

What can Verbs and Adjectives

Tell us about Terms ?

Marie-Claude L'Homme

Département de linguistique et de traduction
Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal (Québec), H3C 3J7, Canada
lhommem@ling.umontreal.ca

Abstract

Corpus processing tools are now an integral part of the compiling of specialized dictionaries and updating of term banks. They have led terminographers to consider terminological data differently, since many regularities and problems are highlighted in a more systematic manner. One linguistic fact more immediately revealed by the use of corpus tools is the relationship between terms in noun form with verbs and adjectives. In this paper, we study two specific types of relationships, namely morphological and syntagmatic relationships. We propose to consider lexical units that have one of these relationships with terms in nominal form. We will demonstrate that verbs and adjectives should be taken into account by terminographers for a number of reasons: some of them provide clues to the meaning of terms, others are morphologically and semantically related to terms in noun form.

Key words : corpus-based terminology, terms, verbs, adjectives, morphological relationship, syntagmatic relationship

1. Introduction

Natural language processing and corpus linguistics have provided terminographers with a set of models and tools to assist them during the compilation of specialized dictionaries and the updating of term banks. These include the use of concordancers (applied on tagged or untagged text), term-extraction, and automatic or semi-automatic retrieval of other relevant terminological data (e.g., cooccurrences, contexts with conceptual information, such as hyperonymy or meronymy).

Some of these technologies have important repercussions on the way terminological research is conducted. Specialists have begun to assess them (e.g., Ahmad and Rogers, 2001; Meyer and Macintosh, 1996, Pearson, 1998).

One consequence is that some tools – and more specifically their output – force terminographers to analyze sets of data they did not necessarily consider before. Automatic processing also highlights characteristics about terms terminographers are not used to taking into account. This is mostly due to the fact that many regularities are captured more systematically when using computer tools. It should also be pointed out that natural properties attached to terms become problematic only when using a computer application.

Another important issue is the fact that machine-readable dictionaries are used in a growing number of computer applications. However, specialized dictionaries only account for a limited number of lexical units that actually occur in specialized texts, namely terms in noun form. Thus, they are not well suited for a number of tasks.

Although these facts have become an everyday reality for practitioners, they have not yet been described formally in terminology. To our knowledge, even though a growing number of courses in terminology include the use of computer applications, very few provide the linguistic background necessary to interpret the data processed by them and presented to terminographers. This paper is an attempt to bridge this gap, in line with previous work in this area.

We will study a specific problem, i.e., terms in noun form, and their semantic interactions with lexical units pertaining to other parts of speech, namely adjectives and verbs. We will consider the problem from the point of view of corpus analysis and, with the help of linguistic arguments (especially Mel'èuk et al., 1995), examine how these relationships can be interpreted in order to improve the description of terms in specialized dictionaries and term banks.

2. Goal of the paper

This paper examines a particular issue raised by the systematic use of corpus processing tools, that is the relation between nominal terms and lexical units that belong to other parts of speech, namely verbs and adjectives.

We will list a number of arguments to show that verbs and adjectives (and potentially adverbs, although we have not studied this part of speech in detail) should be taken into account by terminographers. Generally speaking, we can say that:

- a) some verbs and adjectives provide clues to the meaning of terms; in some instances, they help disambiguate ambiguous lexical forms;
- b) others are morphologically and semantically related to terms in noun form.

To conduct this study, we used a corpus of English texts on computing amounting to approximately 500 000 words¹. This work is based on previous work we carried out on French corpora.

It should be pointed out that the data examined in this paper and the criteria used to analyze it is relevant for terminographical purposes. Even though a computational implementation may be envisaged, it would require further formalization.

3. Nouns and other parts of speech in terminology

Traditionnally, terminographers have considered terms in noun form. This is best observed in specialized dictionaries and term banks in which most entries are nouns (either simple – e.g., *server*, *data*, *cursor* – or complex – e.g., *hypertext link*, *laser printer*, *database management system*). Although there are a few exceptions to this rule, the theoretical models of terminology still exclusively accommodate the description of nouns (and, certain categories of nouns, i.e. nouns that refer to entities), and are not well suited to take other parts of speech into account.

Normally, verbs and adjectives are included in specialized dictionaries if they are not used in other contexts (e.g., the French term *configurer* ‘*configure*’ appears in dictionaries of computing, since it exists only in that field), or if they have a meaning that cannot be described using definitions found elsewhere (e.g., *to surf*, in the field of the

Internet, has a metaphorical meaning that cannot be captured using definitions found in other dictionaries, for instance, general language dictionaries).

