

## Terminological Variation in Source Texts and Translations: A Pilot Study

Sabela Fernández-Silva and Koen Kerremans

Volume 56, Number 2, June 2011

Les corpus et la recherche en terminologie et en traductologie  
Corpora and Research in Terminology and Translation Studies

URI: <https://id.erudit.org/iderudit/1006179ar>

DOI: <https://doi.org/10.7202/1006179ar>

[See table of contents](#)

### Publisher(s)

Les Presses de l'Université de Montréal

### ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

### Cite this article

Fernández-Silva, S. & Kerremans, K. (2011). Terminological Variation in Source Texts and Translations: A Pilot Study. *Meta*, 56(2), 318–335.  
<https://doi.org/10.7202/1006179ar>

### Article abstract

In this article, it is assumed that the choice of terminological variants in specialized source texts is sometimes cognitively motivated and that this motivation is reflected in the choice of equivalents in the target texts. On the basis of a pilot study, we will present a method for comparing the cognitively motivated terminological variants in source texts and their translations. The corpus in the pilot study is composed of three Galician source texts and their English translations. The texts are scientific articles addressing the economic effects of environmental disasters on fisheries. A quantitative study was first carried out in which the number of unique terms in each source text was compared to the number of unique translations of these terms. Next, each unique combination of a source term and its translation equivalent was subjected to a qualitative analysis. A value was manually assigned in order to qualify the “cognitive distance” between the source term and its translation. Based on these values and the frequency of the translation pair in each bitext, we computed the “interlingual variation index” (IVI). Differences in results between the bitexts are linked to extra-linguistic factors related to the translation processes.

# Terminological Variation in Source Texts and Translations: A Pilot Study

**SABELA FERNÁNDEZ-SILVA**

*Pontificia Universidad Católica de Valparaíso, Valparaíso, Chili*  
*Universitat Pompeu Fabra, Barcelona, Spain*  
sabela.fernandez@upf.edu

**KOEN KERREMANS**

*Erasmus University College Brussels, Brussels, Belgium*  
koen.kerremans@ehb.be

## RÉSUMÉ

Dans le présent article, nous soutenons l'hypothèse selon laquelle le choix des variantes terminologiques dans les textes sources est parfois motivé sur le plan cognitif et que cette motivation est reproduite dans le choix des équivalents dans les textes cibles. À partir d'une étude pilote, nous présentons une méthode pour comparer les variantes terminologiques cognitivement motivées dans les textes sources et leurs traductions. Le corpus est constitué de trois articles scientifiques portant sur les conséquences économiques des catastrophes naturelles sur la pêche, écrits originairement en galicien et traduits ensuite en anglais. Nous avons d'abord comparé quantitativement le nombre de termes uniques dans chaque texte avec le nombre de traductions uniques de ces termes. Nous avons ensuite entrepris une analyse qualitative de chaque paire terme source – terme équivalent, dans laquelle nous avons caractérisé la distance cognitive entre ces deux termes à l'aide d'une valeur numérique. À partir de ces valeurs et de la fréquence des paires de traduction dans chaque bitexte, nous avons calculé l'index de variation interlangue (IVI) de chaque texte. Finalement, nous avons mis en relation les différents résultats entre les bitextes avec des facteurs extralinguistiques concernant le processus de traduction.

## ABSTRACT

In this article, it is assumed that the choice of terminological variants in specialized source texts is sometimes cognitively motivated and that this motivation is reflected in the choice of equivalents in the target texts. On the basis of a pilot study, we will present a method for comparing the cognitively motivated terminological variants in source texts and their translations. The corpus in the pilot study is composed of three Galician source texts and their English translations. The texts are scientific articles addressing the economic effects of environmental disasters on fisheries. A quantitative study was first carried out in which the number of unique terms in each source text was compared to the number of unique translations of these terms. Next, each unique combination of a source term and its translation equivalent was subjected to a qualitative analysis. A value was manually assigned in order to qualify the “cognitive distance” between the source term and its translation. Based on these values and the frequency of the translation pair in each bitext, we computed the “interlingual variation index” (IVI). Differences in results between the bitexts are linked to extra-linguistic factors related to the translation processes.

## MOTS-CLÉS/KEYWORDS

variation terminologique, traduction spécialisée, motivation terminologique, index de variation interlangue, distance cognitive  
terminological variation, specialized translation, term motivation, interlingual variation index, cognitive distance

## 1. Introduction

Recent descriptive approaches to terminology share the claim that terms are liable to the same two variation patterns as general words: variation affecting the form or denominative variation, and variation affecting the content, or conceptual variation. In this study, our interest focuses on the various terms that are used to refer to a single object, either within the same language (synonymy) or between different languages (translation equivalents). Throughout this article the term “terminological variation” will be used instead of denominative variation in order to emphasize the fact that the study is focusing only on terminology as a means to express a given thought. Other means of expression are not considered in this study.

Corpus-based research has shown that terminological variation is prominent in specialized communication (Freixa 2002). Several reasons are put forward in the literature to explain why terminological variation occurs (Freixa 2006). An author of a specialized text may decide to use a set of alternative expressions on stylistic grounds in order to avoid having a particular thought always expressed in exactly the same way. This is not necessarily “a random act of defiance or carelessness, but one which is well motivated and useful in expert discourse” (Bowker 1998: 487). The ways of expressing a thought will also differ when for instance a specialist is talking about a topic in his subject field to a colleague as compared to a non-expert. In every new communicative situation, a sender needs to find out which ways of expression seem best to convey a message as clearly as possible to the recipient (Cabré 1995). Established terms may not always seem the best option, either because the thought that he or she wishes to express is slightly different or because the terms that are commonly used to express the thought do not fit his or her way of understanding it (Bowker 1997).

