VOLUMEN MONOGRÁFICO (2012), 33-50

CORPUS ANALYSIS: A PRAGMATIC PERSPECTIVE ON TERM VARIATION

MIGUEL ÁNGEL CANDEL MORA*
MARÍA LUISA CARRIÓ PASTOR
Universitat Politècnica de València

ABSTRACT. *An important issue with regard to terms in linguistics is the detection of term variation, which may be defined as the use of alternative names for the same concept, i.e. synonymy. Its causes include the different practices for the designation of terms employed by language users. For example, the use of alternative words in specific terminology among experts is commonplace, but this practice is not common among general users of language, who prefer to stick to popular terms. The differences may lead to the designation of a concept by means of different terms, a phenomenon known as term variation. As a result, a general user may utilise a term which is different from the one used in the technical corpus. In this paper, we aim to identify the types of term variation that can be found when contrasting specific texts in English and the way in which we can recognise some of the types of variation from a pragmatic perspective, i.e. depending on language use and the purpose of the user. First, we examine the types of variation which commonly occur in corpora compiled from specific texts, explaining the document pre-processing and the methodology followed. In the results and discussion section, we describe the method carried out to extract synonyms and the results are contrasted. Finally we conclude that, from a pragmatic perspective, the specificity of communicative acts plays a vital role in term variation.*

KEYWORDS. *Variation, terms, synonyms, specific texts.*

RESUMEN. *Uno de los temas más relevantes en el análisis de términos en lingüística es la detección de la variación terminológica, que podemos definir como los nombres alternativos que se utilizan para un mismo concepto, es decir, la sinonimia. Este interés es debido, entre otras cosas, a las diferentes formas de designar un término entre los usuarios de una lengua. Por ejemplo, es bastante común usar palabras alternativas en terminología específica entre expertos, aunque no lo sea entre usuarios del lenguaje en general, que prefieren utilizar términos generales. Las diferencias pueden llevar a la designación de un concepto, usando diferentes términos, lo que se conoce como variación terminológica. Como resultado, un usuario general puede utilizar*

un término que sea diferente del usado en un corpus técnico. En este artículo, nuestro propósito es detectar los tipos de variación terminológica que se puede encontrar al contrastar textos específicos en inglés y cómo podemos reconocer algunos tipos de variaciones siguiendo una perspectiva pragmática, es decir, desde el punto de vista del uso del lenguaje y del propósito del usuario. Primero investigamos los tipos de variaciones que aparecen comúnmente en los corpus recopilados de textos específicos, explicando el proceso y la metodología seguida. En la sección de resultados y discusión, describimos el método utilizado para extraer los sinónimos y los resultados obtenidos, comparando los resultados. Finalmente concluimos que, desde una perspectiva pragmática, la especificidad de los actos comunicativos juega un papel importante en la variación terminológica.

PALABRAS CLAVE. *Variación, términos, sinónimos, textos específicos.*

1. Introduction

This study on term variation takes into account, on the one hand, the principles of pragmatics described by Grice (1975) and later by Sperber and Wilson (1986), focusing on studies that contrast the meaning of an utterance and its meaning in context (Beaugrande 1996; Bygate 2005; Bouchet 2010; Larrivée 2011). It is also taken into account, on the other hand, the analysis of the behaviour of language from the perspective of corpus linguistics that is valuable in its ability to reveal patterns of communication and to detect changes in language use. From a functionalist point of view, language is not used in the same way by different speakers, and the choice of terms or structures in specific contexts is a field of study that can bring together the broad areas of pragmatics and corpus linguistics.

Variation in terminology refers to terms in the same language that designate the same concept. While a single term could be described as composed of a lexeme and a set of semantic equivalences, variation takes the form of a standard term (e.g. *study*) and a variant (a synonym, i.e. *research*). In this way, variants are the different terms or synonyms used to name the same concept in discourse (Jacquemin 1996; Bowker and Hawkins 2006; Freixa 2006; Condamines 2010; Compornolle 2011; Tercedor Sánchez 2011). A term variant is a form that is conceptually related to the term that is considered the standard by frequency (Daille et al. 1996; Tercedor Sánchez 2011).

To study the implications of term variation from a pragmatic perspective is the main aim of this research, considering communication from the point of view of the context and purpose of the message. The traditional view claimed that terms should be used unambiguously to refer to clearly defined concepts; nevertheless, it has been argued in some studies that ignoring variation in specialised translations may sometimes be problematic. In the case of medical translations, for instance, Bowker and Hawkins (2006: 80) claimed that "[...] translators may actually over-standardize, creating consistency in places where the use of variants was deliberate and well reasoned". In contrast, it has now been extensively demonstrated through corpus-based research that terminological variation is a common phenomenon in specialised communication, even occurring in highly specialised texts. The selection of a term or its variants in a specific

context depends on the linguistic background and/or the purpose of the communicator. In this sense, the hypothesis of this paper is that variants are preferred rather than standard terms, and that this fact depends on the specificity of the text.

