Lynne BOWKER, Dublin City University, Ireland

Variant terminology: frivolity or necessity?

Abstract

Terms are often supposed not to be prone to variation. Moreover, many standardizing organizations and terminology textbooks take a prescriptive approach to term formation and use in which they disparage variation. However, we believe that variation is not a random act of defiance or carelessness, but rather one which is well-motivated and useful in expert discourse. We hypothesize that multidimensional classification is one of the determining factors behind term choice and we present an empirical study of the influence of multidimensional classification on term use in which we examine variant terms in context in a one-million word corpus in the specialized subject field of optical scanning technology.

Keywords: term variation, term formation, term usage, multidimensional classification, corpus-based terminology

1. Introduction

Terms are the linguistic representations of concepts, and, in the words of Sager (1997:25): "Term formation is the process of naming the concepts required by a particular special language community for the development of cognitive processes and communication". Sager goes on to add that term formation is a conscious activity which differs from the arbitrariness of general word formation because term formation requires a greater awareness of pre-existing patterns and models, and of the social responsibility for facilitating communication and transmitting knowledge. For these reasons, it is often thought that terms are not prone to variation; however, with the increasing availability of electronic corpora comprised of texts written in Language for Special Purposes (LSP), empirical studies (e.g. Daille *et al.* 1996) are beginning to reveal that terminological variation is actually a significant phenomenon. Nevertheless, the use of variant terminology is generally not arbitrary or random, even though it may appear to be so on the surface.

One of the potential causes for this apparent randomness may be multidimensional classification, and in this paper, we explore the impact that it has on term formation and variation. However, because terms can be formed in a great number of ways and because different sets of term formation rules tend to apply to different subject fields and to different languages, we have restricted this preliminary study to the examination of the influence of multidimensionality on the formation and variation of complex noun compounds in the English-language specialized subject field of optical scanning technology. In particular, we will focus on so-called categoric compound terms (Strehlow 1983:31), i.e., compounds whose elements consist of independent characteristics whose order can be altered (e.g. colour flatbed scanner vs flatbed colour scanner).

This paper is divided into three main sections. In section 2, we consider some of the rules and hypotheses surrounding term formation and variation as found in the terminology literature. In section 3, we introduce and explain the concept of multidimensional classification. In section 4, we examine terms in context taken from a corpus in the subject field of optical

scanning technology and investigate how multidimensionality has influenced the formation and variation of these terms.

2. Term formation

The motivational processes of term formation can be viewed in terms of the need to express emerging knowledge, both to assist understanding and to ensure effective specialized communication. This has inspired many terminology researchers to try to prescribe rules for term formation. Such recommendations have been drawn up by both national and international standards bodies (e.g. ISO 704.4 WD 1993), and are proposed in numerous terminology manuals (e.g. Felber 1984:179; Picht and Draskau 1985:113; Rondeau 1984:134; Sager 1990:88). However, many of these recommendations fail to be implemented on a wide scale for two main reasons: 1) the terminologists who compile these general recommendations are often not true subject field experts and therefore they do not fully understand the requirements of the subject fields in question (and these requirements may differ from field to field), and 2) the subject field experts who actually produce and use the specialized discourse are often neither interested nor trained in linguistic issues, and they significantly outnumber the terminologists.

With regard to the nature of the recommendations compiled by terminologists, these recommendations are often either too limited in scope to be of much use, or so highly idealized that they can only be realized in a strictly controlled environment. Some of the most frequently suggested principles deal with issues such as univocity, correctness, transparency and conciseness. But even such apparently logical and straightforward guidelines raise problems. In the case of conflicting principles (e.g. transparency vs conciseness), which should take precedence? Are each of these principles really desirable in all situations? Who decides which characteristics are considered to be essential?

As pointed out by Daille *et al.* (1996:217), when writing a specialized text, a subject field expert who wants to formulate ideas into pre-existing concepts may face a number of difficulties. For instance, on the one hand, the notion that the expert wants to express may be slightly different than the concepts denoted by the terms that he or she knows. On the other hand, the expert may know the correct terminological expression for a precise concept, but he or she may intentionally wish to express a slight shift in the meaning of this concept.