In this article, it is assumed that the choice of terminological variants in specialized source texts is sometimes cognitively motivated and that this motivation is reflected in the choice of equivalents in the target texts (Suárez de la Torre 2004). On the basis of a pilot study, we will present a methodology for comparing the cognitively motivated terminological variants in source texts and their translations. The methodology is derived from two PhD projects dealing with terminological variation. In the project carried out at the CVC research centre of the Erasmus University College Brussels,<sup>1</sup> a method was worked out in order to study and compare terminological variation in source texts and translations (Kerremans 2010). In the project carried out at the IULATERM research centre of the University Pompeu Fabra,<sup>2</sup> a method was adopted to study and describe the cognitively motivated behavior of terminological variation (Fernandez-Silva, Freixa *et al.* 2009).

In section 2 we will summarize both these projects. The methodology worked out in the pilot study will be outlined in section 3. We discuss how the corpus was compiled (section 3.1) and how we identified source terms (section 3.2) and translations (section 3.3). On the basis of the cognitive analysis of source terms and translations, two measures are introduced in this methodology: on the one hand, the “cognitive distance” between source terms and translations and, on the other hand, the “Interlingual Variation Index” of each cluster of terminological variants. These notions will be further explained in section 3.4. After discussing the methodology, we will present the results of our pilot study in section 4. These results will be discussed in section 5 and, finally, in section 6 we present our conclusions.

## 2. Research framework

In this section we will summarize the CVC (section 2.1) and IULATERM (section 2.2) research projects and discuss some essential notions partly derived from these two projects so as to compare how much the cognitively motivated term choices in source texts are reflected in translations.

### 2.1. Terminological variation in specialized translation

The aim of CVC's project was to find out whether certain patterns or trends can be derived from a comparative analysis of terminological variation in source texts and their translations. Translators were thus studied to determine how much their translation practice was influenced by the traditional, prescriptive view in terminology theory which postulates that terms should be used unambiguously to refer to clearly defined concepts (Wüster 1979/1991; Felber 1981). For instance one might expect that translators tend to ignore terminological variation in specialized translations for the sake of term consistency (Merkel 1996). In some studies, however, it has been argued that ignoring variation in specialized translations may sometimes be problematic. Bowker and Hawkins (2006: 80), for instance, claimed that "translators may actually over-standardize, creating consistency in places where the use of variants was deliberate and well reasoned."

Toury (1995)'s law of interference states that elements of the source text tend to be transferred to the target text during the process of translation. This influence of the source language system is not only noticeable at the syntactic level but also at the lexical level. Given the close intertextual relation between a source text and its translation, it would therefore seem reasonable to expect that the set of terms in the source text that designates a common referent is replaced by a set of conceptually equivalent terms in the target language. It is assumed however that the translation of terms is sometimes linguistically more creative due to several linguistic and socio-cultural factors that translators need to take into account when translating a given source term (Durieux 1995). This project aims to find out possible correlations between the degree of terminological variation and some of these factors on the basis of a contrastive study.

The methodology that has been worked out in the Ph.D. project has partly been applied to our pilot study (section 3). Essential in this methodology is the fact that terminological variants are clustered and organized according to a unit of understanding (UoU). The notion of UoU was introduced in sociocognitive terminology theory in order to clarify the inadequacy of classical concept theory for the conceptual structure of most specialized fields (Temmerman 2000). This view – shared by other authors such as Gaudin (2003), Rogers (2004) and Cabré (2008) – acknowledges the flexibility of concepts and conceptual structures and has important implications for the study of terminological variation. In this view, terms are considered part of the same cluster of terminological variants when they point to the same referent that is being introduced and described in a text. This means that, for instance, a term like producción mexilloeira (*mussel production*), its synonym producción mitícola and the hyperonym producción would be considered part of the same cluster or set of variants if they designate the same referent in a text. This textual approach in the study of

terminological variation clearly differs from the traditional, onomasiological view which takes a concept as a starting-point to identify terms having the same meaning. For instance, seen from the onomasiological perspective, only producción mexilloeira and producción mitícola would be considered as part of the same set of variants.

Terms sharing the same referent are also called “co-referents” (Rogers 2007). These terms are identified based on an analysis of the cohesive ties in a text. The resulting cluster of terminological variants corresponds to the UoU in the text. In our methodology, each UoU found in the source texts is assigned a unique identification label (UoU label). An example of such a label is found in Table 1 where the UoU label //AXENTE\_POLUCIONANTE// [//polluting agent//] was created to cluster three Galician terminological variants encountered in one of the source texts in our pilot study. The Galician occurrences in this table show the exact forms in which the terms were encountered in the source text, the lemmas were added afterwards.

TABLE 1

Example of terminological variants appearing in the source text

Unit of Understanding	Galician lemma	Galician occurrence
AXENTE_POLUCIONANTE	axente polucionante	axente polucionante
AXENTE_POLUCIONANTE	axente contaminante	axentes contaminantes
AXENTE_POLUCIONANTE	componente tóxica	componentes tóxicas

## 2.2. Cognitive motivations for terminological variation

The project carried out at the IULATERM research centre aims to describe the cognitively motivated behavior of denominative variation. Starting from the premise that term formation is motivated (Guiraud 1978; Kocourek 1991; Sager 1997 and Myking 2009), it is claimed that terminological variation is the result of multiple motivations that take place in the naming process (Freixa, Fernández-Silva *et al.* 2008). Some of these motivations are situated at the level of the system of terminology while others are situated at the level of use. The type of UoU being named in a domain and the language that is employed are factors influencing term choice at the systemic level (Kageura 2002); but naming is also affected by contextual factors.