In this sense, we aim to detect the types of term variation to be found when contrasting specific texts in English and the way in which we can recognise some of the causes of variation from a pragmatic perspective, i.e. depending on language use and the purpose of the user. Discovering the causes or types of variation is important for both theoretical and practical reasons. From a theoretical perspective, it may help us to determine the extent to which variation is the reflection of the mental processes involved in the selection of one specific term. On a practical level, it could be helpful to researchers and to those involved in disciplines such as translation and terminology.

In short, the main objectives of this study are first, to identify variant patterns in dictionaries; second, to observe whether the standard term or the variant are used in a specific corpus; and finally, to determine the causes that guide the selection of the standard term or the variants.

As mentioned above, terminological variation is quite frequent within specialised communication (Bowker and Hawkins 2006; Fuertes-Olivera and Nielsen 2011). Many of the entries in scientific dictionaries contain cross-references to synonyms or to term variation, although no guidance is provided about the context or genre of the specific use of the term or the variants. The lack of the specific context of term occurrences may produce differences in the use of the main term or of the variants. Term variation should be taken advantage of in order to produce well written texts, although we should be conscious that this is a matter that requires further study, as Bowker and Hawkins (2006: 80) explain:

> It has been suggested by Rogers (1997: 219) that while the use of variants in specialized texts might, on the surface, appear to be arbitrary and by implication sloppy, this is not necessarily the case: "this is an assumption which requires further investigation [...] systematic patterns of variation need to be explained".

Dictionaries provide the standard term but also the variants and so it is the role of linguists to define in which context and how variants should be used. A corpus analysis of the use of term variation may support the recommendations of dictionaries that provide examples to be used by translators and language researchers. Term variation should be defined specifically in order to improve the use of synonyms in specific English. In this sense, dictionaries should include the context of the variants or synonyms in order to help the writer to produce a richer style (Bowker and Hawkins 2006; Fuertes Olivera and Nielsen 2011).

In this study, we take into account a pragmatic standpoint when analysing the selection of a term or a synonym. We view terminology as a tool of specialised communication, entailing the need for standardisation at specialised levels. The use of language varies and the selection of certain terms may change depending on certain factors that need to be determined. The production of specialised texts must be supported by studies based on terminology and therefore, this fact should be accepted by users and

specialists. This procedure is not always considered adequate, and subject field experts choose not to follow the recommendations for term selection and use prescribed by terminologists. From the standpoint of the translator, terminology should be said to be the first reference framework within the text with which to prepare the ground for the translation, analysing the terms and concepts of each domain in order to clearly express the information transmitted. Along with the textual documentation and comparative analysis of this preparatory stage of translation, this may serve to introduce the translator into that area of knowledge.

Terminology cannot be defined without its relation to some concepts from other fields such as information science, linguistics, cognitive science and computer science, among others. Most researchers, such as Temmerman (2000); Resche (2000); Freixa (2006); Condamines (2010) or Fuertes-Olivera and Nielsen (2011), agree that the usage and control of terminology vary according to their application (translation, knowledge representation, information extraction, etc.). Similar studies have been undertaken in other specific fields. As an example, in the field of medicine, Bowker and Hawkins (2006: 83) point out that subject field experts frequently choose not to follow the recommendations of terminologists, who are not generally true subject field experts themselves and may not understand the requirements of the subject fields in question (which may differ from field to field).

The identification of variation has been a matter of interest in terminology. As Ville-Ometz et al. (2007: 36) have said, "The creation of appropriate linguistic tools (Jacquemin 1994) made it possible to appreciate the importance of terminological variation and to study it in corpora (Jacquemin and Royauté 1994)". Corpus analysis plays an important role in this matter and several tools have been developed to detect term variation. Our study has adopted an onomasiological approach to identify term variation, i.e. we start from a concept or standard term that is prior and then locate within a corpus all its linguistic expressions in the sublanguage (e.g., first we consider the concept *able to cause harm* and then we find in the text the terms *hazardous* and *dangerous*).

There are different potential causes behind terminological variation. However, for our study, we focus on linguistically motivated term choices, as proposed by Bowker and Hawkins (2006), which are the most easily identifiable from the observation of a corpus. These authors propose a very concise classification of patterns of linguistic variation into three categories: conceptually motivated term choices, linguistically motivated term choices and socially motivated term choices. Even though Bowker and Hawkins (2006: 92) point out that linguistically motivated term choices are not as important as conceptually motivated term choices with regard to the potential for meaning distortion, they are important nonetheless because they will affect the naturalness or the idiomaticity of the text.

