Terminology Structuring for Learner's Glossaries*

Boyan Alexiev

Dept. of Applied Linguistics University of Architecture, Civil Engineering and Geodesy, Sofia 1046, Bulgaria, boj_fld@uacg.bg

Dr. Boyan Alexiev was born in 1949 in Sofia. He graduated in English Philology from St. Kliment Ohridski University of Sofia in 1975 and is presently associate professor at the Department of Applied Linguistics of the University of Architecture, Civil Engineering and Geodesy in Sofia. His main research interests are focused on terminological metaphor, Language for Specific Purposes (LSP) teaching methodology and terminology structuring for specialized lexicography. He has contributed over 40 publications in terminology, technical translation and English for Specific Purposes (ESP), is the co-author of three bilingual terminological dictionaries and the author of one ESP textbook.

*Acknowledgment: I am very grateful to Prof. Nazarski and Assoc. Prof. Dimov from UACG for their invaluable help with the conceptual analysis for identifying and organizing the glossary terms.

Alexiev, Boyan. **Terminology Structure for Learner's Glossaries.** *Knowledge Organization, 33(2)* 96-118. 24 references.

ABSTRACT: This study presents a methodology for compiling corpus-based learner's glossaries designed for non-specialist translators and Language for Specific Purposes (LSP) learners. The need for such bilingual microglossaries on subsections of a subject field for LSP teaching and translation purposes is emphasized in the Introduction. The concept"learner's glossary' is delimited among other types of terminological collections such as a specialised dictionary, a thesaurus and a term bank in Section 2 where the information categories in an entry are specified (keyterm plus translation equivalent(s), definition of keyterm and exemplary context, narrower terms with synonyms, definitions and translation equivalents; special phrases based on keyterm collocations with translation equivalents and exemplary contexts). Section 3 describes the principles and working methods of modern terminography as well as the source materials available for the glossary compilation. A hybrid term extraction technique is also described in that section and is used to extract the candidate terms with subsequent manual initial processing of the results. In Section 4 the theoretical grounds and methodology for analysing the conceptual relations in a terminological system are presented including the expert validation of the automatically extracted terms as a first phase in that process. Rationale for applying a lexico-semantic analysis in identifying collocational information on the keyterms is provided in Section 5 which is proposed to involve descriptions of the actantial structure of a keyterm for identifying verbal (T+V) collocations and of the paradigmatic and syntagmatic morpho-syntactic relations applicable to a keyterm. Finally, a model for structuring a learner's glossary entry is proposed in Section 6.

1. Introduction

The term "terminology," according to its etymology, is supposed to mean "the study of terms." However, the first usage of that term is recorded as referring to a specialised vocabulary belonging to a special subject field and compiled as a result of a particular terminological activity. In our modern computerized world such an activity is highly automated and based mostly on large machine-readable corpora. This activity is generally performed by following certain theoretical and methodological principles developed within the framework of modern terminology theory. Terminologies come under various forms (indexes, thesauri, termbanks, specialised dictionaries, glossaries etc.) and are designed to meet the needs of translation, Language for Specific Purposes (LSP) teaching, information retrieval, controlled indexing, document consulting and navigation, technical authoring, or merely to help the understanding of technical documents.

The latest achievements in computer-based terminological data processing have particular relevance for both mono- and multilingual terminographic

projects. The implementation of such a project requires a preliminary choice of a suitable theoretical framework and methodology. This choice is determined by two main factors:

1. The intended use of the respective specialised vocabulary.
2. The current situation on the market of specialized vocabularies in the country of publication.

Let us take Bulgaria as an example to illustrate the importance of these two factors for the proper development of terminological and, in particular, multilingual terminographic practices in this country. First, if we consider the terminological dictionaries that have been published here over the last 10-15 years, we will ascertain that most of these are bilingual dictionaries intended mainly for translation purposes. The adopted principle of ordering the information categories in a dictionary entry can be defined as the "alphabetical-nest principle," which, following the Russian lexicographic practice, consists in ordering alphabetically first the generic terms which form "nests," e.g. *fault* (the examples provided are taken from the *English-Bulgarian Dictionary of Mining and Geology* compiled by myself and two subject specialists) and then within the nests, the specific/species terms belonging to the same"nest,' e.g. *compound fault, cable fault, dead fault, earth fault, reverse-slip fault*, etc. Although this hierarchical arrangement looks neat and knowledge-based, one can immediately detect the striking conceptual incompatibility between, let us say, *dead fault* which is a geological concept, and *earth fault* which is an electrical engineering concept. This differentiation has been made explicit in the different translation equivalents provided for the head term *fault* (1. large crack in the Earth's surface; 2. a defect in an electric circuit) but within the "nest" itself the distinction is completely lost. In our opinion, a solution to this problem should be sought in compiling not large bilingual dictionaries but rather, didactically organized bilingual microglossaries containing definitions and highlighting hyponymy, context, collocation, usage as well as grammatical, lexical and semantic information essential to accurate translation. Such a glossary, for instance, will present *fault* only in one of its senses depending on the narrow domain (subfield) chosen to use for term extraction. Moreover, what is very essential for translators is that each entry will contain verbal, nominal and adjectival term collocations whose rendering in the target language often poses serious problems in the course of translating a technical text. For example, it is important for a Bulgarian translator who is grappling with an English text on *Concrete Technology* to be able to consult a bilingual glossary that contains in the same entry a definition plus the following information (taken from existing dictionaries):

– Concrete is "cast/placed/laid/poured" in English but only "laid/poured" in Bulgarian;
– Concrete "sets" in English but "bonds (its constituents together)" in Bulgarian;
– Concrete "bleeds" (note the metaphor!) in English but "cement paste comes out on the concrete surface" in Bulgarian (note the length of the translation equivalent of "bleeding" given in existing bilingual civil engineering and architecture dictionaries!), etc.

Second, if we consider the current situation on the market of specialized dictionaries in Bulgaria, we can conclude that the demand for large bilingual terminological dictionaries has partially been met. Nevertheless, those dictionaries do not only suffer from the defects mentioned above but being also voluminous, they require the efforts of several terminographers to compile. We believe that a solution to this problem can be found if terminographers start a new practice of compiling bilingual microglossaries based thematically on smaller sections of a discipline. For example, several English-Bulgarian microglossaries can be compiled in the field of Structural Engineering, namely, Construction Materials; Timber, Plastic and Steel Structures; Reinforced Concrete Structures; Foundations; Construction Elements; Bridges; Construction Design; Construction Technology; Building Mechanization. These small glossaries will be affordable not only to professional translators but also to civil engineering students who will certainly find them helpful in preparing their assignments in both English and other specialized subjects.

The aim of this study is to propose a methodology for compiling corpus-based learner's glossaries designed for non-specialist translators and LSP students. The tasks envisaged towards achieving the aim involve theoretical inferences and methodological conclusions made on the basis of discussing the following subtopics:

a) Specifications for a learner's glossary;
b) A corpus-based approach to terminography;
c) Conceptual analysis of terminological data;

d) Lexico-semantic analysis of terminological data; and,
e) A model for structuring a learner's glossary entry.

Each of these subtopics will be considered in a separate section below. Before proceeding to these discussions, it is necessary to specify the meaning of the term "terminology structuring." Grabar & Zweigenbaum (2004) propose the following definition: "organizing a set of terms through semantic relations … given a set of terms, obtained from an existing resource or extracted from a corpus, it consists in identifying hierarchical (or other types of) relations between these terms." Within the context of this study, we will use the term "terminology structuring" with the broader meaning of gathering and organizing the vocabulary needed for compiling specialised terminological collections.

2. Specifications for a learner's glossary

As has already been mentioned, we will try to develop a methodology for designing a learner's glossary. The determiner of this complex term suggests unambiguously the intended type of user of that terminological compilation product. The analogy with the term "learner's dictionary" is quite obvious. In this section we will specify the concept "learner's glossary" by first summarising the characteristics of the commonly accepted notion "learner's dictionary." We will then clarify the difference between specialised dictionaries, glossaries, thesauri and term banks. Finally, we will propose the information categories envisaged to be presented in a bilingual terminological learner's glossary.

2.1 Characteristics of a learner's dictionary

Most monolingual learner's dictionaries are designed for advanced learners. They presuppose that learners of a language should gradually move from using a bilingual dictionary to using a monolingual dictionary as they advance in their study of the foreign language. General purpose dictionaries are compiled for native speakers and conceived as being too complex and indeed confusing for their needs. The information that learner's dictionaries usually include is semantic (definitions), grammatical, pragmatic (usage), etc. They emphasize the importance of context and collocation and provide common errors, false friends and so on, which a native speaker knows intuitively. Easy-to-understand pronunciation symbols and a variety of examples are very important components of such a dictionary. On the other hand, these dictionaries do not include etymology and quotations. Such dictionaries are essentially designed to improve the learners' vocabulary and to help them in constructing their own sentences. The power of such a dictionary lies in its complete and clear definition of an entry. They are seen as the sources of the "best" language and provide an authoritative guide to usage.

The brief description of a learner's dictionary presented above suggests that the design of the learner's glossary entry to be proposed in this study should have at least the following main characteristics manifested in any learner's dictionary:

- Complete and clear definitions of the entry items possibly following a pre-designed defining pattern;
- Information on the most typical collocates of a glossary entry; and,
- Contextual examples of the actual usage of a glossary entry.

A bilingual learner's glossary of terms used in a subfield will also have to present additional conceptual (taxonomic, meronymic, associative etc.) and other semantic relationships which can be captured from existing documents and running text corpora by performing conceptual and linguistic analyses described in Sections 4 and 5 below. To better clarify the content and form of the glossary envisaged we will delimit the concept "glossary" within the range of other types of terminological collections such as specialised dictionaries, thesauri and term banks in Subsection 2.2 below.

