wright/ G. Budin 1997: 329).

(ii) From the perspective of socio-cognitive approach to terminology, it is necessary to assume a wider base for distinguishing TM, namely all content-carrying lexical units found in specialised texts, in particular those pertaining to a domain/ field. The point of discussion is therefore shifted from the definition and status of terminological units (i.e. their conceptual/terminological value or preference over

¹⁷ In the narrow sense of the notion.

¹⁸ For example, G. Bedny strongly emphasises the cognitive role of basic terminology in psychology. The scholar claims that “The success of any theoretical and applied research in psychology, as in any science, largely depends on the proper use of basic terminology and fundamental concepts underlying the theory.” (G. Bedny 2015: 75).

other units) to the communicative/ informative value of lexical units found in specialist communication. TM can therefore be defined as a lexical competence of an individual, a specific evaluation tool, a lexical resource and a reference work, representing a selection of specialised vocabulary of a domain/ field. The selection, however, will depend on the final communicative/ (glotto)didactic goal and will reflect usage rather than norm. If constructed correctly, such a reference work can also illustrate the mainstream theoretical and practical approach to a subject-matter considered, even in fields where the conceptual basis is still being discussed. By its nature, the type of TM defined here is capable of reflecting global conceptual, linguistic and textual characteristics of an LSP (such 'registration' role cannot be attributed to prescriptive TM). One of the major imperatives that needs to be observed is that the lexis selected for TM must not distort professional knowledge¹⁹.

As a resource, TM is used as a basis for the development of (glotto)didactic materials (textbooks, course contents, etc.), lexicographic works (descriptive dictionaries, minimum dictionaries, thesauri, etc.), and terminological tools (terminological databases), among other applications.

As a glottodidactic tool, TM is aimed at increasing language competence of speakers in various professional situations. It is most frequently used where LSP is taught, i.e. vocational school, technical secondary schools, corporate language courses, language courses at university level, etc. Yet, it needs to be noted that the glottodidactic role of TM cannot parallel that of LM's, mainly due to the limited scope of lexical units presented in the former. Among its glottodidactic uses, TMs also act as testing means, lexical reference works for use in the classroom (resembling glossaries at the back of a textbook), and self-study materials.

Compared to the prescriptive type, the TM model discussed in this section can be regarded as descriptive, in so far as no standardisation or authoritative goal is attempted. However, with the accuracy of information and specialised knowledge validity maintained, descriptive TMs can also safeguard against knowledge distortion. Undoubtedly, both TM types are dynamic in that they remain open to new items – a feature naturally associated with the development of knowledge/ language. It is also worth noting that the clear (idealised) demarcation line drawn between the two models of TM serves only to conceptualise the issues connected with a rich variety of specialised lexis and general design/construction criteria as well as TM functions vis-à-vis users' needs. Since description and prescription do not have to be mutually exclusive opposites, hybrid TM models are also possible. In particular, TM as a specific reference work or a resource can be modified at macro-, medio- or microstructural level so as to feature both elements (cf. similar solutions in specialised lexicography; N. Guy 2002: 263).

¹⁹ The choice of appropriate sources and the selection procedure are therefore the all-important phases in TM construction. With many competing theories and methodologies, it might be challenging to go beyond frequency parameters (see below).

4.

The quality of any reference work is dependent on all the decisions and steps that precede its release. Accordingly, it is necessary to detail the mechanisms that lead to the creation of a reliable TM²⁰.

S. Gruzca (2004) suggests a universal compilation procedure for terminological minimum. The scholar outlines three general construction stages: 1) Establishment of a list of entries of most frequent terms, 2) Verification and ordering of the terminological material gathered, 3) Addition of missing terms (S. Gruzca 2004: 256).

Although the procedure in its original form is mainly applicable to a descriptive TM, replacement of the frequency parameter for the standardisation requirement, can render the proposal equally useful for the prescriptive TM.

The proposal delineated above may seem simple, yet, it in fact requires a greater number of steps and consideration of a number of specific variables. For example, before any extraction of lexical units (terminology) can take place, a lexicographer/terminographer first needs to identify global and specific design criteria²¹ of the final product as well as all partial tasks connected with its construction, such as identification of universal users' needs or compilation of a quality text corpus. Most importantly, the extraction, verification/ ordering and supplementation processes need to be specified in a clear and methodologically unequivocal manner so as to prevent any erroneous solutions that could negatively influence the quality of the final product. These issues will be discussed below.

4.1.

Since the dominant role of TM as an externalised entity is that of a resource and a reference work, it seems that the design procedure used in specialised lexicography can also be applied in a TM project. The major design elements can be shortlisted to the following points:

- the area/ subject field/ domain to be taken into account, its level of development;
- characteristics of the specialised language (and hence lexis) developed within the area/ subject field/ domain (e.g. primary or secondary or no terminological system; standardisation level; culture-dependent/ culture-independent terminology, etc.);
- potential users and their needs (defining the function on a future TM);
- type of TM to be compiled (e.g. prescriptive/ descriptive);
- availability (and types) of resources for a TM project;
- evaluation of secondary sources (terminographic analysis);

²⁰ Unless specified otherwise, the following discussion will concern TM as a specific resource or reference work.

²¹ My understanding of 'design criteria' is in line with the explanation of the notion provided by E. Debus-Gregor and U. Heid (2013: 1002). According to the scholar, the design criteria are any theoretical and practical aspects that need to be considered in the conception of a TM.

- conceptual continuity to be reflected (vertical and horizontal concept characteristics; level of concept granularity to be presented);
- lexical items inclusion/ exclusion criteria;
- corpus compilation procedure;
- lexical items extraction and verification methods;
- lemma stock complement criteria/ methods;
- data management;
- macro- and microstructure of TM (tailored to the needs of users);
- distribution of the product (e.g. marketing strategies);
- use and user feedback processing.

The foregoing paragraphs discuss some of the variables that need to be addressed in the course of TM design and compilation.

4.2.

Contemporary world has achieved a level of development and degree of specialisation that has never been seen before. These advancements have naturally been reflected in specialised languages, in particular in their lexical layer. Yet, a specific dichotomy can be observed; the most advanced nations and/or languages²² develop more extensive and precise lexical (terminological) systems. Such systems are called primary (=primary terminological systems), as opposed to secondary systems (=secondary terminological systems), which usually borrow concepts/ terms from the former. From the perspective of terminological studies, primary systems usually act as a source of highest quality specialist knowledge and as a specific benchmark against which secondary systems are evaluated. To construct a reliable resource, such as a TM, it may be necessary to examine the nature of a particular system and trace back the origins of both terms and their concepts/ meanings. Needless to say, the boundaries between fields/ domains, and hence their lexical systems, are usually blurred and/or overlapping, not least because (a) it is inherently difficult to distinguish between various realms of research (D.S. Giannoni 2010: 21), and (b) an even greater number of human endeavours are (purported to be) interdisciplinary. This brings to light a question of the thematic scope of primary and secondary texts for TM lexical excerption (see Sections 4.4. and 4.5. for further discussion).

Another issue worth considering is the proportions of various types of lexical units found in specific LSPs, such as between terms and nomenclature names, standardised and non-standardised units, as well as other lexical items found in specialised texts, such as abbreviations. Since some LSP lexicons are highly conventionalised (cf. the lexicon of chemistry), and others are rather far from standardised (cf. the lexicon of linguistics), it may not be feasible to undertake construction of certain TM types in a specific field/ discipline, such as a prescriptive TM in the field of linguistics.

²² The term ‘most advanced languages’ refers here to lingua francas of international professional communication/ knowledge transfer. Despite their worldwide range, such languages may also exhibit a degree of divergence across various (groups of) users.

Finally, conceptual systems of some fields/domains may be specific to a particular (lingo)cultural area or even limited to a geographical region, giving rise to e.g. culture-dependent terms (e.g. legal terms) or lexical/ terminological regionalisms (e.g. terms referring to crafts unique to a cultural minority). This poses a problem of conceptual system comparability in both intra- and interlingual plane, which, in particular, may influence the construction of bi- and multilingual TMs. Contrary to popular view, non-compatibility of conceptual systems can also be observed in culture-independent fields/domains, such as science or technology. G. Bedny (2015: 75) stresses that scientific terms may have different meanings in different languages and may vary from author to author. He discusses the terminological problems in the field of psychology, and more precisely its activity theory (AT), and demonstrates that the AT terminology, which was developed in a very specific sociocultural context in Russia, poses great interpretation and translation problems, in particular in/ into English (G. Bedny 2015: 75, 92–93). In technical fields, variation across technical vocabulary may stem from the development of local technical culture and the influence of local brand names which later become terms proper (e.g. flex).

