Patricia Thomas, "Language International", Nottingham NG1 5AH, United Kingdom

Treatment of compound terminology entries

ABSTRACT: The problem of identifying headwords from multiple lexical entry terms is well documented in both lexicography (covering general language, or LGP) and terminography (special language, or LSP). Recommendations are proposed for assessing the "entry point" or "headword" of terms comprising several lexical units. A cognitive evaluation needs to be made of the part of the unit on which to focus so that the correct descriptive and informative data are compiled to augment the entry term. Elements of phraseology will be addressed and examples given from special languages in English, French, German and Swedish. Comparisons will be drawn with current English LGP lexicography practices.

1. Lexicography and terminography

In the context of this paper, the term "lexicography" refers to the compilation of general language dictionaries, while "terminography" refers to special language lexicography, encompassing the publication in various forms of the language used in special subject fields. Riggs (1989, 89) states that at the functional level, "lexicography has the primary aim of helping readers interpret texts, whereas terminology aims to help writers produce texts". He contrasts the two disciplines at the structural level thus: "lexicography follows a semasiological line, from words to their meanings, whereas terminology adopts an onomasiological model, proceeding from concepts (as defined by a text) to the terms that designate them".

2. Identifying the problem in both LGP and LSP

Special language is used to represent the concepts of a specialised subject field and these concepts are represented by terms, which may be single words, multiple lexical units, phrases, symbols, formulae or graphics. Terms are almost exclusively nouns and frequent recourse is made to compounding, which in English usually takes the form of either noun + noun(s) or adjective(s) + noun(s), often resulting in terms comprising several lexical units, e.g. "Autographa californica nuclear polyhedrosis virus".

Since terms frequently comprise more than one lexical unit, this raises the question of which word should be selected as the "point of entry" for a particular term in the assessment of the terminologist, who is responsible for providing a definition for the term, with additional data, such as bibliographic references, to enable users to retrieve the informa-

tion they require. Consistency in the treatment of multiple lexical units is a vital factor in enabling retrieval of information to be made easily and efficiently. Cowie (1983, 103) has stated that "There are few features of dictionary organisation on which editorial policy differs as strikingly from one dictionary – or lexicographical tradition – to another as the treatment of compounds". The problem is exacerbated in the case of special language dictionaries because the specialists who of necessity must be the compilers of such dictionaries seldom, if ever, have any lexicographic training. Where an editorial policy exists it will usually conform to the requirements of a publishing house, even though in some cases there may be little evidence of a lexicographic tradition. This is borne out by Opitz (1983, 173) who, on the ordering of headwords, or entry terms, has stated that:

> "the lexicographer should take pains to make it a foolproof instrument. His first objective must be consistency, particularly in respect of the listing of compounds. A linguistically aware person may not see much difficulty in deciding which of the components of a compound should be the headword under which the entire expression is listed, but unsophisticated users do unless they can be sure of a CERTAIN PRINCIPLE THAT IS APPLIED THROUGHOUT THE DICTIONARY (my emphasis) and which is at the same time simple enough to be acted upon by all users. To the extent that they are unaware of an established system of nomenclature, their first impulse will indeed be to presuppose the simplest and most consistent alphabetical order, which with Western languages begins with the item on the left-most side of the compound and ends at the right regardless of the semantic or syntactic value of its single components. Yet it is amazing how frequently these simple facts are disregarded by compilers of technical (as well as general) dictionaries who, while generally following the alphabetical principle, superimpose upon it an unrelated conceptual hierarchy within entries or when arranging headwords".

From the foregoing it is clear that guidelines are needed for the compilers of terminographies to help them decide the limits of terms and to choose the headword which will best represent a multi-word term or phrase. It is to be hoped that they will receive satisfactory answers as the result of consultations with subject specialists; nevertheless, at the end of the day the responses may tend to be subjective and are very often intuitive.

