A Corpus-Based Investigation of Variation in the Organization of Medical Terms

Abstract: It has often been suggested that specialized 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, I believe that variation is not due to arbitrariness or carelessness, but rather, that it is well-motivated and useful in expert discourse. Terminologists and translators who are not aware of an expert's motivation for linguistic variation, risk distorting the intended meaning of a term or text. Therefore, the aim of this paper is to present a corpus-based investigation into some of the underlying patterns of terminological variation in medical discourse.

1. Background

The motivation for conducting the study described here stems from an occasion when I found myself faced with the task of translating a text in the subject field of medicine. Although I am a trained translator, I do not specialize in medical translation, and I therefore found myself relying heavily on resources, such as dictionaries. As I consulted these dictionaries, I found that many of the entries contained cross-references to synonyms or linguistic variants. Indeed, as pointed out by researchers such as Pilegaard (1997:174) and Taczalska (1997:425), medical language contains possibly the richest proliferation of linguistic variants of any specialized language. Unfortunately, however, these dictionaries provided almost no guidance with regard to which of these variants should be used in a given context. The high incidence of linguistic variation in medical texts presents a challenge to translators and terminologists, who have typically been trained to embrace terminological standardization, to disparage synonymy in favour of monosemy, and to employ consistency rather than linguistic variation.

The aim of this paper is to investigate what motivates a medical expert to choose one linguistic form over another, and to establish whether the standardizing approach generally favoured by translators and terminologists should be applied to medical discourse. In part 2, we will briefly explore the views on terminological variation as presented in the terminology literature and explain the importance of understanding the factors that motivate term choice. In part 3, we will introduce the notion of “combining forms”, which is the specific type of linguistic variation that is the subject of this study. In part 4, we will describe the results of the corpus-based investigation that was undertaken in order to identify factors motivating terminological variation in the field of medicine.

2. Views on Terminological Standardization and Variation in Translation and Terminology

The fundamental task of a translator is to translate the words in a text, and to do this the translator requires an understanding of the context in which the words operate. Nevertheless, as pointed out by Rogers (1999:104), claims about the importance of context are often relativized when dealing with a language for special purposes (LSP), such as medical language. In such cases, the translation of technical terms is often said to be context-independent. Rogers goes on to note that this view is also found in the discipline of
terminology, which is concerned with the study of specialized terms and concepts and the relations between them. Here, there is a widespread belief that terms are not prone to variation within a particular subject field. For a long time, terminologists have aimed to standardize LSPs, striving to establish a one-to-one match between terms and concepts, thus eliminating synonymy and polysemy.

Nevertheless, as displeasing as it may be to translators and terminologists, there is an undeniable difference between the way terms are used in context and the context-independent way in which they are represented in a dictionary. The role of the context is one that needs to be taken seriously by translators and terminologists, and I fully support Rogers' view that it is grossly oversimplifying matters to merely view language use as, “a bit of a mess which needs to be fixed through standardization” (1999:106).

Subject field experts frequently choose not to follow the recommendations for term formation and use prescribed by terminologists, and often for good reason. The terminologists who compile these recommendations are not generally true subject field experts themselves and therefore they may not fully understand the requirements of the subject fields in question (and these requirements may differ from field to field). With regard to the nature of the recommendations, 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.

Translators and terminologists, however, have typically been trained to respect linguistic standards. They often assume that an author's use of a linguistic variant rather than of a single preferred term is arbitrary and therefore the result of carelessness. They may therefore be tempted to "overrule" the lexical choices made by a subject field expert in order to conform to standards, and this could have disastrous results. As noted by Rogers (1999:109), the terms in an LSP text are typically characterized as the main carriers of specialist information: their accurate translation is therefore crucial. Translators or terminologists who reject deliberately chosen terms in favour of inappropriate standardized terms may inadvertently end up changing the intended meaning of the text.