2. Methodology

For this study, we began with a preliminary identification of variant patterns in isolation as they are presented in dictionaries. Afterwards, upon the identification of the

features and characteristics of the specific terminology, we continued with a corpus-based approach in order to study the patterns of these terms in their actual contexts.

The field of study selected was climate change and environmental engineering. It represents an example of a relatively new area of engineering that has led to the introduction of many new terms. Furthermore, the literature of the field has made an impact on a wide range of target audiences, from researchers and policymakers to the public in general.

However, the effect of the media and international efforts from European and international environmental organisations to publicise environmental issues, and raise awareness amongst society in general, has meant that that some of this terminology has been adapted for a less specialised reader. For this reason, we assumed that a high degree of variation in terminology was to be expected, as the writers might vary their selection of terms depending on the target audience, i.e. a not specialised reader.

In addition, as the field of environmental engineering continues to develop, and the vocabulary required for its study also expands, this study may contribute to a better understanding of its relationship with related disciplines. We divided the method used in this study into two stages. For the first stage, we examined the variants present in a dictionary in order to design a variation grid that could help us to contrast the variation patterns contained within the dictionary with the variation found in the specific corpus compiled for this analysis. In this process, firstly, a dictionary was selected; secondly, we studied the definitions; and thirdly, we analysed the cross-references and, finally, we considered abbreviations and shortened forms.

For the second stage, we compiled a specific corpus and contrasted it with the variation grid designed to detect term variation. The process followed in this study to detect term variation and the changes that terms suffer in use will now be described.

2.1. *Selection of the dictionary*

The *Environmental Engineering Dictionary and Directory* (Pankratz 2001) includes more than 8,000 terms, acronyms, and abbreviations related to environmental engineering and the specific processes and agents involved from different fields such as climate change, wastewater, air pollution, and hazardous waste remediation. We selected this specific dictionary because it provides a wide variety of names, as the author states in the preface: "The most unique feature of this book is the inclusion of more than 3,000 trademarks and brand names" (Pankratz 2001: 1). This will be of special interest in our section devoted to eponymous terms, as many commercial terms for proprietary products or processes are so common that they are used generically.

The terms included in this dictionary correspond to the latest versions seen in use and are assumed to be the preferred form, as books, magazines, dictionaries, glossaries, buyer's guides, catalogues, brochures, and technical papers were all consulted as part of the research process for this dictionary, in order to locate new terms and their definitions.

2.2. *Dictionary observation: methodology*

According to Pavel and Nolet (2002: 26), the nature of definitions varies depending on the subject area. They suggest using definitions as a means to validate the terminology of technical and scientific fields, provided that these definitions are cited in reliable sources.

On terminology management, the guidance issued by the Spanish standards agency, *AENOR*, stresses the importance of the definition as a starting point for the analysis and selection of terms (UNE 166002 EX. 2002: 6). The standard on terminology issued in the *American Society for Testing and Materials* suggests the use of *delimiting phrases*, which consists of delimiting the field of application of a term: "If a term can have different meanings in other technical fields or contexts, the term entry can contain an italicised phrase that delimits the definition to its field of application". (ASTM E1964-98: 2).

Upon processing the 8,207 entries of the dictionary, it was found that there are ninety-one entries with two definitions, nine instances with between three and six definitions, and only one instance of seven definitions, but since this corresponds to an abbreviation, and this field is formed by many different disciplines, it may be regarded as a normal instance. Table 1 shows, as an example, the variants of the abbreviation RA found in the dictionary:

TABLE 1. *Variants of the abbreviation RA.*

(1) Reasonable alternative
(2) Regulatory alternatives
(3) Regulatory analysis
(4) Remedial action
(5) Resource allocation
(6) Risk analysis
(7) Risk assessment

From the list of entries containing two definitions, and therefore showing signs of terminological variation, twelve instances correspond to abbreviations, much fewer than initially expected. Two types of instances were found: definitions which show the full form of the acronym and definitions which provide a cross reference to another entry which defines its alternative full form or a combination of both, as shown in Table 2:

TABLE 2. *Type of definitions for abbreviations*.

Abbreviation	Definitions	
ATP	(1) Average transmembrane pressure.	(2) See "adenosine triphosphate"
BAF	(1) See "biologically active filter."	(2) Biologically activated foam.
EHS	(1) Extremely hazardous substance.	(2) Environmental health and safety.
ERC	(1) Emissions reduction credit.	(2) Environmental Research Center.

