ΜΕΤΑ

Terminology for Translators — an Implementation of ISO 12620

Robert Bononno

Cite this article

Bononno, R. (2000). Terminology for Translators — an Implementation of ISO 12620. *Meta*, *45*(4), 646–669. https://doi.org/10.7202/002101ar

Article abstract

As far as translators are concerned, terminology is primarily an ad hoc affair, more a matter of filling in the blanks in their knowledge than systematically studying a constellation of terms in a given universe of discourse. This article sketches the historical background of terminology and discusses the importance of terminology management for translators. More specifically, it outlines a method by which translators can make use of the data categories discussed in ISO 12620 in their daily work. A specific implementation of these data categories, using off-the-shelf software, is presented.

Terminology for Translators—
an Implementation of ISO 12620

ROBERT BONONNO
New York, United States of America

RÉSUMÉ

Pour le traducteur, la terminologie répond davantage à un besoin ponctuel (comment combler cas par cas les lacunes du texte en cours de traduction) qu'à une étude systématique dans un champ bien précis. Cet article fournit un aperçu historique de la terminologie et souligne l'importance de la gestion de la terminologie pour les traducteurs. On s'attarde, en particulier, à décrire une méthode par laquelle les traducteurs peuvent intégrer dans leur travail quotidien une approche terminologique basée sur les catégories de données couvertes par la norme ISO 12620. Une application spécifique de ces catégories avec un logiciel disponible sur le marché est aussi présentée.

ABSTRACT

As far as translators are concerned, terminology is primarily an *ad hoc* affair, more a matter of filling in the blanks in their knowledge than systematically studying a constellation of terms in a given universe of discourse. This article sketches the historical background of terminology and discusses the importance of terminology management for translators. More specifically, it outlines a method by which translators can make use of the data categories discussed in ISO 12620 in their daily work. A specific implementation of these data categories, using off-the-shelf software, is presented.

MOTS-CLÉS/KEYWORDS

data categories, ISO 12620, standardization, terminology, terminology management, translator workstations

Introduction

The following article describes a fairly simple implementation of a modern terminology standard—in this case the ISO 12620 standard on data categories—that can be used by translators and terminologists. ISO 12620 is designed to promote consistency in the storage and interchange of terminological data through the use of a standard set of data categories for term entries. For those who are unfamiliar with the standard, the introduction to ISO 12620 states:

> "Terminological data are collected, managed, and stored in a wide variety of environments. For purposes of storage and retrieval, these data are organized into terminological entries, each of which traditionally treats information associated with a single concept. Data items appearing in individual terminological entries are themselves identified according to data category. Differences in approach and individual system objectives inevitably lead to variations in data category definition and in the assignment of data category names. The use of uniform data category names and definitions, at least at the interchange level, contributes to system coherence and enhances the reusability of data."

The standard was initially designed to facilitate the exchange of terminological data among existing systems by providing a consistent set of data category names that could serve as a common reference. The standard recognizes that existing termbases have made use of data categories best suited to their needs and that until now there has been little harmonization of data categories on a worldwide basis. ISO 12620 attempts to overcome this by providing standard data categories for *interchange* purposes; they are not, however, mandatory for the design of terminology databases or termbanks[1]. Rather, the large number of categories supplied in the standard should be treated as a registry against which database designers can correlate existing or planned systems. However, there is no reason why these categories can't be used directly in the design of new terminological storage and retrieval systems. In fact, doing so would go far not only in promoting adoption of ISO 12620 but in increasing consistency in the way we store and manipulate terminological data.

The article is divided into two parts. Part 1 provides some necessary background information for non-specialists about terminology and its core principles. It is difficult in an article of this size to provide a comprehensive picture of a field as multilayered as terminology, so I have concentrated on what I believe to be key issues in its historical and theoretical development. My bias is toward the translation community rather than professional terminologists. Hopefully, the current article will provide translators with a better sense of the role terminology plays in their work and the usefulness of standards.

Part 2 is a description of one application of the ISO 12620 data categories using a simple software application that has the virtues of being inexpensive and easy to use. As such it may encourage translators to take a more systematic approach to terminology than they have generally done, whether through lack of tools or training. The current implementation, which is primarily designed for personal rather than organizational use, is a stripped down version of what a large terminology database might require in terms of a database model, number of data categories, and software. In spite of its lack of sophistication, however, the current implementation is a working model that can be used not simply to demonstrate the application of the ISO 12620 data categories but also many of the fundamental principles of applied terminology.

In the past, translators, perhaps because of the nature on their professional activity, have shown limited interest in terminology as a field of activity distinct from translation. There has even been considerable antagonism toward terminology on the part of translators, who often view it as a drain on their time and resources and an impediment to translation proper. In the ideal world that terminologists have presented and translators dream of, terminological "issues" would be handled by terminologists and translation "issues" by translators. Queries would be submitted to terminologists for resolution and the results supplied to translators for integration in a translation product. In large organizations like the United Nations, resources are available to implement such a model, but few companies and certainly no translator can afford the luxury of such a bifurcation of responsibilities. Although there is no easy way to ease the demands made on a translator's time, the following illustrates one way of smoothly integrating terminology into the translation process.

PART 1

1.1. Background

When translators think of terminology they imagine it as a kind of catchall roughly synonymous with a glossary or word list. Terminology is "all the words I don't know and need to find out," the words not found in a dictionary, or the latest jargon. While it is true that current terminological practice touches upon these meanings of the word, all of them misrepresent it in one way or another.

Experienced translators understand the notion of a subject field by virtue of the fact that they either have had to specialize in a specific field of practice or regularly need to make fine distinctions between shades of meaning to match a term to a text. Although generally left unstated, such an approach implies an underlying theory of translation. At the very least such a theory would need to recognize the primacy of context for the text at hand and the importance of function in determining the adequacy of a translation. It also admits of an interpretation that is more sophisticated than traditional word-for-word or "equivalent response" theories, whose limitations have so often been discussed in the literature[2].

Why is this so? Recognition of the existence of subject field specialization in translation itself implies an awareness that there is no such thing as a "general," or generic, text for translation, one which can be approached on purely lexical grounds. Specialization requires that we divide areas of inquiry into increasingly smaller compartments with correspondingly greater focus—which is itself a form of contextualization. One consequence of such specialization is a restriction (voluntary or otherwise) of the types of texts produced, whose structure will in some way reflect their function. Engineering, medicine, law, chemistry, physics, accounting—the list can easily be extended—are broad disciplines with any number of subfields and specializations, each of which produces its own characteristic texts. These texts can be described or analyzed in terms of their formal criteria. They can also be analyzed in terms of their function or purpose as texts: What role do they serve within their respective language communities? What are they trying to convey? Who is the audience for such texts? One of the best models for framing translation in this way is functional theory and *skopos* theory in particular. Those functional theories of translation based on action theory examine the ways in which translations satisfy the *purpose* for which they were intended[3]. Such theories examine the complex interplay among author, translator, and audience and the ways in which a translation satisfies (or fails to satisfy) the requirements of a given context. Their virtue for translation lies in the fact that they give precedence to social norms and conventions in the target language. They are holistic theories that place lexical issues within a larger context designed to meet the needs of a translation client and audience (these may be the same entity). There are two important consequences of this: First, meaning does not inhere in the text itself but is conferred on a text by its readers (Vermeer speaks of the text's producers, senders, and recipients). It is socially determined. Second, and a consequence of the above, there is no single "correct" interpretation of a given text. A given interpretation of a text is governed by a number of extrinsic factors beyond our control.