The contextual factors that have been studied in relation to terminological variation are arranged according to different levels. The cognitive level is related to the perspective from which an expert approaches the UoU in a particular situation, which will determine what features of the UoU the expert puts emphasis on. This has been investigated by authors such as Temmerman (2000) and Fernández-Silva, Freixa *et al.* (2009), who suggested that term choice is influenced by the authors’ domain of specialization. The communicative level involves the circumstances of message production and reception. Freixa (2002), for instance, observed that the level of specialization of a text determines the degree and types of terminological variation. Finally, term variation has also been studied at the discursive level; Collet (2004), for example, has shown that terms are subject to formal and structural transformations when embedded in a discursive environment, giving rise to different types of context-conditioned variants.

In order to describe the patterns and regularities of terminological variation, a methodology has been worked out in which the clusters of terminological variants

referring to the same UoU are analyzed with respect to the features of the UoU they emphasize. This methodology is supported by fundamental works on term formation, which suggest that terms often reflect the UoU's most relevant features (Kocourek 1991; Sager 1997). We have adopted the analysis employed in Kageura (2002) to describe the cognitively motivated patterns of term formation. In this analysis, term formation is seen as the specification of concepts within a conceptual class, as represented by the nucleus, by means of modifications represented by the modifiers. Therefore, the term's content is interpreted as a combination of concepts within the overall conceptual system of the domain. These concepts are reflected in the different constituent elements that make up a term.

Table 2 shows on the basis of an example from a source text in our pilot study how each constituent element of a term is first identified as either 'head' or 'modifier.'

TABLE 2

**Identifying head and modifier**

UoU label	Galician lemmas	Head	Modifier
AXENTE_POLUCIONANTE	axente polucionante	axente	polucionante
	axente contaminante	axente	contaminante
	componente tóxica	componente	tóxica

After the constituent elements are identified, the next step is to link each element to its corresponding concept within the domain. Each concept is characterized according to the conceptual category it belongs to, and the distinctive feature that is reflected in its form. In the example of Table 2, all head elements represent substance concepts, the difference being the distinctive feature chosen to name it –the agentivity in axente (*agent*) or the structure in componente (*component*). The modifiers show either a quality or an activity characterizing this substance (Table 3).

TABLE 3

**Identifying the conceptual information of head and modifier**

Linguistic information		conceptual information	
Head	Modifier	Head	Modifier
axente	polucionante	SUBSTANCE (BY ACTION)	ACTIVITY
axente	contaminante	SUBSTANCE (BY ACTION)	ACTIVITY
componente	tóxica	SUBSTANCE (BY STRUCTURE)	QUALITY

In our pilot study, this method of analysis was adopted to compare the cognitively motivated choices between source terms and their translations. In particular we will look for which aspects of the UoU are given more preference in the choices of terms in the source texts, and we will examine whether or not the same cognitive motivation is respected in translations. The way this analysis is carried out is further explained in section 3.4.

### 3. Methodology

In our pilot study, terminological variation was examined in three Galician scientific articles addressing the economic effects of environmental disasters on fisheries and

their English translations. A quantitative study was first carried out in which the number of unique terms in each source text was compared to the number of unique translations of these terms. Next, each unique combination of a source term and its translation equivalent was subjected to a qualitative analysis in order to examine possible conceptual differences between the source term and its translation.

We will first discuss the criteria that were adopted for compiling the corpus and also provide a brief description of the three texts (section 3.1). We will then explain how the analysis of terms in the source texts was carried out (section 3.2). In section 3.3, we will focus on the analysis of the equivalents of the source terms. We will also show how the source terms and their corresponding translations were annotated semi-automatically in the bitexts. Finally, we will discuss the types of analyses that were carried out on the annotated data (section 3.4).

### 3.1. Corpus compilation

The texts in this pilot study were selected from a bigger corpus used in IULATERM's research project (Fernández-Silva, Freixa *et al.* 2009). In order to allow for comparability, the three selected texts were chosen on the basis of the following criteria:

- **Language:** The source texts were originally written in Galician, and subsequently translated into English.
- **Topic:** All texts deal with the economic consequences of oil spills in the Galician fishing sector.
- **Text type:** All texts are scientific articles, published in conference proceedings or international reviews.
- **Subject field:** all texts are related to the subject field of applied economics.

The reasons for choosing texts from an existing project were also of a practical nature. At the moment of starting the pilot study, we were familiar with the domain and the terminology of coastal fishing and aquaculture, which is necessary to carry out the cognitive analysis (cf. section 3.4.). Furthermore, we had information concerning the genesis of texts and translations, which proved to be of primary importance in correctly interpreting the results, as we will explain in section 5. Relevant textual and extralinguistic information about the source texts is summarized in Table 4.

TABLE 4  
Information about source texts

Reference	Subject field	Text type	Year	Publication format	Audience	Specialization level	Word count
SG	applied economics	scientific article	2004	proceedings	marine economics & marine sciences experts	highly specialized	4166
GN	applied economics	scientific article	2004	proceedings	marine economics & marine sciences experts	highly specialized	8546
DC	applied economics	scientific article	2007	proceedings	field experts, fishermen, businessmen	highly specialized	8960

These three source texts were aligned with their translations at sentence level (wherever this was possible). The resulting bitexts were used to identify and annotate the source terms (section 3.2) and their translations (section 3.3).

### 3.2 Identification of source language terms

Source text analysis and term extraction were carried out with the help of a text analysis tool TextStat.<sup>3</sup> The resulting term list was further complemented with term variants that were manually extracted from the texts. Terms that were considered variants of the same UoU were assigned the same UoU label (section 2.1). The resulting term list as well as the equivalence relations between terminological variants was validated by field experts (Fernández-Silva, Freixa *et al.* 2009).