2.1. Compounding as a method of term formation

Compounds are created when existing words are combined to form a new syntagmatic unit. This unit has a meaning which is independent of the meanings of its constituent parts, and, in the discipline of terminology, this unit must designate a concept relevant to a given specialized subject field. Compound terms have two elements: a determiner and a nucleus. The nucleus of the compound (usually the last element) often indicates the category to which the concept belongs, while the determiner often indicates the criterion for subdivision of the category. This technique of determination gives compounds the capacity to contribute to the building of terminological systems. For example, in the compound term *colour scanner*, *scanner* is the nucleus and it indicates that the concept being referred to is a type of scanner; *colour*, meanwhile, is the determiner and it describes how this type of scanner is different from other scanners, i.e., it can scan colour data and not just black-and-white or greyscale

data. Depending on the nature of the nucleus (e.g. object, property, process, etc.), the determiner can specify different types of features (e.g. purpose, location, function, method, material, etc.).

Thus we can see that compounding achieves two objectives: it narrows a concept's intention, thereby creating a closer determination of the concept, and at the same time, it shows the relation that exists between the new concept and its origin. As pointed out by Sager (1997:30), this can be considered both a strength and a weakness. On the one hand, determination can make a particular dimension of the conceptual structure more transparent. However, on the other hand, by focusing on the expression of one relation, others, which could also be conceptually significant, may be overlooked. As we will see in section 3, this has implications in multidimensional classifications.

3. Multidimensional classification

Classification plays a fundamental role in many disciplines, and its importance to terminology has been well documented in the literature. Essentially, classification is the act of grouping things into classes on the basis of perceived similarities, expressed as characteristics, that are shared by each class member. It follows that what is considered to be like or unlike depends on which characteristic is chosen to be the classifying characteristic. In spite of the tendency of conventional terminology publications and manuals to present one "correct" way of classifying a given concept or subject field, it is commonly accepted that people can "see the same thing in different ways". We use the term *multidimensionality* to describe the phenomenon of classification that occurs when *more than one* characteristic can be used to distinguish between things, and hence those things can be classified in more than one way. A *dimension* represents one particular way of classifying a group of things; a classification with more than one dimension is said to be *multidimensional*.

For example, the concept SCANNER can be classified according to the characteristic 'colour capability' producing the subordinate concepts COLOUR SCANNER, GREYSCALE SCANNER, and MONOCHROME SCANNER. However, there are other ways in which SCANNER can be classified, based on different characteristics that scanners can have. For instance, SCANNER can also be classified according to the characteristic 'design', producing the subordinate concepts HANDHELD SCANNER, FLATBED SCANNER, SHEETFED SCANNER, and OVERHEAD SCANNER. Other classifications are also possible.

Multidimensionality can be caused by a wide variety of factors, including language, culture, relation type (e.g. generic vs partitive), time period, purpose, different levels of perception and cognition, context, scientific schools of thought, and opinion (Bowker 1995:47). The knowledge structure of a subject field can sometimes be reflected in the terms used to describe concepts belonging to that field; however, multidimensionality could mean that not everyone classifies a given subject field in the same way, or even that the same person may view a given subject field from different perspectives at different times. This is bound to have an impact on term formation and use.

The recommendations for term formation mentioned in section 2 raise a number of questions when considered in light of multidimensionality. If a term is meant to contain qualifiers indicative of some property or other essential characteristic, the qualifiers may differ

depending on the dimension under consideration. And if technical communication is a matter of clarity, then might it not be useful for subject field experts to be able to discuss concepts from a variety of perspectives, even if this means using synonyms or terms that might be considered incorrect in the general language?

Modern terminology theory is beginning to adopt a less prescriptive attitude. For example, Sager (1997:25), Rey (1995:66) and Pavel (1993:24) agree that term formation cannot be dissociated from individual creators who are integrated into a specialist community and use the term in discourse to express themselves in a particular situation. Encouraged by these new trends, we decided to undertake a descriptive study of the use of some categoric compound terms in the subject field of optical scanning technology. In particular, we wanted to investigate whether multidimensionality has any impact on term formation and use. In our opinion, meeting the needs of effective specialist communication means adopting and understanding multidimensionality in term formation. This can only be achieved by adopting a more descriptive approach to research into term formation, i.e., by studying the actual functioning of terms in specialized discourse.

Contrary to numerous statements made in the terminology literature, term choice can and does vary in specialist communication. However, we do not believe that this variation can be attributed to carelessness or arbitrariness on the part of the subject field experts. On the contrary, like Patrick (1993:270) and Strehlow (1983:32), we believe that experts formulate their expressions carefully to ensure that the information transmitted is as precise as possible. Our hypothesis that multidimensionality is important to term formation takes root in our view that specialized discourse reveals subtle nuances by means of particular phrasal choices. In other words, we believe that LSP lexicalization is the reflection, in language, of the mental processes involved in concept formation and association.

