Adjectives and collocations in specialized texts: lexicographical implications

Araceli Alonso Campos and Sergi Torner Castells
Universitat Pompeu Fabra, Barcelona, Spain

The General Theory of Terminology (Wüster 1979) states that all terms must be nouns, as the noun is the only category to designate a concept. For this main reason, adjectives and other grammatical categories are not considered as entries in most terminological dictionaries. The Communicative Theory of Terminology (Cabré 1999, 2000, 2002), on the other hand, has recently determined that predicative categories, such as adjectives, verbs and adverbs, can also become specialized lexical units (SLU). However, there are not enough empirical studies at this moment which confirm this hypothesis and examine the main characteristics of these predicative categories when they are used as terms. Specifically, our contribution studies the use of adjectives as terminological units in environmental texts. The study of Environment-related terminology is of special interest, as Environment is a new emerging domain with characteristics different from those of classical domains, such as Medicine or Chemistry. As it has been established in previous works (Alonso 2009, Bracho 2004), many Environment-related words are taken from the general language, but take on a terminological sense when they are used in environmental texts.

Our study focuses on adjectives which form a collocation [N][A]_SAdj_2SN, as this syntactic structure is frequently used in specialized discourse. Our main objective is to determine the ‘terminological value’ of these adjectives and their main characteristics. It is concluded from the data analysis results that the behaviour of adjectives depends mainly on the syntactic-semantic nature of the adjective. It is observed a general tendency to use as terms, either classifying relational adjectives (Bosque 1993, Bosque & Picallo 1996, Picallo 2002), or common qualifying adjectives that adopt a terminological sense in specialized texts. This fact brings about the need of different kinds of treatment for the representation of these adjectives in terminological dictionaries.

1. Introduction

Vocabulary is one of the most important elements in the characterization of domain-specific languages. Most studies on terminology focus on nouns and do not take into account the use of terms in context. The General Theory of Terminology (Wüster 1979) states that all terms must be nouns, as the noun is the only category to designate a concept, to refer to concepts that structure specific domains. For this main reason, Terminology has barely paid attention to the study of adjectives and, as a consequence, adjectives are not considered as entries in most terminological dictionaries.

On the other hand, the Communicative Theory of Terminology (Cabré 1999, 2000, 2002) has recently determined that predicative categories, such as adjectives, verbs and adverbs, can also become specialized lexical units (SLU). However, there are not enough empirical studies at this moment which confirm this hypothesis. Specifically, our contribution aims to fill a gap in terminological studies.

Every subject domain presents different characteristics. Most terminological studies focus on classical domains, such as Medicine, Physics or Chemistry. In an era characterized by the importance of technology and science for the development of societies, with a growing popularization of scientific knowledge, and also an epistemological change in the conceptualization of disciplines, new domains emerge with a specific structure and characteristics different from those of classical domains. In these new fields, the borderline between general language and specific-domain language becomes fuzzy. These aspects are reflected on the vocabulary being used and, as many Environment-related words are also used in general texts, some guidelines must be outlined for a better representations of these units, not only in specialized dictionaries, but also in general dictionaries.
This paper deals with the role of adjectives in Environment-related texts in Spanish in order to observe their use in context. Specifically, the study focuses on adjectives which form a collocation \([N[A_{SA}]}_{SN}\), as this syntactic structure is frequently used in specialized discourse (Estopà 1999, 2000; Alonso & DeCesaris 2005), in order to understand what an adjective is in Terminology, how it is used and what implications there are in relation to its lexicographical representation and treatment. In this paper, three questions are intended to be answered: Should adjectives be included as entries in a terminological dictionary? Which noun collocations with the structure \([N[A_{SA}]}_{SN}\) should be included as entries or subentries of the noun? Must the same principles be applied in the case of general dictionaries? The answer depends on the kind of adjective.