Depending on the function of TM, all the variables discussed above will influence the choice of primary sources as well as vocabulary selection criteria (see below).

4.3.

As has been mentioned in the Introduction (see Section 1 (f)), user studies are at the heart of any lexicographic compilation and influence the final shape of any reference work. As H.K. Simonsen asserts, “the degree of success of any product is heavily dependent on the underlying end-user research conducted”, while “the qualifications and situation of the end-user govern any production process” (H.K. Simonsen 2000: 93). Besides user characteristics (their general knowledge, level of expertise, language proficiency, etc.), also knowledge-related (factual knowledge expansion/ verification), communication-related (text production/ text reception) as well as extralexigraphic user needs should ought to be researched (cf. S. Tarp 2008).

4.3.1.

If TM is to assume a prescriptive terminological role, the primary addressees of the work will be experienced professionals in specific fields, governmental (and other official) bodies, terminologists, linguists (including cognitive linguists), NLP specialists, information science experts, lexicographers (terminographers), translators and possibly educators. Also specialists from related fields can significantly benefit from such TM type, mainly because of its concept defining nature.

If TM is to assume a descriptive role, a catalogue of users widens significantly and includes all groups of LSP users, i.e. specialists, semi-specialists and non-specialists.

In the hands of field professionals, descriptive TM will most likely serve as a resource and reference work, mainly for elaboration of other reference works, such as dictionaries, textbooks and other didactic means (such as e-learning courses, testing tools, etc.), as well as a benchmark for evaluation purposes (e.g. of employees). It

might also be used as a resource for future standardisation of terms; in this way a specific feedback occurs between the two types of TM discussed in this paper.

Definitely, the most numerous group of descriptive TM users is constituted by semi-specialists and non-specialists, and in particular by LSP educators and learners, students of a variety of fields, translators, specialists of other (often related) fields, technical writers, etc. In the area of education, descriptive TM will come in two flavours, depending on the skills developed by their users, namely active and passive (see Section 2.3.). Yet, beyond the most commonly recognised function as a didactic tool, descriptive TM can be a subject-matter guide.

4.3.2.

An example of TM user studies comes from E. Jendrych, who has elaborated on potential addressees of a glottodidactic TM for Business English. The researcher lists the following groups of users along with projected TM functions (E. Jendrych 2009b: 127):

- students (TM used for revision of lexical material, and as self-evaluation tool and a resource for bridging knowledge gaps);
- textbook authors (TM as a benchmark for authors of Business English textbooks and e-learning courses);
- teachers (TM used for the assessment of the degree of specialised vocabulary acquisition by students, and for evaluation of the effectiveness of teaching methods and achieved progress measured by the level of lexical proficiency),
- examiners (TM used for testing language proficiency in the scope of specialised vocabulary);
- employers (TM used for the assessment of language proficiency of interviewees/ employees (e.g. when commissioning a task requiring appropriate level of business language);
- linguists (TM used for determining the level of terminologisation of didactic texts, and for evaluation of textbooks/ e-learning courses).

Expanding her studies, E. Jendrych also has also explored in detail the student group, referring to their professional experience and subject-matter knowledge. According to her studies (2009a: 85–86), student users comprise:

- regular students of economics, finance, and administration, with little professional experience/knowledge;
- extramural students majoring in the above-listed fields, exhibiting considerable professional knowledge;
- students and graduates of language and linguistics departments, who wish to specialise in business English, but who have no knowledge about business;
- participants of closed (corporate) and open Business English courses, who possess various levels of knowledge and professional experience.

It is worth emphasising that the user studies undertaken by the along with a TM usefulness survey (E. Jendrych 2012) are an example of good practice in the area of TM construction.

It seems that in order to define student/ learner characteristics, it is necessary to consider users’:

- language level;
- general knowledge;
- professional experience;
- prior education (i.e. experience with educational materials, setting, etc.);
- learning strategies/ styles;
- motivation;
- specific needs;
- interests.

Undoubtedly, when designing TM as a didactic tool, it is also necessary to take account of the general premises of an LSP curriculum as well as specific variables arising from course objectives, such as: (a) abilities to be acquired for successful communication in occupational settings, (b) proportion between content language acquisition vs. general language acquisition, (c) group characteristics (heterogeneous vs. homogeneous learner group), (d) [possibility of] materials development (K. Gatehouse 2001).

4.3.3.

As a matter of fact, the multitude of variables that define users, their needs, and reference work consultation situations, as outlined in the Function Theory of Lexicography (see above), render any strict user division (e.g. into experts, semi-experts, and non-experts) provisional and somewhat artificial. In fact, the rich list of varied user characteristics results in a diverse array of TM types, some probably of a hybrid nature.

4.4.

Coverage in a TM is a relative matter. Not only is it hardly possible to determine someone’s lexical competence, but it is also difficult to set lexical boundaries in cases where TM is seen as a testing tool, a resource or an educational means. The basic assumption that has to be made is that it is virtually impossible to set any quantitative threshold (or frame) regarding the number of units to be considered/ included in a minimum (see Section 4.9.2.). This is consequent upon the impossibility of establishing conceptual/ terminological boundaries of individual fields/ disciplines as a result of different concept (term) vertical and horizontal characteristics and the interdisciplinarity of modern human endeavours. In some cases it might not even be possible to isolate a leading field/ domain, as disciplines themselves are often remarkably heterogeneous or even internally divided and “the existence of a discipline does not always imply that there is acceptance of an agreed set of problems, objects, practices, theories or methods, not even a shared language or common institutional forms” (A. Barry/ G. Born 2013: 9–10). Therefore, the division into fields/ domains/ disciplines/ etc. is in fact a matter of convention. Accordingly, it might be necessary to use some extralinguistic measures, such as external or internal classification systems, textbook thematic divisions, encyclopaedic coverage of a field, etc. for pre-

selection of texts and text types. Conveniently, specialised texts exhibit certain field-specific characteristics that allow automatic text clustering (see e.g. CLARIN-PL WebStyEn service²³), which might be particularly useful during corpus compilation and text verification.

As has been mentioned above, the frequency parameter is of limited applicability when specifying TM lemma stock, and could possibly only be employed for statistical purposes, yet only where term extraction concerns an extensive set of text and text types of a particular LSP. From this perspective, coverage is understood as a degree to which core concepts of a field/ domain have been represented in a TM, also in cases where some form of lexical material grading is envisaged.

On the microscale, coverage is understood as the completeness of semantic (sub)fields and/or of semantic neighbourhood represented by the lexical units included in a TM. Such completeness is of paramount importance in TMs, since the semantic networks between terms constitute a conceptual backbone of a terminological set, indispensable for the full understanding of any given term. For example, the notion of atom is bound semantically with such notions as electron, neutron, proton, nucleus, chemical element, atomic number, valence electron, electron configuration, electron shell, electron subshells, orbital, mass number, quantum number, excited state, ground state, radical, periodic law, Hund's rules, and Pauli exclusion principle (T. Michta 2009: 109–110). Any term from the above list is linked to (an)other term(s) through semantic attraction, and cannot be fully understood in isolation from its neighbours. Consequently, the use of frequency or rank lists only must be deemed as highly misguided term extraction methods, and if used, should merely constitute the initial step, obligatorily followed by term verification and supplementation on the basis of semantic/ conceptual analysis. Such semantic analyses can currently be performed either manually, through analysis of contexts and verification with external reference works, or automatically, often during term extraction processes when semantic relations between terms are also reconstructed (A. Panchenko 2012, cf. N. Lagutina et al. 2015).

Significantly, when considering the coverage of TM in general terms, a key rule to follow is to maintain conceptual continuity of the lexical set included. This prerequisite will also be applicable to the TM unit inclusion criteria.

4.5.

The quality of any TM as a specific resource or a reference work will predominantly depend on three elements: (i) choice of primary texts for extraction of lexical units, (ii) the selection criteria that will allow the choice of required units and will filter out units that do not conform to a current TM project requirements, and (iii) extraction methods (i.e. their efficiency/ precision). This section discusses primary text selection criteria, with the other two issues referred to in the following sections.