This approach reveals inconsistencies in specialized dictionaries, since many semantically-related lexical units are omitted if they are not nouns. For example, most dictionaries on computing will inform users that *program* is a “set of instructions designed to accomplish a specific task”, and that *programming* is “the set of activities that consist in designing, writing and testing a program”. Some dictionaries will include *programmer*; others might add *to program*; very few will include *programmable*; and even fewer provide an exhaustive list of all the terms we just cited².

Recently, dictionaries of specialized cooccurrents have been offered to users (e.g., Cohen, 1986; Meynard, 2000); others include information on lexical combinatorics in more detailed entries (Binon et al., 2000). These reference works give users access to lists of verbs, adjectives and nouns a given term combines with under a heading represented by the term itself. Table 1 shows an extract from the Meynard (2000) which lists specialized lexical combinations in the field of the Internet.

Browser	Client application that allows the user to view HTML documents on the Web.
	download, install, launch, open, run compatible, complete, external, graphical language of a, settings of a, path of a etc.

Table 1: Extract of an entry (Meynard, 2000).

This work has led terminographers to consider lexical units other than nouns. It also informs users that certain terms are associated with specific activities (e.g., *install*, *launch* a browser) and properties (e.g., *compatible*, *external* browser). In this framework, however, verbs and adjectives are considered as dependent units, i.e., their description becomes interesting only when considered though their combination with a term in noun form. Verbs and

¹ The corpus has been set up at the Laboratoire de linguistique informatique (LLI), Université de Montréal (<http://www.fas.umontreal.ca/lhomme/lli.html>).

² It is worth mentioning that some semantically-related terms might appear in a complex term. For example, *programmable* might not be listed as an entry as such, but can appear in complex terms such as *programmable transistor*. Also, *programming* will appear in entries such as *programming language*, *programming flowchart*. This approach also leads to inconsistencies. Instead of analyzing *programmable* or *programming* as such, they are considered as noun modifiers. This also implies that all these combinations must be listed in the dictionary if the vocabulary of a field of knowledge is supposed to be covered exhaustively.

adjectives are simply listed under an entry in noun form.

Some researchers in computational terminology have developed techniques to take into account terminological variants. Daille (2001), for instance, considers relational adjectives (e.g., *malarial*, from *malaria*), and tries to find the noun they derive from. Jacquemin (2001) considers different types of terminological variants and these include syntactic transformations (e.g., the transformation of a complex noun phrase into a verbal phrase).

Zweigenbaum and Grabar (2000) propose a method for the automatic identification of morphologically-related French terms. These include adjectives. The researchers also make deductions on their conceptual relationship with other lexical units and their contribution to the overall knowledge organization of a specialized field, namely medicine.

Here again, this type of work underlines the importance of considering other parts of speech when identifying terms and analyzing them. It also highlights the fact that this type of data cannot be bypassed when processing specialized corpora.

4. Relationships between nouns, verbs and adjectives in specialized corpora

Corpora and their analysis with corpus processing tools provide several ways to examine different types of relationships between nouns, verbs and adjectives. We will examine two specific relationships in the following subsections: morphological and syntagmatic.

The following subsections simply list a number of observations. We will discuss this data and its consequences on the analysis carried out by terminographers in section 5.

4.1. Morphological families

Terms are related to other lexical units by formal similarity. Browsing through a corpus of texts on computing will inevitably reveal morphological families. Examples (1) show *program* and morphologically-related lexical units, namely *programmer*, *programmable*, *program (to)*, and *programming*.

(1) *In the earliest computer systems with simple operating systems, most programs were executed using serial processing: one at a time, one after the other.*

Programmers spent much time trying to find ways to trim the size of programs so that they could fit into the available memory space.

A programmer need not have an in-depth knowledge of the computer to write application programs.

Many of the support chips in the PC are programmable, which means that their operation can be controlled by software.

Programmable read-only memory (PROM) can be programmed either by the manufacturer or by other companies to meet unique user needs.

Jeppe Cramon and Ingo Guenther responsible for the programming and graphics for the game.

4.1.1. Syntactic derivations

A specific morphological relationship, described in work on terminological variants (Daille, 2001; Jacquemin, 2001), regroups different types of syntactic derivations, i.e. nominalization, adjectivization, etc. In this specific relationship, the related lexical units convey the same meaning, but belong to different parts of speech.

