# What can be an Object of Terminological Description in a Term Bank?

*Igor Kudashev*
*University of Helsinki*

*На первый взгляд, ответ на вопрос «Какие единицы могут быть объектом терминологического описания в терминологических банках данных?» представляется очевидным: это термины. Однако при написании инструкций для пользователей терминологического банка, построенного по принципу Википедии, мы обнаружили, что ответ на поставленный вопрос требует более детальных разъяснений и ответа на ряд смежных вопросов. В статье обсуждается, какие типы единиц ЯСЦ могут быть объектом терминологического описания в терминологических банках данных. Отправной точкой исследования является список единиц, перечисленных в стандарте ИСО 12620:1999 «Применение компьютера в терминологии. Категории данных» в категории «тип термина». Показывается, что принципиальные ограничения, сужающие круг объектов описания ТБД, связаны с особенностями терминологического описания и организации информации в электронных справочных ресурсах. Дополнительные ограничения могут быть связаны с установками проектировщиков ТБД и создателей терминологических коллекций, а также с техническими и организационными трудностями.*

**Keywords**: term bank, terminology management system, terminological description

## 1 Introduction

At the first glance the answer to the question posed in the title seems obvious: term banks are supposed to deal with terminology. However, when writing instructions for the users of a Wikipedia-like platform for collaborative terminology work, we realised that the answer to this question was not that simple and that it required answers to several related questions.

To start with, concepts of term and terminology are defined and interpreted differently even in different ISO standards. According to ISO 1087-1 (2000: 6) *term* is a verbal designation of a general concept in a specific subject field. At the same time, *terminology* is defined as a set of designations belonging to one special language (ISO 1087-1:2000: 10). *Designation* is defined as representation of a concept by a sign which denotes it. The note below the definition says that in terminology work three types of designations are distinguished: symbols, appellations and terms. Thus, terminology ap-

pears to be a broader concept than "a set of terms belonging to one special language" and includes appellations and symbols.

Another standard, ISO 12620:1999 (Computer applications in terminology – Data categories) lists symbols, equations, logical expressions, standard texts, collocations, set phrases, stock-keeping units and part numbers (but not appellations) under the category "term type".

In this article, we would like to discuss whether all of the expressions listed in the ISO 12620:1999 standard can be objects of terminological description in a term bank and whether this list is exhaustive. Our assumption is that the character of description and organization of data in a term bank restrict the array of units which can be objects of terminological description in a term bank.

## 2 Definition of a term bank and terminological description

ISO 1087-2:2000 defines *term bank* as a data bank containing terminological data; *data bank* is defined as a collection of databases including the organizational framework for recording, processing and disseminating data (ISO 1087-2:2000:12).

There can be different opinions about what constitutes an object of a term bank – expression, concept or their combination. We consider a term as a two-sided sign which has a form and meaning. Both form and meaning can to some extent be described independently (for example, it is possible to provide some grammatical information for a form and a definition for a concept) but description of a term is more than just a sum of descriptions of its form and meaning. So in our opinion the primary object of description in a term bank is a term or, more broadly, an LSP designation, while forms and meanings as topics of independent description can only be secondary and auxiliary objects.

The most important type of data in a term bank is terminological descriptions – information about the form, meaning, relations and usage of LSP units in the past, present

and/or as a part of language planning for the future. Terminological description is mostly linguistic in nature although it may contain elements of encyclopaedic description.

**3 Term bank as a reference source**

Terminological descriptions can be presented to the end users in the form of term entries and groups of entries. Being a reference resource, a term bank has to provide a quick and easy access to terminological descriptions. Although data can be organized in many different ways and many different types of access can be used, the easiest and quickest type of access is by some conventional form of the described unit (usually called a canonical form).

The goal of providing LSP designations with terminological description as well as searching, accessing and arranging them by their form implies certain restrictions on the objects of terminological description in a term bank.