3. Identification and designation of LSP terms

Sager et al. (1980, 233) summarise the problems of identifying "extended terminological units" which the special lexicographer, or terminographer, faces when

> "making his decision about the unity of a term with reference to the knowledge structure of a discipline. He has a narrow set of criteria for weighing the evidence and his work may therefore be more prescriptive as a result. He establishes terminological units such as "laterally-reversed, multiple-start screw thread" which are not considered lexicalised in general language. [....] THE IDENTIFICATION OF EXTENDED TERMINOLOGICAL UNITS CAUSES DIFFICULTIES" (my emphasis).

ISO/DIS 1087 (1988) states that the definition of a term may be "any conventional symbol for a concept which consists of articulated sounds or of their written representation (= letters). A term may be a word or a phrase". The formation and recognition of LSP terms has been discussed at length by a large number of workers (see the excellent terminography bibliography compiled by B. Nkwenti-Azeh in Sager 1990, 233-237).

Rapidly developing disciplines, such as biotechnology, give rise to new concepts and hence new terms. These are the terms which are devised by specialists in the discipline; they will initially be descriptive and will be those adopted by students, interpreters and

translators. It is only subsequently that the process of standardisation takes place. There is frequently divergence in the naming process of concepts in industry where each commercial concern seeks to impose its own nomenclature on a new development. Even within the same company there may be divergence, as is the case with pharmaceuticals, where the research chemist uses a different term for the same product from that coined by the advertising department; the latter may also differ from country to country, sometimes because of connotational problems. Later it becomes necessary to standardise the terminology to avoid problems and risks in using the products, particularly internationally.

4. Choice of headword from a multiple lexical unit in LSP

Given that a term represents a concept and that a concept is an atomic unit, the (multiple) term representing it cannot therefore be split. I would argue that the individual concepts which together form another concept need to have some form of "weighting" ACCORDING TO THEIR CONTEXT so that the optimum supporting data may be provided. To give an example, should "pertussis (whooping-cough) vaccine" be entered in that order or under "vaccine, pertussis"? Does the terminographer, when seeking supporting data on the concept of a pertussis vaccine, look to "pertussis" or to "vaccine"? It appears that the order in which the lexical units of the term will be represented will depend on the reason for which the terminology is intended; if it is for a symposium on the outcome of clinical trials on pertussis vaccines (a highly restricted topic), the emphasis will be on the vaccine; if, however, the emphasis is on vaccines for use in children, it is the disease which is being combatted and which will have prior claim. These factors have a great deal in common with the subject indexing of books and it is interesting to ascertain whether frequency of occurrence is a worthwhile indicator of which unit of a term has "priority". In addition to frequency, collocation has been studied to ascertain whether these methods could help in identifying headwords in LSP.

5. Methods for resolving problems: LSP collocations and frequency

In an attempt to find whether collocation could be a factor in identifying the limit of an LSP term, the distance between the number of words constituting a term was examined. Martin et al. (1983, 84) have stated that a

> "significant collocation is one in which the two items co-occur more often than could be predicted on the basis of their respective frequencies and the length of the text under consideration".

Statistical tests have led these workers to the conclusion that more than 95% of all relevant information can be obtained by examining collocates within a range of five words of each other. The terms and phrases studied here all fall within this range.

5.1 Lexical collocations in LSP

It was in lexical rather than grammatical collocations that differences were noted between LGP and LSP (for a discussion of grammatical and lexical collocations, see Benson et al. , 1986, ix - xxviii). The lexical collocations studied in relation to LSP were: compound nouns (noun + noun(s)); compound verbs (i.e. verbalised nouns); compound adjectives; adjective(s) + noun(s); verb + preposition (prepositional verb); adverb + adjective (past participle in LSP); verb + noun idioms (as opposed to verb + noun phrases which are treated in Section 6). The above were analysed and the results are presented briefly below.

5.1.1.Compound nouns
Compound nouns were discussed in the previous section and it appears that context is of primary importance. (See also 5.2.2.)