The pilot study described in section 4 adopts a corpus-based approach to terminology. By being studied in the context of communicative situations, terms are no longer seen as isolated items in dictionaries or as part of a quasi-artificial language that is devoid of any of the functions of other lexical items. It is hoped that observation of usage will lead to a better understanding of the motivation behind the creation of linguistic variants and the role that they play in specialist communication because to date, relatively little serious research exists into these complex processes. Furthermore, if it can be shown that naming patterns, as reflected in complex terms, by means of such devices as determination, derivation, etc., are developed on the basis of the systematic selection of certain properties and characteristics for overt inclusion in the form of a term, then we may actually gain some insight into the mental processes involved in concept formation and association.

4. Corpus-based investigation

Prior to conducting this pilot study, we hypothesized that when an author uses a compound term to describe a particular concept, the structure of that term is often influenced by the dimension that is considered most significant to the discussion at hand. For example, if a particular scanner has the characteristics of having a flatbed design and a colour scanning capability, the author who is primarily concerned with the scanner's colour capability will refer to the concept as a *colour flatbed scanner*, while the author who is primarily interested in the scanner's design will refer to the same concept as a *flatbed colour scanner*.

To test this hypothesis, we first compiled a corpus of over one million words. The corpus was extracted from a CD-ROM called *Computer Select* (Ziff-Davis Publishing, NY). Each disc contains thousands of English-language articles taken from hundreds of journals dealing with a wide range of computer-related topics. The texts used in this corpus were published between 1990 and 1997. For a detailed description of the corpus make-up and compilation procedure, see Bowker (1995:162).

Once the corpus was compiled, we used the concordancing tool WordSmith to extract categoric compounds and their immediate (i.e., three-line) contexts. These contexts were then examined to determine which, if any, of the two dimensions appearing in the term also featured strongly in the context. For example, in the following context for the term *colour flatbed scanner*, we felt that the dimension based on the characteristic 'colour capability' was most evident:

"...green, and white-to-dark-blue as separate 0-to-255 gray scales (also referred to as shades of gray). Thus, this COLOR FLATBED SCANNER interprets up to 256 shades for each of the three primary colors: red, green, and blue."

Meanwhile, in the following context for the same term, we felt that the dimension based on the characteristic 'design' (e.g. flatbed vs handheld) was more strongly represented:

"...while handheld models sell for about $400 to a little over $500. 16685.// The price gap between these and COLOR FLATBED SCANNERS is steadily decreasing, however the HP ScanJet IIcx flatbed, for example, costs as little as $1,000."

Each context was evaluated and placed into one of the following four categories according to the predominance of any given dimension:

1) primary dimension: the dimension based on the characteristic referred to by the *first* element of the compound term was the dimension referred to most predominantly in the context. (In the case of the term *colour flatbed scanner*, this would be the dimension based on the characteristic 'colour capability').

2) secondary dimension: the dimension based on the characteristic referred to by the *second* element of the compound term was the dimension referred to most predominantly in the context. (In the case of the term *colour flatbed scanner*, this would be the dimension based on the characteristic 'design').

3) both dimensions: both the primary and secondary dimensions featured strongly in the context.

4) neither dimension: neither the primary nor the secondary dimensions featured in the context.

Tables 1 through 3 illustrate a sample of some of our findings, which will be discussed in greater detail in section 4.1.

	colour flatbed scanner (75 contexts)		flatbed colour scanner (63 contexts)	
Dimension in context	No. of contexts	exp. as %	No. of contexts	exp. as %
primary	27	36%	14	22%
secondary	3	4%	9	14%
both	1	1%	3	5%
neither	44	59%	37	59%

Table 1. Categorizations of contexts for the categoric compounds *colour flatbed scanner* and *flatbed colour scanner*.

	colour desktop scanner (17 contexts)		desktop colour scanner (20 contexts)	
Dimension in context	No. of contexts	exp. as %	No. of contexts	exp. as %
primary	6	35%	7	35%
secondary	0	0%	3	15%
both	0	0%	0	0%
neither	11	65%	10	50%

Table 2. Categorizations of contexts for the categoric compounds *colour desktop scanner* and *desktop colour scanner*.

	colour handheld scanner (15 contexts)		handheld colour scanner (13 contexts)	
Dimension in context	No. of contexts	exp. as %	No. of contexts	exp. as %
primary	6	40%	5	38%
secondary	0	0%	1	8%
both	1	7%	0	0%
neither	8	53%	7	54%

Table 3. Categorizations of contexts for the categoric compounds *colour handheld scanner* and *handheld colour scanner*.