The assumption that the use of linguistic variants is not the result of a deliberate choice but rather of sloppiness is one that requires further investigation. Systematic patterns of variation need to be explained rather than ignored. I support the notion that the use of variant terms is generally not arbitrary or random, even though it may appear to be so on the surface, and I agree with Pavel (1993:24), who maintains that finding out the causes, the patterns or regularities hidden behind such apparent randomness is one of the new tasks facing terminologists and translators.

3. "Combining forms" in Medical Terminology

It was observed in section 2 that the terminology of medicine contains an abundant number of linguistic variants. Many of these variants consist of elements known as "combining forms" that can be joined together in different ways (e.g., enterogastritis and gasteroenteritis). According to Stedman's Medical Dictionary (1995), combining forms make up 90-95% of medical vocabulary. As pointed out by Pilegaard (1997:170), translators are concerned with communicating the overall meaning of stretches of text, and they begin by decoding the units which carry that meaning: the words. However, in medicine perhaps more than any other LSP, the element that is typically most important to meaning is not the word but the morpheme. Anyone who comes across compound medical terms for the first time is likely to run into what may be referred to as a lexical barrier: the opacity of semantically complex medical terms. This problem is aggravated by the rapid rate of advancement of the medical sciences and the speed at which medical vocabulary is being expanded. Nevertheless, virtually all the complex words of Latin-Greek origin are built up from simpler, smaller parts. If the meanings of the smaller parts are known, it is possible to deduce the meaning of the
complex word itself. For example the term *otorhinolaryngology* can be broken down as follows: *oto* means *ear*, *rhino* means *nose*, *laryngo* means *larynx*, and *logo* means *study*. Therefore, *otorhinolaryngology* is the study of the ears, nose and larynx. However, analyzing a compound medical term is not always as simple as determining the meaning of the separate parts. The same parts can often be combined in multiple ways, which may produce different nuances of meaning. As pointed out by Rogers (1997:219), the arbitrary or non-arbitrary nature of linguistic variation is of particular importance for translators, who have to decide whether the lexical variation in the source text should be retained in the target text. A translator or terminologist who fails to notice that an expert has combined the elements in a particular sequence in order to express a certain meaning may run the risk of altering the intended meaning if he or she recombines the elements in a different order (e.g., in a misguided effort to achieve terminological consistency within a text or to conform to inappropriately prescribed terminology standards).

### 4. Using a Corpus to Identify Patterns of Linguistic Variation in Medical Texts

As noted in section 3, much of the medical vocabulary is comprised of combining forms, and these forms can be joined in different ways. The aim of this study is to investigate what motivates a medical expert to choose one linguistic combination over another. Discovering the motivations behind term choices is important for both theoretical and practical reasons. On a theoretical level, it may help us to determine the extent to which lexicalization is the reflection, in language, of the mental processes involved in concept formation and association. On a more practical level, it will be of great help not only to medical experts who wish to ensure that they communicate their ideas clearly to other experts, but also to those involved in disciplines such as translation, terminology, and library and information science. For these non-medical experts, a clear understanding of the motivations behind term choice will help them to avoid misunderstanding the intended concepts, and thus to avoid misrepresenting or misplacing them in their translations, classifications, etc.

Many of the available resources on medical terminology, such as dictionaries, present users with misleading or incomplete information. For example, *Stedman’s Medical Dictionary* (1995) lists the terms *cardiovascular* and *vasculocardiac* as synonyms; however, an examination of these terms in context (as used by medical experts) reveals that they are in fact consistently used in different circumstances. It appears that the decision about how to "organize" the term (i.e., in what order the elements of the term are combined) is not simply an arbitrary choice, as might be expected in the case of synonyms, but rather a choice which may be highly motivated based on a variety of possible factors.