2.2 Differences between specialised dictionaries, glossaries, thesauri and term banks

It is interesting to note that terminologists, maybe even more so than subject-field specialists, have problems with agreeing on the terminology that they use within their own discipline. One pertinent example in this respect is the inconsistent use of the terms "dictionary" and "glossary." GLOSSARIST, a searchable and categorized directory of glossaries and topical dictionaries, tries to distinguish between a dictionary and a glossary as follows (www.glossarist.com/default.asp, emphasis added):

> Theoretically, a dictionary is a collection of words and definitions about *all subjects* in a par-

ticular language. Having said this, many of the collections of definitions on websites are *called dictionaries but should really be called glossaries as the list of definitions is restricted to specific subject/s*.

On the one hand, this statement emphasizes the major distinguishing characteristic, namely, that a glossary, unlike a dictionary, provides definitions that are restricted to a specific subject. On the other hand, this very general distinction does not take into account the fact that there exists the so-called "terminological dictionary" defined as "Collection of terminological entries presenting information related to concepts or designations from one or more specific subject fields" (ISO 1087-1 2000). One solution to the problem of differentiating between the two closely related concepts is to present a list of various definitions of both concepts and compare them thus identifying the distinguishing characteristics.

2.2.1 Definitions of the concepts "dictionary" and "specialised dictionary"

About 30 definitions can be retrieved from the Web by the search pattern "define: dictionary." We have singled out the following ones that represent the most salient features of this type of lexicographic collection:

1. A reference book containing an alphabetical list of words with information about them.
2. A book containing words alphabetically arranged along with information about their form, pronunciation, functions, etymologies and meanings.
3. A reference source that provides meanings of words and other information. Specialised dictionaries are available for many subject areas.
4. A dictionary is a list of words with their definitions, or a list of words with corresponding words in other languages. Many dictionaries also provide pronunciation information, word derivations, histories, or etymologies, illustrations, usage guidance, and examples in sentences. Dictionaries are most commonly found in the form of a book.

We will add two more definitions taken from authoritative sources, namely the *New Shorter Oxford English Dictionary* and ISO 1087 (1990):

5. A book explaining or translating, usually in alphabetical order, words of a language or languages, giving their pronunciation, spelling, meaning, part of speech, and etymology, or one or some of these.
6. Structured collection of lexical units with linguistic information about each of them.

Google provides only one definition for "specialized dictionary:"

7. A specialized dictionary is a dictionary that covers a relatively restricted set of phenomena. The typical type of specialized dictionary is that which in English is often referred to as a technical dictionary.

It does not offer any definition for either "technical dictionary" or "terminological dictionary" and only one for "LSP dictionary:"

8. A Language for Specific Purposes dictionary is a dictionary that intends to describe a variety of one or more languages used by experts within a particular subject field. The discipline that deals with LSP dictionaries is usually called specialised lexicography and is a branch of lexicography.

Sandro Nielsen is the first lexicographer to suggest a truly lexicographic approach to defining a dictionary in contrast to the traditional linguistic approach. He defines a dictionary in terms of its major features, and a dictionary has three such features: A dictionary is a lexicographic reference work that has been designed to fulfil one or more functions (its pure potential), contains lexicographic data supporting the function(s), and contains lexicographic structures that combine and link the data in order to fulfil the function(s). This definition applies to printed, electronic and Internet dictionaries (Nielsen 1990):

9. Specialized dictionaries (also referred to as technical dictionaries) focus on linguistic and factual matters relating to specific subject fields. A specialized dictionary may have a relatively broad coverage, e.g. a picture dictionary, in that it covers several subject fields such as science and technology (a multi-field dictionary), or their coverage may be more narrow, in that they cover one particular subject field such as law (a single-field dictionary) or even a specific sub-field such as contract law (a sub-field dictionary). Specialized dictionaries may be maximizing dictionaries, i.e. they attempt to achieve comprehensive coverage of the terms in the subject field concerned, or they may

be minimizing dictionaries, i.e. they attempt to cover only a limited number of the specialized vocabulary concerned. Generally, multi-field dictionaries tend to be minimizing, whereas single-field and sub-field dictionaries tend to be maximizing.

Nielsen points out that a description of LSP dictionaries should focus on the number of lemmata (head words) and the number of fields covered by a dictionary. The distinction between maximising and minimising dictionaries, as well as that between multi-field and single-field dictionaries, the latter subdivided into general-field and sub-field dictionaries is important for a number of reasons. First of all, a single-field dictionary is an example of a very specialized dictionary in that it covers only one single subject field. Examples of single-field dictionaries are a dictionary of law, a dictionary of economics, a dictionary of welding, a dictionary of civil engineering etc. The main advantage of single-field dictionaries is that they can easily be maximizing dictionaries, i.e. attempt to cover as many terms of the subject field as possible without being a dictionary in several volumes. Consequently, single-fields dictionaries are ideal for extensive coverage of the linguistic and extra-linguistic aspects within a particular subject field. Secondly, if the lexicographers intend to make a bilingual maximizing single-field dictionary they will not run into the same problems with the space available for presenting the large amount of data that has to be included in the dictionary, cf. a multi-field dictionary. Therefore, the best coverage of linguistic and extra-linguistic aspects within the subject field covered by a dictionary will be found in a single-field dictionary. However, even more extensive coverage is possible in a sub-field dictionary. The learner's glossary we will propose comes very close to the latter type of dictionary in terms of coverage. However, it is envisaged as a glossary and not a dictionary, a distinction to be made further in this study.

We will now analyse the definitions cited above according to a methodology for definition analysis we proposed in Alexiev (2004). For the purposes of analysing the definitions concerned, we will apply the following procedures:

1. Restricting the analysis to the analytical type of definition (Definiendum = Genus + Differentia) formally recognised in terminology theory (Sager 1990, 42);
2. Presenting the conceptual structure of definitions as relations, i.e. predications, assuming that Subject = Definiendum (concept to be defined) and Predicate = Relator + Feature(s). Two basic relations hold between the Definiendum and the other conceptual components of the definition:

   – Definiendum (D) *is a type/part of* Genus (G) (genus predication);
   – Definiendum (D) *is characterised by* Feature (F) (differentia predication)

A predication is here interpreted as a binary structure consisting of a Subject represented by the Definiendum and a Predicate expressed as a judgement made about the Subject. The deep predications in a terminological definition are classified into two types, i.e. genus and differentia predications and the sum total of all identifiable predications constitute a model for the conceptual structuring of a term in line mainly with the Referent-oriented Analytical Concept Theory proposed by Dahlberg (1981, 1988) This model relates the predications constituting the concept to their verbal expression in the term definition thus enabling us to develop a procedure for identifying concept characteristics in terminological definitions.

3. Matching deep predications to the surface structure of definitions for identifying concept characteristics;

This procedure takes into account the main syntactic constituents of a typical terminological definition. The usual definition is represented by a main clause consisting of a subject (definiendum) + a predicate (be + NP/noun phrase/). The NP formula is:

NP = premodifier + N (kernel) + postmodifier

Kernels are usually premodified by adjectives and participles and postmodified by relative clauses or sometimes prepositional phrases, adverbs, etc., which can easily be expanded to relative clauses. First, the *D is a type/part of G* deep predication is matched to the Subject + Kernel surface part of the definition thus determining the genus characteristic. The *type of* relation is often not lexicalised. Then we match the *D is characterised by F* deep predications to the surface premodifier and postmodifier parts of the definition thus determining the differentia characteristics.

By applying the methodology described above we can extract the following genus and species characteristics from definitions 1-6 above:

a) Genus characteristic (variants): book (4 cases), reference source (1 case), collection (1 case); these results show that a dictionary is generally construed to be a type of book, i.e. a book-length lexicographical/terminographical collection.
b) Species characteristics (presented in a generalised manner):
   – contains alphabetically arranged words (lexical units);
   – provides information about: meanings (definitions); form (spelling, part of speech, verbal forms, derivations, etc.); pronunciation; functions; etymologies; usage; context.

For a specialised dictionary we can add the following species characteristics extracted from definitions 7-9 above:

– covers a relatively restricted set of phenomena;
– describes a variety of one or more languages used by experts within a particular subject field;
– focuses on linguistic and factual matters relating to specific subject fields;
– can be: multi-field, single-field or sub-field depending on the number of fields/subfields covered; maximizing and minimizing depending on the number of terms covered.

In summary, a specialised dictionary has the following specifications:

– book-length size;
– alphabetical arrangement of entries;
– coverage of a special field, a subfield or a number of special fields (e.g. dictionary of civil engineering, dictionary of building, dictionary of science and technology, etc.);
– coverage of terms can be comprehensive or limited;
– terms given in one language with definitions (monolingual) or with their equivalents in other languages (multilingual); and,
– coverage of linguistic and extra-linguistic aspects within the subject field (synonyms, grammatical information, usage, context, etc.).

2.2.2 Definitions of the concept "glossary"

The following definitions of "glossary" have been retrieved from the Web:

1. An alphabetical list of technical terms in some specialized field of knowledge; usually published as an appendix to a text on that field.
2. Short list of words related to a specific topic, with brief definitions, arranged alphabetically and often placed at the end of a book.
3. An alphabetical listing of special terms as they are used in a particular subject area, often with more in-depth explanations than would customarily be provided by dictionary definitions.
4. An alphabetical list of terms, limited to a special area of knowledge, with accompanying definitions.
5. An alphabetical list of abstruse, obsolete, unusual, technical, or other terms concerned with a subject field, together with definitions.
6. A glossary is an alphabetical list of words or expressions and the special or technical meanings that they have in a particular book, subject, or activity.
7. An alphabetical list of words and their meanings or interpretations (glosses) in various contexts. In the translation/localization industry, it may refer simply to a bilingual or multilingual terminology list and is often confounded with dictionary.