Generally, the choice of primary texts will be governed by different criteria, depending on the TM type.

²³ <http://ws.clarin-pl.eu/webstyen.shtml> [Accessed: 1.12.2017].

The primary quality of sources selected for a prescriptive TM is their authoritative nature with relation to the terminology set presented. Hence, terminological standards, dictionaries presenting standardised terms, term banks, electronic terminology collections (databases), etc.²⁴ The catalogue of sources is closed and therefore no additional studies are necessary to certify their quality, besides the attestation of their origin (i.e. the publishing institution (usually a standardising body)). If standardised sources are available, the issue of representativeness is irrelevant, as the standardised vocabulary sets already represent an approved selection. Where such sources are unavailable, then some other high quality texts need to be included, with their selection verified by a subject specialist (R. Dubuc/ A. Lauriston 1997: 87). Possibly, inclusion of other non-standardised and/or unverified sources will have detrimental effect on the realisation of the principal function of a prescriptive TM, i.e. that of a cognitive/ terminological standard. Admittedly, the fuzzy notion of ‘quality’ as well as experts’ own subjective decisions may render the entire concept of a prescriptive TM unfeasible in some domains of human activity.

In case of the descriptive TM the list of primary sources is considerably wider, and besides the sources presenting standardised terminology, it also includes original research papers, monographs, textbooks, handbooks, service manuals, technical documentation, etc. In such works usage rather than normative nature of units is of primary importance, and therefore – contrary to the prescriptive TM text selection criteria – any publications of reference nature, such as dictionaries, glossaries, lexicons and other vocabulary lists, should be used solely for cross-checking the completeness of a selection and supplementation of missing terms.

The very notion of a descriptive reference work calls for the representation of a specialised language in its entirety, while the feature most sought for in the representativeness of the text collection, which manifests itself in an appropriate corpus balance (choice of text types and/ or text genres in their natural proportions) and sampling (choice of texts excerpts). In practical terms, it is necessary to establish such natural text proportions in the domain under consideration. It needs to be emphasised that the corpus should by no means consist of a subjective (and often too narrow) selection of texts of secondary nature, such as textbooks, encyclopaedias, dictionaries, etc., mainly because the sources already exhibit a selection made by their respective original authors (cf. E. Jendrych 2009a: 8), and hence constitute secondary or even tertiary selection. This rule must be observed in order to maintain corpus, and therefore TM, representativeness²⁵.

²⁴ Despite the fact that these works are referred to as secondary sources, whose role should mainly be limited to “countercheck” the already chosen set of lexical items, it is argued that in the case of a prescriptive TM, such reference works, provided they present standardised terminology, constitute the primary resources. It is the vocabulary selection procedure that would be of utmost importance in such collections.

²⁵ Even in cases where primary sources in a given language are hardly available (see J. Tomaszczyk 2012: 47). In such cases, construction of a descriptive TM is either impossible or less specialist sources need to be allowed, such as popular science texts for semi- or non-

J. Pearson (1998: 60), referring to a corpus design for a terminological study, lists the features that the texts to be included in the corpus need to possess in order to be of proper value for the project. According to the scholar, the texts should be: (a) written (as spoken texts have to be transcribed, which may lead to deformation of professional message), (b) published (this will validate the reliability of the material, especially if the publishing house ranks high), (c) produced by an acknowledged individual or institution, (d) factual (i.e. representing what is known to exist, or believed to exist).

The proposal seems to be universal in that it can be applied in projects where the compilation/ use of corpora is envisaged or not (cf. source selection for prescriptive vs. descriptive TM). In any case, the 'proper value' can refer to a whole range of variables, conceptual, linguistic, terminological, pragmatic, etc. In fact, the suggested functions of a TM act as a terminographic filter for appropriate text selection (cf. M. Łukasik 2005: 156). For example, in term extraction methods, the source texts should exhibit high level of terminologisation (understood as a relatively high number of terms compared to the number of all tokens in a given text/ corpus). Yet, some highly technical texts may be lexically poor (i.e. exhibiting few unique terms), which stems from the terminological consistency rule (i.e. avoidance of synonymous expressions and thus frequent repetition of terms). M. Gotti gives an example of Russian theoretical mechanics, where as few as 480 words form up to 91% of texts in the discipline (M. Gotti 2005: 34). The scholar adds that in such disciplines, the type-to-token ratio (TTR) may be very low, ranging between 0.043 and 0.135 (ibid.). Accordingly, the degree of terminologisation, based on frequency measures, cannot be the only parameter behind text selection criteria, if only on account of the varied vertical and horizontal characteristics across terms. Yet, quantitative matters, including the corpus size, are all-important prerequisites in automatic term recognition (see Section 4.8.).

Currently, more and more texts are obtained from the Internet, some of which with the use of automatic corpus compilation software (e.g. BootCaT online service or WebBootCaT module in the Sketch Engine). This calls for more rigid quality check (e.g. at the level of seed or URL selection); online texts may be linguistically highly idiosyncratic, poorly written, may present unsubstantiated proposals or may be faulty translations of other original works. Also, texts presenting new ideas in a field should be treated with caution as new concepts may be vague and named only tentatively, often with the use of sometimes erroneous analogies, metaphor or transferred meaning (W. Lange 1995: 416). Unless a specific aim is attempted, e.g. a study of neologisms, such texts should be discarded, in particular as sources for a TM.

4.6.

As has been mentioned above, secondary sources are used for verification of term candidates and supplementation of missing terms in TMs. Among the most universally cited resources are encyclopaedias, specialised lexicons/ glossaries/

experts. Alternatively, sources in other languages, in particular those, in which the terminology of a field is more developed (i.e. representing the primary terminological system), could be included, yet the descriptive status of a TM is thus extended beyond a particular national LSP.

dictionaries, terminological databases/ banks, thesauri and TMs in other languages, in particular in a language representing the primary terminological system. Equally useful are monographs, textbooks, documents, etc. not included among the primary sources. The catalogue of such works depends on the specificity of a field/ domain, prevalent publishing tradition, preferred professional communication setting, etc.

It is of utmost importance to attest quality of such reference works. Regarding terminological dictionaries, quality refers to the terminological coverage of a field/ (sub)discipline. Yet, such coverage should be evaluated by means of both quantitative and conceptual measures against the benchmark set by other, more comprehensive works, specialised thesauri or specialised corpora. Such studies are necessary in light of an extensive heterogeneity of specialised reference works. G. Norman notices that “a very significant proportion of the items contained in scientific dictionaries of terms are not nomenclatural items, but rather non-formalized field-specific lexis” (N. Guy 2002: 268). Accordingly, such works cannot be used for terminological verification indiscriminately.

Decisions concerning the choice of secondary sources can be supported by extensive terminographic analyses or professional reviews. Results of terminographic analysis can provide a global evaluation of terminographic practice in a given scope, along with assessment of individual works (see, e.g. M. Łukasik 2007, 2015)²⁶. Otherwise, TM authors should seek opinions of professional terminologists or field experts.

4.7.

Despite reservations presented by some scholars (see below), an electronic specialised corpus seems to be a well-suited source for the extraction of terms and other data necessary for the compilation of most TM types of descriptive, and in individual cases, also of prescriptive nature (cf. A. Dörre 2010: 103ff). However, specialised corpora are scarce and only a few are widely available (see M. Weisser 2017). Moreover, most of such corpora will be made accessible through a website interface, and therefore thorough quantitative studies may not be possible. Also, most specialised corpora offer a narrow selection of texts in specific subject fields/ domains (e.g. the GENIA corpus²⁷ in clinical biology or even the PERC corpus²⁸ incorporating texts from 22 domains). These confinements render such corpora of limited applicability as primary TM sources.

It may therefore be necessary to compile a project-specific specialised corpus. Its usefulness, however, will largely depend on a number of decisions made at the design, text gathering, corpus construction, and data extraction stages.

(i) It first has to be borne in mind that for a special purpose corpus, one “cannot simply adopt the criteria used to design LGP corpora” (L. Bowker/ J. Pearson 2002:

²⁶ Interestingly enough, by entering a given terminographic market, TMs themselves become an object of terminographic analysis, global and/or detailed (see M. Łukasik 2012).

²⁷ GENIA Corpus Website: URL: <http://www.geniaproject.org/> [Accessed: 1.12.2017].

²⁸ PERC Corpus Website: URL: <https://scn.jkn21.com/~percinfo/> [Accessed 1.12.2017].