With regard to the rest of the 79 instances, there is no observable regular pattern as to the reason of variation. At first sight, it may seem to come from the fact that the dictionary includes terms from a few disciplines closely related to environmental engineering, such as in the examples below in Table 3:

TABLE 3. *Variants from different disciplines*.

Terms	Definitions	
Capillary	(1) A slender hair-like structure or a very fine, small bore tube.	(2) A blood vessel with very fine openings that joins the smallest arteries with the smallest veins.
Detritus	(1) Decaying organic matter such as root hairs, stems, and leaves usually found on the bottom of a water body.	(2) Grit or fragments of rock or minerals.
Flux	(1) Flowrate per unit area.	(2) Heat transfer rate per unit area.

A more detailed study reveals that the inclusion of several of the definitions registered does not necessarily entail different disciplines as can be seen in Table 4:

TABLE 4. *Definitions from the same area of knowledge*.

Term	Definitions	
Adulterated	(1) Any pesticide whose strength or purity falls below the quality stated on its label.	(2) A food, feed, or product that contains illegal pesticide residues.
Black water	(1) A condition in drinking water that results from the presence of excess oxidized manganese.	(2) Water that contains animal, human, or food waste.
Compaction	(1) The reduction of the bulk of solid waste by rolling and tamping.	(2) The reduction in thickness of a filter medium or membrane as a result of pressure.
Corona	(1) The layer of ionized gas surrounding the sun.	(2) A sometimes visible electric discharge resulting from a partial electric breakdown in a gas.

As mentioned above, in addition to contributing to terminological standardisation, from the point of view of natural language processing applications, especially indexing and machine translation, the use of consistent spelling is particularly important. As expected, from the 8,000 entries, the dictionary only identifies seven terms that may have a different spelling. We used the search engine Google as a corpus in order to detect the frequency of the entries with different spelling. A Google search on the web confirms that both spellings are frequent, although the variant may be even more frequent than the main entry, as can be observed in Table 5:

TABLE 5. *Different spellings in dictionary entries and number of Google hits*.

Main entry	**Google hits**	**Also spelled**	**Google hits**
Acidophil	92,600	Acidophile	107,000
Amoeba	9,590,000	Ameba	887,000,000
Amoebic Dysentery	356,000	Amebic dysentery	585,000
Amoebicide	17,200	Amebicide	64,200
Basophil	507,000	Basophile	209,000
Salinization	447,000	Salination	599,000
Savanna	26,800,000	Savannah	148,000,000

However, a further search of common collocations of one of these terms, *salinization/salination*, reveals that the choice of the exact term used differs greatly according to the usual context they are usually found in, as can be seen in Table 6.

TABLE 6. *Common collocations for different spellings*.

Term	**Google hits**
Salinization plant	848,000
Salination plant	106,000
Water salinization	536,000
Water salination	1,060,000

In this way, we observed that the number of hits in Google can be useful for the observation of variant preferences and, although this is not the main aim of this study, this information may be relevant for the interpretation of the results.

2.3. *Dictionary observation: cross-referencing*

A cross-reference in a dictionary usually refers to related or synonymous information within the same work. Thus, it is an excellent indicator of the existence of different words for the same concept and its connections.

As stated above, the study of terms as they are recorded in a dictionary can provide evidence for terminological variation. Undoubtedly, cross-references or variants of terms within the entries of a dictionary provide quick and reliable clues as to the degree of acceptance of variants by field experts. These variants are commonly introduced using the following keywords:

See
Commonly known as
Also known as
Known as
Also called

The entries in the dictionary on environmental and climate change include synonyms or linguistic variants, although, as Bowker and Hawkins (2006: 80) point out, there is no guidance with regard to which variant should be used, which in turn presents a challenge to translators and for indexing.

In our corpus, it was found a total of 231 entries that include a cross-reference to other entries that can be classified into two large categories: cross references to another entry within the same dictionary, and cross-references directly to a variant of the term being described. Examples of the first category are usually introduced with the expression *see* and followed by the reference to the other entry as can be seen in:

Amoebiasis	See "amoebic dysentery." Also spelled "amebiases."
Anhydrite	See "calcium sulfate."
Anti-degradation clause	See "prevention of significant deterioration (PSD)."

However, in many other instances, the entry is defined as usual and the last part of the definition includes the cross-reference to the variant:

Bittern	The bitter liquid remaining after the crystallization of salt from brine. See also "mother liquor."
Chelating agent	A compound that is soluble in water and combines with metal ions to keep them in solution. See also "sequestering agent."
Purified water	Bottled water that has undergone significant treatment prior to bottling. See also "USP-purified water."