The following two definitions have been taken from authoritative print sources:

8. A list of words or terms and their definitions or other explanation of their meanings (De Bessé et al 1997); and,
9. A glossary is essentially a list of terms in one or more languages. The amount of information contained in glossaries can vary greatly, and the level of detail in any glossary will usually depend on the purpose for which it is intended. Thus, at one end of the spectrum, the most basic glossary will simply contain list of terms and their equivalents in one or more foreign languages. … At the other end of the glossary spectrum, you will find richly detailed glossaries containing definitions, examples of usage, synonyms, related terms, usage notes, etc. (Bowker and Pearson 2002, 137-38).

By applying the same analytical procedure as that used to identify the characteristics of the concept "dictionary," we obtain the following genus and dif-

ferentiae (differentiating characteristics) of the concept "glossary:"

a) generalised genus characteristic: list of terms (all 9 cases);
b) generalised species characteristics:
   – alphabetically arranged entries;
   – in a specialised field or usually subfield;
   – with definitions or other explanation;
   – often published as an appendix to a text;
   – can be multilingual; and,
   – can contain rich semantic and pragmatic information (examples of usage, synonyms, related terms, usage notes, etc.).

These specifications emphasize two basic differences between a specialised dictionary and a glossary. The first one lies in the size of the respective type of terminological collection. Glossaries are definitely much smaller in size as compared to dictionaries because they usually cover basic terms from a narrow subfield. The second one consists in the proportion of the different type of information contained. Dictionaries in general and specialised dictionaries in particular tend to present more linguistic data as compared to glossaries which are more semantically and pragmatically biased.

2.2.3 Definition of the concept "thesaurus"

This type of terminological collection is easy to distinguish from the other two types discussed above because unlike a simple list of accepted keywords, a thesaurus is a hierarchical list. That is, it displays not only the terms but also the relationships to other terms - Broader, Narrower, or Related. A thesaurus also provides cross references from synonyms to the "official" terms. About 30 definitions of the term "thesaurus" can be found in the Google search engine but we will specify its meaning by following a different path to the one followed in order to arrive at the characteristics of the concepts "specialised dictionary" and "glossary" above. We will resort to the *Transportation Research Thesaurus* (*TRT*) which covers all modes and aspects of transportation. The *TRT*'s purpose is to provide a common and consistent language between producers and users of transportation information. It is available online at http://trt.trb.org/trt.asp. We will refer to *TRT* again in Section 4 where it will be used as a valuable source of information required for the conceptual analysis of the terminological data to be processed so as to be used in constructing the learner's glossary envisaged. Now we will focus our attention on the definition of the term "thesaurus" given in the Glossary appended to the *TRT*:

> A controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators should be employed reciprocally.

The terms "controlled vocabulary" and "relationship indicator," in turn, are defined in the following way:

1. A list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary must have an unambiguous, non-redundant definition
2. A word, phrase, abbreviation, or symbol used in thesauri to identify a semantic relationship between terms.

To make the specification of "thesaurus" clearer and unambiguous, we will provide some examples from the subfield of Construction Materials:

Efflorescence
- **Broader** terms: Surface defects
- **Related** terms: Weathering, Leaching, Staining ….

Lightweight aggregates
- **Narrower** terms: Polystyrene beads, Perlite, Vermiculite …

Exposed Aggregate Concrete
- **Used for**: Aggregate transfer method
- **Broader** terms: Architectural concrete
- **Related** terms: Concrete finishing, Concrete finishes, Decorative aggregates

Flowable Fill
- **Use**: Controlled Low-Strength Materials (CLSM)

Microsilica
- **Use**: Silica fume

Young's modulus
- **Use**: Modulus of elasticity

2.2.4 Term banks

A term bank is "an automated collection of the vocabularies of a subset of specialised knowledge created to serve a particular user group" (Sager, 1990). A typical entry contains the term itself and its synonyms, together with definitions, explanatory notes, references, etc. For the purposes of this study we will present in summary the basic data categories that have been internationally agreed upon as essential for a sound terminological record:

ENTRY TERM/ET (the full form of a term or expression)

Conceptual specification:

a) Definition (links the entry term to the concepts which it represents, can be in a style specific to the term bank/TB);
b) Relationships (indicate the most obvious broader term (BT) to the entry term (ET) and also the type of relationship that exists between these two terms, e.g. generic, partitive, other (e.g. NT=narrower term, RT=related term, etc));
c) Subject field (general field, subfield. *Note*: Terminology is divided by subject field before it is ordered in any other way); and,
d) Scope note (a further specification of subject or register and intended to indicate a special field of application of the ET, e.g. a term specific to one particular model of a motor car, or a process which is tied to a particular type of machine).

Linguistic specification:

a) Grammatical information (spelling/s/, pronunciation/s/, gender of nouns /m, f, n/, parts of speech /n, v, adj, adv/, principal parts of verbs /inf., past, pp), transitivity (tr., i.,) special plural or other forms, etc.);
b) Language (a two-letter language code of ISO is used, e.g. BG, followed, if necessary, by a slash and the country code); and,
c) Parallel information categories to the entry term (e.g. spelling variant(s), full synonyms/ full substitutes for ET), abbreviated form.

Pragmatic specification:

a) Context (exemplification of the ET usage in a segment of running text); and,
b) Usage note (information about the ET usage in context that cannot be provided in the form of examples). The following markers are usually used: colloquial (usually spoken language but also found in documents); obsolete (no longer in current usage); slang (spoken in only very restricted usage of great familiarity and casualness of situation); mandatory (prescribed usage for the text type), firm-specific (used exclusively by the firm, organization, etc.); standardized (as generally prescribed by a national or international standard or other authoritative body for a particular usage in which case the authority should be cited); preferred/deprecated variant; translation (coined only as a translation equivalent without any claim to general acceptability).

Source reference specification: the sources of definitions, contexts, translation equivalents, synonyms are usually recorded in term banks.

Administrative *data:* ET record number, author + date of record, information about updates.

2.3 Scope and data categories proposed for a bilingual learner's glossary

In the previous subsection we specified the essential differences between the major types of terminological collections depending mainly on the information categories contained in each type. This differentiation makes it possible to locate the learner's glossary we propose among those collections by specifying its parameters. We propose that the latter, in view of the prospective users of the glossary (non-specialist translators and LSP learners) could be determined as follows:

1. Number of entries – up to 70 terms;
2. Thematic scope – a narrow subfield;
3. Information categories:
   – keyterm
   – synonyms of keyterm (if any)
   – translation equivalent(s) of keyterm
   – keyterm definition
   – exemplary context(s) for keyterm
   – narrower terms (types) to keyterm
   – synonyms of narrower terms to keyterm (if any)
   – definitions of narrower terms to keyterm
   – exemplary context(s) for narrower terms to keyterm
   – keyterm collocations with variants

– translation equivalents of keyterm collocations
– exemplary contexts for keyterm collocations

3. A Corpus-based approach to terminography

The concept "corpus" is used in modern linguistics to refer to both running text and lists of lexical items excerpted from running text for various, including terminographic, purposes. In this sense, when we talk about a corpus-based approach to developing the terminographic project that we will propose and name provisionally *English-Bulgarian Learner's Glossary of Concrete Terms*, the corpora used for extracting the relevant terminological data will include both printed materials (textbooks, dictionaries, etc.) and electronic sources of any kind relating to the chosen subfield. In this section we will first present briefly the principles and working methods of modern terminography. (In fact, such a bilingual learner's glossary would be too narrow in scope for our teaching and translation purposes and is suggested here only as a model glossary for exemplifying our methodological approach. Although glossaries of concrete terms actually exist (see below), a realistic terminographic project for our conditions would be *An English-Bulgarian Learner's Glossary of Construction Materials*.) Then we will propose an approach to corpus collection for learner's glossary compilation. Finally, we will describe a term extraction tool for initial processing of the terminological data contained in the textual corpus collected.

3.1 Principles and working methods of modern terminography

Terminographic activity involves "the study and practice of describing the linguistic, conceptual and pragmatic properties of terminological units of one or more than one language in order to produce reference works in printed or electronic form"(De Bessé et al 1997). It cannot be performed by individual specialists but is governed by internationally agreed recommendations.

3.1.1 Principles of terminography

Cabré (1999, 115-116) formulates certain theoretical principles of terminography which we will summarise as follows:

1. Terminology work consists in gathering the designations used by specialists to refer to concepts and proposing alternatives for inadequate ones;
2. Terms for a glossary should be collected from real texts and not invented or created;
3. Terminological gaps in a subject field should be filled in by neologisms;
4. Terminography is guided by the principle that terms are indivisible units of form and content; and,
5. There are guidelines and recommendations published as standards by international committees for unifying designations and concepts as well as for the methods to be applied for the presentation of terminological data.

A terminographer who undertakes a terminological project should determine the type and content of the materials to be used for extracting the relevant terminological data and process them according to established working methods. We will present in summary the types of source material generally used and the working methods applied in terminography, following mainly Cabré (1999) but also referring to some recently published ISO International Standards concerning the recording, maintenance and retrieval of terminological information.

3.1.2 Materials used in terminographic projects

The type and content of materials used for extracting terminological data necessary to build up a terminological collection is an essential element in the overall implementation of a terminographic project. Cabré (1999, 116) distinguishes between three types of source material generally used for terminology processing: reference works, specific documents and support materials. These are:

(a) Reference works, i.e. documents for obtaining background information on theoretical, methodological, practical or bibliographical aspects on the subject field. The information contained in these documents refers to the conceptual system of a given special domain, the corresponding system of designations as well as additional aspects of the professional activity associated with that field. The reference materials may involve terminological works on the same or related topic, dictionaries, glossaries, thesauri, terminological databases, etc. covering the terminology in question, handbooks and other background materials, etc. Other very important reference works available for terminography are the internationally agreed documents on the research method and

presentation of work, i.e. ISO standards on terminology.