4.1. Discussion of data

In all cases, the clear majority of contexts featured neither of the dimensions in question. This is not particularly surprising given the different types of contexts that exist (e.g., defining, explanatory, associative, usage example (Rondeau 1984:81)). In a specialized text, it is likely that a given term would only need to be defined or explained once or twice (and possibly not at all if it is presumed to be well-known to the target audience and is not the actual focus of the document in question). In contrast, however, this term would likely appear in a

considerably higher number of associative or usage example contexts, which would contain few references to characteristics or indications of dimensions.

It was also the case for each set of terms that the number of contexts featuring both dimensions was very small. Again, this is not too surprising because, as stated above, the number of defining contexts, in which references to multiple characteristics or dimensions would be likely to occur, will probably be limited.

However, of more significant interest is the general trend in term choice shown by experts when the context contained reference to only one of the two dimensions featured in the term itself. In this situation, experts showed a preference for using the term that had as its *first* element, the characteristic which reflected the dimension featured in the context. For example, if the context was focusing primarily on the 'colour capability' of a scanner, then the expert was more likely to use the term in which *colour* was the first element (e.g. *colour flatbed scanner*, *colour desktop scanner*, *colour handheld scanner*) rather than the second element (e.g. *flatbed colour scanner*, *desktop colour scanner*, *handheld colour scanner*).

4.2. Other considerations

The data presented in tables 1 through 3 was obtained from a small pilot study and while it may be used to indicate a general trend, we acknowledge that it cannot be deemed wholly conclusive. There are numerous other factors that must be taken into consideration. We have attempted to address a number of these factors to some degree, but others have proved more elusive and challenging.

4.2.1. General vs idiosyncratic usage
It is important to establish whether a term in the corpus is used by only one particular author or whether it is a generally accepted term in the field. In the case of the pair of categoric compounds *flatbed colour scanner* and *colour flatbed scanner*, we were able to establish that in our corpus, each of these terms were used in more than a dozen different texts written by different authors.

4.2.2. Life cycles of terms
It is a well-known fact that a term has a life-cycle. According to Sager (1990: 59), most terms are initially provisional until the equation *term-definition-concept* becomes widely accepted and incorporated into the lexicon of a particular LSP. At this stage, one of a number of rival terms gains a higher status and the rivals will gradually lose ground. Again, for the pair of categoric compounds *flatbed colour scanner* and *colour flatbed scanner*, we were able to establish that in our corpus, each of these terms appeared in texts spanning the full eight-year period covered by the corpus.

4.2.3. Purpose of the text
Another factor which could have an impact on term choice is the purpose of the text. It is generally accepted that term choice can vary depending on the audience in question (e.g. a technical term may be used for an expert audience and a more general term for a lay audience), but this type of distinction is not relevant for two different categoric compounds such as *colour flatbed scanner* and *flatbed colour scanner* as they belong to the same register. However, other reasons for term variation may be more relevant. For instance, deliberate variation may occur if the term is part of a marketing document where it is important for the

concept (and hence the term used to designate it) to be different and somehow stand out from the crowd. In a large, semi-automatically compiled corpus such as the one used for this study, it would be very difficult and time-consuming to determine the specific purpose of each text. Therefore, we were unable to take this factor into account in our pilot study.

4.2.4. External alterations to the text
Another factor worthy of consideration is whether or not the text has been subjected to external alteration after it left the author's hands. In some cases, the text may have been examined and possibly altered by an editor who is not a subject field expert and who may have changed terms in order to achieve consistency without realizing that the author may have been using variant terminology to express a nuance of difference. This factor would also be very difficult to determine to any degree of certainty in our corpus, and we did not take this into account in our pilot study.

4.2.5. Other means for expressing dimensional relevance
Finally, we must acknowledge the possibility that experts may use other means to express the relevance of certain dimensions which may not be reflected in texts. For example, it is possible that, when speaking, experts may use verbal emphasis to stress the importance of a particular dimension; i.e., an expert may say "colour *flatbed* scanner" putting an emphasis on *flatbed* even though it appears in the secondary position. Once again, it is impossible to determine this type of behaviour from our corpus. Perhaps as technology improves, it will become easier to compile the type of spoken-language corpora that would make this type of study more feasible.

5. Concluding remarks

Based on the evidence presented in the corpus, we suggest that subject field experts are not careless or arbitrary in their choice of terms. Rather, we suggest that an expert's choice of term can actually be driven by multidimensionality; in other words, an expert does not choose a term based on principles of univocity or grammatical correctness, but rather based on factors such as the purpose or the focus of the communication situation at hand. An expert may consciously choose to use one term to emphasize a particular dimension at one time, and another term to focus on another dimension at another time, even though both terms are referring to the same concept. Moreover, we submit that this process is both valid and necessary in specialist communication, where there is an emphasis on precision.