The term list is used to automatically find and highlight terms in the source language contexts of the bitexts. Each matching term is placed between identification tags. The identification tag provides two types of information: it links the term to a UoU by showing the unique UoU label. It also shows a unique number for each term per sentence. The number is used to locate the translation of the source term in the target sentence (section 3.3). This is crucial if a UoU is expressed more than once in a source sentence.

The following example illustrates how source terms in the corpus texts were annotated semi-automatically by means of a program written in Perl that was developed in the framework of the CVC research project (section 2.1). This sample is taken from the ‘GN’ source text (Table 3).

[...] pode ver-se retardada a velocidade de migración, con todo o que iso implica na ||2-PESCA||pesca||2-PESCA|| e supervivencia da especie, xa que se o migrante se desvia do seu lugar habitual de freza, por mor dun ||1-AXENTE\_POLUCIONANTE||axente polucionante||1-AXENTE\_POLUCIONANTE||, a poboación pode verse afectada dun xeito esaxerado.

### 3.3 Identification of translation equivalents

After annotation of the source terms, translations of these terms were searched for in the target sentences in each bitext. A translation received the same identification tag as its corresponding term in the source text, as is shown in the following sample taken from the ‘GN’ text (Table 5).

TABLE 5  
Annotated fragment of source and target aligned texts

<b>Galician</b>	pode ver-se retardada a velocidade de migración, con todo o que iso implica na   2-PESCA  pesca  2-PESCA   e supervivencia da especie, xa que se o migrante se desvia do seu lugar habitual de freza, por mor dun   1-AXENTE_POLUCIONANTE  axente polucionante  1-AXENTE_POLUCIONANTE  , a poboación pode verse afectada dun xeito esaxerado.
<b>English</b>	The speed of migration may be reduced, with all that implies to   2-PESCA  fishing  2-PESCA   and to the survival of the species, since if the migrant strays from its usual spawning ground due to a   1-AXENTE_POLUCIONANTE  polluting agent  1-AXENTE_POLUCIONANTE  , the population may find itself affected to an extreme degree.



In case the source term was not translated in the target text, the unique identification tag of the source term was added at the end of the target sentence. In that way the translation unit was highlighted as a “zero translation,” resulting from the process of deletion.

The combination of a UoU label, the source term and its translation equivalent is called a “translation unit.” Examples of translation units derived from one of the bitexts in our corpus are shown in Table 6.

TABLE 6  
Examples of translation units

Unit of Understanding	Galician lemma	Galician occurrence	English lemma	English occurrence
AXENTE_POLUCIONANTE	axente polucionante	axente polucionante	polluting agent	polluting agents
AXENTE_POLUCIONANTE	axente contaminante	axentes contaminantes	pollutant	pollutants
AXENTE_POLUCIONANTE	componente tóxica	componentes tóxicas	toxic component	toxic components

### 3.4 Analysis of terminological variation

From a list of 629 terms referring to 484 UoUs, only those UoUs appearing more than one time in any of the texts were retained because our analysis only focused on UoUs that were characterized by terminological variation in the source texts. Translation units were automatically extracted from the annotated corpora and copied to a table format. The format is illustrated by means of Table 7 in which the examples were taken from the ‘DC’ text. Each translation unit in the table was further complemented with a specification of the lemmas in both source and target languages. Apart from that, it was also computed how often a particular translation unit appeared in each text.

TABLE 7  
Example of output data

UoU label	Galician lemma	Galician occurrence	English lemma	English occurrence	Frequency
ESPECIES_COMERCIAIS	especies comerciais	especies comerciais	affected species	affected species	1
ESPECIES_COMERCIAIS	especies comerciais	especies comerciais	market resource	market resources	1

The results in the table were then used to carry out three types of analyses. In a first quantitative analysis we examined terminological variation appearing in the source texts. The aim of this analysis was to verify how many UoUs in the source texts were characterized by terminological variation. In order to examine whether and how this variation was also reflected in the specialized translations, two additional analyses were carried out: on the one hand, a quantitative analysis and comparison of the number of expressions (for each UoU) found in both source and target texts and, on the other hand, a qualitative cognitive analysis of the translation units.

The cognitive analysis presented in section 2.2 allowed us to compare the conceptual information between source terms and translations. An example of such a comparison is shown in Table 8. Note that the columns are arranged according to the position of heads and modifiers in Galician, resulting in an inversion of elements in the English adjective + noun compounds. The example shows that in two out of three translations the conceptual pattern was reflected literally, which means that the translator respected the cognitive motivation of the author of the source text. The source term *axente contaminante* and its translation *pollutant* differ in terms of conceptual information: whereas in the source term the activity is expressed in the modifier and the head emphasizes the active role of the substance, the translation names the substance after the activity, leaving out its role of agent.