(b) Specific materials for terminographic work involve all kind of textual or reference sources used in specialist communication as it is usually the specialists who create terms and thereby introduce designations considered suitable to the system of their language, into their special domain. Cabré (1999, 121) formulates the following requirements for high quality source materials used in terminography:
   – They should be representative of the subject matter in accordance with the objectives of the task and the delimitation of the topic;
   – They should be comparatively up-to-date both regarding the designations that experts really use and the topic;
   – They should be explicit enough to allow retrieval of information at any point in time.

(c) Support materials are the records used in systematic terminological searches, which can be classified into extraction records, terminological records and correspondence records. An extraction record normally has the following fields: an entry in ist "canonical" form, its grammatical category, the context in which it appears and the complete reference of the source document. Terminological records contain all the relevant information about a term, including the information extracted from the extraction records. International Standard ISO 12616 2002-Translation-oriented terminography (ISO 2002) recommends that the following general data categories should be recorded in a terminological entry which in translation-oriented terminography will contain one main entry term for each language:
   1. Data categories for terms and term-related information;
   2. Data categories related to concept description; and,
   3. Administrative data categories.

In some bilingual or multilingual databases in which the information is stored on separate records by language, a correspondence record can be used to correlate all the designations for a single concept.

3.1.3 Working methods used in terminography

The main characteristics of a systematic approach to terminographic work can be summed up as follows:

– Systematic is a combined onomasiological and semasiological approach, i.e. first going from concepts to terms and then, from terms to concepts;
– Terminological work should be carried out with two languages considered equally valid for naming technical and scientific realities, even though each one of these languages reflects a particular division of reality. Recognizing the terminological allomorphy between two different languages necessitates separate lexicographic descriptions of their terminological universes;
– At the final stage of terminological work the two descriptions will be brought together, when equivalents between terms and concepts will be established; and,
– If the target language has a gap in designations, it will be filled in by means of neology.

The most important advantage of this systematic approach to terminography is that it avoids qualitative gaps in the terminologies of the languages in question.

3.2 Source materials for learner's glossary compilation

As has already been mentioned in the previous subsection, the first phase of implementing a terminographic project of the type we envisage should involve collection of source materials. Taking into account Cabré's distinction between three types of source materials (see 3.1.2 above) as well as the specificity and limitations of the task we undertake in this study, we will propose a slightly different classification of the source materials required particularly for learner's glossary compilation. For our purpose, we propose three types of source materials for compiling learner's glossaries:

a) Reference terminological collections in both print and electronic format (mono-/multilingual terminological dictionaries, glossaries, thesauri, term banks);
b) Extraction corpora of running text in both print and electronic format; and,
c) ISO standards on terminology with recommendations on the presentation of terminological data.

We will present a list of the reference terminological collections which are accessible to us and can be referred to in the process of terminology structuring for the glossary compilation:

Printed terminological collections:

- Alexiev B. et al. 1998. *English Bulgarian dictionary of mining and geology*. Sofia: RATIO-90 Publishing House.
- Delev, K. 2001. *English-Bulgarian construction dictionary*. Sofia: ABC TECHNIKA.
- MacLean, J. H. and Scott, J.S. 1995. *The Penguin dictionary of building*. London: Penguin Books.
- Phillipova, M. et al. 1990. *English-Bulgarian dictionary of civil engineering and architecture*. Sofia: TECHNIKA State Publishing House.
- Phillipova, M. and Ivanov, L. 1998. *English-Bulgarian dictionary of civil engineering and architecture*. Sofia: VEZNI-4 Publishing House.
- Phillipova, M. & Ivanov, L. 1999. *Bulgarian-English dictionary of civil engineering and architecture*. Sofia: VEZNI-4 Publishing House.
- Scott, J. 1991. *Dictionary of civil engineering*. London: Penguin Books.
- Harrison, T. 2003. *Guidance on the use of terms relating to cement and concrete*. Berkshire: The Concrete Society. The British Cement Association.

Online terminological collections:

- *Glossary of concrete terms* http://*www.*moxie-intl.com/glossary.htm
- *Transportation Research Thesaurus* http://trt.trb.org/trt.asp (Links to follow: Materials-classes of materials-building materials – building materials by properties – concrete)
- *EuroDicAutom* http://www.europa.eu.int/eurodicautom/Controller
- *Termium* http://www.termium.gc.ca/

We used the following extraction corpora of running text, taken from both printed and Internet source materials to extract the candidate terms for our glossary. In our opinion, the materials used for extracting the terminological data required for compiling our glossary meet all three criteria for high quality source materials formulated by Cabré (see above):

- Marotta, T.W. 2005. *Basic construction materials.* (university textbook), Chapter 4 (Portland Cement Concrete)–scanned text, number of words – 27,487
- *Concrete basics*. An Internet Guide to Concrete Practice / Cement Concrete and Aggregates, Australia (www.concrete.net.au/pdf/concretebasics.pdf) number of words – 11,976
- *Cement applications concrete*. Holcin Cement Institute (www.hlci.lk) – number of words – 6221
Total number of words – approx. 46 000.

Three *ISO standards on terminology* and one *European standard on Concrete* can be used as reference tools for the presentation of the terminological data in the learner's glossary envisaged:

- **ISO 1990:** ISO 1087. *Terminology – Vocabulary.*
- **ISO 2000:** ISO 1087-1. *Terminology – Vocabulary-Part 1:Theory and application*.
- **ISO 2002:** ISO 12616. *Translation-oriented terminography*.
- **Concrete BS EN 206-1** (Bulgarian version)

3.3 Automatic term extraction

The compilation of specialised terminological collections is nowadays considered impossible without using term-extraction tools for identifying terms in domain-specific corpora. The term-extraction techniques can generally be classified into linguistic, statistical and hybrid. Below we will describe briefly the hybrid term extraction technique *TermoStat* designed by Drouin (2003) which we found to be suitable for the initial processing of our specialised extraction corpora. Then we will present the results of our extraction corpora processing and analyze them.

3.3.1 The TermoStat *term-extraction tool*

The software tool identifies not only complex (multi-word) terms such as *reinforced concrete, cement paste*, etc. but also simple (single-word) terms (also known as uniterms) such as *concrete*, *admixture*, etc. The tool extracts corpus-specific lexical units using a statistical technique that compares frequencies in a technical and a non-technical corpus. A virtual corpus, called global corpus (GC) is built at run time from a reference corpus (RC), i.e. a non-technical corpus and an analysis corpus (AC), i.e. a domain-specific corpus. The behaviour of the lexical units inside the dynamically built GC is compared and the

items specific to the AC are identified. The RC is composed of 13,746 articles taken from *The Gazette*, a Montreal-based newspaper containing a total of approximately 7,400,000 tokens (separate words) which correspond to roughly 82,700 word forms. Domain-specific terminology is brought out by statistical comparison of the frequencies observed in the two corpora (RC and AC). The statistical measure takes into account the frequencies as observed in both corpora thus quantifying the deviation from a normal distribution. This simplified description of *TermoStat* is deemed sufficient for the purposes of this study. What is important to emphasize here is the need for evaluating the quality of the output of *TermoStat*. For the purpose, Drouin proposes a two-step validation process. The first step involves automated validation by comparing the identified subset of the lexicon with a list of terms found in a terminology database dedicated to the field/subfield in question. The second step consists in human validation by subject specialists in terms of two main criteria: 1) the candidate term(s) is/are representative of the domain; and 2) the candidate term(s) is/are representative of the main topic of the corpus.

Drouin has tested *TermoStat* with three analysis corpora by using the two types of validation and has obtained an overall precision of 81% (the ratio of relevant items retrieved to the total number of irrelevant and relevant items). He concludes that by using a purely statistical approach some relevant lexical items will always remain unidentified. Therefore, he suggests that an additional level of tagging should be used that could take meaning into account. The results reported confirm the need for additional manual text processing.

3.3.2 Results of the extraction corpora processing

The three corpora of running text collected for the purpose of extracting terminological data were grouped in one file, namely, "concreteall" and submitted in a txt format to the software tool which analysed it within 1-2 min. We have to point out one very important feature of *TermoStat*, viz. its capacity to provide contexts on the candidate terms extracted which can be used as sources of definitional and collocational information. If we are interested in analysing collocations with some term, e.g. "concrete," we can arrange the items alphabetically by clicking once on the hyperlink "Candidate (root form)" thus facilitating the localisation of the "concrete" collocates. Then we can gain access to contexts on each candidate term collocation to be included in the respective entry in our learner's glossary.

These contexts provide different types of information that can accompany the respective collocation. For example, a click on the collocation "concrete placement" (see Sample 3 below) retrieves a number of different contexts, among which "For concrete placements during warm weather, a retarder is generally used" provides additional grammatical information, namely, the use of the nominalization "placement" in the plural. This function of *TermoStat* obviates the need for a concordancer, a software tool for identifying the occurrences of a particular word in its immediate contexts.

3.3.3 Manual processing of results

The candidate terms extracted by *TermoStat* are either nouns or nominal phrases since they have been previously POS-tagged. The selection of candidate terms (CTs) was made by moving top-down starting from the CT that has obtained the highest score. Two manual operations were performed while selecting the items to be subjected to further validation. The first operation, which will be called "noise" removal, consisted in removing the unwanted items retrieved during the corpus processing, such as "chapter," "psi," e.g. "concrete basics" etc. The second operation involved restricting the number of items to be extracted from the table for further processing. For the purposes of compiling the *English-Bulgarian glossary of construction materials* envisaged as a realistic terminographic project, we need a limited number of terms and terminological collocations specific to the *Concrete* topic. Items designating very broad special concepts that belong to a wide range of special domains are considered irrelevant and were consequently excluded. Among these are: *percent, surface, type, colour, material, test, sample, particle, standard deviation, chemical composition, cubic yard, unit weight, diameter,* etc.

The manual selection of the items to be submitted to a subject specialist for further "fine" selection yielded 128 candidate terms. The results of the "fine" selection will be reported in the next section because we consider the expert assistance in terminology as part of the conceptual analysis of the terminological data collected for a terminographic project.