45). Therefore, such corpus parameters as representativeness, balance, size, etc. are viewed differently in specialised corpus projects.

One of the corpus features/ qualities that needs to be considered at the design stage is representativeness. Its importance is underlined by A. Kilgarriff, who claims that the “quality of the frequency list, and thus the validity of all that follows, is premised on the composition and representativeness of the corpora” (A. Kilgarriff 1997: 136).

D. Biber provides an overarching definition of representativeness, emphasising that it is “the extent to which a sample includes the full variability in a population” (D. Biber 1993: 243). In other words, a text corpus (=sample) is representative if it exhibits features (e.g. linguistic ones) found in a general pool of texts considered (=population)²⁹. Moreover, representativeness may concern any linguistic (such as text type/ genre, lexical or grammatical features (e.g. degree of terminologisation), pragmatic/ situational parameters, etc.) or extralinguistic feature, such as demographic parameters or density of encyclopaedic information.

Taking text type/ genre as an illustrative example, it needs to be noted that a corpus is said to be representative as regards text type/ genre if it reflects the real (authentic) text proportions and textual features found in a generalised population of such texts. To achieve representativeness in this case, various text types/ genres are sampled and, ideally, constitute a corpus in their natural proportions. One issue that arises here is lack of comprehensive text type/ genre analyses in general and specialised languages. Viewed from this perspective, representativeness is a relative measure and indeed some form of approximation. Therefore, some scholars rightly point out that statistically speaking a corpus is representative only of the texts it is composed of; a corpus of abstracts of medical papers will be representative of these very abstracts of medical papers, and only if it has been compiled according to strict corpus compilation methods, e.g. in the area of corpus sampling.

Accordingly, the unattainability of corpus representativeness is probably more pronounced in specialised-text corpora, in which only a selection of texts is considered (cf. Ch. Enguehard 2005: 972). Some scholars have therefore put forward solutions that could bring closer the ideal of representativeness. In particular, two such proposals can bring about interesting results in the area of higher corpus quality, namely the parameter of corpus closure/saturation and a cyclic corpus compilation procedure.

Corpus “closure/saturation for a particular linguistic feature (e.g. size of lexicon) of a variety of language (e.g. computer manuals) means that the feature appears to be finite or is subject to a very limited variation beyond a certain point [of corpus size]. To measure the saturation of a corpus, the corpus is first divided into segments of equal size based on its tokens. The corpus is said to be saturated at the lexical level if each addition of a new segment yields approximately the same number of new lexical items as the previous segment [...]” (T. McEnery/ R. Xiao/ Y. Tono 2006: 16). It seems that the parameter can also be successfully applied to account for features important

²⁹ An explanation provided in this paper is a simplified version of in-depth considerations provided by D. Biber (1993).

from the perspective of a particular study, such as the degree of corpus terminologisation in the compilation of a specialised corpus (J. Pearson 1998: 45; for the application of corpus closure/ saturation parameter see R. Shams/ A. Elsayed/ Q.M-Z. Akter 2010: 71–72).

A direct consequence of the application of corpus saturation/ closure parameter is the relativisation of the corpus size parameter. Construction of large corpora need not be the primary goal in specialised projects. However, appropriate corpus size may play a role in registering low-frequency items/ language phenomena. Furthermore, automatic term extraction processes require corpora of considerable size to allow statistics-based algorithms to run properly (cf. C. Fantinuoli 2006: 173).

One of the compilation methods said to account for corpus representativeness is a cyclic corpus compilation procedure, whereby the corpus is built/ extended in several steps. In such a scheme, a design stage is followed by addition of an appropriate body of texts, i.e. such that is in line with the design criteria. The corpus is then tested, design criteria verified against the research/ practical objective, new texts – fulfilling the changed set of parameters – are added to the corpus, and the process repeats (D. Biber 1993: 255–256). This method allows progressive approximation to the desired threshold of resource acceptability for a particular use. Clearly, almost any feature can be parametrised in this approach. When combined with the notion of corpus closure/ saturation, the procedure may be a robust approach to follow. Yet, the final quality depends on the verification methods applied, and probably also on the type of data obtained from the corpus (some data may be difficult to evaluate/ quantify). Moreover, some researches may, possibly unwittingly, make an incorrect choice of the features analysed or adjust the compilation criteria in order to collect texts that will prove their hypotheses.

An important element of corpus methodology is annotation, i.e. an additional descriptive layer of metadata added to a text or text collection, such as part-of-speech, semantic, prosodic, etc. tagging and/or corpus mark-up, that allows more extensive corpus studies. Despite reservations voiced by some scholars that annotation invalidates corpus authenticity (J. Sinclair 2004: 191) and hinders objective analyses (J. Sinclair 2003: 83–84, 2007: 26), addition of such metadata is regarded as advantageous in most studies (T. McEnery 2003: 454–455), in particular for automatic term extraction methods, automatic lexicon generation, among other applications in natural language processing (T. McEnery/ A. Wilson 1996/2001: 142). Nowadays the accuracy of automatic taggers (part-of-speech, semantic, etc.) is relatively high (minimum 90%) and acceptable for most LSP projects (P. Rayson/ D. Archer/ S. Piao/ T. McEnery 2004: 7, Ch.D. Manning 2011: 171).

Depending on the predominant function of the TM, the choice of texts will vary accordingly. Admittedly, the most comprehensive collection will be required in a descriptive TM to be used as a resource. Since TM is supposed to present a selection of specialised lexis of a particular specialised language, in most cases compilation of a monolingual specialised corpus will be envisaged. If a bi- or multilingual TM is planned, a comparable bi- or multilingual corpus should be constructed in order to provide for consistency at the conceptual/ terminological level across texts in various

languages. Comparable monolingual corpus will be used in the study of culture-dependent (or anisomorphic) terminological sets/ systems, such as the legal one. Creation of comparable corpora is in particular important, since, to date, terminological resources based on the description of the concept systems have long neglected the use of terms in context, while it is the context that reveals the syntagmatic behaviour of terms, which – next to systemic term analysis – can help in choosing terminological equivalents (J. Pimentel 2015: 427).

(ii) Despite the large popularity of corpus-based studies and corpus applications in lexicography, some objections to the method have also been voiced. In particular, scholars developing the ‘Function Theory of Lexicography’ denounce the ‘corpus-only’ approach to lexicography as ‘misguided’, and advocate the use of a variety of sources for the compilation of specialised dictionaries (P. Fuertes-Olivera 2014: 29–30). They claim that corpora are predominantly used for secondary lexicographic tasks, such as attaching frequencies to some lemmas, crafting examples or indicating language patterns. Meanwhile, the most essential tasks in specialised dictionary-making, such as defining, attachment of grammatical data and translated texts, adding cultural remarks or offering proscriptive remarks require the use of other sources and “can only be done if an expert and a lexicographer are working side by side” (ibid.). In this light, corpora are seen as auxiliary rather than primary sources.

(iii) All things considered, it seems safe to claim that an electronic corpus constitutes a relatively reliable tool to extract the required data, especially for a descriptive TM. One way of approaching the problems associated with corpora, such as its size or representativeness³⁰, is to approach such parameters from a relative perspective and define separately for each specific project. Accordingly, corpus design stage should never be omitted, while steps should be taken to develop the most reliable methods for its construction.

4.8.

Parallel to the discussion on the sources for TM, another widely debated issue concerns the inclusion criteria of lexical units. First and foremost, any discussion should start with the acknowledgement of the dual nature of specialised vocabulary units, namely that of a sign of a concept (defined more or less precisely), and that of a linguistic unit. This assumption points to the possible selection criteria of lexical units to TM as a specific resource and reference work.

In light of the definitions of the two fundamental types of TM considered in this paper, the two groups of lexical units taken into account are: a) standardised terms, in

³⁰ The notion of representativeness is nevertheless a convenient guide when it is necessary to ensure a certain standard of a corpus. Yet, to fulfil its role, the parameter should possibly be seen as a guideline rather than a strict parameter and a goal in itself.

the case of a prescriptive TM, b) all units characteristic of specialised texts considered, in the case of a descriptive TM³¹.

(i) Selection of prescriptive TM units does not pose significant challenges because all standardised units constitute the basis for headword list in such TMs. Extraction of standardised terms amounts to inclusion of selected terms from various authoritative sources (e.g. documents, terminological databases or online lists of items approved of specific committees). This in particular concerns terminological standards, which usually present only a selection of the entire standardised terminology pool. In all other cases where the so-called informal (i.e. usage based) standardisation is claimed, only descriptive TM can be considered.