2. The domain of Environment and environmental discourse

Environment as a field of knowledge emerges due to the awareness of societies about the deterioration of nature mainly as a consequence of industrial and technological development. It is a specific domain that is relatively new as an academic discipline; it is dynamic, as it changes through time; it is socially interesting, as it concerns everybody; it is formed by different systems (physical, biological, social, economical, political and cultural) in which people and other organisms interact; it studies all factors that affect this interaction, and, therefore, it draws on many other disciplines. As Myerson & Rydin (1996) have pointed out ‘environment belongs to every discipline and to none.’ Environment does not fit into the way classical fields have been represented as a hierarchical structure. Instead, it can be seen as a network of interconnected topics in relation to other specialized areas which can be studied from many different perspectives.¹ The conceptualization of Environment as a field of knowledge influences how the study of Environmental discourse evolves.

Interest in the study of Environmental discourse is relatively new and there are barely studies on Environmental discourse from a linguistic perspective, especially in Spanish – see Alonso (2009) and Bracho (2004) for more information –. Taking into account the few studies based on the analysis of environmental texts, and as a result of this field’s own peculiarities, it can be stated that the language used to talk about Environment has its own specific characteristics which differ from those of other classical specialized domains, such as Medicine or Physics,² and which are reflected in its vocabulary: being a current field, Environment-related words have not been standardized and there are many variants for each form. Due to its interdisciplinary and dynamic character, many terms are borrowed from other specialized areas, either maintaining or changing their meaning. Furthermore, it is a domain that is limited sociologically and geographically, the characteristics of which are closely related to the geography of a specific environment and the people who live in that environment; this is also reflected in the vocabulary used. Since it is a field with a great social impact, many Environment-related terms migrate to general language, becoming part of a speaker’s idiolect. In general terms, previous studies – Alonso (2003, 2008 & 2009), Alonso & DeCesaris (2005, 2006) – determine that there is a lack of delimitation and precision at using environmental words and that the distinction between the specialized lexical unit and non-specialized lexical

¹ This idea is taken from Myerson & Rydin (1996: 7) who explains the language of Environment from a rhetorical point of view by means of the metaphor of ‘the enironet’, a network making linkage upon linkage between the Environment. It seems that this metaphor may also be used for determining how this domain is structured as a field of knowledge.

² For more detailed information, see Rodríguez & Garriga (2006) and Gutiérrez (1998).
unit becomes fuzzy. This translates into a non-systematic representation of these words in both general and specialized dictionaries.

3. Methodological aspects for the analysis of adjectives

In order to analyze the role of adjectives in environmental texts, and taking into account that a lexical unit is used with a specialized sense according to its context (Cabré 1999), a corpus-based methodology is needed. In Alonso (2009), where it was determined how Environment-related lexical units are used in context, techniques from corpus-based lexicography were employed and a methodology based on the combination of elements from the Theory of Norms and Exploitations and Corpus Pattern Analysis (Hanks 2004) and the proposal of collocational networks (Williams 1998) was applied. That study focused on the use of Environment-related nouns and their representation in dictionaries. In this paper, this methodology is considered, though adapted to the study of adjectives.

The corpus named AquaCorp is a subcorpus extracted from the Corpus Técnico del IULA. Assuming that Environment is a network composed of topics (Myerson & Rydén 1996: 36), the topic related to water is considered. The corpus is formed by 20 texts in Spanish from different sources related to water issues – a total of 347,051 words. In order to establish the level of balance in the corpus, all texts are classified according to the following criteria: language, subject (subareas), genre and type (argumentative, expositive, etc.), speakers (layperson, semi-layperson or specialist), and specialized level (high, medium or low level of specialization).

By using bwanaNet and Jaguar tools a list of the most frequent nouns is extracted. A total of 4,346 nouns with 8,446 occurrences is selected. From the list, it can be observed that by means of frequency, many terminological units would be discarded, as they show a low frequency in the corpus – a total of 1,522 are hapax legomena. Frequency seems not to be enough to determine the core vocabulary used in AquaCorp. However, by combining frequency measure with dispersion and use measures (Juilland & Chang-Rodríguez 1964), a list of the most relevant, nuclear and frequent nouns in the corpus can be established. A total of fifty nouns are selected according to their relevance inside the corpus. By means of Sketch Engine, the most significant words which accompany the selected nouns, the word sketches for each noun are extracted. Taking into account these word sketches and, by applying the method proposed by Williams (1998), a collocational network for each noun is built. The networks show the most significant collocations used in the corpus. It can be observed that, one of the most significant constructions is the use of the noun modified by an adjective ([N][A]SN).