4. Conceptual analysis of terminological data

The next phase in compiling the learner's glossary requires the assistance of an expert or experts who

can validate the automatically extracted terms thus restricting the number of possible glossary entries and give advice on their organization. The final list can be specified by analysing the conceptual relations between the candidate terms using available reference materials and term definitions. These sources can also be used to identify some additional terms worth including in the glossary which are related hyponymically or meronymically to the keyterms (terms designating key concepts in the respective subdomain and used as entries in the glossary) but for some reason do not occur in the automatically processed data. In this way we can provide a reliable set of narrower terms (types) specified as a necessary information category in the glossary (see Subsection 2.3) as well as other terms expressing concepts that enter into partitive relations to the key concepts. All definitions that are to be used within the entries have to be designed in a unified manner following a number of recommendations and in compliance with certain conceptual and formal requirements.

4.1 Expert validation of automatically extracted terms

Two top civil engineering experts were asked to check the validity of the candidate terms extracted automatically by putting a question mark next to the candidate terms that they considered unsuitable for the project. The aims of our terminographic undertaking were explained to the experts by showing them a model of a glossary entry (see Section 6). They were also asked to specify which terms could be keyterms, i.e. terms to be represented as entries with the maximum possible range of information categories determined in Subsection 2.3. The results of the expert validation are presented in the table below:

Accelerator	*Mixer*
Admixture (*keyterm*)	Modulus of rupture
Aggregate (keyterm)	*Moisture*
Aggregate finish	Mould
Air entrainment	Permeability
Alkali	*Plain concrete*
Angular *aggregate*	Plastic *concrete*
Batch	Plastic sheet
Bicarbonate	Plastic *shrinkage*
Bleedwater	Plastic *shrinkage* crack
Broom?	Poker?
Calcium chloride	Poker vibrator?
Calcium hydroxide	Porous *aggregate*
Cement (*keyterm*)	*Portland cement*
Cement concrete	Power trowel
Cement mortar	*Precast concrete*
Cement paste	Pullout test
Cement placement	Pump line
Cement powder	*Ready mix*
Cement *mix*	Rebound hammer?
Clinker	*Reinforcement* (*keyterm*)
Coarse aggregate	*Reinforcement* bar
Coloured concrete	*Retarder*
Compacted concrete	Screed
Compaction	Scrub?
Compressive strength	Segregation
Concrete (*basic keyterm*)	Shear force
Concrete mix	Shotcrete
Concrete mix design	*Shrinkage*
Concrete pump	*Shrinkage* crack
Concrete strength	Silica fume
Concrete structure	*Slab*
Control joint	Slag
Crack	*Slump*
Crushed rock	*Slump* cone
Design *strength*	*Slump* plate
Dry *concrete*	*Slump test*
Durability (*keyterm*)	Sodium sulfate
Durable concrete	Specimen
Fine aggregate	Splitting *tensile strength*
Finishing	Stain
Flexural *strength*	*Steel* (*keyterm*)
Fly ash	*Steel reinforcement*
Formwork	Stiff *concrete*
Free lime	Stirrup?
Freezing	*Strength* (*keyterm*)
Fresh concrete	Subgrade
Ground *cement*	Sulfate
Hardened concrete	*Superplasticizer*
Hardening	*Tensile* force
Heavyweight concrete	*Tensile strength*
Hopper?	Thaw resistance
Hydration	Transit *mix truck*
Impurity	*Truck agitator*
Joint	*Truck mixer*
Kiln	Void system
Lightweight aggregate	*Water*
Lightweight concrete	*Water* reducer
Mesh	*Water*tightness
Microsilica	Wear resistance
Mineral *admixture*	Wet *concrete*
Mix	White *cement*
Mix design	*Workability*
Mixed concrete	*Workability* of cement

Table 1: Items selected from TermoStat *results*

4.2 Identifying conceptual relations between terms

The next phase in processing our terminological data should involve analysis of the conceptual relations between candidate terms in order to provide a sound basis for structuring our learner's glossary in an effective way following certain cognitive principles. We will first discuss the theoretical and methodological grounds for analysing conceptual relations between terms and then we will propose the respective analytical procedure.

4.2.1 Theoretical grounds for conceptual analysis of term data

The terms pertaining to a given subject field, subfield or even a topic within that subfield (cf. the "concrete" topic within the subfield of "building materials" as part of the field of Civil Engineering) are characterized by both internal and external systematicity (Popova 2005), the first type relating to the internal structuring of the terms in a terminological system and the second one to the communicative function of that system. A good terminographic project is necessarily based on a careful analysis of the internal (inherent) systematicity of the set of terms envisaged as entries in the respective terminological collection. Different terminologists propose different models for describing that systematicity. The representatives of the traditional Vienna School of Terminology focus primarily on taxonomic and meronymic relationships and are often criticized by proponents of alternative approaches for overlooking the multifaceted and multidimensional nature of terms whose relationships can also be described by using linguistic models, i.e. within lexico-semantic frameworks. We will propose such a framework for processing part of our term data in Section 5. However, we consider the conceptual analysis to be indispensable for structuring terminologies which can lay a solid foundation for identifying the proper entries for any type of terminological collection.

A conceptual analysis of terms for terminographic purposes should start with adopting a certain typology of conceptual relationships. This is not a very easy task since, as we have already mentioned above, there is a large variety of typologies proposed. For example, Felber (1980, 120) presents the following basic types of relationships (we do not present the subtypes): 1) Logical relationships; 2) Ontological relationships; and 3) Relationships of effect.

Another interesting and consistent typology of systematic relations between terms is proposed by Popova. Based on the conception of the two types of links between concepts, she postulates two types of systematicity (scheme of relations) among terms: implicational and classificational. The former consists of two subtypes, viz. partitive/meronymic, i.e. whole-part relations and associative, i.e. relations of contiguity between entities participating in a real situation semantically represented as a predicative "scene" where referents perform semantic "roles" (agent, object, result, purpose, etc.) assigned by the predicate. A similar actantial structure, but based on a different theory, will be used in our lexico-semantic analysis in Section 5 below. Therefore, with the view to preserving the distinction between the two types of analyses, we will adopt the classification of conceptual relations most frequently used in terminology (Sager 1990, 30-37) which we present below with certain modifications:

1. Generic (hyperonymic and hyponymic) relationships which establish a hierarchical order; a broader (generic) concept is superordinate to the narrower (specific) concept(s) and, conversely, the respective narrower concept is subordinate to the generic concept. It is important to note here that in certain cases it is necessary to indicate the criterion by which types have been declared. Such type indicators are known in information science as "facets." For example, *building materials* can be classified by properties: ceramics, composites, plasticizers, etc.; or by function: abrasives, adhesives, coatings, insulating materials, etc.
2. Meronymic/partitive relationships also referred to as "whole-part" relationships which indicate the connection between concepts consisting of more than one part and their constituent parts. For example, *cement* is a fundamental ingredient in *concrete*.
3. *Complex relationships* such as: cause-effect; material-product; material-property; material-state; process-product; process-instrument; process-method; process-patient; phenomenon-measurement; object-counteragent; object-container; object-material; object-quality; object-operation; object-characteristic; object-form; activity-place. For example, *aggregate*, *cement* and *water* are mixed (process) to produce *concrete* (product).

4.2.2 *Methodology for analyzing conceptual relations in a terminological system*

As a matter of fact every system of terminological units is structured around a basic keyterm from which all other terms in the system originate through complex branching of its characteristics in a certain hierarchical order. Hence, a conceptual analysis of the basic keyterm can be expected to yield the basic candidate terms to enter a glossary or any other terminological collection envisaged to cover that topic. Since concepts consist of characteristics, the analysis the conceptual structure of a term should involve specification of these characteristics. The conceptual characteristics of a term can be deduced from both its meaning (definition) and its contents (description). Term contents (detailed explanations or descriptions) can be found in some single-field maximizing monolingual dictionaries such as our reference terminological collections, *The Penguin dictionary of building* and *dictionary of civil engineering* which actually complement each other. The basic entry terms in these dictionaries are provided with italicized terms within the explanations which also occur as separate entries. Definitions can also be extracted from online glossaries, *EuroDicAutom* or other term banks. Some narrower terms can be specified by consulting a thesaurus such as the *Transportation Research Thesaurus* (see Subsection 3.2 above).

The next step in the analysis involves extraction of the characteristics which we will carry out by applying the simplified procedure for identifying concept characteristics in terminological definitions described in Alexiev (2004). Let us recall that the procedure (see Subsection 2.2.1 above) consists of three steps. The first is developing a classification scheme of the general aspects of the basic concept. In this respect the powerful search machine *Google* can be of much help. For example, we used the search pattern *materials science basics* to reach the Webpage of the MIT Course Catalogue of the Department of Materials Science and Engineering. In a description of a BSc course we can read: "All aspects of materials are considered, including their structure, properties, processing and performance." With the assistance of our expert advisors these aspects for the building material *Concrete* were reduced to *composition*, *properties*, *technology* and *use*.

One very important aspect of all concepts is their typology, i.e. their potential to be further subdivided into types which yield the narrower terms. The latter have been proposed as essential information categories to be presented in our bilingual learner's glossary. Some extended terminological definitions (explanations) actually give some types which can be supplemented by referring to a thesaurus (see above). We suggest that typology should be included as the first species characteristic to be identified since it is a major aspect of concepts and, in particular, of material concepts such as "concrete," etc. Hence, for our analysis we will refer to five general aspects of "concrete:" types, composition, properties, technology and use.

The second step is presenting general aspects as deep predications: concrete *is a type of* X (genus predication); concrete *is characterised by* TYPES, COMPOSITION, PROPERTIES, TECHNOLOGY, USE (species predications). And the third step is matching deep predications to the linguistic structure of definitions. In other words, the species characteristics will be identified by the five aspects specified above and presented as generalised (from all definitions) characteristics arranged in a hierarchical order.