The second type of instances, with a more direct reference to the term variant, are introduced by revealing expressions such as *also known*, *commonly known* or *also called*. It has not been possible to identify a regular pattern as to the decision to include a variant. As in the section on definitions, the variant does not always correspond to a more common technical term, to avoidance of Latin terms, or to the provision of a shortened version of the main term, but a combination of all these methods. Some examples are:

Psychoda flies	A small, dark-colored fly that creates a nuisance by breeding in trickling filter beds. Commonly known as "filter flies."

Sludge drying bed	A partitioned area consisting of sand or other porous material upon which sludge is dewatered by drainage and evaporation. Also known as "drying bed" or "sand drying bed."
Carbonaceous biochemical oxygen demand (CBOD)	The portion of biochemical oxygen demand where oxygen consumption is due to oxidation of carbon, usually measured after a sample has been incubated for 5 days. Also called "first-stage BOD."

2.4. *Abbreviations and shortened terms*

One of the main features of this type of technical text is the widespread use of abbreviations and shortened forms, with a reduced representation of the terminological unit, with just one or several of its letters. In fact, it should be noted that, at some points, there is an excessive use of acronyms, which not only hinders a fluent reading of the text, but can pose challenges to comprehension for users unfamiliar with all the issues involved in the management of climate change. Some examples can be seen in Table 7:

TABLE 7. *Examples of full and shortened terms.*

Term	Shortened term
Belt filter press	Belt press
Biochemical oxidation	Bio-oxidation
Biological filter	Biotower
Algal bloom	Cloom
Clinoptilolite	Clino
Cogeneration	Cogen
High efficiency particulate air filter (HEPA)	HEPA filter
Cement kiln dust	Kiln dust
Percolation test	perc test
Physical-chemical treatment	Phys-chem
Tropical rain forest	Rain forest
Oxidation-reduction potential (ORP)	Redox potential
Saline water	Salt water
Siliceous gel zeolite	Gel zeolite
Physical-chemical treatment (PC)	Phys-chem
Electron beam irradiation	E-beam

Thus, the dictionary includes inconsistent criteria for the inclusion of abbreviations in terms that refer to chemical substances, units of measurement, treatment processes, brand names, associations, etc., to mention just a few examples.

Acronyms have also been included in the dictionary when they are registered trademarks or commonly used abbreviations. On the other hand, nonregistered product trademarks that are the same as the name of a company are not always included, as Pankratz (2001: 10) explains. Some examples from the dictionary are:

A2/O®	Biological treatment process for phosphorus and nitrogen removal by USFil- ter/Krüger.
AAPCO	American Association of Pesticide Control Officials.
ABA8000®	Alumina oxide for fluoride removal by Selecto, Inc.
ACBM	Asbestos-containing building material.
ACCU®	Air sampler by Rupprecht & Patashnick.
AHS	See "aquatic humid substances (AHS)."

Once the patterns to be considered in the detection of variants were established, we focused on the analysis of the specific corpus compiled for this study in order to apply the designed method. In order to study terms in their actual context, the corpus used in this study contained approximately one million words and was comprised of documents from the European Environment Agency and the United Nations Framework Convention on Climate Change. Finally, we extracted and analysed the results obtained from applying the designed method to a real corpus and drew our conclusions. We also calculated the confidence interval (P) to determine the significance of the data obtained.

3. Results and discussion

The statistical data of the corpus compiled for the analysis of this study is shown in Table 8:

TABLE 8. *Statistical data climate change corpus*.

Tokens (running words in text)	2,403,677
Tokens used for word list	2,234,680
Types (distinct words)	19,578
Type/token ratio (TTR)	1
Standardised TTR	30.59

We extracted these data using *Wordsmith Tools 5.0* in order to obtain a general idea of the lexical density of the corpus selected. The types showed that the corpus was adequate for our purposes. The most frequent words were chosen from the corpus compiled for this study. From the general results, we extracted the key words that we had previously selected from the dictionary, in order to find term variation in the real data corpus of this study. The terms under study were selected. Table 9 shows the list of the

key entries or standard terms, the number of occurrences and the percentages detected after the analysis of the corpus. These entries were previously identified as having a variant in the dictionary:

TABLE 9. *Number of occurrences in the corpus of the list of terms identified in the dictionary as having a variant.*