5.1.2. Compound verbs
In the case of compound verbs consisting of noun + verb or noun + noun, the first word denotes the entry point, e.g. "to phase-separate". Where the first word has an adverbial function, e.g. "to down-load", then the second word, denoting generic action, is the recommended point of entry. Cross-referencing is advisable in both cases.

5.1.3. Compound adjectives
With compound adjectives such as "methionine-deficient" which represents a single concept, preference should be given to the first word.

5.1.4. Adjective + nouns
In the adjective + noun construction, e.g. "carrot yellow leaf virus", again the first word is preferred.

5.1.5. Verb + preposition
The verb will invariably take precedence in a verb + preposition (prepositional verb) construction, e.g. "oxidised by".

5.1.6. Adverb + adjective (in LSP = past participle)
The adverb + adjective construction, such as "genetically- engineered", should be placed under the adverb which serves to make the term more specific.

5.1.7. Idiom (verb + noun)
With a verb + noun IDIOM, which changes its meaning when split into its components, e.g. "to map genes", it is advisable to cross-reference both parts.

Three forms which were not found in the LSPs studied but which are found in LGP were: verb + adverb particle (phrasal verb), simile and metaphor; however, similes and metaphors appear frequently for impact in popular technical areas which are a combination of LGP and LSP, such as the advertising of household machinery, cameras and cars.[1]

5.2 Comparison of LGP and LSP lexical collocations

Four instances of differences in lexical collocations between LGP and LSP were identified:

5.2.1. Adjective + noun
In LGP the noun invariably predominates, e.g. "cold DAY", whereas in LSP the adjective is often a predominant past participle, e.g. "INACTIVATED vaccine".

5.2.2. Noun used adjectivally (attributively) + noun
In LGP, entry is usually at the second noun, e.g. "house PARTY", unless the combination forms an entirely new concept which cannot be deduced from its components, e.g. "birthday suit". However, in LSP the entry depends on the use for which the work, e.g. a specialised dictionary, is intended. The following example invokes the use of frequency.

In a book on "Rabies", the term "rabies" appeared 172 times and vaccine(s) 97 times. The collocation/concept "rabies vaccine(s)" appeared 16 times. Since the generic concept "rabies" not surprisingly appears more frequently than "vaccine(s)", it is the term "vaccine", being specific to a chapter in the book, which would be the entry term, say in the subject index of the book or in a glossary of terms on rabies. Conversely, in dealing with a dictionary on vaccines, the specific part of the term would in this instance be "rabies" and it is under this term that the concept of rabies would be entered.

5.2.3. Noun + verb.
In LGP this collocation usually appears under the noun, e.g. "ELEPHANTS trumpet". However, a cross-reference to the collocating verb "trumpet" would be of help to foreign language learners of English. In LSP, where the verb truly belongs to the domain, such as "viruses MUTATE", then the entry should be the verb. However, since a large number of LGP verbs are adopted by LSP where they acquire a specific meaning, cross-referencing is advisable, as in "particles disintegrate".

5.2.4. Group nouns
These are more likely to feature the specific term in LGP, e.g. "school of PORPOISES", whereas LSP favours the generic, e.g. "GENUS Morbillivirus".

Collocation in LSP terminology leads to phrases and phraseology. It can be difficult to differentiate between an idiom and a phrase and there may be some overlap in definitions; probably the most notable difference is that an idiom is semantically based, whereas a phrase is more syntactically oriented.

6. LSP phrases

LSP phrases have been referred to as terminological phrases, phrasemes, phraseological units or phraseological terms. A definition of a "terminological phrase" given by Arntz and Picht (1989, 34) is as follows:

"'Fachsprachliche Wendung' oder kurz 'Fachwendung' ist das Ergebnis der syntaktischen Verbindung von mindestens zwei fachsprachlichen Elementen zu einer Äusserung fachlichen Inhaltes, deren innere Kohärenz auf der begrifflichen Verknüpfbarkeit beruht." (An LSP phrase results from the syntactic linking of at least two LSP elements to form a phrase or expression which has an LSP content, the inner coherence of which is based on the ability of the elements to combine conceptually.) These elements are often a verb + noun clause, e.g. "avkunna en dom", "the tide ebbs and flows" (see also Schlomann, 1928; Warner, 1960).