In the case of a descriptive TM, additional selection criteria can be imposed, depending in the specific function (and hence type) of a TM, and may include such variables as frequency, grammatical category, type of terminological unit (e.g. terms v. nomenclature names), word/ term formation potential, etc. It is also possible to envisage a TM of general scientific/ technical vocabulary, listing lexical units considered either as universal or interdisciplinary terms (found in many specialist fields/ domain, not specific to one particular field/ domain) or as hypoterms³² (general language words used in specialist texts, such as researcher, study or result).

(ii) Extraction of specialised lexical units from a corpus brings to light a number of issues, in particular connected with the determination of term candidates and delineation of term boundaries. It has been proposed that the criterion which determines whether a particular lexical item is a term or a (general language) word is the nature of its conceptual reference within the discipline concerned, i.e. general (referring to general knowledge) or special (referring to specialist knowledge) (J.M Pugh/ J.C. Sager 2004: 1924). Others have suggested a list of features of a term that should help differentiate the term from general words. The most frequently-cited characteristics include univocity, lack of expressive-emotional load, strong lexical and morphological correlations, conventionality and meaning independence (see e.g. A. Reformatskij 1959, S. Czerni 1977: 18, J. Lukszyn 2005: 45). There are, however, several problems involved with such propositions, mainly stemming from a considerable degree of subjective assessment regarding both the reference between a lexical item and human knowledge and the features that terms are purported to exhibit. Any such verification procedures are also rather difficult to apply in a modern corpus-based approach.

In view of the proponents of the most recent (i.e. sociocognitive) approach to terminology, terms primarily exist in texts, and therefore any extraction of terms should start with specialised texts (S.E. Wright 1997: 13). It is claimed that texts themselves offer the necessary context that makes it possible to reconstruct the

³¹ Plainly against the prevailing trend of focussing on nouns and noun phrases only, as advocated in the traditional approach to terminology, modern terminologists and terminographers acknowledge a wider range of lexical categories in terminological studies, namely verbs, adjectives and adverbs, among other categories (J. Pimentel 2015: 428, N. Guy 2002: 259, M.-C. L'Homme 2003: 403–404).

³² Also called sub-technical vocabulary, a term coined by Ronayne Cowan (L. Trimble 1985: 128).

linguistic and semantic/ conceptual features of a term. This point is emphasised by R. Dubuc and A. Lauriston:

It is the presence of conceptual features relevant to the term that determines the extent of the context. The context thus plays a double role: it first provides living proof that the term is used in the field of reference and then allows a conceptual content to be associated with the term being studied. The second role is the more important. It makes it possible to determine the specific relationship between a term and its subject field through concept identification (R. Dubuc/ A. Lauriston 1997 :81).

Accordingly, the whole process proceeds in a semasiological—onomasiological manner; terms, or term candidates, are extracted from authentic text collections (corpora), then their conceptual value (meaning/ position in a conceptual system) as well as linguistic and pragmatic relevance is determined and the list is verified (= non-terms or terms not complying with a particular TM project are removed from the list) according to term inclusion/exclusion criteria. The list is later supplemented on the basis of refined primary sources (see cyclic compilation procedure of a corpus) or secondary sources (textbooks, encyclopaedias, dictionaries, etc.). The assistance of field specialists is also highly recommended at this stage, while the methods applied are mainly based on frame semantics (J. Pimentel 2015). Alternatively, the core-periphery paradigm can be applied to evaluate the prototypicality of terms. Most importantly, the inclusion/exclusion criteria should reflect potential users' needs (linguistic, conceptual, pragmatic), and therefore the function of TM.

(iii) One of the practices that can ensure a greater degree of objectivity in the process of term extraction is to use automatic term identification/ recognition/ extraction tools. Such tools use a number of statistical and/or linguistic methods (e.g. part-of-speech patterns³³) to extract term candidates (L. Bowker 2003: 60). Admittedly, one of the prerequisites for using statistical methods is their language-independent character (Ch. Enguehard 2005: 971). However, simple frequency-based corpus tools, such as plain frequency/rank lists or n-gram lists do not suffice in the case of LSP texts and terms, and are considered unsystematic approach (H. Bergenholtz/ S. Tarp 1995: 88). Not only does specialised lexis come in a variety of types, but it also represents different conceptual paradigms. Frequency lists alone cannot account for the inevitable dispersion of conceptually-bound units in specialised texts (see above)³⁴. Moreover, high-frequency units usually exhibit greater ambiguity (cf. A. Östling 1991: 194).

Automatic term recognition and extraction has recently seen a considerable progress. This is also a result of more robust algorithms in the extraction of conceptual relations from texts. It seems that providing a wider context is indispensable in term

³³ In fact, part-of-speech annotation is used to statistically extract syntagmatic term schemata.

³⁴ This is one of the major differences between lexical minimum and terminological minimum; in the design and composition of the former, frequency can play an important role, even if frequency-based vocabulary lists need to be supplemented with secondary selection items (see e.g. M. Warchał 2013: 495).

processing tasks, also at a cognitive level. The contextual frame may, however, vary, depending on the degree of scattering of semantic information.

Also, in automatic term extraction, two salient parameters are taken into account, namely unithood and termhood. Unithood can be defined as the degree of strength or stability of syntagmatic combinations or collocations of a term candidate, while termhood is a degree to which a linguistic unit (i.e. term candidate) is related to domain-specific concepts. The latter can be estimated through term frequency and the so-called bias of frequency³⁵, which improves the precision rate (K. Kageura/ B. Umino 1996).

As has been mentioned above, particularly promising results in term extraction are achieved when statistical methods, such as association measures, are coupled with linguistic filters, such recurrent part-of-speech patterns identification. Most tools, however, are language-dependent; even if methods for ranking term candidates are universal, grammar rules used for selecting terminology concepts from texts differ across languages (it is necessary to develop the so-called extraction grammar, which is language-specific; M. Marciniak/ A. Mykowiecka/ P. Rychlik 2016: 2278). New methods are also being developed to account for the phenomenon of nested terms (terms within terms), which pose a specific problem in highly inflected languages, such as Polish (M. Marciniak/ A. Mykowiecka 2014)³⁶.

(iv) Extraction of specialised lexical units needs to take a view of the type of such units, derived from a combination of such variables as: textual (e.g. distribution across texts; genre-specific variations), and/or linguistic (e.g. part of speech), and/or terminological (e.g. culture-dependent/ culture-independent terms), on top of their cognitive value (e.g. degree of concept granularity represented by a term/ set of terms) and pragmatic significance (e.g. context-dependent term use). Accordingly, in automatic term recognition, evaluation of term candidates is the key procedure necessary to ensure extraction quality. Such evaluation can be performed automatically or manually by an expert.

Evaluation based on experts/ terminologists is by far the easiest to implement and process. However, such experts need to have access to modern tools, such as comprehensive specialised corpora and other machine-readable terminological and linguistic (re)sources to observe different occurrences of a term.

The most common approaches to manual evaluation of term candidates extracted automatically are based on (i) comparison with a reference list, and (ii) judgement of an expert/ terminologist. There are several issues connected with either method.

In case of the former, it has to be noted that:

(1) terms often vary on a linguistic plane (i.e. have different linguistic form) and such variations are not usually included in reference lists (e.g. such lists usually provide only one form of a term and do not include spelling variations, not to mention synonymous expressions); moreover, context-dependent terminological variation is

³⁵ i.e. inverse document frequency which points to words that occur rarely and tend to be omitted from a number of other texts.

³⁶ See TermoPL project: URL: <http://zil.ipipan.waw.pl/TermoPL> [Accessed: 20.11.2017].

usually not considered (i.e. situations in which a lexical unit may be considered term in one context, and a general-language word in another context);

(2) the reference lists, even if they are based on pre-established set of terms (e.g. a terminology bank) or a special corpus-based list drafted by terminologists, can be far from complete. Term candidates for descriptive TM can be evaluated against terminological registers, i.e. published lists of (presumably) all terms found in a particular LSP within a defined time period (usually a year). Unfortunately, such resources are scarce, difficult to obtain and usually outdated.