Another consequence of specialization is the recognition that, for a translation to be "successful," the translator must not only look beyond purely lexical issues but must look beyond the text itself for insight into its interpretation. This goes far beyond

the traditional focus on research in translator training programs, with its reliance on dictionaries and reference material. Indeed, at its most sophisticated, there is a blurring of the distinction between translator and specialist, whose functions are combined in a single individual possessing linguistic and extra-linguistic knowledge at a fairly high level. An integral component of such an approach is, of course, the recognition of the technical vocabulary of the field of specialization.

Modern theories of terminology are based upon a clear distinction between the general lexicon, the words we use in everyday speech to express a full range of feelings and ideas, and special languages, subsets of the global language stock that are generally used for communication among experts. Already, by the mid-nineteenth century scientists and philosophers were aware of the need for a clear and concise language to describe the growing understanding of the natural environment. As early as 1847, William Whewell (1847: I) wrote: «It has been shown in the *History of Science*, and has further appeared in the course of the present work, that almost every step in the progress of science is marked by the formation or appropriation of a technical term. Common language has, in most cases, a certain degree of looseness and ambiguity; as common knowledge has usually something of vagueness and indistinctness. In common cases too, knowledge usually does not occupy the intellect alone, but more or less interests some affection, or puts in action the fancy; and common language, accommodating itself to the office of expressing such knowledge, contains, in every sentence, a tinge of emotion or of imagination. But when our knowledge becomes perfectly exact and purely intellectual, we require a language which shall also be exact and intellectual;—which shall exclude alike vagueness and fancy, imperfection and superfluity;—in which each term shall convey a meaning steadily fixed and rigorously limited. Such a language that of science becomes, through the use of Technical Terms.» Whewell's statement is prescient and demonstrates that the broad outlines of contemporary terminology theory had already been established long before they were formalized by Wüster and other modern theorists. The day-to-day (or "common") language referred to by Whewell is now often referred to as general language; any specialized (or "technical") language is referred to as a special language. Terminology as a field of study is, by definition, restricted to such special languages and there is no theoretical limit to their number or size. "The motivation provided by the dual movement towards the division of knowledge into subject fields and professional specialization of user groups leads to the establishment of special subject languages. These sublanguages are often defined in terms of the further constraints of the general language, occupying a narrower range of linguistic forms available and a narrower set or areas of usage and purposes of use" (Sager 1994). Sager's comment[4] reflects the fact that as cultures develop and new discoveries are made, there is a need to identify and organize the concepts reflected in those discoveries. The rapid and seemingly uncontrolled growth of terminology in the fields of information technology and telecommunications (acronyms, jargon, slang, etc.) is only one (perhaps most obvious) example of this phenomenon.

Regardless of their level of difficulty or obscurity, words in the general lexicon are not the kinds of things that fall within the purview of terminology[5]. Most of the general mono- and bilingual dictionaries found in the translator's toolchest have a rather relaxed approach to this distinction, although they do contain "technical" terms. The problem is not limited to general language dictionaries, however. Putting

aside the problem of "filler" words (which translators rightly complain about) culled from the general lexicon, the majority of specialized dictionaries incorporate terminologies from a variety of special languages within a broad field. Our construction glossaries, multimedia and telecommunications dictionaries, medical dictionaries, etc., incorporate terms from several areas of specialization, several sublanguages, often without distinction. The better dictionaries may provide a subject field label (mechanical engineering, organic chemistry, finance), within the entry but such categories are too broad for practical utility as special language markers.

Translators work with words, although they do not necessarily "translate words." This is not mere rhetoric but a critical distinction about the nature of translation. Now, the terms found in terminologies are also "words," but terminology, unlike translation, is not about words at all. This is strange when you stop to think about it, since most sources of terminology seem to contain words, and the more the merrier. The fact of the matter is that terminology (unlike lexicography) is constructed around concepts and concept systems and cares little for the actual signs used to represent them. (I am exaggerating a little. Terminologists, especially those working into a language of limited diffusion, are greatly concerned with the morpho-syntactic structure of terms. After all, transparency and motivation are legitimate—and justifiably so—concerns for anyone who is serious about term creation and management. The problem takes on even greater significance in the context of standardization. In industry prescriptive terminology serves not only to foster better communication but can promote efficiency and safety.) Exactly what a concept might be is a matter of some contention and far beyond the scope of this article. Although no completely satisfactory definition has yet been supplied by the theory of terminology[6], for our purposes it may be safe to say that a concept is a mental state or particular and that when we have a specific concept (DOG, CAT, etc.), we are simply in a specific mental state or rather in a specific mind–world relation[7]. Unfortunately, we have no way of representing concepts directly, we can't see, hear, touch, or smell them. Terminology gets around this problem by assigning labels, or designations, to concepts as a convenient way of manipulating and organizing them. Such labels are referred to as "terms." Definitions and descriptions are provided for concepts as a way of explicating them. Concepts, terms, and definitions are the fundamental components of any terminology.

Earlier I mentioned the relation between specialization and extra-linguistic knowledge. Such knowledge reflects more than the ability to mimic a formal text structure in a given discipline, it is the ability to accurately and efficiently conceptualize our knowledge of a given field. Those same core concepts and their interrelations, when they are limited to a given discipline or subdiscipline, are the objects of study of terminology—indeed they are its raw materials. This conceptual relationship serves as a bridge between terminology and translation, and will hopefully allay translators' fears that the two disciplines have little in common. If anything, it helps emphasize the fact that terminology and translation are, at bottom, two sides of the same epistemological coin. We make use of conceptual knowledge during translation to disambiguate a text; it enables us to correctly structure our response and arrive at a reasoned interpretation of "authorial intent." (The scare quotes are intentional. I am not trying to resurrect the notion that we can discern the author's "intent" through

the veil of text, simply that we make certain assumptions about what we believe the author intended during the course of a translation. Those assumptions need to be tested against other factors.) Conceptual knowledge bridges the gaps in a text that words only serve to highlight. This knowledge, when organized, can be thought of as a multidimensional knowledge space in which concepts can be mapped to intersecting axes representing semantic features (Sager 1990). These concepts are in turn symbolized by the terminology of the field—terminology here being understood as a collection of terms in a specific field of knowledge.