TABLE 8  
Comparing the cognitive information

Terms (SL)	cognitive pattern [HEAD + MODIFIER]	Terms (TL)	Cognitive pattern [HEAD + MODIFIER]
axente polucionante	[SUBSTANCE (BY ACTION) + ACTIVITY]	polluting agent	[SUBSTANCE (BY ACTION) + ACTIVITY]
axente contaminante	[SUBSTANCE (BY ACTION) + ACTIVITY]	pollutant	SUBSTANCE (BY ACTIVITY)
componente tóxica	[SUBSTANCE (BY STRUCTURE) + QUALITY]	toxic component	[SUBSTANCE (BY STRUCTURE) + QUALITY]

Based on this type of comparison, we were able to manually assign a value to each translation unit in order to qualify the cognitive distance (i.e., the differences in conceptual information) between the source term and its translation. A value of '0' indicates 'no cognitive distance' in the sense that there is no difference in conceptual information between the source term (e.g., *axente polucionante*) and its translation equivalent (e.g., *polluting agent*). A value of '0.5' means a 'partial cognitive distance.' This value is assigned to a translation unit if there is a partial overlap in the conceptual information between the source term and its translation. The example of *axente contaminante* and its translation *pollutant*, for instance, would be qualified as such. When a translation unit is qualified as '1,' it means that on the basis of the conceptual information there is a minimum degree of correspondence between the source term and its translation. An example would be *axente económico* (*economic agent*) which was translated in the English 'DC' text as *fishermen*. In this case, the only similarity is that concepts in the head position belong to the same broad category of humans (animate entities), but the distinctive feature as well as the activity related to the professional is not coincident.

Based on the cognitive distances of the translation units and the frequency of the translation unit in each bitext, we computed the interlingual variation index (IVI). The IVI is a measure between '0' and '1' that indicates how a UoU in the source text is transferred to the target text. If this measure is close to '0' it means that overall a more direct or literal translation was adopted with respect to the translation of a particular UoU. An IVI measure which is closer to '1' indicates a more free translation. The IVI measure is an average. It is the result of the sum of the weighted cognitive distances (i.e., the cognitive distance of a translation unit multiplied by the frequency of the translation unit) which is then divided by the total number of translation units for each UoU. This is illustrated by means of the example in Table 9.

TABLE 9  
Calculating the Interlingual variation index (IVI)

UoU	Lemma (ST)	Lemma (TT)	Frequency	Cognitive Distance	Weighted Cognitive Distance
AXENTE_ POLUCIONANTE	axente polucionante	polluting agent	1	0	0
	axente contaminante	pollutant	1	0,5	0,5
	componente tóxica	toxic component	1	0	0
<b>Interlingual variation index (IVI)</b>					0.17

The IVI allows us to assess the cognitive consequences of the equivalent choices and better describe the differences between source texts and translations. This measure does not only allow us to see whether the variation pattern for each UoU in the source text is reflected in the translation. It also takes into consideration to what extent the translator respected the cognitive point of view reflected in the author's choice of terminology in the source text.

#### 4. Results

The results presented in this section are only based on those UoUs that appear more than once in each text (section 3.4). Some of these UoUs were referred to by one single term only, while others were characterized by multiple denominations. Since UoUs appearing only once were excluded from our analysis, only 41 of the 59 UoUs that we found in the 'DC' text were retained (i.e., 69.5%). In the 'GN' text 48 out of 74 UoUs were considered relevant (i.e., 64.9%) and, in the 'SG' text, 40 of the 53 UoUs (i.e., 75.5%) were retained.

##### 4.1 Terminological variation in the source texts

In this section we examine how much terminological variation is encountered in the source texts. The results of the quantitative analysis are shown in Table 10. Category 'All' refers to all UoUs appearing at least two times in the same source text. Category '= 1' is a subset of Category 'All' and represents the number of UoUs that are characterized by only one term. Category '> 1' shows the subset of UoUs that are characterized by more than one term in the source text.

TABLE 10  
UoUs appearing more than once in the texts

	Text occurrences			Lemmas		
	All	= 1	> 1	All	= 1	> 1
DC	41	19	22	41	20	21
GN	48	14	34	48	15	33
SG	40	9	31	40	10	30

We can derive from the results in Table 10 that 22 UoUs out of the 41 appearing more than once in the 'DC' text are characterized by more than one term (cf. '> 1').

This corresponds to roughly 53.7%. After lemmatizing the terms, this number corresponds to 51.2%. In the case of the ‘GN’ text, we find that 70.8% of the UoUs are characterized by more than one term (68.8% after lemmatization). Finally, in the ‘SG’ text this amount corresponds to 80% (75% after lemmatization). These results show that the ‘DC’ source text is characterized by less terminological variation than the other two texts.

The small differences in numbers between terminological variation based on text occurrences and lemmas show that morpho-syntactic variation (i.e., singular vs. plural word forms) does not seem to be the primary cause of variation, either in the Galician source texts, or in the English translations. Most of the UoUs that are characterized by multiple denominations consist of terms that partly or completely differ from one another in terms of surface realization. We find UoUs that have been named by a more general term and one of its hyponyms. The UoU //ECOSISTEMA// for instance is characterized by *ecosistema* (*ecosystem*) and the more specific term *ecosistema mariño* (*marine ecosystem*) in the ‘DC’ text. In other cases, we find UoUs in which one constituent element of a term remains while the other one varies. Examples are *efecto da catástrofe* (*effect of the catastrophe*) and *efecto do derramo* (*effect of the spill*) in the ‘SG’ text or *zona afectada* (*affected zone*) and *zona polucionada* (*polluted zone*) in the ‘GN’ text. In all other cases, the entire term is replaced by another, either by a linguistic synonym or by a term that is used as such in the specific context. This is for instance the case in the UoU //ACTIVIDADE\_PESQUEIRA// which is characterized by the following lemmas in the ‘GN’ text: *actividade pesqueira* (*fishing activity*), *captura* (*catch*), *explotación* (*exploitation*) and *pesca* (*fishing*).

#### 4.2 Terminological variation in source texts and translations

A comparison of intralingual variation in source and target texts is based on the results in Table 11. Note that the results of category ‘> 1’ (Table 10) are taken as a starting-point for comparison because we decided to focus only on UoUs that were characterized by multiple denominations in the source text.