All the adjectives are extracted from the collocational networks in order to be analyzed and observe the most frequent patterns of these adjectives in relation to the nouns they...

---

3 For more information on the Corpus Técnico del IULA, see Bach et al. (1997), Cabré et al. (2006), Vivaldi (2009).

4 bwanaNet and Jaguar are both corpus tools developed at the Institute of Applied Linguistics at Pompeu Fabra University for the exploitation of the Corpus Técnico del IULA. For more information, see http://bwananet.iula.upf.edu/indexen.htm and http://melot.upf.edu/cgi-bin/jaguar/jaguar.pl?lnt=En.

5 Sketch Engine is a corpus tool which generates, amongst other things, the ‘word sketches’ (one-page automatic, corpus-based summaries of a word’s grammatical and collocational behaviour) for the words of the corpus being used. For more information, see Kilgarriff et al. (2004). http://www.sketchengine.co.uk.
accompany. In this way, not only the most frequent adjectives are taken into consideration, but those which play an important role once they are used with the most significant nouns used in the corpus.

4. Data analysis

A basic assumption from the *Communicative Theory of Terminology* is that specialized lexical units (SLU) do not differ in nature from other lexical units stored in the mental lexicon by the speaker of that language, but they acquire ‘terminological value’ (Cabré 1999) once being used in specific-domain contexts. This assumption brings about two important consequences for the terminological study of adjectives. In first place, mechanisms in which adjectives modify nouns should not be exclusive to specialized texts. Secondly, some adjectives may be used both in general and in specialized texts; when used in a specialized text, however, they take on a terminological sense. Both hypotheses seem to be validated according to our data.

4.1. Semantic analysis of adjectives

A commonly accepted analysis of adjectives is the distinction proposed by Bolinger (1967) between two classes of adjectives: relational adjectives and qualifying adjectives. The classes show a different syntactic behaviour, and the distinction can be made according to three formal tests. Firstly, relational adjectives are neither gradable nor part of a comparative sentence – as in (1) –, as opposed to qualifying adjectives – as in (2):

(1) a. Es un arquitecto técnico.
   (‘It is a technical architect’)
   b. *Es un arquitecto muy técnico.
      (‘It is a very technical architect’)
   c. *Es un arquitecto más técnico que el otro.
      (‘It is an architect more technical than the other’)

(2) a. Es un arquitecto mediocre.
    (‘It is a mediocre architect’)
   b. Es un arquitecto muy mediocre.
      (‘It is a very mediocre architect’)
   c. Es un arquitecto más mediocre que el otro.
      (‘It is an architect more mediocre than the other’)

Secondly, relational adjectives show more restrictions than qualifying adjectives at being used in copulative sentences, as it is shown by comparing examples (3) and (4):

(3) *Este arquitecto es técnico (non-grammatical in the sense of ‘It is a technical architect’).
    (‘This architect is technical’)

(4) Este arquitecto es mediocre.
    (‘This architect is mediocre’)

---

6 Recent studies on adjectives establish a third class of adjectives, the so-called adverbial adjectives (cf. Demonte 1999). However, in our study they have not been taken into consideration, as there are not samples in our corpus.

7 On the differences between classes of adjectives, see Demonte (1999).
Finally, relational adjectives in Spanish cannot go before the noun:

(5)  
*técnico arquitecto (versus mediocre arquitecto)  
(‘technical architect’ versus ‘mediocre architect’)

The distinction between these two classes is quite relevant for a study on the terminological sense of adjectives. As Estopà (2000) and Estopà et al. (2002) state, there is a tendency to use adjectives with a terminological sense, as relational adjectives or as qualifying adjectives being used as relational ones.