If *a* terminology is not merely a collection of terms (even though that is what we are accustomed to seeing), then what is it? A "terminology" is a systematic arrangement of concepts within a special language. Concepts, not terms. Systematic, not alphabetic. This notion of an organized arrangement of knowledge accessed through a system of signs is common to the classificatory sciences in general. Terminology's specificity lies in its use of concept systems and relations as an organizing feature. "Terminology signifies the collection of terms, or technical words, which belong to the science. But in fixing the meaning of the terms, at least of the descriptive terms, we necessarily fix, at the same time, the perceptions and notions which the terms are to convey; and thus the Terminology of a classificatory science exhibits the elements of its substance as well as of its language" (Whewell 1847). The "substance" Whewell refers to here is the content of the concepts, or "notions," that form the core of a special language.

Concepts can be arranged in various ways and exhibit certain types of relations among themselves (genus-species, part-whole, etc.) but we need to bear in mind that the relations are ordered, not arbitrary. Dictionaries and glossaries, which are arranged alphabetically, provide no indication of systematic arrangement and consequently no information about the relations between concepts in a given field. These ordered relations, although difficult to establish for the non-expert, provide valuable information about interrelations and dependencies within a subject field. A partitive, or part-whole, arrangement, for example, can provide vital clues about both the function and location of a physical object (a throwout bearing in a clutch mechanism, for example, or a bone in an anatomical structure). The importance of such an arrangement for the translator should be obvious. As translators, we commonly encounter source-language terms without (descriptive) contexts, without definitions, abstracted from any referential environment other than the text itself (which can be quite illuminating or infuriatingly vague). We may have enough information available to know what a term isn't but not necessarily what it is. In such cases it can be extremely helpful to know that the object in question is part of something else or related to something else in a particular way. If you know that X is part of a clutch mechanism or valve or pump, you are that much closer to identifying X accurately and finding a target-language equivalent than if all you know is that it's simply a mechanical part.

Conventional technical dictionaries, at least the better ones, often provide labels to indicate the field of application or domain for a headword. This may get you within striking distance of the term you need but not much closer. Such labels, although essential, provide merely a top-level organization for terms and can even be misleading (how much information do we gain by learning that a term is part of the field of electrical engineering, for example?). If we know, however, that X is a compo-

nent of a chuck assembly on a horizontal lathe, the probability of identifying its equivalent in a target language is greatly improved[8]. But a general technical dictionary won't indicate this. At best you will learn that X is used in the field of industrial engineering or machine tools, which may not tell you more than you already know. There are, of course, many specialized technical dictionaries on the market, restricted to fairly narrow areas of practice. The problem is that, with notable exceptions, the majority of such references are not arranged thematically and do not always provide a means for making fine distinctions among concepts.

1.2. How not to Do Terminology

As I noted above, translators, at least those who bother to think about terminology, even in a vague way, generally associate it with collections of unknown words or terms, the critical entries that dictionary makers always seem to miss. So "terminology" tends to be thought of as parallel columns of words: source term on one side, target term on the other. Tables are wonderful for creating this sort of structure. It's easy, it's fast, you can do it right in your word processor, and, unless you're still using an early version of some obscure software application, you can usually exchange files with others. If you're ambitious, you can add a third column for client comments or corrections.

Some of what passes for terminology gets stored in spreadsheets such as Microsoft Excel. Now, personally I have nothing against Excel or its makers, but it was never designed to store text, much less terminological data. It simply makes it easy to create big tables without going through the trouble of sizing and formatting them. But it was designed to crunch numbers, not words. And it's not a database, so there is no organizational structure, no way to search or specify a field or category, no way to create entities and attributes. Although both methods (tables and spreadsheets) are common in translation today, neither can be recommended for serious terminological research. They produce fairly large files in proprietary formats generated by applications designed for purposes that have nothing to do with terminology. If these methods are commonly used, it is because they are readily accessible and there are few terminology management programs being marketed to the individual translator.

There are other problems with such makeshift solutions, including the lack of management tools, inadequate search and retrieval tools, impoverished content (the record is reduced to term pairs), lack of efficiency, lack of good import/export tools, etc. But the biggest problem, and the one that can't be so easily overcome, is the fact that such methods fail to reflect the most basic requirements of termbase design. Not only that, they fail to reflect the most fundamental premises of terminology.

1.3. Designing for Reusability

One of the key factors in terminology management is the ability to retrieve, exchange, and manipulate data. Putting aside for now the question of just what needs to be compiled and stored, it should be obvious that the value of this information is dependent on several factors: the ability to store new data and retrieve existing data, overall ease of use (configurability, interface design), and the ability to exchange data

efficiently. Integration with other software programs is useful, and sometimes essential, but none of the requirements for terminology management are well served by word processors and spreadsheets, which result either in a disorganized mass of large numbers of small files or a single, unwieldy, monolithic file.

Typically, structured information of this sort is stored in a database of one kind or another. A conventional "flat-file" database is designed around entities and attributes and stores data in records, which are broken down into fields[9]. Modern relational databases distribute entities among separate tables (which can be thought of as individual files) that are linked by a common key. This key enables the system to identify data associated with a single entity instance distributed across several tables. There is no inherent reason why terminological data should be stored in a relational database rather than a conventional flat database, since by database standards the organization of most terminological data is fairly simple (it is also primarily text-based rather than numeric). However, the large number of data categories available to modern terminology management applications are frequently broken down into groups of entities[10] (term-related, concept-related, administrative, etc.), which lend themselves to the relational model. Since the design and structure of a relational database model are advanced topics beyond the scope of this article, I will confine my comments to general recommendations about data categories and organization.

Most translators work independently and do not as a rule exchange terminology except on an as-needed basis. If terminology is stored at all, it is stored in a word processor file or occasionally an ASCII file. Word processors, however, are not designed to provide rapid access to information, they are writing and layout tools. Although tables are convenient and easy to set up, they are primitive tools when it comes to storing data and make reuse of information problematic. They also tend to be relatively large because of the formatting codes they contain. If terminology is stored in this way on a project-by-project basis, the end result is a collection of small files scattered across a hard drive with no easy means of retrieving information from them. To find something, you have to know where it is beforehand, which somewhat defeats the purpose of storing it in binary form in the first place. Using a word processor's Find by Content command is no panacea. It forces you to use the word processor you used to create the file to locate the data, and when there are a number of files or if files are large, it is slow and inefficient.

A more intelligent way of handling terminological data is to design the system for reuse and exchange from the start. Records should be designed so that they are easy to read, easy to update, and serve as a common, centralized repository of terminological data. Terminology databases, or termbases, should provide the user with easy access to all data at all times. This is not as easy as it sounds and has become a growing problem with the abandonment of legacy operating systems and applications, where data may have been stored in proprietary or currently unsupported formats. The problem is further exacerbated by the use of different computing platforms (Windows and Mac OS primarily) and software applications. In spite of claims for seamless transfer, import/export tools routinely fail to capture the exact formatting of the source application. To add insult to injury translators tend to treat terminological data piecemeal, a problem to be solved in the context of a specific text to be translated, rather than as a growing corpus that can be developed on its own. If translators were to adopt a long-term view of terminological data,

the need to include reuse and exchange as primary design criteria would become obvious.