We will then try to outline the configuration of the system by following the model of systematicity that Popova (2005) proposes for organizing a given set of terms in a system which has been successfully applied in compiling a Dictionary of Basic Terms (Volume 1 – Natural Sciences) (in Bulgarian). This model is based on two essential concepts, namely, a system-structuring characteristic and a system-structuring term. Popova defines the former as the characteristic which corresponds to a term in a given terminological set and the latter as the term to whose semantic structure this system-structuring characteristic belongs. The conceptual analysis then logically starts with the basic term in the system which we called "basic keyterm" since it gives rise to the other terms of the system in a hierarchical order vertically downward. The analysis then proceeds according to the adopted typology of systematic relations between terms. As we have mentioned above, our analysis will be performed on the basis of the most-frequently used classification of conceptual relations leading to identification of hyperonymic, meronymic and, where necessary, complex relations extracted from definitions and descriptions of the basic keyterm first, and then its hierarchically related terms. This will help us construct a terminological network of the basic keyterm "concrete" and compare our analytical results with the initially processed *TermoStat* results thus enabling us to make a final selection of our glossary entries, their possible synonyms and the narrower terms within these entries.

4.2.3 Analytical procedure

Step 1 – Extracting definitions and descriptions of the basic keyterm from reference sources

The following definitions and descriptions of "concrete" were extracted from our reference terminological collections (see Subsection 3.2 above):

1. A composite material which consists essentially of a binding medium within which are embedded particles or fragments of a relative inert filler in portland cement concrete, the binder is a mixture of portland cement, possibly additional cementitious materials such as fly ash and water; the filler may be any of a wide variety of natural or artificial, fine and coarse aggregates; and in some instances, an admixture. (Glossary of concrete terms)
2. A composition of cement, sand, gravel (aggregate) etc. which after mixing with water has the property of hardening into a stone-like solid, used in construction. (EuroDicAutom)
3. A mixture of water, sand, stone and a binder (nowadays usually Portland cement) which hardens to a stone-like mass. See aerated concrete, air-entrained concrete, lightweight concretes, prestressed concrete, reinforced concrete, vacuum concrete, creep, cube test, water/cement ratio, workability. (*Dictionary of civil engineering*)
4. A mixture of coarse and fine *aggregate* (sand and stone), *cement,* and *mixing water*, plus any *admixtures*, that sets within a few hours of mixing. Dense concrete is a versatile, cheap material, strong in compression, and with low *moisture movement*. Wall and floors have good *fire resistance* and no *fire hazard*. Concrete is highly alkaline, which prevents the rusting of the steel *reinforcement*. It is attacked by *sulphates*, many acids (acid rain, lactic acid), as well as by *salt air* or seawater. *Carbonation* can lower the alkalinity that prevents rusting. Ordinary concrete is unsuitable for severe exposure. Normal structural concrete for *in-situ* work is usually *readymix*, delivered fresh to enable *placing* in *formwork* before it sets, although silos containing dry mix, with a screw-feed water mixer, are also used in Germany. The use of *plasticizers*, a low *water/cement ratio*, and good compaction improve *durability* while reducing concrete's irreversible *drying shrinkage* and *creep*. The gain in strength with time is fairly slow; it needs *curing* and protection while it is *green*. (*The Penguin dictionary of building*)

Step 2 – Identifying system-structuring characteristics and terms

Genus characteristics (hyperonyms):
- composite building material;
- composite material;
- composition;
- mixture.

The four genus characteristics obtained form the four definitions exemplify the hyperonymic relations the basic keyterm "concrete" enters in. The fact that the concept "concrete" relates to more than one genus concept (e.g. "composite material" is at a higher level of abstraction than "composite building material" and "mixture" may not be a material at all) can be accounted for by the multidimensionality of concept systems. This problem will be solved in the next Subsection 4.3 where only one genus will be selected for representation in the glossary entry definition according to certain criteria for unification of terminological definitions.

Generalised species characteristics (identified by aspects):

1. Types (hyponyms):
   - *aerated concrete* (Glossary);
   - *air-entrained concrete* (TermoStat);
   - *lightweight concrete* (TermoStat);
   - *prestressed concrete* (Glossary);
   - *reinforced concrete* (Glossary);
   - vacuum concrete;

Some more types can be added by consulting the *Transportation Research Thesaurus*
- *cast-in-place concrete*;
- *cellular concrete* (Glossary gives synonyms – *aerated concrete, foam concrete, gas concrete*);
- expansive concrete;
- *fresh concrete* (*TermoStat*);
- high performance concrete;
- lean concrete;
- *mass concrete* (Glossary);
- microconcrete;
- porous concrete;
- *Portland cement concrete* (*TermoStat*);
- *precast concrete* (*TermoStat* and Glossary);
- pumped concrete;
- *ready mixed concrete* (*TermoStat*);
- roller *compacted concrete* (*TermoStat*);
- self-compacting concrete;
- tremie concrete.

It is evident that all types of concrete listed above cannot be used as narrower terms in our glossary entry "concrete" because of its limited size. Therefore, we have to look for some criterion for reducing their number. One reliable criterion can be their frequency and degree of usability that we can establish by checking them against the automatically extracted candidate terms and the *Glossary of concrete terms*. The results of our check are given in round brackets after the terms occurring in these collections.

Additional candidate terms for our glossary can be provided by oppositions defined by Popova (2005) as a partial case of positional coordinated relations:

– *plain* vs. *reinforced concrete*;
– *in-situ* vs. *ready-mixed* vs. *precast concrete*;
– *lightweight vs. heavyweight concrete* (*TermoStat*).

Finally, by consulting the available English-Bulgarian dictionaries of civil engineering and architecture and *Termium* (the Canadian Term Bank) we identified some more synonyms such as:

– *plain concrete = ordinary concrete;*
– *reinforced concrete = ferroconcrete* (obsolete) = *steel concrete;*
– *in-situ concrete* (Br.)=*cast-in-place concrete* (Am.).

Now we have a good number of reliable candidate terms designating "concrete" types (all terms given in italics) *(*we consider all italicized candidate terms from the various schemes presented to be suitable for our glossary) that can be used as narrower terms to the basic keyterm *Concrete*. These terms have also been italicized in the *TermoStat* results which we use as a control list of candidate terms of proven high frequency of occurrence in running text.

2. Composition (meronyms)
  – *cement binder*
    – *Portland cement* (*TermoStat*, Glossary)
    – *cementitious materials* (Glossary)
    – *mixing water* (*TermoStat*)
    – *fly ash* (*TermoStat*, Glossary)
  – *aggregate* – fine & coarse (*TermoStat*, Glossary)
    – *sand* (Glossary)
    – *gravel* (stone)
  – *admixture* (*TermoStat*, Glossary)
    – *accelerator* (*TermoStat*, Glossary)
    – *retarder* (*TermoStat*, Glossary)
    – *calcium chloride* (*TermoStat*)
    – *air-entraining agent* (*TermoStat*, Glossary)
  – *additive*
    – *plasticizer* (*TermoStat*)
    – *superplasticizer* (*TermoStat*)

As can be seen above, meronyms can be ordered hierarchically on the basis of the information we can extract from the definitions and explanations but the assistance of expert advisors in this respect is necessary, at least for the final approval of the classification scheme proposed. We assume that all the partitive terms to "concrete" identified are system-structuring terms and are suitable for representing in our glossary as either separate entries or items within the entries. A check against the Glossary and *TermoStat* confirmed our assumption (see Table 1). Despite the fact that the terms "cement binder" and "additive" do not figure in these lists, they are important nodes in the overall classification by composition thus worth being included as suitable learner's glossary entries.

3. Properties
  – *durability* (*TermoStat*, Glossary)
  – low *water/cement ratio* (*TermoStat*, Glossary)
  – good *compaction* (*TermoStat*, Glossary)
  – high *compressive strength* (*TermoStat*, Glossary)
  – *workability* (*TermoStat*, Glossary)
  – highly alkaline
  – low moisture movement

A comparison with the Glossary and *TermoStat* singled out the property terms which we have italicized.

4. Technology
  – aggregate, cement and water are *mixed*
  – mixture *sets* within a few hours of mixing
  – concrete dries after *mixing* and *placement*
  – cement *hydrates* and eventually mix *hardens* into a stone-like material

From *TermoStat* and Glossary we can add:

  – concrete *bleeds*
  – concrete is *cured*
  – concrete is tested (*slump test*)

5. Use
  – concrete is used in construction
  – for foundations
  – in buildings for structural walls, columns and *slabs* (*TermoStat*)
  – for fire encasement
  – as a background for plaster

Some of these applications of concrete and the terms used to express them will occur in some of the exemplary contexts to be provided for the terms in the glossary. Otherwise, they belong to the external (functional) systematicity of the "concrete" terminological system (cf. Popova 2005) and are not relevant for the analysis we have undertaken.

Step 3 – Selection of the final list of glossary entry terms

We have already selected a considerable portion of the final list of entries suitable for our glossary (see all italicized terms above). This list can be supplemented by performing the analysis at a more detailed level that can yield some complex relations typical of the specific terminological subsystem we are concerned with. For example, the multi-word terms "concrete mixer" and "concrete truck," where "concrete" is a determiner, can be analysed conceptually in the following way:

– *concrete mixer* and *concrete truck (syn. truck mixer/agitator)* → relationship of product-manufacturing equipment.

4.3 Unified terminological definitions for learner's glossaries

ISO standard 1087 defines the concept "definition" as: "statement which describes a concept and which permits its differentiation from other concepts within a conceptual system." When preparing the definitions for our learner's glossary, we have to take into account certain conventions and general principles such as (a) the capacity of definitions to differentiate the defined concept from similar concepts in the same or in different special fields and (b) their appropriateness for the aims of the project in which they are presented. With regard to their form of expression, definitions should (a) adhere to the formal standards for writing definitions; (b) use language that is suitable for the intended readership; (c) be in the form of a single sentence.