Above all, however, expert opinions may differ, and hence more extensive panels of specialists would need to participate in the evaluation process. Undeniably, such process is time consuming and costly, and still not devoid of subjectivism. Therefore, automatic term candidate evaluation is advocated.

(v) Among other important sources that could be incorporated in term candidates evaluation are thesauri, ontologies, including prototypical ontology representations (K. Ahmad/ L. Gillam 2002: 4), and semantic networks. These resources are significant tools in knowledge representation, and therefore are primarily used in information retrieval systems and word sense disambiguation, among other NLP applications. In particular, ontologies and semantic networks have been used in word sense disambiguation (for Princeton's WordNet see X. Li/ S. Szpakowicz/ S. Matwin 1995; for Polish SłowoSieć, also known as plWordNet³⁷, see P. Kędzia/ M. Piasecki/ M.J. Orlińska 2015), including in specialised domains (R. Prokofyev et al. 2013), while thesauri have been employed in document indexing and searching, identification of terms, and presentation of semantic content. Term extraction/ recognition and validation is a by-product of such extensive systems.

On a final note, evaluation procedures will to some extent depend on the future use of the terms extracted, whether for indexation, thesaurus design, improvement of automatic translation systems, semantic disambiguation, etc. (Ch. Enguehard 2005: 972) Consequently, each specific type of TM will possibly require a separate set of evaluation criteria.

4.9.

4.9.1.

Since TMs are specific terminographic works, they share some characteristics with other dictionaries, for example structure (i.e. macro-, medio-, and microstructure) or design criteria. The structure of TM as a resource or a reference work will depend on a number of variables elaborated on at the design stage. Yet, as has been mentioned in the introductory part of this paper, the structure of LM/TM should be kept simple, rather than become universal. There have been some proposals for the minimum set of data in a terminological work. H. Felber and G. Budin suggest that such entry should encompass (a) indication of the subject field, (b) language code, (c) terminological unit, (d) source, (e) usage notes, (f) concept description/ context (H. Felber/ G. Budin 1994: 188). Other scholars limit a minimum entry to (a) an

³⁷ <http://plwordnet.pwr.wroc.pl/wordnet/> [Accessed: 1.12.2017].

indication of the subject field, (b) the term itself, and (c) the [indication of] source of the term (F. Mayer 2013: 1466). From the perspective of a correctly compiled TM, a minimum entry consists of (a) the term, and (b) an indication of the sources, possibly in a form of a collective bibliography. From the terminographic perspective, in a work such as TM, the mere fact of inclusion of a lexical unit in a lemma list is highly informative, since it proves the value of the unit for use as a representation of a quantum of specialist knowledge. An additional indication of a subject field (or subfield) should be included in reference works encompassing a choice of vocabulary units from a macrofield or a number of related fields. In a descriptive TM, linguistic variants, synonyms as well as syntagmatic neighbourhood of the entry term should be provided for. Moreover, disambiguation notes, short descriptions, frequency count and terminological notes are other optional elements of an entry. In bi- and multilingual TMs, language equivalents are the obvious elements of the microstructure, while language indexes are the important elements of the outside matter.

The difference between a TM with a minimum set of microstructural elements and a terminological register lies in the fact that the latter either catalogues all terms of a specific specialist field or new terms in a field within a specified time period, while the former always presents a selection of vocabulary (following a rigorous and scientific compilation method).

4.9.2.

Considerations regarding the nature of specialised vocabulary and the functions of TM, as presented throughout this paper, prove that any attempt to (pre)define the number of lexical units to be included in a TM must be viewed as futile. This state of affairs primarily results from the non-quantifiable nature of human competence, including the lexical one. Also, it is impossible to definitely and accurately measure lexical stock of an LSP, since the feature that primarily determines the number of headwords in a TM is the conceptual integrity of the field/ domain/ subdomain to be represented in the work. Omission of terms that are semantically/ ontologically/ conceptually related within the same field/ domain must be considered a methodological error. Therefore, if any figures regarding the number of TM entries are provided, they can be used for comparative studies (i.e. used to relate the sizes of various TMs) or as a marketing tool.

4.9.3.

One of the most important (and marketable) features of modern terminological resources, including TM, is their currency (up-to-dateness). This stems from the need to reflect (codify) rapid developments in the majority of fields, and the resulting discoveries and changes in theoretical paradigms and practical solutions. From the practical perspective, the currency of any terminographic work is relative and highly dependent on the up-to-dateness of the primary sources as well as compilation techniques (traditional vs. electronic online resource). For a TM to retain its high value, it needs to be updated regularly, which calls for an electronic form of the

reference work, with a potential of distant (online) updates. The financial aspect of a TM project is probably a decisive factor in long-term TM management.

4.9.4.

In discussions on the shape of TM as a didactic means, it is suggested that the vocabulary included in such work should represent a B2/C1 language level (E. Jendrych 2009a: 126, O.L. Yaroshenko 2014: 104). Evidently, scholars focus only on the language proficiency level, somewhat disregarding the conceptual side of terms. While it is true that in order to comprehend/ discuss (specialised) professional matters, it is necessary to attain a certain level of general-language proficiency (e.g. to be able to understand/ formulate meaningful text structures/ grammatical constructions/ pragmatic frames, etc.), it is worth emphasising that any vocabulary grading is conventional and only partially corresponds to language proficiency. For one thing, vocabulary is only one element of the entirety of means/ skills constituting the human language, and for another, terminology represents a specific subcategory of lexical means, in so far as it reflects more precisely the cognitive side of terms (i.e. concepts/ units of understanding). In fact, the choice of terms on the basis of specific-level conceptual hierarchy and/or a prototypical differentiation, supported by frequency counts and appropriate dispersion (range) measures, possibly constitute the major axis of the compilation method. In educational contexts, the overarching goal of TM consists in enabling users to achieve a higher level of specialised language proficiency, while the final aim is to develop student autonomy, so that they can further develop their own specialised/ professional competence on the basis of unabridged materials³⁸.

4.9.5.

As has already been hinted, similarly to general language lexical minimums, or core vocabularies, terminological minimums have a great potential of being used in NLP applications, including machine translation (H. Lehmann 1991, A. Östling 1991). To this end, the resource needs to assume a machine-processable form.

The importance of computerised terminography have long been recognised, mainly in light of a vast vocabulary stock to be dealt with by terminologists (H. Felber 1982: 322). It has been demonstrated in many parts of this paper that the modern TM should become a specific piece of software, with the terminological data stored in a database or generated on the fly. In fact, the software could become an independent tool, integrating most, if not all, of the above automatic systems, such as automatic corpus creation, term extraction, and verification/ selection, and, in particular, the functionality of generating TMs according to user-specified criteria. In the glottodidactic context, for example, electronic TMs could provide the educators and learners with statistical data, various types of terms extracted from the corpora, disambiguation notes/ contexts, collocations, foreign language equivalents, etc., on top of their ability to generate lexical educational materials, graded tests, self-study

³⁸ In such cases, TM resembles defining vocabulary used in learner's dictionaries.

materials Moreover, modern digital world requires that such database should be interconnected with other similar resources, including databases, term banks, dictionary systems, etc. Similarly to a lexical minimum, TM could be a stand-alone downloadable (with remote access to the database or automatic retrieval system), accessible online or be part of other lingware, such as translation or terminology management systems.

5.

Although the notion of terminological minimum is predominantly associated with an educational tool, this paper has attempted to demonstrate that the concept should be viewed from a much wider perspective and discussed from a variety of angles. From the ontological perspective, TM is primarily a specific lexical competence of a person. As a specific tool, TM is viewed as a terminological resource and a teaching/ learning aid. Accordingly, theoretical and practical (methodological) discussions related to TM should involve findings of several fields, including (cognitive) linguistics, quantitative linguistics, glottodidactics, terminology, information/ computer science, etc.

TM as a terminological resource can play a significant role in maintaining conceptual integrity of a field/domain, mainly through presentation of base terms extracted on the basis of the strongest conceptual/ semantic relations binding the concepts/terms in question. Such a resource will become a valuable tool in terminology standardisation, translation, as well as a basis for LSP materials development.

TM as a specific glottodidactic means aids specialised vocabulary teaching/ learning process, parallel to its function as a standard-setting and testing tool. In fact, TM should be the first resort lexical reference work in LSP courses.