TABLE 11

A comparison of variation in source and target texts

	Lemmas (> 1)			
	UoU	Ga < En	Ga = En	Ga > En
DC	21	6	12	3
GN	33	6	18	9
SG	30	8	15	7

In this table, UoUs having more unique source terms as compared to unique translations were classified as ‘Ga > En.’ If more translations were found in the target text, the UoU was classified as ‘Ga < En.’ Finally, a UoU was classified as ‘Ga = En’ if the same number of unique expressions were found in the source and target texts.

We derive from the results in Table 11 that most UoUs tend to have the same number of lexicalizations in source and target texts: resp. 57.1% (‘DC’ text), 54.5% (‘GN’ text) and 50% (‘SG’ text). This shows that, overall, the variation occurring in the source texts is also present in the target texts. The lower number of UoUs that

fall under the ‘Ga > En’ category in the ‘DC’ text (14.3%) when compared to the ‘GN’ text (27.3%) and the ‘SG’ text (23.3%) may imply that the translation of terms in this text seems to deviate more from its source as compared to the other two translations. This should be further explored in the qualitative analysis (section 5.3).

For some UoUs having more variants in Galician than in English (‘Ga > En’), situations of graphic variation are attested. For instance, in the ‘DC’ text we encountered the terms *producción mexilloeira* and *produzón mexilloeira* which were both translated as *mussel production*. The existence of a higher amount of spelling variants in Galician texts than in English texts can be due to the fact that the use of Galician as a language for specialized communication is not widespread, and therefore, the experts writing in Galician may not be familiar with the correct spelling of some terms in their field. Moreover, the orthographic variation in the ‘DC’ text is due to the presence of two orthographic norms for the Galician Language, the Reintegrationist Norm and the Official Norm.<sup>4</sup>

Some UoUs having more English variants than the Galician source texts (‘En > Ga’) pertain to realities that are specific to the Galician fishing sector, and might not have a direct and full equivalent in other languages (Table 12). For example, the notion //MARISQUEO// which concerns the activity of shellfish harvesting, is a very important activity in Galicia and it is legally considered as a modality of fishing. In other countries, however, this activity has been abandoned and shellfish is only produced by cultivation, therefore considered as a modality of aquaculture. This explains the variety of expressions found in target texts for translating the activity //ACTIVIDADE\_MARISQUEIRA//, the sector //SECTOR\_MARISQUEIRO// or the professional //MARISCADOR//. In the UoU //ACTIVIDADE\_PESQUEIRA// [//fishing activity//] we also observe the strategy used by the translator to include shellfish harvesting in the target text. Other culturally-rooted notions of inaccurate equivalence are the mussel culture sector //SECTOR\_MEXILLOEIRO// or the specific markets where fish is sold on the auction //LONXA//.

TABLE 12

**Translation of culture-specific UoUs in the ‘Ga > En’ category**

UoU	Occurrences Galician texts	Occurrences English texts	Text
ACTIVIDADE_MARISQUEIRA	marisqueo	shellfish collecting, shellfish-collecting, shellfish-collecting activity	DC
ACTIVIDADE_MARISQUEIRA	actividade marisqueira, marisqueo	shellfish harvesting activity, harvesting, shellfish harvesting	GN
SECTOR_MARISQUEIRO	marisqueo	shellfish collecting, shellfish-collecting	DC
SECTOR_MARISQUEIRO	marisqueo, subsector de marisqueo	harvesting, shellfish harvesting, subsector of shellfish harvesting	GN
MARISCADOR	mariscadores, produtores	persons engaged in mariculture, shellfish harvesters, producers	SG
ACTIVIDADE_PESQUEIRA	actividade pesqueira, captura, explotación, pesca	fishing and shellfish harvesting activity, capture, catch, fishing or harvesting, harvesting, fishing	GN
SECTOR_MEXILLOEIRO	sector mexilloeiro, subsector de miticultura	mussel culture sector, the mussel harvesting sector, subsectors of mussel culture	GN
LONXA	lonxa	market, auction market	DC

Some cases of dissymmetry between the number of source terms and translation equivalents are explained by regular processes of morpho-syntactic variation existing in each language. They are employed mainly for stylistic reasons or writing preferences. Therefore, the translator might feel free to reduce or provoke this type of variation. The following table shows some of these situations:

TABLE 13  
Morpho-syntactic variation in source and target texts

UoU	Occurrences Galician texts	Occurrences English texts	Text
CRIA	cria	seeding, seeds	GN
MADURACIÓN	maduración	maturation, maturing	GN
SECTOR_DA_PESCA	pesca	fisheries, fishing	GN
ZONA_DE_PROHIBICIÓN	zonas de prohibición, zonas prohibidas	prohibited zones	SG
PRODUCCIÓN_DE_MEXILLÓN	producción de mexillón, produción do mexillón	mussel production	SG

Other dissymmetric uses are due to the employment of full synonyms in the source or target language (Table 14). A comparison of these terms in each language does not show any changes with respect to the conceptual information (section 4.4).

TABLE 14  
Full synonyms in source and target texts

UoU	Occurrences Galician texts	Occurrences English texts	Text
ACIDENTE	catástrofe	catastrophe, disaster	DC
CRU	fuel	fuel, oil	DC
SOBRECAPTURA	sobrecaptura	excess catch, over-catch	GN
ZONA_AFECTADA	zona afectada, zona polucionada	affected area, affected zone, zone affected, area polluted	GN
EMPREGO	emprego	employment, jobs	SG
FORTE_CONTAMINACIÓN	forte contaminación	heavy pollution, massive pollution	SG
ZONA_AFECTADA	zonas afectadas	affected zones, areas affected	SG

#### 4.3. Analysis of the cognitive distance and interlingual variation index

In Table 15 we only show the average IVI for each text.