Some of these principles which are relevant for our project can be briefly formulated as follows:

– genus terms should be of the same grammatical category as the entry term (e.g. *Concrete (noun) is a building material (noun)*);
– if terms are used in the definition, they should be defined in the same glossary (e.g. *Concrete is a composite material composed of coarse and fine aggregate*, etc."aggregate' should be an entry in the glossary as well);
– definitions should not be circular (e.g. *durable: having relatively high durability*); terms should not be defined by negation (e.g. *lightweight concrete: not heavyweight concrete*).

Kolkovska (2005) gives practical advice on how to process source definitions that can be used for a given terminographic project. A basic requirement is to check whether the processed definition reflects what Kolkovska calls the "classification characteristics" of the term subsuming both genus-species and partitive characteristics. Particularly relevant for our project is the recommendation to choose the nearest genus as the most suitable genus characteristic. For example, from the four possible genus characteristics of "concrete," viz. *composite building material*, *composite material*, *composition* and *mixture*, we have to choose *composite building material*. Another useful recommendation that Kolkovska makes is the addition of a classification characteristic, especially a species one. In Section 6 we will show how we can use the results of the conceptual analysis to involve maximum amount of information within a single-sentence definition.

4.4 Specifying the translation equivalents of glossary entries

Finally in this section we will discuss briefly our approach to specifying the translation equivalents of the terms to be used as entries in our glossary. Although the envisaged glossary is bilingual, since it is designed primarily for Bulgarian English for Specific Purposes (ESP) learners and non-specialist translators, we do not intend to involve any Bulgarian definitions, explanations or contexts. The Bulgarian text in an entry will contain only translation equivalents of terms and their collocations. The latter can be specified by a contrastive analysis based on available multilingual reference tools or by consulting experts. Since most terminological units that we plan to include in the project represent basic concepts, their Bulgarian translation equivalents have long been established and fixed in existing dictionaries. Nevertheless, there are still some uncertainties that require the assistance of a specialist.

For example, the translation of the verbal collocation *concrete bleeds* (mentioned in the Introduction) can be derived from the verbal noun *bleeding* whose Bulgarian translation equivalent in the *English-*

Bulgarian dictionary of architecture and civil engineering is given in the form of a long phrase. We were pleasantly surprised when our expert advisors proposed a much shorter translation equivalent, viz. *concrete release** (literally) that sounds similar to *gas release* in Bulgarian. This fact confirmed our belief that engineering specialists are not less concerned than we, language specialists are, about the economy of expression in their specialised communication and a unified effort in this direction will definitely yield positive results.

5. Lexico-semantic analysis of terminological data

In the previous section we applied a conceptual analysis for identifying the candidate entry terms for the provisional English-Bulgarian Learner's Glossary of Concrete Terms and the narrower terms within these entries. As we have already mentioned, translators of technical texts very often encounter difficulties when translating not the terms themselves but the words they usually co-occur, i.e. their collocates. This is why we have decided to include term collocations as an information category in our glossary. For the purpose, we searched the literature to find an approach to capturing such information data from running text. Our search efforts led us to the Canadian lexico-semantic approach to terminology structuring, developed by the terminology research group at the University of Montreal. Below we will discuss the theoretical and methodological premises of that approach and will show how it can be adapted to analyse our corpus aiming to capture collocational information on the entries we have already specified.

5.1 Theoretical premises of the lexico-semantic approach to terminology structuring

The theoretical basis for the lexico-semantic approach to structuring terminological data for terminographic purposes is provided by the Explanatory and Combinatorial Lexicology/ECL which is the lexicological component of the Meaning-Text Theory/MTT. This theory proposes a formalized model of natural language, i.e. a Meaning Text Model/MTM representing a system of rules which simulate the linguistic behaviour of humans. An MTM is designed to perform the transition from meanings in general (any information/content a speaker transmits by using natural language) to texts (physical manifestation of speech) and vice versa. ECL, in turn, proposes an apparatus, namely, lexical functions (LFs) for capturing semantic relations between lexical units. LFs are a means for a systematic description of the so-called "institutionalised" lexical relations. Some simple examples of institutionalised lexical relations are those between *attention* and *pay*, *wolves* and *pack*, etc. From our "concrete" terminological microsystem we can provide the following examples: *concrete* and *mix*, *concrete* and *set*, *concrete* and *harden*, *concrete* and *batch*, etc. LFs are based on Saussure's dichotomy of paradigmatic vs. syntagmatic relations. Paradigmatic relations can be defined as all contrast and substitution relations holding between lexical units in specific contexts. Syntagmatic relations are relations holding between lexical units that can co-occur, i.e. appear together in the same phrase or clause. Mel'čuk (1996, 39-40) explains that the term "function" in the theory is used in its mathematical sense $f(x) = y$ where f is the function, x is the argument and y is the value expressed by the function when applied to a given argument. There is no doubt that this theoretical framework has had and will have important repercussions for a broad variety of lexicological endeavours. For the purposes of our particular project we are interested in the extent to which these theoretical assumptions can be used for analysing terminological data for terminographic purposes.

Comparing the two different approaches to terminology, viz. the conceptual one and the lexico-semantic one, L'Homme (2004) points out their advantages and shortcomings. She argues that truly conceptual approaches do not allow a flexible integration of terms and relationships between terms. On the other hand, lexico-semantic approaches are more compatible with data gathered from corpora. For the lexico-semantic analysis of the computer term "program" L'Homme applies lexical functions to formalize the following relationships "program" enters in:

- synonym: **Syn** ($program_1$) = computer ~ ;
- agent of program: $\mathbf{S_1}$ ($program_1$) = programmer;
- create a program: $\mathbf{CauseFunc_0}$ ($program_1$) = create [DET ~], write [DET ~];
- cause a program to function: $\mathbf{CauseFact_0}$ ($program_1$) = execute [DET ~];
- the program stops functioning: $\mathbf{FinFact_0}$ ($program_1$) = [DET ~] ends, [DET ~] terminates.

In our opinion, this analytical procedure shows clearly two disadvantages of that approach. On the one hand, the LF notation is very complicated and will obviously have to be simplified in order to be

conveniently applied to the analysis of terminological items. On the other hand, the specificity of the terminological system may require the postulation of new specific lexical functions that have not been considered in the Explanatory and Combinatorial Lexicology. For example, there is no LF and notation, respectively, for the so-called "self-running natural processes" expressed by verbs such as "set," "harden," "bleed," which collocate with our basic keyterm "concrete" (see below). Therefore, for the lexico-semantic analysis of our corpus in view of extracting and consequently presenting in our keyterm entries useful collocations, we will use a methodology that relies on the general principles of ECL for performing the analysis along both the paradigmatic and syntagmatic axes but does not involve the standard lexical functions.

5.2 Methodology for lexico-semantic analysis of term data

We will first present the methodology L'Homme & Bae (2006) propose for developing multilingual resources for terminology. It is divided into five steps, namely: 1) compilation of the corpora; 2) selection of terms; 3) sense distinction; 4) definition of the actantial structure; 5) listing of semantic relationships. The methodology is currently being applied in building lexical resources, i.e. dictionaries of basic French and Korean terms belonging to the fields of computer science and the Internet, which provide detailed lexico-semantic information on terms in those fields.

Now if we consider our approach to structuring the terminological data required in order to compile our learner's glossary and compare it with the steps proposed above, we will ascertain that we have actually performed the first two steps in Sections 3 and 4 above using the principles of conceptual analysis. Sense distinctions (step 3) are necessary to be made in large maximizing dictionaries like the one L'Homme & Bae are compiling (e.g. the different senses of "program" within different collocations) but for a minimizing learner's glossary where each keyterm is clearly specified, such a procedure is again irrelevant. We will make use of the methodology for describing specialised predicative units that combine with terms in order to identify keyterm verbal collocations. We will also analyse paradigmatic and syntagmatic morpho-syntactic relations applicable to a keyterm. Both analytical procedures are truly lexico-semantic and generally comply with the ELC without using its LF apparatus. Hence, in view of the aim pursued in this study, we will modify steps 4 and 5 as follows:

Step 4: Analysing the actantial structure of a keyterm for identifying verbal (T+V) collocations
Step 5: Analysing paradigmatic and syntagmatic morpho-syntactic relations applicable to a keyterm

5.2.1 Analysing the actantial structure of a keyterm for identifying verbal (T+V) collocations

The ECL postulates lexical functions such as S_i (noun for actant), S_{instr} (noun for instrument), S_{loc} (noun for location), S_{res} (noun for result), etc. which can be used for capturing specific relations between lexical units. However, L'Homme (2003) proposes a model for analysing contexts containing verbs by making use of their actantial structure. She admits that the class system and labels she resorts to can be related remotely to concepts introduced in, e.g. Fillmore and Atkins (1998) ('frame elements'). In L'Homme and Bae (2006), the lexico-semantic analysis of the actantial structures of predicative terms (verbs) is exemplified by representing the term browse in a tabular form (the original examples are in French).

AGENT	LOCATION	INSTRUMENT
user	Internet	browser

As can be seen, the actantial structure gives the position of actants and explains them in terms of actantial roles.

We will follow a similar procedure to identify the verbal collocations of the keyterms "cement" and "concrete" (T+V collocations), leaving aside adjectival/nominal (A/N+T) collocations that we have already identified by the conceptual analysis in the previous section as most of these actually designate generic or partitive relations.

The special collocations with the keyterms "concrete" and "cement" have been extracted from contexts provided by *TermoStat*. The specialised lexical combinations with these terms analysed below are selected because they have specialized meaning within the field of construction (L'Homme 2000), e.g. the meaning of "cure" (make a person or animal healthy again) is altered within the specific combina-

Natural self-running process	Means	Object	Location
bleeds/compacts/ hardens/sets	admixture, additives	concrete	formworks

Table 2.