For this study, I adopted a corpus-based approach, which made it possible to study terms in actual contexts, rather than in isolation as they are presented in dictionaries. I began by compiling a 500,000-word electronic corpus of articles taken from the Medline database (1995-2000). Using the WordSmith Tools concordancer, which is a software package that allows users to quickly and easily retrieve from a corpus all the occurrences of a particular search term and its immediate context (see Table 1), I then proceeded to examine the different combinations of combining forms in context in order to identify possible motivations behind term choice. This study has revealed that terms which appear on the surface to be very dynamic and variable can actually be shown to be quite stable or predictable because they tend to be used regularly in a restricted set of circumstances.

The following sections describe a number of potential motivations behind term choice in the LSP of medicine. They are grouped into three broad categories: conceptually-motivated choices, linguistically-motivated choices and socially-motivated choices. In the context of this paper, we have focused primarily on the first two categories and will discuss the third category only in very general terms.
Table 1. A sample concordance for the search pattern *cardiopulmonary* generated using WordSmith Tools.

4.1 Conceptually-Motivated Term Choices

In the context of translation and terminology, conceptually-motivated term choices are the most important because if the translator or terminologist fails to recognize them, and instead chooses to recombine the elements in a different order, the intended meaning of the term will almost certainly be changed. Two related patterns of conceptually-motivated term choice were revealed in my corpus: patterns of directionality and patterns of origin and consequence of a pathology.

Patterns of directionality occur when two elements are combined in an order that shows movement from point A to point B. For example, the combination *craniocaudal* indicates movement from a point closer to the head [cranium] towards a point closer to the tail [cauda], whereas *caudocranial* indicates movement in the opposite direction. Similarly, the term *ventrodorsal* is used to describe movement from front [ventral] to back [dorsal], whereas *dorsoventral* describes movement from back to front. Frequently, both possible combinations can be used to modify the same nouns. For instance, the corpus reveals that *atrioventricular* and *ventriculoatrial* can both be used to modify *conduction*, but in the case of *atrioventricular conduction*, the cardiac impulse is travelling forward from the atria to the ventricles, whereas in the case of *ventriculoatrial conduction*, the cardiac impulse is moving backward from the ventricles to the atria.

A similar type of conceptually-motivated term choice focuses on the originating location and resulting consequences of a given pathology or condition. For example, the corpus provides evidence that a *gastroenteric lesion* is one that originates in the stomach and erodes into the small bowel, which results in acidic stomach contents leaking into the bowel, whereas an *enterogastric lesion* is one that erodes from the bowel into the stomach and results in feces leaking into the stomach.

In cases where an expert has chosen a term on the basis of directionality or on origin/consequence of a pathology, a translator or terminologist who combines these terms in the wrong order will end up presenting a meaning quite different to the one that was intended by the author.

4.2 Linguistically-Motivated Term Choices

Linguistically-motivated term choices are not as important as conceptually-motivated term choices with regard to the potential for meaning distortion, but they are important nonetheless because they will affect the naturalness or idiomaticness of the text. One of the most common criticisms levelled at translations is that they do not “sound right”. In other words, even if the translator has not distorted the meaning of the text, he or she may have chosen expressions that are not idiomatic in the LSP in question and reading the text may therefore be an unsettling experience for readers because they are forced to work harder in order to understand the text. Based on the evidence in the corpus, it appears that experts make linguistically-motivated term choices on the basis of factors such as collocation, semantic prosody and abbreviations.

Collocations are generally understood to be groups of words that “hang around together” (i.e., words that appear together on a regular basis). Although it has already been
established that combining forms can be joined in different ways, a particular sequence may have become established as part of a collocation or fixed expression. Once such a pattern has become established, it would sound odd to someone familiar with a given LSP if that pattern were to be changed. For example, the terms *cardiopulmonary* and *pulmonary-cardiac* are shown by the corpus to collocate with very different sets of nouns. *Cardiopulmonary* collocates frequently with *bypass* and *resuscitation*, but there are no instances of *pulmonary-cardiac* collocating with these nouns. Instead, *pulmonary-cardiac* collocates with *deficiency* and *insufficiency*. These collocations also reveal that these two terms have different semantic prosodies. As defined by Louw (1993:157), semantic prosody refers to the connotation (whether positive, negative, or neutral) that is expressed by a lexical item in association with its collocates. *Cardiopulmonary* seems to have a more positive semantic prosody, as expressed by *bypass* (a lifesaving operation) and *resuscitation* (bringing someone back to life), whereas the semantic prosody of *pulmonary-cardiac* seems to be more negative, as expressed by *deficiency* and *insufficiency*.