Unlike a word processor, a database application is designed from the start for efficient data storage and rapid retrieval. In addition to the ability to store vast quantities of data, database applications can be sorted quickly and complex searches can be conducted on specific fields, such as a source or target term. Modern relational databases also provide support for graphics of various sorts (sometimes referred to as binary large objects, or BLOBs). This enables designers to include images (the database stores a link or pointer to the actual graphic file) in a terminology record to provide visual information about a concept. Although the initial design and development of a database may turn out to be onerous for the individual translator, its use will pay off in the long run.

1.4. Categorization and retrieval

One of the key factors frequently overlooked by translators when researching and compiling terminology is the need to define and segregate terms by domain. All terminology is stipulatively restricted to a subset of the general lexicon. These subsets, or special languages, are often carefully circumscribed and highly restricted in terms of their application. A top-level domain—chemistry or engineering—provides only a vague and imprecise indication of the field of application. A term such as "cycle" can have a range of related meanings depending on the area of specialization. For example, a recent ASTM standard[11] includes the term "cycle" in a terminology of fatigue measurement. In this case the term is specifically restricted to test methods for fatigue measurement by appending the words "in fatigue" to the term, alerting the user to the fact that the definition is domain specific and may not be appropriate in other contexts.

Modern industries, such as automotive and chemical engineering, aeronautics, and medicine, comprise a range of technologies populated by experts with narrow specializations. This makes the question of deciding how best to categorize a term a decidedly non-trivial one for the translator. In *A Practical Course in Terminology Processing*, Sager (1990: 28) writes: "To date there is no single generally accepted or acceptable classification for concepts which goes beyond relatively small and well established subject areas. While this is no problem for isolated terminological collections, it is a serious problem for term banks." I have tremendous respect for Sager's work but would argue that the problem is equally acute for the individual working with a small collection of data, an "isolated terminological collection" *par excellence*. Large terminology databanks are confronted with the problem of processing—entering, analyzing, and validating—large volumes of data from a range of subject fields. The individual translator treats far smaller volumes of data but the problems of classification are just as real. Moreover, the individual generally does not have access to outside experts who can confirm a specific conceptual structure or the time to research a specific body of knowledge in depth.

The problem is further exacerbated by the fact that there are several ways of partitioning an area of knowledge and that many subject fields are interdisciplinary. Sager lists well over a dozen possible types of relationship along with the more com-

mon generic and partitive relations. In addition to broadly exclusive (or inclusive) categorizations, facetted[12] classification systems can be used to subdivide an area of knowledge by some common characteristic: property, function, shape, method, etc. The problem for the classifier (in this instance you, unfortunately) as Sager notes is that "the total number of possible concept divisions and, hence, possible facets is given by the diversity of subjects in existence." Should metal testing or fatigue measurement in the aviation industry be considered subfields of aeronautics, engineering, metallurgy, metallography, materials testing, or physics? Should fatigue measurement be listed as a subdivision of metal defects or testing methods? If we encounter a term relating to the crystal structure of a metal, how should it be categorized: physics, chemistry, crystallography, metallography? Should a caching method used to improve access to Web sites be classified under software, hardware, information technology in general, telecommunications, networking, or something else? Texts submitted for translation are rarely precategorized by the client to facilitate terminological research. Even highly specialized texts such as patents and standards can include terms that refer to objects outside the scope of the text under consideration. Obviously, these are questions for experts and the best the individual translator can hope to do is provide an intelligent guess for their categorization—or become an expert.

Although it is better to provide a general subject field than no subject field at all, such broad categorizations may be less than helpful the next time the entry is retrieved or when fine distinctions need to be made, as with the term "cycle" above. One of the key issues in the design of any termbase is the accuracy of the information it contains. By randomly entering terms into a database (or even a table for that matter) without a proper conceptual framework, you reduce their value enormously and greatly increase the risk of error. Clearly, translators who limit the subject matter of the texts they translate (who specialize for want of a better term) stand a better chance of accurately identifying and categorizing their terminology. But given the scope and proliferation of terminology in science and technology, is it realistic to assume that any one individual will be able to keep abreast of even a small part of it?

The need to keep accurate records is crucial. While translators are generally attentive to detail because of the nature of their work, they have a tendency to view terminology as nothing more than a collection of bilingual word lists—a short-term solution to a short-term problem. Although, in principle, terminology should be the client's responsibility (for example, corporate titles and departmental names can vary from company to company) and provided along with the source text, it is the rare client who will take the trouble to compile and provide a bilingual glossary, especially with smaller projects. Inevitably the work falls to the translator, who does what can be done to deliver the translation on time.

If I were to design a flowchart of the translation process, the compilation of terminology would be one of the first steps in the process. How would a terminologist approach the problem? A hypothetical situation might include the following steps:

1. Examine the source material
2. Identify and extract terms and contexts
3. Enter provisional source terminology and related data into a database

4. Research the individual concepts and locate appropriate descriptions (usually a formal definition)
5. Submit data to an expert for confirmation or approval
6. Make any needed changes to the terminology records
7. Load validated records and index and sort data

The above summary approach describes a monolingual situation and the above steps would have to be repeated for any target-language data. We also need to consider the additional difficulty of determining whether or not source- and target-language concepts match (and the extent to which they do) and if there is an appropriate target-language designation for them. Obviously, nothing in the above scenario guarantees that the text being scrutinized will be restricted to a single narrow domain or subject field. Terminologists do not ordinarily research general subjects, however. They try to identify the relevant terms in a predetermined subject field, and this subject field will have been fairly narrowly defined (human resources, for example, or endocrinology) from the start. Words from the general lexicon and common scientific terms are, therefore, eliminated from the search.

In an ideal situation, the term (more strictly speaking, concept) record will contain many pieces of information. Step 2 above would, therefore, involve cataloging not just the term, but a *useful* context[13] (a part name in a list or a random acronym is not very illuminating as contextual information), gender, part of speech, any usage notes or synonyms, bibliographic information about the source of the data, status of the term, and any appropriate administrative information (who prepared the record, when was it prepared, status of the record, etc.). For the record to be useful, however, we need more than a designation (the term), we need a valid definition or description of the concept it symbolizes. In most situations it is not the terminologist's (much less the translator's) responsibility to prepare definitions. They should be researched from authoritative sources or prepared by an expert if no print reference can be found. The source must be accurately identified. The same process must also be followed for the target-language term group and at some point the terminologist will have to determine the degree of correlation between source and target concepts.

Once the information has been compiled, it can be printed and submitted to an expert for approval and any needed changes made. Only after these steps have been followed will the data be loaded into the system and indexed (assuming it is part of a large terminology collection). Obviously, such a picture bears little relation to the image of the translator frantically looking up entries in a technical dictionary. It is a separate process from translation and requires a different, though perhaps related, skill set. And if done correctly, it is also time consuming—something translators should be aware of before volunteering to "donate" terminology to their clients. In spite of the labor-intensive nature of the work, every translator appreciates the value of accurate and reliable sources of terminology, and I would encourage translators to consider implementing at least some aspects of the above scenario in their own work. By doing so you ensure not only that you can recycle terminology with some degree of confidence but you will have created a valuable resource you can exchange with others.