TM for communication purposes could/ should also include LGP units to enable uninterrupted flow of information (in fact, each LSP text includes some percentage of general-language and/or general-technical/scientific set of vocabulary). Alternatively, it could be advocated that in certain uses/ contexts, such as in various communications settings and in LSP courses (especially where foreign language instruction is in place), LM and TM could be combined.

In the digital age, the only practical shape for TM is a specific type of lingware, possibly incorporated in other tools, such as electronic didactic platforms or translation software. It could also be available as a downloadable and/or an online resource. In its simplest form, TM takes a shape of a continually updated database, from which data chunks can be obtained for a number of purposes. An advanced variety envisages an automatically created TM database, designed according to user-defined criteria, offering a number of different output data for a variety of purposes (e.g. terminological resources, self-study materials, etc.). Such model will most likely be descriptive in its nature, unless it is also connected to other standardised databases.

The above proposals are not guided by wishful thinking of the author – the technology and methodology are out there, waiting to be combined in a useful end-product for a growing number of prospective users and uses.

References

- Ahmad, K./ L. Gillam (2002), *Sharing the knowledge of experts*. In: Fachsprache 24 (1), 1–19.
- Alekseev, P.M. et al. (1996), *Častotnyj anglo-russkij fizičeskij slovar'-minimum*. Sankt-Petersburg.
- 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.
- Barry, A./ G. Born (2013), *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London/ New York.
- Bedny, G. (2015), *Application of Systemic-Structural Activity Theory to Design and Training*. Boca Raton/ London / New York.
- Bergenholtz, H./ S. Tarp (1995), *Manual of Specialised Lexicography: The preparation of specialised dictionaries*. Amsterdam/ Philadelphia.
- 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. (2003), *Terminology tools for translators*. In: H. Somers (ed.), *Computers and Translation. A Translator's Guide*. Amsterdam/ Philadelphia, 49–65.
- Bowker, L./ J. Pearson (2002), *Working with Specialized Language. A Practical Guide to Using Corpora*. London/ New York.
- Brezina, V. / D. Gablasova (2013), *Is There a Core General vocabulary? Introducing the 'New General Service List'*. In: *Applied Linguistics* 36, 1–22.
- Cao, D./ X. Zhao (2008), *Translation at the United Nations as Specialized Translation*. In: *The Journal of Specialised Translation* 9, 39–54.
- Charalabopoulou, F. et al. (2017), *Building corpus-informed word lists for L2 vocabulary learning in nine languages*. In L. Bradley/ S. Thouësný (eds.), *CALL: Using, Learning, Knowing*, EUROCALL Conference. Gothenburg, 49–53.
- Czerni, S. (1977), *Słowniki specjalistyczne*. Warszawa.
- Danchev, A.Ī. et al. (1982), *Lexical minimum of English for Bulgarian learners*. Sofia.
- Debus-Gregor, E./ U. Heid (2013), *Design criteria and 'added value' of electronic dictionaries for human users*. In: R.H. Gouws/ U. Heid/ W. Schweickard/ H.E. Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography (HSK 5.4)*. Berlin/ Boston, 1001–1013.
- Dörre, A. (2010), *Terminologische Mindestwortschätze an der Grenze zum allgemeinsprachlichen Wortschatz und ihre Einbindung in elektronische Wörterbücher. Textkorpusstudie am Beispiel der russischen Wirtschaftssprache*. Frankfurt am Main.
- Dubuc, R. / A. Lauriston (1997), *Terms and Contexts*. In: S.E. Wright/ G. Budin (comp.), *Handbook of Terminology Management Vol. 1. Basic Aspects of Terminology Management*. Amsterdam/ Philadelphia, 81–87.

- Enguehard, Ch. (2005), *Terminology*. In: R. Köhler/ G. Altmann/ R.G. Piotrowski (eds.), *Quantitative Linguistics. An International Handbook (HSK 27)*. Berlin/ New York, 971–987.
- Fantinuoli, C. (2006), *Specialized Corpora from the Web and Term Extraction for Simultaneous Interpreters*. In: In. M. Baroni/ S. Bernardini (eds.), *WaCky! Working papers on the Web as Corpus*. Bologna, 173–190.
- Felber, H. (1982), *The terminological data elements as derived from the general theory of terminology and their recoding in machine-readable form*. In: W. Nedobity (ed.), *Terminologies for the Eighties (Infoterm Series 7)*. München, 322–366.
- Felber, H./ G. Budin (1994), *Teoria i praktyka terminologii*. Warszawa.
- Fuertes-Olivera, P.A (2014), *Designing Online Dictionaries of Economics: Two Opposing Views*. In: *Hermes – Journal of Language and Communication in Business* 52, 25–40.
- Fuertes-Olivera, P.A./ S. Tarp (2014), *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin/ Boston.
- Hanks, P. (2003), *The Probable and the Possible: Lexicography in the Age of the Internet*. In: *Studies in Lexicography* 11.1, 7–36.
- Hartmann, R.R.K./ G. James (2002), *Dictionary of Lexicography*. London/ New York.
- Hirsh, D./ P. Nation (1992), *What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure*. In: *Reading in a Foreign Language* 8(2), 689–696.
- Gajda, S. (1990), *Wprowadzenie do teorii terminu*. Opole.
- Giannoni, D.S. (2010), *Mapping Academic Values in the Disciplines: A Corpus-Based Approach. Linguistic Insights 124*. Bern.
- Gouadec, D. (2010), *Translation as a Profession*. Amsterdam/ Philadelphia.
- Gatehouse, K. (2001), *Key issues in English for Specific Purposes Curriculum Development*. In: *The Internet TESL Journal* VII, 1–10.
- Gordeeva, E. Ju. (2015), *Professional lexical minimum for customs students*. Sankt-Petersburg.
- Gotti, M. (2005), *Investigating Specialized Discourse*. Bern/ Berlin/ etc.
- Grucza, F. (1983), *Zagadnienia metalingwistyki. Lingwistyka – jej przedmiot, lingwistyka stosowana*. Warszawa.
- 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.
- Guy, N. (2002), *Description and prescription in dictionaries of scientific terms*, *International Journal of Lexicography* 15 (4), 259–276.
- Jendrych, E. (2009a), *Minimum terminologiczne dla potrzeb glottodydaktyki języków specjalistycznych (na materiale angielskich tekstów biznesowych)*. Warszawa. (Unpublished PhD Thesis).
- Jendrych, E. (2009b), *Minimum terminologiczne angielskiego języka biznesu dla potrzeb glottodydaktyki*. In: M. Górnicz/ M. Kornacka (eds.), *Języki specjalistyczne 9. Wyraz – tekst – interpretacja*. Warszawa, 119–136.

- Jendrych, E. (2012), *Przydatność minimum terminologicznego angielskiego języka biznesu w ocenie jego przyszłych użytkowników*. In: *Linguodidactica* 16, 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* 5, 27–35
- Kageura, K./ B. Umino (1996), *Methods of Automatic Term Recognition: a Review*. In: *Terminology* 3 (2), 259–289.
- Kędzia P. / M. Piasecki/ M.J. Orlińska (2015), *Word Sense Disambiguation Based on Large Scale Polish CLARIN Heterogeneous Lexical Resources*. In: *Cognitive Studies | Études cognitives* 15, 269–292.
- Kilgarriff, A. (1997), *Putting frequencies in the dictionary*. In: *International Journal of Lexicography* 10(2), 135–155.
- Kühn, P. (1979), *Der Grundwortschatz. Bestimmung und Systematisierung*. Tübingen.
- Kühn, P. (1991), *Das Grundwortschatzwörterbuch*. In: F.J. Hausmann/ O. Reichmann/ H.E. Wiegand/ L. Zgusta, *Dictionaries* (eds.). *An International Encyclopedia of Lexicography. Volume II (HSK 5.2.)*. Berlin / New York, 1353–1356.
- L’Homme, M.-C. (2003), *Capturing the lexical structure in special subject fields with verbs with verbs derivatives. A model for specialized lexicography*. In: *International Journal of Lexicography* 16 (4), 403–422.
- Lagutina, N. et al. (2015), *An approach to automated thesaurus construction using clusterization-based dictionary analysis*. In: *Conference of Open Innovation Association, FRUCT*, 104–109.
- Lange, W. (1995), *The special lexicon and problems of EST/ESP – course design or the box of the Pandora*. In: R. Dirven/ J. Vanparrys (eds.), *Current Approaches to the Lexicon. A Selection of Papers Presented at the 18th LAUD Symposium, Duisburg, March, 1993*. Frankfurt am Main, 413–432.
- Lehmann, H. (1991), *Towards a Core Vocabulary for a Natural Language System*. In: *Proceedings of the fifth conference on European chapter of the Association for Computational Linguistics*. Berlin.
- Li, X./ S. Szpakowicz/ S. Matwin (1995), *A WordNet-based Algorithm for Word Sense Disambiguation*. In: *International Joint Conferences on Artificial Intelligence Organization Vol.2*. Montreal, 1368-1374. Date of access: 15.11.2017 <http://www.ijcai.org/Proceedings/95-2/Papers/045.pdf>
- Lukszyn, J. (ed.) (2005), *Języki specjalistyczne. Słownik terminologii przedmiotowej*. Warszawa.
- Łukasik, M. (2005), *Teksty terminoosne – filtry terminograficzne – stopień dokładności tezaury*. In: J. Lewandowski/ M. Kornacka (eds.), *Teksty specjalistyczne w kontekstach zawodowych i tłumaczeniach (Języki Specjalistyczne 5)*. Warszawa, 155–164.
- Łukasik, M. (2007), *Angielsko-polskie i polsko-angielskie słowniki specjalistyczne. Analiza terminograficzna*. Warszawa.