First, described units have to be LSP units, i.e. designations of objects, properties and relations specific for a special domain of discourse. The nature of the described object should allow providing them with a terminological description. Pragmatically, there should also be a practical need for such description.

Second, it should be possible to include described objects into a terminological collection as headwords and search for them. In practice it means that they can only contain characters supported by Unicode.

Third, if the contents of a terminological collection are supposed to be presented in a linear way, headword by headword or entry by entry, it should be possible to sort the headwords in a way which is predictable for the users.

In the sections to follow we provide examples of LSP units which can be objects of terminological description in a term bank and then examples of units from the ISO 12620:1999 list which can hardly be such objects.

## 4 Examples of LSP units which can be objects of terminological description in a term bank

It is hard to classify LSP units within a single classification scheme, so we will classify them according to several facets, the first of which will be the type of concept denoted by an LSP unit.

### 4.1 LSP units according to the type of concept

According to the definition cited above, *term* is a verbal designation of a general concept in a specific subject field (ISO 1087-1:2000: 6). We consider this definition too broad and suggest that a refinement proposed by V. Leičik (2006: 32 and earlier works) should be applied. According to this refinement, terms denote general concepts which reflect a particular theory, i.e. they serve as elements of rather well-organized concept systems. Designations like *galaxy* or *particle accelerator* are terms.

While terms denote general concepts, i.e. concepts the extension of which consists of two or more objects, *appellations* denote individual concepts, i.e. concepts which correspond to only one object. Designations like the *Andromeda Galaxy* or the *Large Hadron Collider* are appellations.

Many companies would gladly include *nomenclature* in the term bank, in particular names of objects of mass production, such as "Dreamland Soft" mattress, "Delux Beauty Relax" pillow or "Ecomoods Fabia" lamp, as there is a need to deal with them in a multilingual environment (for example, to translate and localize names of a company's products).

Classes of nomenclature are formed on the basis of *necessary* conditions which, however, are not *sufficient* for unambiguous distinction of the class, especially in diachrony.

Classes of nomenclature are *primitive classes* with more or less detailed description; they can not be assigned a genus-species definition (cf. Kandelaki 1973: 63; Bereznikova 1976: 88). While defining a revolver is easy because revolvers have a special principle of operation that distinguishes them from other pistols, describing a difference between *Smith-and-Wesson model 13* and *Smith-and-Wesson model 27* requires comparison of their specifications: design, technical characteristics, etc. (cf. Leičik 1974: 20–21). Although the intention of concepts denoted by names may be somewhat blurry, their extension is usually quite clear thanks to different identification systems and specifications.

LSP designations denoting "raw" and "vague" *notions* which lack a stable place in a concept system are called in Russian terminological tradition *terminoids* (e.g. Grinev 1993: 49; Leičik 2006: 77; the term was coined in Haûtin 1972: 30, 104). Terminoids are abundant in branches of special knowledge that lack solid theoretical basis. Designation *ghost* is an example of a terminoid (cf. "this term denotes only the apparition of a deceased person, and is not sufficiently precise for use in psychical research" – "Glossary of key words frequently used in parapsychology", available at http://www.parapsych.org/glossary_e_k.html#g, accessed 31.3.2010).

From the point of view of terminology work, distinguishing of the above mentioned LSP units is important for two reasons. First, the users of a term bank must realize that in addition to "classical terms" they can also add appellations, nomenclature and terminoids. Second, it is important for them to know that appellations, nomenclature and terminoids can not be assigned a genus-species definition and require a different kind of description.

### 4.2 LSP units according to their typographical characteristics

LSP designations may contain characters of a given national language, other national languages, digits, punctuation, special symbols as well as inline formatting, such as italics, upper and lower indexes, etc. For example:

- β-carotene
- 20 %-flexibility clause
- N,N'-di-2-butyl-1,4-phenylenediamine
- x-coordinate
- CO2 laser.

Special characters and formatting can present a problem for the software developers but users have the right to expect that a modern terminology management system supports the whole range of Unicode symbols and processes correctly the elements of inline formatting.