PART 2

2.1. Background

What do you do once you have decided to take terminology seriously? Even with the best intentions in the world, there are at present no tools available to the translator for automating the terminology process[14]. To make matters worse, there are few affordable and customizable terminology management programs being sold commercially. Translator's Workbench, a translation memory program from Trados, includes Multiterm, an integrated terminology management program. The Workbench was selling for approximately $1,000 for standalone use at the time this article was written. If we exclude word processor tables and spreadsheets from the mix, this leaves translators with few precious tools for their work. The only alternatives are to design a terminology management system from a relational database, use a simpler "flat-file" or free-form database, or store information in text files.

You can create a kind of rudimentary database with records and fields in a word processor with which you can produce nicely formatted hardcopy output. But indexing and finding information in such a file is painfully slow, the files tend to be large, and you lock yourself (and your data) into a proprietary format. Cross-platform exchanges (remember the Macintosh?) also become problematic and word processor files may need to be converted. I know, because I've used this method to store terminology (using *WordPerfect* 5.1 merge files). It works, sort of, but the only way to make it functional is through the use of tools such as full-text indexers, which can index and search large volumes of text data. Incidentally, this is a method I would urge translators to consider for legacy data. Modern search-and-retrieval software is fast and designed to work with extremely large amounts of data, far more than the average translator could generate in a lifetime. Typical search/retrieval tools include Isys (Odyssey Development), ZyIndex (ZyLabs), dtSearch, and several high-end systems designed primarily for client—server environments (Verity, PCDocs, Excalibur, etc.). The disadvantage of any such tool is that it does not reflect structural relationships in the underlying data (assuming they exist). A termbase is a structured environment for storing text data and, as such, the structure has meaning. Termbases can also be designed with different degrees of granularity, depending on the amount and kind of information being stored and the level of discrimination that is needed. Search/retrieval tools index raw text. Whether it is structured or not, the text corpus is treated as a continuous unstructured data stream. The indexing tool is used to quickly and efficiently locate and present text strings, which can occur anywhere in the stream.

There are alternatives, but not many. For the past year or two, I have been working with a simple free-form database for the Macintosh called Infogenie[15]. You can use it to easily create structured database files, however. Such files use the conventional fields and records found in an ordinary flat-file database. You can perform a full-text search (which will search on all fields without discrimination) or a field-specific search. The size of the database is limited in principle to the amount of storage space and memory in the computer. Because the database is memory resident while active, searches are very fast. In spite of the lack of high-end tools common to relational database management systems (data integrity control, SQL support, report

writer, screen painter, etc.) the application is small, inexpensive, and functional. There is no limit to the number of fields per record and—one of the nicer features of the program—you can add or reorder fields after the database has been created and can sort records on two fields. Although I wouldn't recommend the program to anyone trying to implement a sophisticated multi-user terminology management system with hundreds of thousands of records, it can serve the individual translator well. In addition to the small file size and ease of use, you can import and export data in a variety of delimited text formats.

2.2. Design

One of the biggest problems facing designers of terminology management systems today is the number and types of data categories that will be included in a term record. Although ISO 12620 goes a long way toward organizing and defining these categories, because of its ambitious scope it is an unwieldy resource for the individual. Where it does help is by providing a very full listing of categories, arranged in groups, which termbase designers and terminologists can implement as needed. Two of the standard's authors, Alan Melby and Sue Ellen Wright, have attempted to address these criticisms in a recent article. They write: «As noted above, the significant factor in the proliferation of data categories is that various users or user groups have very specific information management needs, and terminology management is being applied in increasingly different environments, from language planning to inventory control, from safety to quality assurance, and from monolingual to bilingual to broadly multilingual venues. Each unique combination of language and information management needs is coupled with resource management considerations that dictate either a simple or a more granular entry, resulting in custom-tailored sets of labels designed for the storage and retrieval of information in that specific environment[16].»

Translators and terminologists should feel free to use 12620 as a resource and guide in selecting data categories for inclusion in a termbase. Such a move will not only promote consistency on an industry-wide level, it will facilitate data exchanges in the future. The standard itself does not specifically address the question of termbase design—this is not its purpose—but it does provide a full list of data categories that designers can refer to when creating a database model.

The data categories in 12620 are considered meta-categories. Terminologists needn't worry about using the same data category names as long as the contents of their categories reflect the contents of the categories that appear in the standard. For interchange purposes the same data category names should be used, however. The question remains concerning just what *you* should include in a termbase. While pretty much everyone agrees on the basic data categories or attributes—term, definition, context, subject field—the meeting of minds ends there. Since translators are working with two or more languages, separate terminology information groups (TIGs) will be needed for each language. The record (information about a single entity) as a whole represents the concept, which has the peculiar quality that it can only be represented indirectly in the form of a description of some sort (definition, context) or a designation (term, alphanumeric ID, etc.). Since the termbase is concept oriented, information is—or should be—grouped in clusters that help identify individual concepts and relate them to one another. What is frequently lost sight of

in practice is that terminologies, groups of terms, reflect a concept *system*, that is, a group of concepts that are related to one another in some systematic fashion, not an arbitrary collection of words and phrases (something you might find in a dictionary). There are ways of relating concepts to one another and we can identify the nature of the concept system (partitive, genus-species, etc.), the link between concepts (part-whole, genus-species, temporal, spatial, etc.), and their relative position to one another (superordinate, subordinate). Even though computers now enable us to model fairly large concept systems, none of this is easy to visualize or implement in practice.

The following illustration is a screen shot of an unfinished bilingual (English and French) term record created with Infogenie. Information about an English language term (eterm) has been entered with no equivalent French data (fterm, fdef, fcontext, etc.). The data categories follow the ISO 12620 registry very closely. In the illustration, the following data categories are concept related: domain, subdomain1, subdomain2, subdomain3, fdef, fcontext, edef, econtext, fsuperordinate, fsubordinate, esuperordinate, and esubordinate. Term-related categories are represented by fterm, fabbrev_form, fsyn, eterm, eabbrev_form, esyn. The fnote and enote categories correspond to a separate note data category in ISO 12620. The fields for input_date, mod_date, ftsource, etsource are covered by 12620's Administrative information categories (subgroup 10).

Figure 1. A Sample Term Entry

2.3. Concept-related categories

ISO 12620 is broadly divided into three groups of data categories: term and term-related, concept-related, and administrative. This breakdown reflects the fact that terminologies are arranged thematically by groups of concepts. In print media, an alphabetized index is generally provided to facilitate access to the data; it is the first place a translator will look for an unknown term. Thematic structures are extremely useful when we know (or have an idea) what something is but not what it's called. A carefully arranged and comprehensive concept-oriented system can help pinpoint a concept ("what do we call an object that …") and an associated designation ("we call it an X"). If we know what something does or what purpose it serves (not always obvious from a given context), we can often determine its name. For example, the *Pictorial Handbook of Technical Devices* includes chapters on "Machine Technology, Magnetics and Electronics, Light and Optics," etc. The chapter on Machine Technology is broken down into sections on "Machine Elements, Gearing, Mechanical Transmissions, Mechanisms and Kinematics, Fastening and Fasteners." If we know that a foreign term/concept is a type of fastener, we can turn to that section and (hopefully) identify the equivalent term/concept in our target language. I realize that there are many things that can go wrong with this approach in practice—there are relatively few thematic dictionaries, they are not sufficiently comprehensive, they are not current, etc.—but the principle is sound. The real problem for translators is the inability to accurately describe or delimit the foreign concept in the source language in the first place.