Another linguistically-motivated factor that appears to have an impact on term choice is the presence of an accepted abbreviation. For example, the term *cardiopulmonary resuscitation* is so well known by the abbreviation CPR that it would seem odd to encounter the term *pulmonary-cardiac resuscitation*, even though the intended meaning of this term would not be difficult to understand. In fact, the establishment of a widely accepted abbreviation may contribute to the acceptance of one term over another, even if the form of the accepted term is not the most logical from a pathophysiological standpoint. For instance, a condition exists whereby food and saliva pass from the esophagus to the trachea [windpipe]. One might therefore expect this condition to be known as an *esophagotracheal fistula*; however, the presence of the commonly used abbreviation TEF appears to have contributed to *tracheoesophageal fistula* becoming the more widespread term. The fact that the abbreviation ETF is already used to refer to the term *electron transfer flavoprotein* may also be a contributing factor to this trend.

These examples demonstrate that although linguistically-motivated factors do not have serious implications for the distortion of meaning, translators or terminologists who ignore these factors will run the risk of alienating the reader by creating a text that is not idiomatic with regard to the LSP in question.

Other linguistically-motivated factors which have not been investigated here include morphological and phonological considerations. With regard to morphology, it should be noted that LSPs typically tend to abide by the same rules as general language with regard to what sequences of letters are permitted in a given language. For example, the combination “qfr” is not considered to be a “legal” sequence in everyday English and this holds true for medical English also. It is therefore possible that the order of certain combining forms is governed by the morphological constraints of the language in question and this would be an interesting area for future investigation.

Ease of pronunciation appears to be another linguistically-motivated factor that is used to determine term choice in an oral setting, but unfortunately this cannot be investigated using a written-language corpus. Expert testimony has indicated however that in situations where conceptually-motivated factors are not at play, ease of pronunciation can sometimes be a factor when choosing the order in which to combine terms.

4.3 Socially-Motivated Term Choices

As previously mentioned, the category of socially-motivated factors affecting term choice will only be discussed here in very general terms. This is primarily because many of these factors are difficult to glean from the type of corpus used for the investigation described here. In order to be able to determine the degree of influence of these factors, it would be
necessary to have access to more information than is typically provided in most corpora. Socially-motivated factors could include the following: evolution of language or of medical knowledge over time, prototypical or popular usage (frequency), which of the variants was first learned (habit), influence of standards or other references (e.g. MeSH), geographic variation (e.g. British vs American usage), native vs non-native speaker variation, etc. These would all be interesting areas for future investigation but would require a much more detailed and rigourously designed corpus.

5. Concluding Remarks

Language, and particularly specialized language, cannot be completely random or people would not understand one another; however, it does admit a greater degree of variability than previously thought. It is important for translators and terminologists to realize that variation is not necessarily the result of carelessness on the part of the subject field expert. While some degree of non-systematic variation is inevitable, it appears that medical experts generally make an effort to formulate their expressions carefully based on a combination of conceptually-, linguistically- and socially-motivated factors. Translators and terminologists must learn to recognize the patterns that lie beneath linguistic variation so that they do not inappropriately standardize terminology and thereby distort the intended meaning of the text.

Acknowledgements

I am grateful to Dr. Sanjay Rao of the Civic Hospital (Ottawa, Canada) for his constructive feedback on some elements of this paper. The research reported here was funded in part by the Social Sciences and Humanities Research Council of Canada.

References


WordSmith Tools: http://www.oup.co.uk/elt/catalogu/multimed/4589846/4589846.html