Agent	Action (on)	Object	Result/ Product	Location	Equipment
builder	places/pours/cures/levels/sprays	concrete		form(works)	
builder	mixes	cement, water, aggregate	concrete		mixer/truck mixer

Table 3.

tion "concrete is cured." Two types of activities can be captured by the methodology described above, namely, self-running natural processes during concrete manufacturing and actions performed on "cement" and "concrete." Some of the results of the analysis are presented in Tables 2 and 3.

The analytical results confirm L'Homme's conclusion that "senses that can be observed in specialized domains are likely to be limited in number" and "semantic classes in a given syntactic position could be used to discover typical 'frames'" (L'Homme 2003). In other words, a terminographer doing a research into a terminological system or subsystem with the view to identifying collocations is very likely to be forced by circumstances to "discover" (definitely with the help of specialists) new actantial structures typical of the particular specialised discourse. We have to point out that the collocational relations we have established are in fact syntagmatic relations. Other syntagmatic relations (morphological families and syntactic derivations) generalised as "syntagmatic morpho-syntactic relations" are also relevant for our project and hence will be considered briefly below.

5.2.2 *Analysing paradigmatic and syntagmatic morpho-syntactic relations applicable to a keyterm*

The following paradigmatic and syntagmatic morpho-syntactic relations applicable to "concrete" have been identified:

A. Paradigmatic relations (synonyms, antonyms, etc.)

An interesting case can be observed with the collocation "place (v) concrete:" Our Corpus and *Termium* give only one synonym "pour concrete" for "place concrete." The *English-Bulgarian dictionary of civil engineering* for the same Bulgarian equivalent gives four English equivalents without any differentiation, namely, "place," "pour," "cast," and "lay." Corpus and *Termium* give "cast" only as a premodifier to "concrete" usually prefixed ("precast concrete") or as a syntactic terminological phraseme ("cast-in-situ/cast-in-place concrete"). Neither of the two sources give *"lay concrete" or *"concrete is laid." Therefore, an additional bilingual contrastive analysis can help to specify predicative (T+V) relations. All these real occurrences should be reflected in the "concrete" entry.

B. Syntagmatic morpho-syntactic relations (morphological families, syntactic derivations, etc.)

If we consider again the collocation "place (v) concrete," we can find the following morphological and syntactic derivations in the terminological data processed by *TermoStat*:
- "placing of concrete"
- "concrete placing"
- "placement of concrete"
- "concrete placement"
- "concrete placements" (pl.)

In fact, all these are true synonyms and are suggested to be represented with particular contexts from the Corpus, with "concrete pouring" given also as a synonym of "concrete placing/placement"(cf. *Termium*).

6. A model for structuring a learner's glossary entry

Summarizing the findings from the two types of analysis made above, we can propose the following model for structuring a typical dictionary entry for the provisional English-Bulgarian Learner's Glossary of Concrete Terms:

CONCRETE – Bulgarian equivalent/BE

DEFINITION: A composite building material composed of coarse and fine aggregate (sand, gravel, crushed rock, etc.) held together by a hardened paste of hydraulic cement and water with added admixtures, which is characterised by durability, high compressive strength and compaction, low water/cement ratio and workability and is used in building foundations, structural walls, columns, slabs, etc.
E.g. The composition of concrete is determined initially during mixing and finally during placing of fresh concrete. The type of structure being constructed as well as the method of construction determine how the concrete is placed and therefore also the composition of the concrete mix or mix design.

TYPES OF CONCRETE:
- PLAIN CONCRETE /syn. ORDINARY CONCRETE – BE (consists of the same aggregate and cement, without admixtures); e.g. *Plain concrete is used when the structure or structural member in question is unlikely to develop tensile stresses in service*
- REINFORCED CONCRETE /syn. FERROCONCRETE (obsolete), STEEL CONCRETE - BE (hardened onto imbedded metal, usually steel); e.g. *In reinforced concrete, the tensile strength of steel, supplementing the compressive strength of the concrete, provides a member capable of sustaining heavy stresses of all kinds over considerable spans.*

SPECIAL PHRASES:
1. CONCRETE SETS/HARDENS - BE – e.g. *Concrete sets faster or slower according to how much gypsum is added to the mixture.*
2. CONCRETE BLEEDS – BE (mixing water emerges from freshly placed concrete)
3. CONCRETE IS PLACED/POURED (Variants: CONCRETE PLACING/PLACEMENT(S)/POURING; e.g. *(a) Place concrete as near to its final position as possible; (b) Prestressed concrete requires the application of a load to the steel before concrete placement.*
4. CONCRETE IS CURED – BE (Variants: CONCRETE CURING – Maintaining the humidity and temperature of freshly placed concrete for a definite period to assure satisfactory hydration of the cementitious materials and proper hardening)

This model exemplifies the variety of information items that can be represented in only one glossary entry. We should note that this particular entry will be the largest in size because it represents the basic keyterm. The other entries (keyterms) are normally expected to contain less information. We also have to emphasize the length of the definition of "concrete" composed by following the principles and requirements for structuring unified terminological definitions given in Subsection 4.3 above. The underlined terms in the definition are cross-references, i.e. they will appear in the glossary as separate entries. Again it is necessary to note that the definitions of the other keyterms will certainly be shorter.

7. Conclusions

Two final conclusions can be drawn: First, terminology building cannot be considered as a full automated process but rather as a cooperative task between terminological tools and terminologists. Identifying terms in a technical domain is a matter of word usage and expert agreement. Second, the findings reported here provide evidence for the feasibility of a combined conceptual and lexico-semantic approach to terminology structuring for multilingual terminographic purposes. As the conceptual analysis of terminological data requires knowledge-based expert information concerning the special domain in question, which is available in existing reference documents, its use can save effort and time but should preferably be performed by a language specialist in collaboration with a domain expert. It is necessary to stress that the two approaches can be considered compatible provided that preliminary decision is made as to which type of analysis is appropriate for the particular type of information category presented in the entry term.

The assets of the richly detailed learner's glossaries proposed can be summarised as follows. It contains valuable information such as definitions, examples of usage, synonyms, related terms, usage notes, etc. Translators and ESP students can use them to understand terms, identify equivalents, learn how to use terms and term collocates, identify synonyms if they exist, identify related terms and their equivalents. We believe that the content and structure of the learner's glossary we propose will facilitate the specialised bilingual communication within the respective discourse community and assist in initiating its prospective users (LSP learners and translators) to that community.

References

Alexiev, B. et al. 1998. *English-Bulgarian dictionary of mining and geology.* Sofia: Ratio-90 Publishing House.

Alexiev, B. 2004. Towards an experientialist model of terminological metaphorisation. *Terminology* 10: 189-213.

Alexiev, B. 2006. A lexico-conceptual approach to multilingual terminology structuring. *Contrastive linguistics* (forthcoming).

Bowker, L. and Pearson, J. 2002. *Working with specialized language*. London and New York: Routledge.

Brown, L. ed.. 1993. *New shorter Oxford English dictionary*. London: Ocford University Press.

Cabré M.T. 1999. *Terminology: theory, methods and applications.* Amsterdam/Philadelphia: John Benjamins Publishing Co.

Dahlberg Ingetraut. 1981. Conceptual definitions for INTERCONCEPT. *International classification* 8: 16-22.

Dahlberg, Ingetraut. 1988. Concept and definition theory. In *Classification theory in the computer age: conversations across the disciplines*. Proceedings of a Conference Nov.18-19, 1988. Albany, NY. School of Inform.Sci. & Policy SUNY, pp. 12-24.

De Bessé et al. 1997. Glossary of terms used in terminology. *Terminology* 4: 117-56.

Drouin, P. 2003. Term extraction using non-technical corpora as a point of leverage. *Terminology* 9: 99-117.

Felber, H. 1984. *Terminology manual*. Paris: UNESCO and Infoterm.

Fillmore, C. and Atkins, B.T.S. 1998. FrameNet and Lexicographic Relevance. In *Proceedings of the First International Conference on Language and Evaluation (LREC) Granada, Spain*, pp. 417-23.

Grabar, N. and Zweigenbaum, P. 2004. Lexically-based terminology structuring. *Terminology* 10: 23-53.

ISO 1087-1. 2000. *Terminology – vocabulary- part 1:theory and application*.

ISO 12616. 2002. *Translation-oriented terminography*.

Kolkovska, S. 2005. Some aspects of term standardization in lexicographic practice. *Bulgarian language* 1: 34-47 (in Bulgarian).

L'Homme, M.C. 2000. Understanding specialized lexical combinations. *Terminology* 6: 89-109.

L'Homme, M.C. 2003. Capturing lexical structure in special subject fields with verbs and verbal derivatives: a model for specialized lexicography. *International journal of lexicography* 16: 403-22.

L'Homme, M.C. 2004. A lexico-semantic approach to the structuring of terminology. In *Computerm 2004*, dans le cadre de Coling 2004, Université de Genève, Genève (Suisse), 29 août 2004, pp. 7-14.

L'Homme, M.C. & Hee Sook Bae. 2006. A methodology for developing multilingual resources for terminology. *LREC 2006, Language Resources and Evaluation. Proceedings*, Genoa (Italy). 22-27 August 2006. Available at: http://www.ling.umontreal.ca/lhomme/docs/LREC-2006-Lhomme-bae.pdf

Mel'čuk, I. 1996. Lexical functions: a tool for the description of lexical relations in a lexicon. In Wanner L. ed. *Lexical functions in lexicography and natural language processing*, Amsterdam/Philadelphia: Benjamins, pp. 37-102.

Nielsen, S. 1990. Contrastive description of dictionaries covering LSP communication. In *Fachsprache/International Journal of LSP* 3-4: 129-36.

Popova, M. 2005. Internal systematicity of the basic terms in Linguistic Pragmatics. *Bulgarian language* 1: 5-19 (in Bulgarian).

Sager, J.C. 1990. *A practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins Publishing Co.