In the term record shown in the illustration, I have (somewhat arbitrarily) provided data categories for up to four subject fields, or domains. There is no hard and fast rule about the number of subject field categories needed in an actual record, but I find that four provides a sufficient degree of granularity. If you have client-specific collections of terms, you can include a company or client name as a subdomain and sort or select records on that key to create data subsets. If you prefer to use a separate data field, ISO 12620 includes a subset identifier known as "customer subset" for this purpose. Determination of subject field is far from obvious beyond rather broad categorizations (medicine, law, etc.), however. In the example shown above, the term "information qualification" is categorized as follows: pharmacy → regulatory affairs → drug synthesis. Obviously the concept in question is not an element in the drug synthesis process, it is an ancillary concept that refers to the conformity of the development process to certain industry standards. To the extent that these requirements are legally mandated, couldn't the concept be just as easily classified as part of a system of concepts belonging to law or government? For example, government → FDA → pharmaceutical industry → etc.

ISO 12620 does not include data categories for specific languages but makes use of the language categories provided in ISO 639 and 639-2. ISO 12620 provides a language symbol category that can be used to identify the language associated with a term and any associated data. A term and its related data can be organized into a language-specific "terminology information group," or TIG, but 12620 itself does not address this issue[17]. To identify language groups I have defined separate, concept-related data categories for French and English as follows: fdef, fcontext, edef, econtext, fsuperordinate, fsubordinate, esuperordinate, and esubordinate. The values

for these should be obvious from the field names I have supplied. Here, fdef and fcontext are truncated versions of French_definition and French_context. I selected the names for brevity and convenience, and since the system is designed for my personal use, there is little possibility of confusion. A better method would be to use the ISO two- or three-letter language codes from 639-2, however. Someone who wanted to store data about French, Finnish, and English terms, for example, would have problems with the above category names. The field names correspond to the ISO 12620 categories as follows, the descriptions are taken from the standard.

Table 1. Concept-related Data Categories

Infogenie data category name	ISO 12620 data category name	ISO 12620 Description
fdef	definition	A statement that describes a concept and permits its differentiation from other concepts within a system of concepts.
edef	definition	Same as above.
fcontext	context	A text or part of a text in which a term occurs. *Bear in mind that not all contexts are created equal. The goal is to find one that provides relevant information about the concept in a fairly succinct*
definitions.	*fashion. The most useful*	*contexts approximate*
econtext	context	Same as above.
fsuperordinate	superordinate concept	A concept in a hierarchical system that can be subdivided into a number of lower-ranking concepts.
esuperordinate	superordinate concept	Same as above.
fsubordinate	subordinate concept	A concept in a hierarchical system that can be grouped together with at least one more concept of the same level to form a higher ranking concept.
esubordinate	subordinate concept	Same as above.

The standard itself allows for much greater granularity than the above system would imply. In 12620 definitions and contexts can have permissible instances, depending on the type of definition or context included. A description of these types is beyond the scope of this article, but interested readers are encouraged to read the standard or any of the many available references on terminology.

2.4. Term-related categories

The following data categories are associated with the ISO 12620 term and term-related subgroups: fterm, fabbrev_form, fsyn, eterm, eabbrev_form, and esyn. These correspond to the ISO 12620 categories as shown below.

**Table 2. Term-related Data Categories**

Infogenie data category name	ISO 12620 data category name	ISO 12620 Description
fterm	term language by a linguistic expression.	A designation of a defined concept in a special expression.
eterm	term	Same as above.
fabbrev_form	abbreviated form of term	A term resulting from the omission of any part of the full term while designating the same concept.
eabbrev_form	abbreviated form of term	Same as above.
fsyn	synonym	Any term that represents the same or a very similar concept as the main entry term in a term entry.
esyn	synonym	Same as above.

As with the concept-related categories, ISO 12620 doesn't define separate fields for individual language groups; the Infogenie categories are designed for my personal use and the number of languages can be extended to suit your needs. The abbreviated form of a term can assume several specific forms in 12620, none of which are broken out into separate categories here: abbreviation, short form, initialism, acronym, clipped term. Once again, whether you need this degree of granularity in your database is a matter of personal choice. If you are using an RDBMS, however, it would be possible to define an abbreviated form type field and use the above types to populate a drop-down menu. You could then select the appropriate value when entering the term. You may have noted that I haven't included any categories for grammatical information. In case you're worried ISO 12620 does indeed provide data categories for part of speech, gender, number, animacy, noun class, adjective class, etc. Even within the above structure, you can see that it would be fairly simple to define such fields for both language groups. Whether or not you need to store this information is basically a design choice and will depend on your needs and the amount of time you have to keep the records up to date.

ISO 12620 also includes several categories for degree of equivalence and directionality. Degree of equivalence refers to the extent of overlap between two concepts (source and target language concepts) and the standard lists several permissible instances of equivalence for use. Supplying this type of information is indeed work, but all translators are familiar with the problem of anisomorphic concepts, concepts that have no exact equivalent in a foreign language. For example, although French terms such as *juge d'instruction* and *notaire* can be translated into English, there is no one-to-one correspondence between the two languages. The French concepts are embedded in the French legal system, which is based on rather different foundations than our own case law system. In such situations an equivalence field could be provided, indicating the degree of equivalence between these concepts. Additionally, ISO 12620 provides a directionality category with two possible values: bidirectional and monodirectional. This is useful for those situations where a term can be translated in one direction only. Directionality is important when trying to equate a narrower

concept in one language with a broader concept in another. The French term *précision* is frequently (and incorrectly) used to refer to both "accuracy" and "precision," an endless source of frustration to translators. The correspondence between *précision* and "accuracy" is monodirectional, however, since the English term "accuracy" is (or should generally be) translated as French *exactitude* or *justesse* (which term pairs are bidirectional).

Figure 2. Example of Monodirectional and Bidirectional Equivalence.
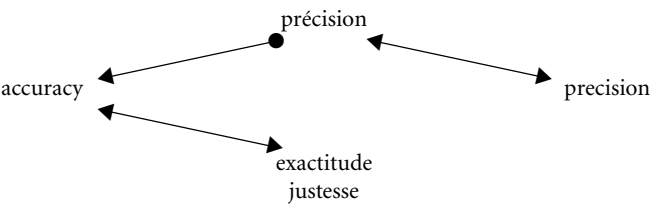
There are two note fields in the record shown in figure 1, one for each language group. You can use these to enter any supplemental information about the term/concept that might be relevant.

2.5. Administrative Data Categories

Figure 1 also illustrates the use of several fields that correspond to the ISO 12620 administrative data categories. These are shown in the table below.

Table 3. Administrative Data Categories.