To summarize, they may allow

- internal disjuncture -fixed word combinations but not fixed word order
- formulae, symbols, graphics, provided there is one linguistic element (Galinski, 1990)

They may not allow "free" adjectives or adverbs, that is, adjectives and adverbs which do not form a restricted collocation with their associating noun or verb. Permissible examples would be "mutant virus" and "hermetically sealed"; non-permissible examples would be "cold/warm day", "happily/miserably employed".

It should be borne in mind that a multi-word term represents a single concept, whereas a phrase (which comprises other parts of speech in English, one of which is usually a verb) contains more than one concept. Moreover, it appears that the choice of words constituting a phrase may be arbitrary when studied multilingually, as has been noted in contrastive translation studies, and may be semantic rather than syntactic. Compare for example:

(en) "the Court of Justice shall give its ruling in camera" with
(fr) "la Cour de justice statue à huis clos" and
(sv) "domstolen skall träffa avgörande inom stängda dörrar"

One form of LSP phrase has been analysed, that of verb + noun phrase. Studies were undertaken on corpora of monolingual virology texts containing c. 47,000 words in English and c. 53,000 in French to assess whether internal disjuncture is a feature which needs to be considered in verb + noun phrases because of the way it could affect the way in which phrases are indexed. As would be expected in a scientific LSP, the preponderance of the passive form of the verb in both languages and the use of the reflexive in French, were notable. In French too, those adjectives which came before nouns were mostly numbers. Articles, adjectives and prepositions were the most common forms of internal disjuncture. However, when the following examples are considered,

(en) "to transmit/isolate/reactivate/attenuate virus"
"to elicit antibody"
"to induce the uptake of virus"

(fr) "sélectionner les rayons excitateurs"
"procurer une couverture immunitaire"
"provoquer une maladie maligne"
"inactiver en 30 min."

it may be concluded that in a highly restricted subject field, internal disjuncture is rare and does not pose too great a problem in the placing of phrases. Less restricted subject fields are not so rigorous and may pose more of a problem in the positioning of headwords if the range of their collocations exceeds five words.

Collocation clearly has great importance in the ordering of LSP phrases, particularly for people producing papers in a special subject field, such as students and foreign language learners. Translators and interpreters find such information invaluable and there is a strong case for compilers of monolingual specialist dictionaries to ensure that examples of collocation are included in their work.

7. The choice of headwords from LSP phrases

Previous workers have formulated principles for the ordering of phrases. Eismann (1979, 192) advocates a "strikt formales grammatikalisch-alphabetisches Prinzip" which, it will be seen, can be used for both LGP and LSP:

> "Enthält die phE (phraseologische Einheit) ein Substantiv, wird sie unter diesem erklärt und mit Belegen versehen; bei zwei und mehr Substantiven unter dem jeweils ersten. Enthält die phE kein Substantiv, aber ein Adjectiv, so wird sie unter diesem erklärt. Die weitere Reihenfolge ist: Verb, Adverb, Pronomen, Numerale, Präpositionen. Jede phE wird so oft aufgeführt, wie sie bedeutungstragende Wörter enthält, und mit einem Verweis auf den Ort ihrer Erklärung versehen."
>
> ("If the phraseological unit [phE = phraseologische Einheit] contains a noun, it will be explained under this entry, and examples will be given. If two or more nouns are present, the phrase will be entered under the first. If the phE has no nouns, but has an adjective, it will be entered under this. The order after this is verb, adverb, pronoun, numeral, preposition. Any phE will be entered as many times as it contains separate words, and a reference will be given to the location of this description.")

