Lynne Bowker
School of Applied Language and Intercultural Studies, Dublin City University, Dublin, Ireland

Peering Through the Linguistic Keyhole: What Can Term Choice Tell Us About Knowledge Organization?

Abstract: It has often been suggested that terms are not 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 due to arbitrariness or carelessness, but rather that it 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.

1. Introduction

Humans have a long-established propensity for ordering things. Order helps us cope with the world around us. In many cases, imposing order entails prioritizing certain elements, and those things which are accorded the highest priority are generally the things that «come first». This pattern of associating priorities with primary positions can be seen all around us: in the entertainment world, where the priority is entertaining the maximum number of people (and thereby making the maximum amount of profit!), the songs or movies that are the most popular achieve «number one» status on the charts or at the box office; on a list of things to do, the action which is considered the most urgent or important is often at «the top of the list»; when travelling, if comfort and good service are your top priority, then «first class» is the way to go. Given this predilection for putting important things «first», why then does it seem so unreasonable to accept that a similar situation might arise in the context of term formation and use?

Terms are the lexical items used to describe concepts in highly specialized subject fields. Because terms are used in specialist communication, where there is an emphasis on clarity and precision, it is often thought that they are not prone to variation. However, in this paper, we will present the results of an empirical pilot study which indicate that term choice can and does vary, and that experts tend to apply the heuristic of «most important element first» when forming and using certain types of compound terms.

Unlike many terminology researchers, who frown on terminological variation, often putting it down to sloppiness on the part of the expert, we support the views of Patrick (1993:270) and Strehlow (1983:32), who believe that experts formulate their expressions carefully to ensure that the information transmitted is as precise as possible. Our hypothesis that multidimensional classification 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 term formation and use is the reflection, in language, of the mental processes involved in concept formation and association. Furthermore, if it can be shown that naming patterns, such as those reflected in complex terms for example, 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 gain further insight into the mental processes involved in concept formation and association.

This paper is divided into three main sections. In section 1, we consider some of the notions concerning term formation and variation as found in the terminology literature. In section 2, we introduce and explain the concept of multidimensional classification. In section 3, we examine terms in context taken from a corpus in the subject field of optical scanning technology and investigate how multidimensional classification has influenced the formation and variation of these terms.

2. Term Formation

Terms, which are the linguistic representations of concepts, are often supposed not to be prone to variation. Sager (1997:25) identifies term formation as the process of naming the concepts required by a particular special language community for the development of cognitive processes and communication, adding 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.

The motivational processes of term formation can therefore be viewed in terms of the need to express emerging knowledge, both to assist understanding and to ensure effective specialist communication. Some researchers feel that the most effective way to meet these needs is to prescribe recommendations for the formation and use of terms. One of the primary recommendations promotes univocity (i.e., a concept should be designated by only one term and a term should be assigned to a single concept only); others call for transparency and concision.

Although they seem logical on the surface, many of these recommendations fail to be implemented on a wide scale because they 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. 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). For example, 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.

According to Pavel (1993:24): «Finding out the causes, the patterns or regularities hidden behind such apparent randomness is one of the new tasks facing terminologists.»

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 the formation and variation of terms. 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).
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 MONOCROME 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.

There are a wide variety of factors that can lead to multidimensional classification, 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 univocity principle discussed in section 1 raises concern when considered in light of multidimensionality. If specialized communication calls for clarity and precision, 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 terminological variants?

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.

The pilot study described in section 3 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.

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 wants to prioritize the scanner’s colour capability will refer to the concept as a \textit{colour flatbed scanner}, while the author who wants to prioritize the scanner’s design will refer to the same concept as a \textit{flatbed colour scanner}.

In order to test this hypothesis, we began by compiling a one-million word corpus which we extracted from a CD-ROM called \textit{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 typical contents and compilation procedures associated with specialized corpora, 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- or four-line) contexts. These contexts were then examined to determine which, if any, of the two dimensions appearing in the term was also prioritized in the context. For example, in the following contexts for the term \textit{colour flatbed scanner}, we felt that the dimension based on the characteristic ‘colour capability’ was given priority:

«...more than 8 bits? Most \textsc{COLOUR FLATBED SCANNERS} claim to capture 8 bits of information per colour for a total of 16.7 million colours, but high-end scanners use 12 bits per colour internally, sampling down to 8 bits for output. Others offer 9 bits (such as La Cie Ltd’s SilverScanner II) or 10 bits (such as the Arcus Plus from Agfa) per colour.»

Meanwhile, in the following context for the same term, we felt that the dimension based on the characteristic ‘design’ was prioritized:

«...Canon Computer Systems Inc. this week will ship a \textsc{COLOUR FLATBED SCANNER} that the company claims will be the smallest on the market. The IX-4015, which Canon has targeted at the small office and home office markets, is about 11 x 16 x 3½ and can scan documents up to 8.5 x 14½ at a hardware resolution of 400 dpi in 20 seconds.»

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

1) \textbf{primary dimension}: the dimension based on the characteristic referred to by the \textit{first} element of the compound term was the dimension prioritized in the context. (In the case of the term \textit{colour flatbed scanner}, this would be the dimension based on the characteristic ‘colour capability’).

2) \textbf{secondary dimension}: the dimension based on the characteristic referred to by the \textit{second} element of the compound term was the dimension prioritized in the context. (In the case of the term \textit{colour flatbed scanner}, this would be the dimension based on the characteristic ‘design’).

3) \textbf{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 3.1.

<table>
<thead>
<tr>
<th>Dimension in context</th>
<th>Colour flatbed scanner (75 contexts)</th>
<th>Flatbed colour scanner (63 contexts)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of contexts</td>
<td>exp. as %</td>
</tr>
<tr>
<td>primary</td>
<td>27</td>
<td>36%</td>
</tr>
<tr>
<td>secondary</td>
<td>3</td>
<td>14%</td>
</tr>
<tr>
<td>both</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>neither</td>
<td>44</td>
<td>59%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dimension in context</th>
<th>Colour desktop scanner (17 contexts)</th>
<th>Desktop colour scanner (20 contexts)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of contexts</td>
<td>exp. as %</td>
</tr>
<tr>
<td>primary</td>
<td>6</td>
<td>35%</td>
</tr>
<tr>
<td>secondary</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>both</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>neither</td>
<td>11</td>
<td>65%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dimension in context</th>
<th>Colour handheld scanner (15 contexts)</th>
<th>Handheld colour scanner (13 contexts)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of contexts</td>
<td>exp. as %</td>
</tr>
<tr>
<td>primary</td>
<td>6</td>
<td>40%</td>
</tr>
<tr>
<td>secondary</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>both</td>
<td>1</td>
<td>7%</td>
</tr>
<tr>
<td>neither</td>
<td>8</td>
<td>53%</td>
</tr>
</tbody>
</table>

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 although we felt it may indicate a general trend, we acknowledge that it cannot be deemed wholly conclusive. There are a number of 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 such as that of univocity, but rather based on factors such as the purpose or the focus of the communication situation at hand. In other words, an expert may consciously choose to use one term to prioritize a particular dimension at one time, and another term to prioritize another dimension at another time, even though both terms are referring to the same concept. Furthermore, 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.

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

The author would like to thank Jennifer Pearson of Dublin City University, and Angela Chambers, Jean Conacher and Tadhg Ohflearnain, all of the University of Limerick, Ireland, for their constructive feedback on an earlier version of this paper.

References


Advances in Knowledge Organization, Vol.6(1998)