4.3 LSP units according to part of speech

Lexical LSP designations can be nouns, verbs and adverbs, for example:

- a programme
- to programme
- programmatically.

Adjectives and participles (such as programmatic, programmed) normally do not function as independent LSP designations but rather serve as term elements (discussed below). Substantivized adjectives and participles are an exception to the general rule.

4.4 LSP units according to their morphological structure

This characteristic is dependant on the morphological classifications adopted in/for different languages. Below is an example of such classification for the English language:

- simple word (word containing only one root): sound, light;
- complex word (word formed from a simple word by the addition of one or more derivatives): accountability, partnership;

- phrasal verb (combination of a verb and a preposition, a verb and an adverb, or a verb with both an adverb and a preposition, any of which are a part of the syntax of the sentence, and so are a complete semantic unit): take off, tank up;
- compound word (word containing two or more roots): know-how, airstrike;
- abbreviated form (designation formed by omitting any parts from a longer form): DNA, adj., flu, radar;
- multiword expression (phrase consisting of two or more words): cross-boarder co-operation, direct employment effect.

### 4.5 LSP units with usage restrictions

LSP designations can be restricted in use by many factors, the most common of which are the following:

- geographical area (e.g. en-US, en-GB, dialect expression);
- time period (e.g. obsolete, neologism, used during WWII);
- register (e.g. official term, informal term, professional slang);
- organization (e.g. term used in/by Nokia, Microsoft, UN, WHO);
- proprietary relations (e.g. trade mark and trade name);
- scientific school/theory (e.g. Newtons's physics, Einstein's physics; Danish structuralism in linguistics);
- professional group (e.g. physicians, nurses, medical assistants);
- normative regulations (e.g. standardized, preferred, recommended, non-recommended term).

As many terminology projects impose restrictions on such units, users of a term bank may decide that units with usage restrictions do not belong in a term bank at all. In practice there are no obstacles for including such units. However, it is a good idea to provide them with usage labels.

## 4.6 Means of formal notation in LSPs

Special concepts and objects may be referred to with the help of means of formal notation, such as

- special symbols: §, €, °, ∑, ∞;
- formulae: H2O, [As@Ni12As20]3−;
- international scientific names: "*Salix starkeana* subsp. *cinerascens*";
- codes: "ABHD12", "C20orf22", "DKFZP434P106", "dJ965G21.2", "BEM46L2", "ABHD12A" are code names of a gene with an official name "abhydrolase domain containing 12";
- catalogue names: "Messier 31", "M31", "NGC 224" are catalogue names of the Andromeda Galaxy.

Formal notations are often used in LSP texts interchangeably with the corresponding appellations and terms and sometimes they are the only existing designation of a special object or concept.

If there is a parallel lexical designation, means of formal notation usually become a part of description. They can also be made searchable by including them as headwords of reference articles. If there are no parallel lexical designations, means of formal notation become an object of description in their own right.

## 4.7 Lexicalized units

In addition to lexical units, a term bank may also contain *lexicalized LSP expressions*. Lexicalized LSP expression is a single-word or multi-word expression which has a relatively stable form and function in a particular LSP or special area of application. Below are a few examples of lexicalized LSP expressions:

- instructions, such as "handle with care" or "this end up" (ISO 12620:1999: 10);
- military commands, such as "Stand at ease!", "Eyes right/left!", "Double march!";

– set phrases used in radio and signalling: “More to follow”, “How copy?”, “Solid copy!”;
– software menu commands, such as “Save as…”, “Print…” or “About…”.

Lexicalized expressions can not be “defined” in a traditional way because they do not denote a special concept; instead, they can be provided with a description of their function, i.e. situations in which they are used. For example, “Double march!” can be defined as “a drill command: an order to jog in time”.

In many respects lexicalized expressions are alike lexical units. In some domains and spheres of activities lexicalized expressions play a very important role. Due to this they can and should be objects of terminological description in terminological reference resources.