However, this is at variance with the policy of Siemens' TEAM term bank described by Schulz (1980, 223), which will relate to LSP. He states that terms are stored in their basic form, as in a dictionary: as a rule, nouns are in the singular, verbs in the infinitive and so on. This also applies to complex terminology units, such as multi-word terms and compound names; these and phraseology units are recorded in their NATURAL WORD ORDER (my emphasis), i.e. when a German adjective precedes the noun it modifies, e.g. "symbolische Adresse", it would appear under "symbolische". Clearly this method is easy to adopt because of the facilities of computer retrieval; however, purely grammatical ordering does not take account of the semantic relationships between elements of phrases, nor of the changing nature of rapidly developing subject fields.

8. Conclusion

Since parts of speech other than the nouns and adjectives which comprise the majority of LSP terms are included in LSP phrases, it is clearly important for a consistent formula to be devised for the compilation of works containing special language phrases, to avoid much time being wasted in futile searches. The use of frequency is clearly an important factor but the element of subjectivity needs to be taken into account. There is need for more research into what constitutes and limits LSP phrases so that publishers are able to give consistent advice to the producers of special language dictionaries and similar works.

Endnote

1 These recommendations have been discussed in far greater detail in relation to multilingual terminology data banks and will appear in a chapter in "Terminology and Translation Studies: Introduction to Terminology", eds. W. B. Sonneveld and K. L. Loening, to be published by Elsevier Science Publishers B.V., Amsterdam.

Bibliography

Arntz, R. and Picht, H. (1989): Einführung in die Terminologiearbeit. Georg Olms Verlag, Hildesheim.

Benson, M., Benson, E. and Ilson, R. (1986): "The BBI Combinatory Dictionary of English: a Guide to Word Combinations. John Benjamins, Amsterdam and Philadelphia. Collocations and Idioms".

Cowie, A. P. (1983): "The Treatment of Collocations and Idioms in Learners' Dictionaries". In: Lexicography: Principles and Practice. Ed. by R. R. K. Hartmann. Academic Press, London.

Eismann, W. (1979). Einige Überlegungen zur Erstellung zweisprachiger phraseologischer Wörterbücher, Auf der Grundlage des bulgarisch-russischen phraseologischen Wörterbuchs von A. Koselev und M. Leonidova. In: Anzeiger für slavische Philologie 9-10/1978-79, 189- 199.

Galinski, C. (1990): Terminology and phraseology. J. Int. Inst. for Terminology Research 1 (1-2), 70-86.

ISO/DIS 1087, (1988): Terminology Vocabulary.

Martin, W. J. R ., Al, B. P. F. and van Sterkenburg, P. J. G. (1983): "On the Processing of a Text Corpus in Lexicography". In: Principles and Practice. Ed. by R. R. K. Hartmann. Academic Press, London.

Opitz, K. (1983): "On dictionaries for special registers: the segmental dictionary". In: Lexicography: Principles and Practice. Ed. by R. R. K. Hartmann. Academic Press, London.

Riggs, F.W. (1989): Terminology and Lexicography: Their Complementarity. Int. J. Lexicog., 2 (2) 89-110.

Sager, J. C. (1990): A Practical Course in Terminology Processing. John Benjamins Publishing Company. Amsterdam/Philadelphia.

Sager, J. C., Dungworth, D. and McDonald, P. F. (1980): English Special Languages: Principles and Practice in Science and Technology. Brandstetter Verlag, Wiesbaden.

Schlomann, A. (1928): Illustriertes Technisches Wörterbuch in sechs Sprachen: D-E-R-F-I-S. Ed. by A. Schlomann. Vol. 2 Elektrotechnik und Elektrochemie. Verbesserte, vervollständigte und stark vermehrte Auflage.

Schulz, J. (1980): A Terminology Data Bank for Translators (TEAM). META XXV (2), 211-229.

Warner, A. (1960): Internationale Angleichung fachsprachlicher Wendungen der Elektrotechnik. Versuch einer Aufstellung phraseologischer Grundsätze für die Technik. Beihefte der ETZ. Heft 4, s.p. VDE-Verlag, Berlin.

KEYWORDS: special language headwords, terminological entry terms, special language phrases