TABLE 15  
Results of the Interlingual Variation Index (IVI)

	UoU	Cognitive distance '0'		Cognitive distance '0.5'		Cognitive distance '1'		Total nº of translation units	IVI
		Nº	%	Nº	%	Nº	%		
DC	21	92	62,1	22	14,9	34	22,97	148	0.304
GN	33	303	83,24	46	12,64	15	4,12	364	0.104
SG	30	187	85,78	22	10,09	9	4,13	218	0.092

As expected, the proportion of equivalents scoring 0 is higher in the three texts, which confirms our hypothesis that translations tend to be consistent and reflect as close as possible the term choices in the source texts. Similarly, the number of equivalents with a cognitive distance of 0.5, corresponding to translations in which the cognitive information partly overlaps with that of the source term, is also quite homogeneous among the three texts. However, the proportion of variants differing notably from the original text is remarkably high in the 'DC' text. As a matter of fact, its IVI is three times higher as compared to the other two texts. This clearly shows that the 'DC' text tends to deviate more from the original source text. The translator of the 'DC' text adopted a more free translation style as compared to the translators of the other two texts. This is for instance shown in the following example:

TABLE 16

**Example of free translation in the 'DC' text**

UoU	Occurrences Galician texts	Occurrences English texts	Freq	Cognitive distance
EFECTO_DA_CATASTROFE	efectos da marea negra	oil spill effects	1	0
	efectos do vertido	effects of the disaster	1	0,5
	efectos do vertido	spill	1	0,5
	impacto da marea negra	ecological catastrophe	1	1
	impacto da marea negra	oil spill	1	0,5

We observe that three source terms have been translated by five different expressions. Moreover, the recurrent cognitive pattern in the source terms [RELATION OF CONSEQUENCE + CHANGE OF STATE] is only respected in the first occurrence. The terminological choice in the translation shows a clear shift in perspective. The translator emphasizes the dramatic ecological consequences of the incident: the source term *impacto da marea negra* (*oil spill effects*) is translated as *ecological catastrophe*.

## 5. Discussion

In general, the results observed in the previous section confirm our hypothesis about the use of terminological variation in the specialized texts and translations. The results in section 4.1. confirm that variation is a typical phenomenon of specialized communication. The degree of variation and the types of variants might vary according to the author's preferences or style, as the percentage of UoUs having more than one variant differs from text to text. But in general, it seems that terminological variation is commonly used by experts.

Concerning the use of variation in translations, the quantitative results in section 4.2 show that the UoUs that are characterized by terminological variation in the source texts, are also characterized by variation in the target texts. For the majority of UoUs in each text, the same number of variants in source and target texts was encountered.

The results of the IVI (section 4.3), enable us to confirm our initial hypothesis that overall the cognitively motivated term choices in the source text are reflected in the translations. In the 'DC' text, however, the higher proportion of translation units in which the cognitive distance between the source term and its translation were

marked as '1' (22.97%) is remarkable when compared to the 'GN' text (4.12%) and the 'SG' text (4.13%). From these results we can derive that in the 'DC' text a much more free translation strategy was adopted, whereas for the other two texts the translation was much more direct or literal. The level of knowledge of the professional translators with respect to the topic addressed in the source texts, their knowledge of the target language, their familiarity with the terminology of the subject field, the availability of terminological resources during the translation process and the possible existence of a translation policy with respect to the translation of terminology into English are some of the extra-linguistic factors that need to be taken into consideration in order to account for the differences between the 'DC' text and the other two texts.

For instance, what we know about the translations in our corpus is that the 'SG' and 'GN' texts were translated by professional translators whereas the translation of 'DC' text was done by the author of the source text. This may explain why a more free translation strategy was adopted in the latter. Also the fact that the translations of the 'SG' and 'GN' texts appeared in the same proceedings of 2004 as the original source texts, may explain why a more direct translation strategy was applied by the professional translators. The translation of the 'DC' text appeared in a journal *International Journal of Oceans Affairs*, 2 years after the publication of the source text in the conference proceedings in 2007. In contrast to the proceedings which mainly targeted field experts, fishermen and businessmen, the readers of the international journal are mainly marine economists and marine sciences experts.<sup>5</sup> This may also explain why a more free translation strategy was applied in the 'DC' text with respect to the translation of source terms.

## 6. Conclusion

We presented a methodology that allowed us to examine whether and how terminological variation in specialized source texts resulting from cognitively motivated term choices is reflected in translations. Essential in this methodology is that source terms having co-referential status are clustered and considered part of the same UoU (section 2.1). The UoU label that identifies each UoU or cluster of terminological variants is used to annotate source terms and their translation equivalents in the bilingual corpus. The annotation is carried out semi-automatically. This means that the program which was developed for this task (section 3.2) first asks feedback from a user before it automatically places the source terms and translations between the correct identification tags. This program will be further refined in CVC's research project (section 2.1) in order to speed up the analysis of terminological variation in future source and target texts.

A first quantitative analysis was carried out in order to identify the UoUs that were characterized by terminological variation in the source texts. A comparison between the number of source terms and translations for each UoU shows that the translators of the texts in our study did not follow the consistency rule put forward in prescriptive terminology. Based on the results of the cognitive distances between source terms and translations and the interlingual variation index for each UoU, we were able to conclude that the translations tend to be consistent and reflect as close as possible the term choices in the source texts. The fact that in one of the texts a more free translation strategy was applied to the translation of certain source terms



was explained on the basis of extra-linguistic factors related to the translation processes (section 5). This illustrates once more the importance in corpus-based studies of having information available related to the genesis of each text. On the one hand, it is important to know more about the people involved in the writing or translation processes: e.g., what is their native language and how familiar are they with the subject matter and terminology? On the other hand, it is also relevant to know more about the translation policy, the resources that were consulted during the writing or translation processes (e.g., terminology lists, existing documents, general or specialized dictionaries) or the tools that were used (e.g., translation memories, technical writing tools). Such information is usually not provided when corpora are being developed and it is often difficult to gather after the corpus is released. Nevertheless such information was felt crucial in this pilot study for understanding the particular choices with respect to source terms and translations.