Recent studies in lexical semantics (Bosque 1993, Bosque & Picallo 1996, Picallo 2002) distinguish two different uses of relational adjectives. On the one hand, the so-called argumental or thematic adjectives fill an argument slot of the modified noun. For instance, constitucional (‘constitutional’) in (6) fills the argument slot of theme for the noun reforma (‘reform’), while policial (‘related to police’) in (7) fills the argument slot of agent for the noun respuesta (‘answer’):

(6)  
a. Los partidos de la oposición piden una reforma constitucional.  
(‘Opposition parties ask for a constitutional reform’, i.e., ‘a reform of the Constitution’)
b. La rápida respuesta policial evitó un desastre.  
(‘The rapid response by the police could avoid the disaster’)

On the other hand, the so-called classifying adjectives classify the noun as part of a specific group:

(7)  
análisis vectorial (‘vector analysis’), compuesto químico (‘chemical compound’), agua residual (‘residual water’)

Kornfeld & Resnik (2000) observe terminological uses of both types of relational adjectives, though, as can be seen in our data, they make different contributions to the way the lexical unit is set as a term.

4.2. Adjectives in Environment-related texts

In the case of our corpus, 85 adjectives forming nominal collocations are detected in collocational networks. A total of 18 are not used in a terminological sense, but rather with a non-terminological sense, i.e. as general lexical units: for instance, comparative adjectives (alto ‘tall’, bajo ‘short’, inferior ‘lower’, superior ‘higher’, etc.), forms used to structure discourse (siguiente ‘following’), and adjectives such as importante (‘important’) or necesario (‘necessary’), among others. These adjectives have been excluded from our study. A total of 67 adjectives, then, are analyzed. All these adjectives are part of terminological noun collocations.

By observing these 67 adjectives, three main groups can be established. Firstly, a small group of adjectives used as thematic relational adjectives can be identified. In this case, the nucleus of the noun phrase is either a deverbal noun or an eventive noun, and the adjective usually fills the argument slot of theme – as in (8a-b) –, and occasionally the argument slot of agent – as in (8c):

(8)  
a. abastecimiento urbano (‘urban supply’), desarrollo urbano (‘urban development’)
b. impacto ambiental (‘environmental impact’)
c. actividad humana (‘human activity’)

876
Kornfeld & Resnik (2002) propose that, in such cases, the adjective is not used with a terminological sense. Instead, it modifies the noun, which is the terminological unit. This kind of adjective can be usually replaced by a prepositional phrase (el desarrollo de las ciudades ‘development of the cities’). The adjective alternates with other collocations with the same noun, the same meaning, and the adjective does not vary. However, our data show more complex behaviour for this kind of adjectives. While phrases in (8a) fit the same behaviour described by Kornfeld & Resnik (2002), phrases in (8b-c) have been lexicalized with a particular terminological sense.

Secondly, a bigger group which comprises classifying relational adjectives or qualifying adjectives used as classifying relational adjectives is detected:

(9) planta depuradora (‘sewage plant’), planta acuática (‘water plant’), medio acuático (‘aquatic habitat’), bacteria aerobia (‘aerobic bacterium’), ácido carbónico (‘carbon acid’), carácter estacional (‘seasonal character’)

These collocations contain either adjectives which a low frequency in general texts (anaerobio ‘anaerobic’, fecal ‘fecal’, freático ‘phreatic’, piezométrico ‘piece metric’) or adjectives frequently used in general discourse (agrícola ‘agricultural’, ambiental ‘environmental’, manual ‘manual’).

McNally & Boleda (2004) propose an analysis of the semantics of this kind of adjective, in which adjectives do not reveal a property of an entity, but a kind of that entity. That is, common nouns designate kinds of entities, and the referents of the noun phrases are instantiations. Relational adjectives modify the kind of entity which designates the noun being modified by the adjective. In simple terms, the meaning of a noun phrase such as planta depuradora (‘sewage plant’), therefore, is not the result of the intersection between the total of plants and the total of cosas depuradoras (‘things for sewage treatment’). The adjective depuradora specifies the kind of entity designated by the noun plantas (‘plants’). If this analysis proves to be adequate, it would not be surprising to observe a high frequency of collocations with structure ‘Noun + classifying relational adjective’ in specialized texts, taking into account that terms play an important role in the conceptual structure of a domain. Relational adjectives influence the concept designed by the term, but not the entity to which it refers.