Standard terms	**Occurrences (%)**
CO2	694 (66.09%)
Conversion	108 (10.28%)
Sequestration	70 (6.66%)
Tonne	68 (6.47%)
Recycling	28 (2.66%)
Mass balance	28 (2.66%)
CFC	17 (1.61%)
Sewage	15 (1.42%)
Potable water	5 (0.47%)
Saline water	4 (0.38%)
Rejection	2 (0.19%)
Rain forest	2 (0.19%)
Bloom	2 (0.19%)
Sludge digestion	1 (0.09%)
Rise rate	1 (0.09%)
Privy	1 (0.09%)
Ponding	1 (0.09%)
Entrainment	1 (0.09%)
Domestic wastewater	1 (0.09%)
Cogen	1 (0.09%)
Total	**1,050 (100.00%)**

The next step was to identify the words that were catalogued as variants of the terms included in Table 9. The number of occurrences and the percentages found of the variants of the key entries or standard terms of the dictionary are displayed in Table 10:

TABLE 10. *Number of occurrences in the corpus of the list of terms identified in the dictionary as term variants*.

Variant terms	**Occurrences (%)**
Carbon dioxide	269 (50.01%)
Wastewater	95 (10.07%)
Recovery	50 (10.04%)
Drinking water	14 (2.81%)
Metre	13 (2.61%)
Cogeneration	10 (2.00%)
Conditioning	10 (2.00%)
Salinization	10 (2.00%)
Schistosomiasis	6 (1.20%)
Flue gas	5 (1.00%)
Heavy metals	3 (0.60%)
Metric ton	3 (0.60%)
Tropical rain forest	2 (0.40%)
Water recycling	2 (0.40%)
Digestion	1 (0.20%)
Algal bloom	1 (0.20%)
Cryptosporidium	1 (0.20%)
Distillation	1 (0.20%)
Zone of saturation	1 (0.20%)
Chlorofluorocarbon (CFC)	1 (0.20%)
Total	**498 (100.00%)**

Next, we contrasted the occurrences found in the corpus selected. The key entries of the dictionary were contrasted with the cross-referenced term or variant included in the dictionary. Table 11 shows the selected terms and the number of occurrences and percentages:

TABLE 11. *Comparison of occurrences in corpus of terms vs. variants proposed in dictionaries.*

Number of occurrences in corpus	**Key entry**	**Cross-referenced term**	**Number of occurrences in corpus**
694 (66.09%)	CO2	Carbon dioxide	269 (50.01%)
108 (10.28%)	Conversion	Recovery	50 (10.04%)
70 (6.66%)	Sequestration	Chelation	0 (0.00%)
68 (6.47%)	Tonne	Metric ton	3 (0.60%)
28 (2.66%)	Recycling	Water recycling	2 (0.40%)
28 (2.66%)	Mass balance	Material balance	0 (0.00%)
17 (1.61%)	CFC	Chlorofluorocarbon (CFC)	1 (0.20%)
15 (1.42%)	Sewage	Wastewater	95 (10.07%)
5 (0.47%)	Potable water	Drinking water	14 (2.81%)
4 (0.38%)	Saline water	Salt water	0 (0.00%)
2 (0.19%)	Rejection	Salt rejection	0 (0.00%)
2 (0.19%)	Rain forest	Tropical rain forest	2 (0.40%)
2 (0.19%)	Bloom	Algal bloom	1 (0.20%)
1 (0.09%)	Sludge digestion	Digestion	1 (0.20%)
1 (0.09%)	Rise rate	Overflow rate (OFR)	0 (0.00%)
1 (0.09%)	Privy	Outhouse	0 (0.00%)
1 (0.09%)	Ponding	Pooling	0 (0.00%)
1 (0.09%)	Entrainment	Mist eliminator	0 (0.00%)
1 (0.09%)	Domestic wastewater	Sanitary wastewater; domestic sewage	0 (0.00%)
1 (0.09%)	Cogen	Cogeneration	10 (2.00%)
0 (0.00%)	Stack gas	Flue gas	5 (1.00%)
0 (0.00%)	Sludge conditioning	Conditioning	10 (2.00%)
0 (0.00%)	Saturated zone	Zone of saturation	1 (0.20%)
0 (0.00%)	Salination	Salinization	10 (2.00%)
0 (0.00%)	Meter	Metre	13 (2.61%)
0 (0.00%)	Distill	Distillation	1 (0.20%)
0 (0.00%)	Crypto	Cryptosporidium	1 (0.20%)
0 (0.00%)	Chromium	Heavy metals	3 (0.60%)
0 (0.00%)	Bilharzia	Schistosomiasis	6 (1.20%)
1,050 (100.00%)	**Total**	**Total**	**498 (100.00%)**

During the initial analysis of the results obtained from the keyness and the concordance analysis with *WordSmith tools*, it seems that a number of occurrences need to be excluded from the overall count as they are terms originating from general vocabulary which acquire specific meaning in technical fields, labelled by Alcaraz (2000: 43) as *semi-technical* or *sub-technical* vocabulary and by Sager (1990) as *re-designated* general language items.