We therefore support the move away from the prescriptive rules which cause conflict and confusion between subject field experts and terminologists, and towards a more descriptive approach to terminology which will help terminologists to appreciate and understand what motivates the formation and use of terms by subject field experts.

5.1. Further research

The possibilities offered by computer-assisted analysis of large quantities of linguistic data are significant and far from mapped out. Concordances and statistical data will no doubt help terminologists gain considerable insights into the naming techniques and comprehension of complex terminological units in many different subject fields, as well as into the way we represent knowledge by linguistic means.

This particular study represents only a preliminary exploration of the influence of multidimensionality on term formation and use. It would be interesting, for example, to see if the same kind of categoric term variation occurs in other types of specialized subject fields, and to investigate why or why not. We have already collected some data which indicates that this type of variation occurs in other computer-related subfields (e.g. *colour laser printer* vs *laser colour printer*; *17" SVGA monitor* and *SVGA 17" monitor*), but it would be even more interesting to see if the same type of variation exists in fields such as law, medicine, arts, social sciences, etc. It would also be interesting to determine if the same trends in term choice apply when terms contain elements which make specific reference to three or more dimensions, such as: *flatbed 36-bit colour scanner; 39-bit desktop flatbed scanner; three-pass 24-bit colour scanner; 24-bit colour desktop flatbed scanner;* and *single-pass 24-bit colour 400-by-800-dot-per-inch scanner.*

The difficulty here is in finding a sufficient number of contexts for these terms. Often, an author will use the complete term early on in the text and then later refer to it by a shortened version (e.g. simply as "the scanner"). Our corpus would have to be much larger in order to gather enough contexts to allow us to make reliable observations about such terms.

Finally, it would be interesting to explore whether or not this trend occurs in languages other than English. We have already undertaken some preliminary work in French, where we have encountered similar variations including *scanner à plat A4* vs *scanner A4 à plat* and *scanner à main couleur* vs *scanner couleur à main.*

In closing, it is worth considering the following comment by Sager (1990:62):

"Any attempt to discover regularities in term formation must, however, be fully aware of the limited usefulness of this enterprise and of the circumstances in which term formation occurs. Most new terms are formed as and when new concepts are created in such instances as new discoveries, restructuring of existing knowledge, incidental observations or planned industrial developments. In each of these cases, the new concept to be named is seen in a particular light in relation to other concepts around it. It is therefore not surprising that the linguistic sign for a concept can be quite arbitrarily chosen and often is".

The study presented here does have limitations, but in spite of these, it is hoped that this work will provide a basis for future research on the theme of terminological variation influenced by multidimensional classification.

Acknowledgements

6. References

Bowker, Lynne (1995). *A Multidimensional Approach to Classification in Terminology: Working Within a Computational Framework.* PhD Thesis, University of Manchester Institute of Science and Technology.

Daille, Béatrice, Habert, Benoît, Jacquemin, Christian, and Royauté, Jean. (1996). "Empirical Observation of Term Variations and Principles for their Description," *Terminology.* Vol. 3(2), 197-257.

Felber, Helmut. (1984). *Terminology Manual.* Paris: Unesco.

ISO 704.4 WD. (1993). "Terminology Work: Principles and Methods". Geneva: International Organization for Standardization.

Pavel, Silvia. (1993). "Neology and Phraseology as Terminology-in-the-Making," *Terminology: Applications in Interdisciplinary Communication.* H. Sonneveld and K. Loening (eds). Amsterdam: John Benjamins, 21-34.

Patrick, Y.L. (1993). "The Amazing World of Compound Adjectives," *Translation – The Vital Link* (Proceedings of the 13th FIT Conference). London: Institute of Translation and Interpreting, 267-271.

Picht, Heribert and Draskau, Jennifer. (1985). *Terminology: An Introduction.* Guildford: University of Surrey.

Rey, Alain. (1995). *Essays on Terminology.* Translated and edited by Juan C. Sager. Amsterdam: John Benjamins.

Rondeau, Guy. (1984). *Introduction à la terminologie (2e édition).* Québec: Gaëtan Morin.

Sager, Juan C. (1990). *A Practical Course in Terminology Processing.* Amsterdam: John Benjamins.

Sager, Juan C. (1997). "Term Formation," *Handbook of Terminology Management.* S.E. Wright and G. Budin (eds). Amsterdam: John Benjamins, 25-41.

Strehlow, Richard A. (1983). "The Varieties of Compound Terms and their Treatment," *Standardization of Technical Terminology: Principles and Practices.* C.G. Interrante and F.J. Heymann (eds.) Philadelphia: American Society for Testing Materials, 26-33.