Finally, a limited number of qualifying adjectives can be observed. In general, those are very common adjectives which also modify nouns being frequently used with a non-terminological sense. Although in the case of environmental texts, they form part of a noun collocation which is lexicalized as a terminological unit:

(10) zona árida (‘arid area’), suelo húmedo (‘humid soil’), material sólido (‘solid material’), agua libre (‘free water’)

5. Preliminary conclusions: lexicographical implications

Terminological dictionaries do not usually include adjectives as entries. Adjectives are only included when they form part of a noun collocation which is displayed as a subentry of the noun. This treatment is adequate for most but not all cases. Different kinds of adjectives need

---

8 Formal questions on this analysis are extremely complex to be explained in this paper. For more detailed information, see cf. McNally & Boleda (2004).
different kinds of treatment in lexicographical terms. In this sense, our analysis suggests ways to establish more precise patterns for the representation of terminological adjectives in terminological dictionaries. The same treatment may also be applied for those Environment-related adjectives also used in general texts to be compiled in general dictionaries.

The answer to the questions outlined in the introduction depends on the kind of adjective, as the three different subclasses of adjectives being analyzed show different semantic behaviour and are also different from a terminological point of view. Most adjectives documented in our study used as terms are relational adjectives. Specifically, thematic relational adjectives tend to modify nouns which are terms related to the specific domain of Environment, but these adjectives neither are terms nor do they frequently form a noun phrase which is used as a term. Only in some few cases, the combination [N[A]SAdj]SN has been lexicalized as a terminological unit. On the contrary, classifying relational adjectives usually form collocations with the modified nouns. These collocations are usually terminological units. Some of these adjectives are rarely used in general texts; they are terminological units. Finally, some not specialized qualifying adjectives also modify nouns which are terms. Most of them are used with the same meaning as when being used in general texts. In these cases, their use is not terminological. Nevertheless, they are sometimes used in specialized texts with a meaning which partially differs from the meaning when being used in general texts. This meaning is well defined in the specialized domain, and for that reason these adjectives should be considered as terminological units.

Taking into account the behaviour showed by these three kinds of adjectives from a terminological point of view, it is stated that a different treatment of these adjectives in terminological dictionaries should be applied:

a) Thematic relational adjectives: In general, noun collocations with this kind of adjective should not be compiled as subentries of the modified noun, because only the noun has a terminological value. The semantic function of the adjective modifying the noun is just to indicate the theme, or occasionally the agent, of the event that the noun denotes. For instance, in abastecimiento urbano (‘urban supply’), the adjective urbano (‘urban’) indicates the theme of the action of supplying. As a consequence, the lexicographical entry for abastecimiento (‘supply’) in a terminological dictionary of Environment should not include abastecimiento urbano. Just in the case that the noun collocation has been lexicalized with a terminological sense of a specific domain, such as in impacto ambiental (‘environmental impact’), the collocation must be considered to be included as a subentry of the noun entry.

b) Classifying relational adjectives: Noun collocations with this kind of adjective should be compiled as subentries of the modified noun, as in this case these collocations are used as terms in specialized texts. In those cases in which adjectives are used with a terminological sense (adjectives that are not used in general texts), they must be compiled as independent entries in the dictionary, as these adjectives are considered terminological units. For instance, a terminological dictionary related to the Environment should not include as an independent entry the adjective acuático (‘aquatic’), but it should include as subentries of the noun entry the noun collocations planta acuática (‘aquatic plant’) and medio acuático (‘aquatic habitat’). On the contrary, adjectives as anaerobio (‘anaerobic’), freático (‘phreatic’) or piezométrico (‘piezometric’) should be included as independent entries.

c) Non-terminological qualifying adjectives: Adjectives such as húmedo (‘humid’), sólido (‘solid’) or libre (‘free’) should not be included in specialized dictionaries, as they are
common adjectives taken from general language, which are used in specialized texts with a general sense. However, we advocate that general qualifying adjectives should be included as independent entries in the dictionary in those cases they have acquired a terminological sense in specific domains. For instance, in environmental texts, *húmedo* (‘damp, humid’) does not mean ‘containing a great deal of water or vapour water’, but rather it refers to specific, well-defined climate conditions in this domain.
References


Estopà, R. et al. (2002). ‘El rol de los adjetivos en los textos especializados’. In *VIII Simposio Iberoamericano de Terminología: La Terminología: entre la globalización y la localización.*