## 4.8 Term elements

Term elements are productive components of terms which have a relatively stable meaning in a given LSP. Here are a few examples of term elements used in the domain of medicine:

– prefixes “a-“, “an-“ mean an absence of something (apathy, analgia);
– suffix “-ac” means “pertaining to something” (cardiac);
– root “aur(i)-” means “pertaining to the ear” (aural).

Words (in particular adjectives and participles) can also function as term elements. When described as objects in their own right, term elements can be provided with a more comprehensive description of their meaning, etymology, usage as well as term formation models.

## 5 Designations which are unlikely to be objects of terminological description

In our opinion, several types of LSP units listed in the ISO 12620:1999 data category inventory as “term types” are unlikely to become objects of terminological description in a term bank.

The first group of such objects are statements like equations (e.g. E=mc2) and logical expressions (e.g. $x \neq y$) and standard texts (e.g. force majeure clause) as they are not designations of objects or concepts. These units can be a part of description but not an object of description in a term bank. Extensive recurrent chunks of text are a natural object for translation memory systems.

Collocations (e.g. "immunization against smth.") demonstrate combinatory restrictions of LSP units and should be a part of description, not its object.

*Stock-keeping units* and *part numbers* are unique (within a given context) alphanumeric designations assigned to objects in a manufacturing or inventory control system. These designations have some common features with nomenclature but there is also an important difference: stock-keeping units and part numbers refer to *consignments* of products rather than products themselves. Variable characteristics, such as size, colour, dimensions, number of pieces in one package, which are encoded in skus and part numbers are essential only for a particular consignment of goods and usually non-essential for the description of the product as such.

Being based on non-essential, variable characteristics, “concepts” designated by material management categories have a very short life and their scope of usage and functions are extremely limited. Material management categories are not supposed to be managed as linguistic signs: they can not be altered, inflected, translated or localized. Linking between a material management system and a terminology management system is possible but material management categories are very unlikely to be objects of terminological description in a term bank.

## 6 Conclusion

Summing up, we can answer the question ”What can be an object of terminological description in a term bank?” in the following way: the most natural object are lexical or lexicalized LSP units and term elements which can function as headwords, be searched, sorted in a predictable way and provided with a terminological description.

The array of units which can be objects of terminological description in a term bank is restricted by the character of description and organization of data in a term bank. In practice additional restrictions may originate from technical and organizational problems or be imposed by the designers of a term bank or creators of individual terminological collections.

## References

Bereznikova, R. E. (1976). Podača nomenov v slovarâh raznyh tipov. In: *Problematika opredelenij terminov v slovarâh raznyh tipov*, 83–91. Leningrad: Nauka.

Grinev, S. V. (1993). *Vvedenie v terminovedenie*. Moskva: Moskovskij Licej.

Haûtin, A. D. (1972). *Termin, terminologiâ, nomenklatura: Učebnoe posobie*. Samarkand: Samarkandskij gosudarstvennyj universitet.

Kandelaki, T. L. (1973). K voprosu o nomenklaturnyh naimenovaniâh. In: *Voprosy razrabotki naučno-tehničeskoj terminologii*, 60–71. Riga: Zinatne.

Leičik, V. M. (1974). Nomenklatura – promežutočnoe zveno meždu terminami i sobstvennymi imenami. In: *Voprosy terminologii i lingvističeskoj statistiki*, 13–24. Voronež: Voronežskij tehnologičeskij institut.

Leičik, V. M. (2006). *Terminovedenie: predmet, metody, struktura*. Moskva: KomKniga.

ISO 1087-1:2000(E/F). Terminology Work – Vocabulary – Part 1: Theory and Application. Geneva: ISO.

ISO 1087-2:2000(E/F). Terminology Work – Vocabulary – Part 2: Computer Applications. Geneva: ISO.

ISO 12620:1999(E). Computer Applications in Terminology – Data Categories. Geneva: ISO.