Infogenie **data category name**	ISO 12620 data category name	ISO 12620 Description
ftsource	source identifier graphic entry identified	The code assigned to a document in a terminological collection and used as both the identifier for a bibliographic entry and as a pointer in individual term entries to reference the biblio with this code.
etsource	source identifier	Same as above.
fdefsource	source identifier	Same as above.
edefsource	source identifier	Same as above.
fconsource	source identifier	Same as above.
econsource	source identifier	Same as above.
input_date	input date etc.) is input into a data	The date on which an element (field, record, entry, collection
mod_date	modification date	The date when a field, record, etc. is edited or otherwise modified.

The illustration includes source identifiers for French and English term, definition, and context entries. These identifiers are pointers (or would be if they were actually linked to a table in an RDBMS) to fields in a separate bibliographic database. It is of critical importance that you include sources for all entries. If a source term comes from a text to be translated, that text should be labeled (you can use the filename if it's an electronic document). Source contexts will, in most cases, although not exclusively, come from the same source as the term. Definitions should be drawn from authoritative sources: standards, manuals, industry experts, etc., whenever possible, and their references indicated in the database. Failure to enter this information will leave you with a collection of random terms and no means of determining the accuracy of the equivalent designations at a later time. As with the term "cycle" mentioned earlier, we need to consider whether or not a term (and its equivalent) will have any usefulness in the future if it is not accompanied by a definition and relevant sources. The sources not only provide critical information about the term and its description, they enable you to validate the information in the data collection. A term that is the product of an intelligent guess by a translator may be useful as a kind of ad hoc solution, but it needs to be flagged as such with a reliability code of some sort. Consider the situation where you disagree with a client about the translation of a term. You may have excellent reasons for your choice, but unless you can substantiate it somehow, the ultimate decision becomes a question of your word against theirs. In such situations, an accurate record (source, definition, context, etc.) can go a long way toward bolstering your case.

A word about sources. As I remarked above, I use a source identifier in my records. This identifier is simply a key or pointer to a value in another data collection, viz., a bibliographic database. For example, the value provided in the source fields (etsource and edefsource) in the above record corresponds to the bibKey value in the bibliographic database record shown below. If you're using a flat-file database such as the above, you may want to consider a separate database for bibliographic data, which allows you to enter fuller bibliographic records. In a modern, relational DBMS, however, bibliographic information can be stored in a separate table, linked to the term entry tables by a common key (such as the bibKey value discussed here). The alternative would be to include the bulk of the data shown below in the main term entry, which will increase the size of the entry considerably and add a great deal of redundancy to the database, since the same bibliographic source may be referenced many times. Although size is no longer a critical issue in today's computing environment, redundancy significantly increases the possibility of error and is considered poor database design. Here I've simply used the ISSN number as a unique identifier for the record. Because not all sources will have ISBN or ISSN numbers associated with them (gray literature, expert testimony, advertising or marketing material, information on the Web), you can create a systematic code for source material. Given the enormous volume of possible sources out there, trying to develop a meaningful mnemonic code is hopeless. But you can develop a system and try to use it consistently. The key to making it work is that it be short, unique, and easy to generate. Obviously, an ISBN number is hardly transparent when trying to determine the nature of the source, but that's what a bibliographic database is for.

Although full term records are considerably more work to set up and maintain than a table, they provide a number of benefits in everyday use. Perhaps the most

important feature of database structures is that the term record reflects the single-concept structure that underlies terminology, where each record represents a unique concept. A database structure also enables you to include as many fields as you need for your purposes and modify them if necessary. Moreover, since most data entry is done on a computer, contexts (and occasionally definitions) can simply be cut and pasted into the data category entry field, which saves time and eliminates errors.

Figure 3. A Sample Bibliographic Entry

The above data fields are modeled on the ISO 12620 data categories as shown below (author's comments appear in italic).

Table 4. Bibliographic Data Categories

Infogenie **data category name**	ISO 12620 **data category name**	ISO 12620 Description
article_title	article	An independent text forming a part of a publication.
author	author	A person or corporate body responsible for the intellectual or artistic content of a document.
bibKey	source identifier	The code assigned to a document in a terminological collection and used as both the identifier for a bibliographic entry and as a pointer in individual term entries to reference the bibliographic entry identified with this code.
category	category	Type of publication. *These publications can be individual articles, glossaries, standards, gray literature, laws, reference works, etc. The standard provides a number of possible category types for use.*

city	city	City in which a document is published.
copyright	copyright *itself or how it should be* *was copyrighted.*	*The standard provides a definition of the term "copyright" rather than a description of the category used. In the above record, the copyright field is used to store the date the work*
country	country	Country in which a document is published.
edition	edition	Whole set of copies of a document produced from one composition or from a single copy used as a master.
editor	editor	An organization or person responsible for the preparation or publication of a document from the point of view of its intellectual content.
ISBN	ISBN	Internationally recognized unique standard number assigned to each edition of a book or other monographic publication for identification purposes.
ISSN	ISSN	Internationally recognized unique standard number assigned to a serial for identification purposes.
issue	issue number	Number of an element of a document published over a period of time.
LoCnum	Library of Congress card number	Number assigned by the U. S. Library of Congress for the purpose of providing access to a complete catalog record for a work.
notes		*Any relevant additional information or comments about the bibliographic entry.*
origin		*Source of a document regardless of its type. This category can be used to record the source of privately published works, unpublished papers, expert testimony, etc.*
publicationdate	date of publication	Indication of the year, and if necessary, of the month and day of publishing.
publisher	publisher name	Name of person or organization responsible for the production and dissemination of a document.
serial_title	serial title	The title of a monographic series.
title	title	A word or phrase, usually appearing on the document, by which it is convenient to refer to it, which can be used to identify it, and which often (though not invariably) distinguishes it from any other document.
translator		*I have included a data category for the translator of a work. The need for such a category should be obvious, and its inclusion parallels Library of Congress usage.*
volume	volume identifier	Identifier for a material unit assembling a certain number of leaves under one cover to form a whole or part of a set.

Once again, the actual selection of data categories is a function of your needs and the purposes for which the termbase will be used. Regardless of the contents of the records or their ultimate purpose, there is an obvious value in trying to adhere to standardized data category names. Aside from facilitating the exchange of information, it provides users and designers of terminological databases with a common language. The fact that the standard is international in scope means that users and creators of terminology around the world now have some hope of being able to design relatively transparent systems. Although ISO 12620 doesn't address the interchange of terminological data directly, it is tightly correlated to ISO 12200 (MARTIF), which does.

Use of an application like Infogenie to create termbase files isn't all a bed of roses, however. One of the problems with this system is that individual files are not linked to one another. There is no way, from within the termbase itself, for example, to jump to a source record stored in the bibliographic file (although it could possibly be done with a script). There is also no way to create sophisticated hardcopy output from these files. You can easily print records. You can even format them to some extent, but you do not have total control over the appearance of the final copy. There is also no way to integrate data from separate files into a single output file. Almost any RDBMS (Access, FileMaker, Oracle, DB2) will enable you to do this. Whether it's worth your while to implement such a system for personal use is a very different issue and one that depends on a variety of factors. For the average user, designing a RDBMS application will probably be more trouble than its worth. The beauty of simple information managers like Infogenie is that they are easy to set up, easy to work with, and provide several of the advantages of a more sophisticated system.