Some morphologically-related lexical units discussed in the previous section are syntactic variants. We provide other examples below (2) to (4):

Verbs and nominalizations

(2) *... the touch event data into mouse events, essentially enabling the sensor panel to "emulate" a mouse. than can be achieved with software emulation.*

Nouns and adjectivizations

(3) *IBM introduced MCA in April 1987, forsaking its older architecture for a new 32-bit design. Digital will manage the architectural process to ensure architectural consistency, and will continue to develop future Alpha designs. Financial terms were not disclosed.*

Adjectives and nominalizations

(4) *Insight Manager's only drawback is its interface, which isn't as user-friendly as it could be. The concern for user friendliness has overflowed into the development of other computers, too.*

4.1.2. Morphologically- but not semantically-related

Morphologically-related lexical units are not necessarily semantically-related. For instance, *application*, identified as a term in the field of computing ("program"), is morphologically related to the verb *apply*. However, no semantic relationship was found in the corpus analyzed. This is shown in examples (5):

(5) *When the user double-clicks on an attachment, most systems are configured to start the application associated with the file type. The trouble is that these applications will also execute any macros within the received file, thus enabling the virus to infect.*

Of course, the above remarks do not always apply. ... some rules do apply. In some cases ...

The *application, apply* pair shows a first case in which morphologically-related lexical units are never semantically-related. In other cases, one of the units may be polysemic. This second case is illustrated in examples (6). The verb *to address* is polysemic: a first meaning is related to the term address; the second one is not.

(6) *You can assign pointers to one another, and the address is copied ...
...it can perform computations and address memory 32 bits at a of 16 bits at a time.
When you have completed addressing and filling in the necessary message*

4.2. Combinatorics

Terms are also related to verbs and adjectives via syntagmatic relationships. In (7) and (8), we provide a short list of adjectives and verbs related to *program* found in the corpus on computing.

Adjectives

(7) *... but almost any language that can create executable programs can be used ...
There are specialized programs a user can utilize to perform a specific function
it depends on the cooperation of the active program to share its resources*

Verbs

(8) *...depending on which programs are running ...
In order for you to start writing computer programs in a programming language called Java ...
Then the program ends.
Once it's loaded into memory, the program ...*

5. What this data tell us about terms in noun form

In this section, we will examine how these relationships can be analyzed by terminographers when building entries in specialized dictionaries or term bank. We will see that morphologically-related lexical units and cooccurents can be used to:

- a) Make semantic distinctions;
- b) Analyze the meaning of terms;
- c) Build classes of terms.

We demonstrated elsewhere (L'Homme, 1998) that some verbs should be considered as terminological units, and provided a list of arguments to support our view as well as a methodology to describe them. We also think that adjectives lend themselves to the same kind of analysis. However, we will not consider this aspect here and focus our examination on the information verbs and adjectives yield on the meaning of terms in noun form. But it should be gathered that both issues are interrelated.

5.1. Making semantic distinctions

Examining lexical units that are related morphologically to a nominal term under examination helps confirm semantic distinctions. Let us consider the two series of contexts in (9) extracted from the corpus composed of texts on computing.

(9) *You can enter the address of the location you wish to visit and the browser...
At the machine level that location has a memory address. The four bytes at that address are known to you, the programmer, as I, and the four bytes can hold one integer value.*

Only the occurrences in the second series are semantically-related to *addressable* ("that can be addressed, that can have an address"), as shown in (10).

(10) *The 386 has a huge amount of addressable memory compared to the 286.*

The memory is *adressable*, but not the location (see (9)).

Verbal and adjectival cooccurents also provide clues to differentiate the meanings of a polysemic nominal term. For example, the examples (11) show that *configuration* has two different meanings.

(11) *The basic configuration includes a 166-MHz Pentium processor ...
You have now completed your configuration of your Newsreader.*

The first context refers to the list of characteristics of a computer. It may refer to the list of characteristics of another computing device. The second context shows the verbal meaning of *configuration*: the act of setting up a piece of hardware or software. The verbal meaning is incompatible with *include*; similarly, the "list"

meaning is incompatible with a process verb such as *complete*.

5.2. Analyzing the meaning of terms

The analysis of related lexical units helps better circumscribe specific meanings, and consequently, produce more accurate definitions. For instance, considering all the lexical units related to *program* (1) refines the analysis, since they refer to one another. The *programmer* is “the person who writes programs”; *to program* is “to design programs”; *programming* is “the act of designing programs”; and *programmable* qualifies “something that can be programmed”.