These terms make up most of the specialised language in any discipline, as their use is not limited to scientific and technical texts, but come from the general language and denote different concepts when they are used within a particular context or specialised field. From the point of view of translation, they are the main source of translation problems, because the semi-technical equivalent in another language does not necessarily come from the general language and may actually be a specific technical term. In the field of translation, Resche (2000: 628) refers to this concept of semi-technical terms as *interface terminology* and, in applied linguistics Lerat (1997: 52) defines it as *supporting vocabulary*.

Looking at our results, the term *conversion* and its variant *recovery* exemplify this situation. Their definition in the dictionary is: "In reverse osmosis processes, recovery indicates the amount/percentage of product water recovered from the feed stream." It is clearly stated that their designation applies only to reverse osmosis processes, while the instances of *recovery* in the corpus include: *economic recovery*, *energy recovery from biogas*, *data storage and the recovery of registry services*, *landfill methane recovery, waste-water management*, and *disaster recovery services* to mention just a few of the fifty total occurrences in the corpus, without any mention of osmosis whatsoever.

The same happens with the occurrences of *conversion*, with instances such as *conversion technologies*, *land conversion or degradation*, *human-induced conversion of non-forest land*, *most conversion processes based on second law*, *installed energy conversion capacities* and *the conversion of organic matter to carbon dioxide by organisms other than plants*, from the total of one hundred and eight occurrences.

The case of *rejection* and *salt rejection* is very similar to the situation described above, although in this case *salt rejection* has zero occurrences and *rejection* has two. The definition in the dictionary is: "In reverse osmosis, the percentage of dissolved solids removed from the feedwater as it passes through the semipermeable membrane". However, as can be seen below, neither of the occurrences found in the corpus refers to *salt rejection*:

> […] accordance with the procedures referred to in paragraph 39 above regarding the *rejection* or alteration of requests for registration or issuance; Requests the…
>
> You are here, we are together, to tell the rest of the world that hatred and *rejection* of the "other", confrontation and the perversions of terrorism and extermination…

The results shown in Table 11 demonstrate that the key entries such as *CO2, conversion, sequestration, tonne, recycling, mass balance* and *CFC* are more frequently used than the cross-referenced terms. Interpreting these results, it is evident that

abbreviations are preferred in specialized texts while the alternative, less opaque, semi-technical terms are selected more often by writers of semi-specialised texts. The rest of the cross-referenced terms analysed in the corpus present equal or higher occurrences than the standard terms. These results indicated that the writers prefer to use variants instead of the standard words. It was calculated the statistical significance of the results; the value P of the occurrences found was lower than 0.05; consequently, it meant that the results obtained were statistically significant.

4. Concluding remarks

We observed in this analysis that terms are not used in a random way by technical writers. The use of the terms depends on the pragmatic perception of the writer, who adapts a text for a particular target audience. After the results obtained in Tables 9, 10 and 11 we believe that the selection of a standard term or its variants in a specific context depends on the audience, the communication channel and the context. In this paper, we used a corpus-based study to show that term variation occurs depending on the pragmatic implications of the text, as Bowker and Hawkins (2006) explained; but also, the intention of the writer and the audience are important in order to decide whether a standard term or a variant is more appropriate. With the identification of the variation patterns, it is possible to design a grid that can help translators and specific writers to detect term variation and the frequency of terms in context.

These aspects are relevant for the identification of the variant to be used in a certain context and also for the avoidance of the underuse of certain terms that are considered specialised, but are already being used in non-specific contexts. We should stress that the results found in this study apply to the genre analysed herein; nevertheless, further research may reveal similar results in other genres.

Regarding the objectives set in this study variant patterns were identified in dictionaries, as can be seen in the Methodology section. It was also shown that the selection of the variant is more frequent in specialised contexts in the corpus analysed. The cause why the authors select variants may be that the texts are intended to be read by specialists who are interested in this field. The possible motivations behind term choice could be conceptual, linguistic and social as pointed out by Bowker and Hawkins (2006: 101). In this study, we consider that social and conceptual motivations are the most frequent when choosing variants. In this sense, the pragmatic implications derived from our analysis are evident. The sender and receiver of the message play a vital role in term variation. The writer chooses variants to transmit specialised concepts that distinguish this kind of discourse from a more general one.

It is important to note that the research into patterns of variation and the motivations for choosing one variant or another could be helpful to researchers and to those involved in disciplines such as translation and terminology. Furthermore, non-native writers may also be able to choose a variant depending on context and audience, following linguistic patterns that may improve terminological richness.