Regardless of the method you use, I hope I've been able to convince you that it is worthwhile applying a systematic approach to the storage and retrieval of terminology, even for personal use. Admittedly, the initial selection of data categories and their incorporation into an existing software application may appear difficult. But once the initial structure is created and a storage method decided upon, such a system will become increasingly easy to use and provide significant long-term benefits.

NOTES

1. "It is recommended that designers of terminology databases and other collections ensure that the content of data categories used in their systems conform to the content defined in these data category specifications. Terminological data prepared for interchange shall conform to the data category names and descriptions specified in annex A." See ISO 12620, p. 3.
2. See Gentzler (1993) for a good summary of contemporary translation theories.
3. This is not to say that *skopos* theory is the only or even the most suitable model for explaining translation, simply that it accommodates my thesis that the way we approach a translation in practice implies a theory. See Vermeer (1996) for a fuller discussion of skopos theory.
4. Pierre Lerat criticizes Sager's description of special languages as being too narrow because it restricts communication to experts. Lerat writes: "The concept of a special language is more pragmatic. It is a natural language considered as a vector of specialized communication."
5. Unabridged general dictionaries like *Webster's Third New International Dictionary* include many scientific and technical terms that are not part of day-to-day use. The fact that such terms can be found in a dictionary of this sort does not imply that they are not "terms" or part of a special language.
6. The ISO definition of a concept as "A unit of thought constituted through abstraction on the basis of properties common to a set of objects" is merely a convenient fiction that fails to take into account contemporary developments in fields such as philosophy, psychology, and cognitive science.

Juan Sager wisely sidesteps this difficult issue completely by referring to concepts as undefined "mental primitives."

7. See Fodor (1998) and Margolis (1999) for an excellent contemporary treatment of concepts and concept acquisition.

8. The Oxford-Duden pictorial dictionaries are a well-known example of this arrangement. Although encyclopedic in scope and largely out of date, they are arranged systematically rather than alphabetically and provide detailed illustrations with numbers keyed to bilingual labels. Like other visual dictionaries the Oxford-Duden attempts to describe how things work and what they are called. The problem for translators is that, given the degree of specialization of contemporary texts, to be really useful the Oxford-Duden would have to be the size of the *Britannica* and updated on a regular basis.

9. This is a greatly simplified version of current database design. Relational databases store entity data in separate tables, which are linked to one another in specific relationships. Each table has a set of attributes and tables are linked by a common key.

10. See for example ISO 12620, *Computer Applications in Terminology—Data Categories*, for a complete list.

11. ASTM standard E 647–95a, "Standard test method for measurement of fatigue crack growth rates," ASTM, 1998.

12. A facet is an "exhaustive set of properties of a similar kind," for example, a specific attribute or function by which we categorize the members of a collection (by product, by purpose, agent and patient, etc.). The number and kind of facets will vary with the objects being studied. One advantage of a faceted classification is that it provides a means for additional search strategies in information retrieval systems and the construction of expert systems applications. See Ekholm.

13. Four types of context can be defined: Definitional context: The context provides a good description of the term but one that does not possess the rigor of a well-formed definition. Encyclopedic, or explanatory, context: The context provides information about usage, history, development, etc., or otherwise explains something about the term without defining it. Associative context: The context provides information that relates the term to a domain but nothing more. Linguistic context: The context provides information about the way a term is used as a syntactic element, provides information on usage and possibly collocations. See Dubuc (1997) for examples and further discussion of the first three types of context.

14. Automatic terminology extraction programs have been developed but such systems are predicated on the availability of a large corpus in machine-readable form. The output from such programs also needs to be edited and refined by a trained terminologist. Unfortunately, in terminology as in other walks of life, there is no free lunch.

15. From Casady and Greene, Inc., 22734 Portola Drive, Salinas CA, 93908-9920 (408-484-9228).

16. Alan Melby and Sue Ellen Wright (1998): "MARTIF—Putting complexity in perspective. 3. Conflicting views on the complexity of the MARTIF standard," http://www.ttt.org/clsframe/termnet/termnet4.html.

17. See ISO 12200 for a discussion of the Machine Readable Terminology Interchange Format (MARTIF). This standard is designed to address the problems associated with exchanging terminological data between systems and how that data might be structured prior to exchange. MARTIF uses SGML as its markup language and defines a DTD for generating and validating terminological data structures. A MARTIF representation can include separate blocks of information for every language group (term group) associated with a concept.

REFERENCES

ASTM (1998): *1998 Annual Book of ASTM Standards: Volume 03.01. Metals—Mechanical Testing; Elevated and Low-Temperature Tests; Metallography*, West Conshohocken (Penn.), ASTM.

Dubuc, R. (1997): *Terminology: A Practical Approach*, adapt. E. Kennedy, Québec, Linguatech.

Ekholm, A. (1996): "A Conceptual Framework for Classification of Construction Works," http://itcon.fagg.uni.lj.si/~itcon/.

Fodor, J. A. (1998): *Concepts: Where Cognitive Science Went Wrong*, Oxford, Oxford University Press.

Gentzler, E. (1993): *Contemporary Translation Theories*, London, Routledge.

ISO 639-2 (1996): *Codes for the representation of names of languages—Part 2: Alpha-3 code*, Geneva, International Organization for Standardization.

ISO 1087 (1990): *Terminology–Vocabulary*, Geneva, International Organization for Standardization.

ISO 12200 (1998): *Computer Applications in Terminology—Machine-Readable Terminology Interchange Format (MARTIF)—Part I: Negotiated Interchange*, Geneva, International Organization for Standardization.

ISO 12620 (1998): *Computer Applications in Terminology—Data Categories*, Geneva, International Organization for Standardization.

Lerat, P. (1995): *Les langues spécialisées*, Paris, Presses universitaires de France, coll. «Linguistique Nouvelle».

Margolis, E. and S. Laurence (1999): *Concepts: Core Readings*, Cambridge (Mass.), MIT Press.

Peirce, C. S. (1903): *The Essential Peirce*: *Selected Philosophical Writings*, vol. 2 (1893-1913), Bloomington, Indiana University Press (1998).

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

Sager, J. (1994): *Language Engineering and Translation: Consequences of automation*, Amsterdam/Philadelphia, John Benjamins. Twenty-two points, plus triple-word-score, plus fifty points for using all my letters. Game's over. I'm outta here.

Schwartz, O. B. and P. Grafstein (1971): *Pictorial Handbook of Technical Devices*, New York, Chemical Publishing Co.

Vermeer, H. (1996): *A Skopos Theory of Translation*, Heidelberg, TextconTEXT-Verlag.

Whewell, W. (1847): *The Philosophy of the Inductive Sciences Founded Upon Their History*, 2 vol. [repr. (1967): New York, Johnson Reprint Corp.].