Syntactic derivatives can be considered together, since they convey the same meaning. They offer a means to capture a larger number of contexts. In addition, the verbal meaning of a noun should be considered together with the verb itself; or the adjectival meaning of a noun with the corresponding adjective. For example (12), the verbal sense of *configuration* can be analyzed looking at contexts containing *configuration* itself and *configure*. These contexts will most likely reveal the common semantic features these lexical units have, such as the same cooccurrences.

(12) *To configure the computer.*
you are ready to configure your browser for running email
Barring severe machine and power problems,
there should be no reboots to configure software or hardware,

You have now completed your configuration of your Newsreader.
However, IDE can provide a reasonably acceptable performance and does not require any software configuration

This strategy will also help distinguish verbal meanings from others, for example the two distinct meanings of *configuration* (11).

Cooccurrences also help making finer-grained semantic distinctions. For instance, we saw that *configuration* (in 11) has two different meanings, and that these meanings could be distinguished according to verbal cooccurrences.

These meanings are seldom distinguished in specialized dictionaries. These can retain only one of the meanings, or provide a vague definition that encompasses both meanings.

5.3. Building classes of nominal terms

Verbal and adjectival cooccurrences can help reveal groups of terms that are semantically-related or, to put it in terminological terms, conceptual classes.

The contexts in (13) show that the verb *run* combines with several terms – program, operating system, routine, application – all terms that refer to a type of “program”.

(13) *To run the program, type samp (or, on some UNIX machines, ./samp).*
... it must be able to run more than just an operating system.
and it will automatically run the install routine.
... the operating system interprets the user's instructions, handles input and output, runs applications ...

run : program, routine, application, operating system

The examples in (14) show that adjectives can also be used to build semantic classes. The adjective *editable* cooccurs with terms referring to “text”.

(14) *Maybe you have to scan some documents and convert them to editable text with OCR software.*
Another feature of WordPerfect is its editable formatting codes ...
Newssoft's Presto Wordlinx OCR software for turning scanned print into editable text files.

These contexts can help terminographers identify groups of terms that could, for instance, be defined using similar sets of characteristics.

6. Conclusion : Considering adjectives and nouns during terminological analysis

In a nutshell, the data discussed in the previous sections shows that adjectives and verbs should be considered when analyzing nominal terms. They provide several clues on the meaning of noun terms and criteria to support an analysis performed by terminographers.

However, they can be used differently according to the application at hand.

First, they can simply be used during the analysis of noun terms to support semantic distinctions or build conceptual classes.

Second, they could be included in specialized dictionaries. Here, two methods can be envisaged. They can be listed in an entry whose heading is a noun (as in Binon et al, 2000). An explanation of the specific relationship linking them to the head noun can be provided.

They can also be considered as entry themselves. But here again, it would be important to clarify the relationship between noun terms and others in cases in which the meanings are related.

7. References

- Ahmad, K. and M. Rogers, 2001. Corpus Linguistics and Terminology Extraction. In Wright S.E. and G. Budin (eds.). *Handbook of Terminology Management*, Vol 2. 725-760.
- Binon, J., S. Verlinde, J. van Dyck et A. Bertels, 2000. *Dictionnaire d'apprentissage du français des affaires*, Paris : Didier.
- Cohen, B., 1986. *Lexique de cooccurrents: Bourse, conjoncture économique*, Brossard (Québec): Linguatech.
- Daille, B., 2001. Qualitative terminology extraction: Identifying relational adjectives. In Bourigault, D., C. Jacquemin and M.C. L'Homme (eds.). *Recent Advances in Computational Terminology*, 149-166, Amsterdam / Philadelphia: John Benjamins.
- Frawley, W. 1988. New Forms of Specialized Dictionaries, *International Journal of Lexicography* 1(3): 189-213.
- Jacquemin, C., 2001. *Spotting and Discovering Terms through Natural Language Processing*, Cambridge: MIT Press.
- L'Homme, M.C., 1998. Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie* 73(2): 61-84.
- Mel'èuk, I., A. Clas and A. Polguère, 1995. *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve: Duculot / Aupelf-UREF.
- Meyer, I. and K. Macintosh, 1996. The Corpus from a Terminographer's Viewpoint, *International Journal of Corpus Linguistics* 1(2): 257-268.
- Meynard, I., 2000. *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français-anglais*, Brossard (Québec): Linguatech.
- Pearson, J. (1998). *Terms in Context*, Amsterdam/Philadelphia: John Benjamins.
- Sager, J.C., 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Zweigenbaum, P. and N. Grabar, 2000. Liens morphologiques et structuration de terminologie. In *IC 2000. Ingénierie des connaissances* : 325-334.