Further research could include other genres or other variants that show that there are different ways of conveying the same reality. We tend to forget that different writers do not perceive concepts or express thoughts in the same manner. This future research may prove that communication depends on the linguistic, cognitive, cultural and social background of the writers and, moreover, the rules for how language functions are not as general, as fixed or as evident as we suppose. The interpretation and production of language is not a one-way street: it has different paths that allow us to express our ideas in a wide variety of forms.

NOTES

* Correspondence to: Miguel Ángel Candel Mora. Departamento de Lingüística Aplicada. Universitat Politècnica de València. Camino de Vera, 14. 46022. Valencia. E-mail: mcandel@upvnet.upv.es.

REFERENCES

Beaugrande, R. de. 1996. "The 'pragmatics' of doing language science: The 'warrant' for large-corpus linguistics". *Journal of Pragmatics* 25: 503-535.

Bouchet, D. 2010. "Pragmatics of intercultural communication. The bounded openness of a contradictory perspective". *Pragmatics and Society* 1 (1): 138-154.

Bowker, L. and S. Hawkins. 2006. "Variation in the organization of medical terms. Exploring some motivations for term choice". *Terminology* 12 (1): 79-110.

Bygate, M. 2005. "Applied Linguistics: A Pragmatic Discipline, a Generic Discipline?" *Applied Linguistics* 26 (4): 568-581.

Compornolle, R. A. van. 2011. "Developing a sociocultural orientation to variation in language". *Language & Communication* 31: 86-94.

Condamines, A. 2010. "Variations in terminology. Application to the management of risks related to language use in the workplace". *Terminology* 16 (1): 30-50.

Daille, B., B. Habert, C. Jacquemin and J. Royauté. 1996. "Empirical observation of term variations and principles for their description". *Terminology* 3 (2): 197-257.

Dubuc, R. and A. Lauriston. 1997. "Terms and contexts". *Handbook of Terminology Management*. Eds. S. E. Wright and G. Budin. Amsterdam: John Benjamins. 80-88.

Freixa, J. 2006. "Causes of denominative variation in terminology. A typology proposal". *Terminology* 12 (1): 51-77.

Fuertes-Olivera, P. A. and S. Nielsen. 2011. "The dynamics of terms in accounting. What the construction of the accounting dictionaries reveals about metaphorical terms in culture-bound subject fields". *Terminology* 17 (1): 157-180.

Grice, H. P. 1975. "Logic and conversation". *Syntax and Semantics*, Volume 3. Eds. P. Cole & J. Morgan. New York: Academic Press. 41-58.

Jacquemin, C. 1996. "What is the tree that we see through the window: A linguistic approach to windowing and term variation". *Information Processing and Management* 32 (4): 445-458.

Jacquemin, C. 1994. "FASTR: A Unification-based Front-End to Automatic Indexing". *RIAO*: 34-48.

Jacquemin, C. and J. Royauté. 1994. "Retrieving terms and their variants in a lexicalized unification-based framework". *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*. Heidelberg: Springer-Verlag. 132-141.

Larrivée, P. 2011. "The role of pragmatics in grammatical change: The case of French preverbal *non"*. *Journal of Pragmatics* 43: 1987-1996.

Lerat, P. 1997. *Las lenguas especializadas*. Barcelona: Ariel.

Pankratz, T. M. 2001. *Environmental Engineering Dictionary and Directory*. London: CRC Press-Lewis Publishers.

Pavel, S. and D. Nolet, 2002. *Manual de Terminología*. [Available at http://www.btb.gc.ca/btb.php?lang=eng&cont=705].

Resche, C. 2000. "An Approach to Interface Terminology: The Example of Environmental Economics in English as a Foreign Language". *Meta* XLV (4): 628-645.

Sager, J. C. 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Sperber, D. and D. Wilson. 1986. *Relevance. Communication and cognition*. Oxford: Blackwell Publishing.

Temmerman, R. 2000. *Towards new Ways of Terminology Description: The Socio-cognitive Approach*. Amsterdam: John Benjamins.

Tercedor Sánchez, M. 2011. "The cognitive dynamics of terminological variation". *Terminology* 17 (2): 181-197.

UNE 166002 EX. 2002. *Gestión de la I+D+I: Terminología y definiciones de las actividades de I+D+I*. Asociación española de normalización y certificación. Madrid: AENOR.

Ville-Ometz, F., J. Royauté and A. Zasadzinski. 2007. "Enhancing in automatic recognition and extraction of term variants with linguistic features". *Terminology* 13 (1): 35-59.