- Łukasik, M. (2012), *Terminological dictionary as a comprehensive cognitive and linguistic tool*, In: *Language in Different Contexts: Research papers = Kalba ir kontekstai* 5 (1), 98–108.
- Łukasik, M. (2015), *Quality of Multilingual Specialised Lexicography in Poland*. In: *Cognitive Studies | Études cognitives* 15, 79–95.
- Łukasik, M. (2017a), *Defining an Entry Unit in a Dictionary of Non-Standard Professional Vocabulary*. In: *Polilog. Studia Neofilologiczne* 7, 301–314.
- Łukasik, M. (2017b), *Lexical Minimum Re(Defined)*, In: *Lingwistyka Stosowana* 23 (3), 47–63.
- Mamet, P. (2002), *Relacja pomiędzy kompetencją językową a kompetencją merytoryczną na przykładzie języka biznesu*. In: J. Lewandowski (ed.), *Języki specjalistyczne 2. Problemy technolingwistyki*. Warszawa, 141–151.
- 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.
- Marciniak, M./ A. Mykowiecka/ P. Rychlik, P. (2016), *TermoPL - a Flexible Tool for Terminology Extraction*. In: N. Calzolari et al. (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*. Portorož, 2278–2284.
- Marciniak, M./ A. Mykowiecka (2014), *Terminology extraction from medical texts in Polish*. In: *Journal of Biomedical Semantics*, 1–14.
- Mayer, F. (2013), *Models for the representation of terminological data on the computer: Terminological databases*. In: R.H. Gouws/ U. Heid/ W. Schweickard/ H.E. Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography (HSK 5.4)*. Berlin/ Boston, 1461–1480.
- McEnery, T. (2003), *Corpus linguistics*. In: R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford, 448–463.
- McEnery, T./ A. Wilson (1996/2001), *Corpus linguistics*. Edinburgh.
- McEnery, T./ R. Xiao/ Y. Tono (2006), *Corpus-Based Language Studies. An Advanced Resource Book*. London/ New York.
- Micha, T. (2009), *O słowniku systemowym na materiale terminologii chemicznej*. In: M. Łukasik (ed.), *Publikacja jubileuszowa I. Na drodze wiedzy specjalistycznej*, Warszawa, 99–111.
- Oxford English Dictionary*. Date of access: 31.10.2017, <https://en.oxforddictionaries.com/English/>.
- Östling, A. (1991), *A Swedish Core Vocabulary for Machine Translation*. In: *Proceedings of Nodalida*, 187–198.
- Panchenko, A. (2012), *AUTHECO: AUtomatic THEsaurus COstruction. Information extraction for automatic construction of thesauri and semantic networks from text corpora*. Date of access: 10.11.2017. <http://cental.fltr.ucl.ac.be/projects/autheco/>
- Pearson, J. (1998), *Terms in Context*. Amsterdam/ Philadelphia.

- Pimentel, J. (2015), *Using frame semantics to build a bilingual resource on legal terminology*. In: H.J. Kockaert/ F. Teurs (eds.), *Handbook of Terminology*. Amsterdam/ Philadelphia, 427–450.
- Plawska, M. (2016), *English-German-Polish Learner's Dictionary of Terminology Related to Gastronomy and Catering Services*. Słupsk. (An unpublished Master's Thesis)
- Prokofyev, R. et al. (2013), *Ontology-Based Word Sense Disambiguation for Scientific Literature*. In: *Advances in Information Retrieval*, 594–605.
- Pugh, J.M./ J.C. Sager (2004), *Terminology in special languages*. In: G. Booij et al. (eds.), *Morphology: An International Handbook on Inflection and Word-Formation*, Vol. 2. Berlin/ New York, 1924–1928.
- 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.
- Reformatskij, A. (1959/ 1961), *Čto takoe termin i terminologija*. In: *Voprosy terminologii*. Moskva.
- Sager, J.C. (1990), *A Practical Course in Terminology Processing* Amsterdam/ Philadelphia.
- Schnörch, U. (2002), *Der zentrale Wortschatz des Deutschen: Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung*. Tübingen.
- Shams, R./ A. Elsayed/ Q.M-Z. Akter (2010). *A Corpus-based Evaluation of a Domain-specific Text to Knowledge Mapping Prototype*. In: *Journal of Computers* 5(1), 69–80.
- Simonsen, H.K. (2000). *Design, Development and Compilation of a bilingual multifunctional Intranet-based Differential Telecom Lexinome at a major Danish Telecoms Group*. In: U. Heid/ S. Evert/ E. Lehmann/ Ch. Rohrer (eds.), *Proceedings of the Ninth Euralex International Congress, EURALEX 2000 Vol. 2*. Stuttgart.
- Sinclair, J. (2003), *Corpus Creation*. In: G. Sampson/ D. McCarthy (eds.), *Corpus Linguistics. Readings in a widening discipline*. London/ New York, 78–84.
- Sinclair, J. (2004), *Trust the text: language, corpus and discourse*. London/ New York.
- Sinclair, J. (2007), *Language and computing, past and present*. In: K. Ahmad/ M. Rogers (eds.), *Evidence-based LSP: Translation, Text and Terminology*. Bern, 21–51.
- Strack, R. (2015), *Grundwortschatz für Pflegeberufe*. Stuttgart.
- Tarp, S. (2008), *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen.
- Temmerman, R. (1997), *Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology*. In: *Hermes – Journal of Linguistics* 18, 51–90.
- Temmerman, R. (2000), *Towards new ways of terminology description. The sociocognitive approach*. Amsterdam/ Philadelphia.

- Tomaszczyk, J. (2012), *Nauczanie angielskiego języka specjalistycznego informacji naukowej i bibliotekoznawstwa*. In: *Praktyka i teoria informacji naukowej i technicznej XX* (1), 44–54.
- Trimble, L. (1985), *English for Science and Technology. A discourse approach*. Cambridge.
- Tsakona, V. (2007), *Bilingualisation in practice: Terminological issues in bilingualising a specialised glossary*. In: *International Journal of Lexicography*, 20 (2), 119–145.
- UNESCO, *Glossary of basic terminology on disaster risk reduction*. (URL <http://unesdoc.unesco.org/images/0022/002257/225784e.pdf>). [Accessed: 12.10.2017].
- 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.
- Weisser, M. (2017). *Specialised corpora*. Date of access: 11.10.2017, http://martinweisser.org/corpora_site/spec_corpora.html.
- West, M. (1953), *A General Service List of English Words*. London/ New York.
- Wright, S.E./ G. Budin (1997), *Handbook of Terminology Management Vol. 1. Basic Aspects of Terminology Management*. Amsterdam/ Philadelphia.
- Wright, S.E. (1997), *Term Selection: The Initial Phase of Terminology Management*. In: S.E. Wright/ G. Budin (comp.), *Handbook of Terminology Management Vol. 1. Basic Aspects of Terminology Management*. Amsterdam/ Philadelphia, 13–23.
- Yaroshenko, O.L. (2014), *Principles of lexical minimum selection for teaching general technical English to engineering students*. In: *Advanced Education 2*, 104–110.
- Zgółkowa, H. (2013), *Słownik minimum języka polskiego*. Kraków.