

Contribution of terminological resources in learning specialized languages

Isabelle Carrière
Observatoire de linguistique Sens-Texte (OLST)
Université de Montréal

Studies (L'Homme 2004; Carrière 2006) have shown that adjectival derivatives of nouns (e.g. *viral*, derived from the noun *virus*, or *adhérent*, derived from the verb *adhérer*) provide a great deal of information regarding the nature of the text in which they are used. These adjectives have been observed to carry a richer semantic content than a simple reference to their nominal base, the meaning indicated in most specialized dictionaries. For example, we have found that the adjective *abdominal* (*noun base is abdomen*), when combined with *abcès* (abcess), means “which affects the abdomen,” but when combined with *chirurgie* (surgery) means “which is performed on the abdomen.” Existing medical dictionaries that include adjectives generally give definitions such as “related to the abdomen” (my translation).

In recent years, linguists and terminologists have shown a growing interest in the role and the place of adjectives in specialized texts (Bae 2006; L'Homme 2004; Maniez 2001; and Zweigenbaum, Hadouche and Grabar 2003) and their automated processing (Daille 2001; Grabar and Zweigenbaum 2003; and Normand and Bourigault 2001). These studies have helped us better understand adjectival derivatives, the role of adjectives in text and the complexity of their translation. In our description of the various meanings of the adjectives, we take these analysis a step further.

This presentation describes the contribution of terminological resources such as *MédiTerm*, along with the analysis of adjectival derivatives of nouns, as a learning tool for specialized languages (in this case, medicine). First, we will present what we mean by adjectival derivative of a noun, and the way we express the various meanings of these adjectives. Secondly, we will present *Méditerm*, a database that reflects the encoding we chose to access the results of the analysis of medical adjectives. Finally, we will present the multilingual aspect of this work.

1— Adjectival derivative of noun

This section presents the main characteristics of adjectives derived from nouns along with the place they occupy in noun phrases. In order to explain the role of the adjectives in the text, we present noun phrases according to the following syntactic pattern: *head noun + semantic relation + base noun* (e.g. *abdominal surgery*: surgery + which is performed on + the abdomen). Using this pattern, we introduce the head noun and present its role in the noun phrase as well as its impact on the meaning carried by the adjective analysed. With the syntactic pattern as a starting point, we explain the paraphrasing principles used to explain the semantic relation that links the head noun of the noun phrase to the base noun of the adjective. In doing so, we will explain the methodology that leads to the identification of the base noun.

2— Encoding of our terminological database: *MédiTerm*

This section presents the encoding of the data gathered during the analysis of the adjectives and the information that can be accessed in *MédiTerm*. Starting with the query form, we describe the various criteria that can be used for searching, identifying the pertinence of each criterion for finding terminological information to assist in learning specialized languages. After this outline, we will simulate a search to show not only the terminological data, but also the phraseological data (e.g. collocations) contained in the defining paraphrases used to describe the meaning of the adjective. Finally, we will present the model adopted for the term records. This overview of our resource allows us to present its terminological content as well as the additional information available to the user learning specialized languages.

3— Multilingual approach of *MédiTerm*

This third section presents the current status of the work and our intention to expand *MédiTerm* to other languages, starting with English. The terminology record is already designed to contain multilingual information. We are also now exploring various techniques that will allow us to find equivalent terms based on their meanings. We will then quickly present the paraphrase system created to guide the phrasing of our defining paraphrases and the methodology we will try to apply in order to work around the various differences in the terminology and in the syntactic structure of French and English. This should allow us to update the existing data and develop the English section of the terminological resource.

By describing adjectival derivatives of nouns and their role in the text, and by presenting a terminological encoding rich in semantic and phraseological information, we believe that a terminological resource, such as *MédiTerm*, is an interesting asset for learning specialized languages. At the moment, the database is a French dictionary of medical adjectives; by using a paraphrase system model, as we intend, we can look ahead to the development of a multilingual dictionary, starting with the addition of the English adjectives. Focusing on the constant evolution of our resource, we hope to expand the multilingual aspect of the database by eventually adding Spanish adjectives and possibly by exploring some emerging languages, including Asian languages such as Korean. Finally, reflections on the various languages may lead to other possibilities for the use of our semantic approach, for example raising the question of whether such an approach can be used in domains other than medicine, such as in chemistry or even physics.