The results that emerged from this study also support the idea that from the point of view of translation the notion of terminological equivalence may differ from the traditional, onomasiological point of view in which terminological equivalence is confined to terminological variants (synonyms or translation equivalents) that name the same concept. Translators apply different translation techniques in order to establish equivalence between a message in the source language and its translation into the target language. This sometimes results in translation units in which the source term and its translation reflect different conceptualizations. An example is for instance the Galician term *especies comerciais* [commercial species] which was found translated in the 'DC' text as *affected species*, *commercial stock* and *market resources*. It may be interesting to examine how the results of a study of terminological variation in source and target texts could further complement the information found in electronic specialized translation resources and in this way improve the quality of such resources. We believe this issue to be an interesting challenge for future corpus-based research on terminological variation in specialized translation (Kerremans forthcoming).

#### ACKNOWLEDGEMENTS

The authors would like to thank prof. Rita Temmerman (CVC, Erasmus University College Brussels) for her comments on an earlier draft of this article.

#### NOTES

1. <<http://cvc.ehb.be>>, visited on 13 May, 2011.
2. <<http://www.iula.upf.edu/iulaterm/tpresuk.htm>>, visited on 13 May, 2011.
3. More information about this tool can be found here: <<http://neon.niederlandistik.fu-berlin.de/en/textstat/>>, visited on 13 May, 2011.
4. The Reintegrationist Norm claims the common origin of Galician and Portuguese, and adopts some orthographic conventions from Portuguese language. The Official Norm, created in 1982 by the Royal Academy of Galician language in the framework of the Linguistic Normalisation Plan follows the orthographic conventions of the Spanish Language.
5. The target audience is explicitly stated in the presentation page of the journal: <[http://www.elsevier.com/wps/find/journaldescription.cws\\_home/30453/description#description](http://www.elsevier.com/wps/find/journaldescription.cws_home/30453/description#description)>, visited 13 May, 2011.

## REFERENCES

- BOWKER, Lynne (1997): You say “flatbed colour scanner,” I say “colour flatbed scanner”: A descriptive study of the influence of multidimensionality on term formation and use with special reference to the subject field of optical scanning technology. *Terminology*. 4(2):275-302.
- BOWKER, Lynne (1998): Variant terminology: frivolity or necessity? *In: 8th EURALEX International Congress*. Liège: University of Liège, 487-495.
- BOWKER, Lynne and HAWKINS, Shane (2006): Variation in the organisation of medical terms: Exploring some motivations for term choice. *Terminology*. 12 (1):79-110.
- CABRÉ, Maria Teresa (1995): On diversity and terminology. *Terminology*. 2(1):1-16.
- CABRÉ, Maria Teresa (2008): El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en Terminología (I). *Ibérica*. 16:9-36.
- COLLET, Tanja. (2004): Esquisse d’une nouvelle microstructure de dictionnaire spécialisé reflétant la variation en discours du terme syntagmatique. *Meta*. 49(2):247-263.
- DURIEUX, Christine (1995): Texte, contexte, hypertexte. *Cahier du CIEL*. (1994-1995):214-228.
- FELBER, Helmut (1981): The Vienna school of terminology, fundamentals and its theory. *Infoterm Series*. 7:69-86.
- FERNÁNDEZ-SILVA, Sabela, FREIXA, Judit and CABRÉ, Maria Teresa (2009): Multiple motivations in the denomination of concepts: the case of “production area” in the terminology of aquaculture in French and Galician. *Terminology Science and Research*. 20. Visited 13 May, 2011, < <http://lipas.uwasa.fi/hut/svenska/iitf/tsr2005/vol20/>>.
- FREIXA, Judit (2002): *La Variació terminològica: Anàlisi de la variació denominativa en textos de diferent grau d’especialització de l’àrea de medi ambient*. Barcelona: Institut Universitari de Lingüística Aplicada/ Universitat Pompeu Fabra.
- FREIXA, Judit (2006): Causes of denominative variation in terminology: A typology proposal. *Terminology*. 12(1): 51-77.
- FREIXA, Judit, FERNÁNDEZ-SILVA, Sabela and CABRÉ, Maria Teresa (2008): La multiplicité des chemins dénominatifs. *Meta*. 53(4):731-747.
- GAUDIN, François 2003. *Socioterminologie: une approche sociolinguistique de la terminologie*. Bruxelles: De Boek Duculot.
- GUIRAUD, Pierre (1978): *Les mots savants*. Paris: Presses Universitaires de France.
- KAGEURA, Kyo (2002): *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Amsterdam/Philadelphia: John Benjamins.
- KERREMANS, Koen (2010): A Comparative Study of Terminological Variation in Specialised Translation. *In: Carmen HEINE and Jan ENGBERG, eds. Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*. Aarhus: Aarhus School of Business - Aarhus University.
- KOCOUREK, Rostislav (1991): *La langue française de la technique et de la science: vers une linguistique de la langue savante*. Weisbaden: Oscar Brandstetter.
- MERKEL, Magnus (1996): Consistency and Variation in Technical Translations – a Study of Translators’ Attitudes. *In: Proceedings from Unity in Diversity, Translation Studies Conference*. Dublin. Visited 13 May, 2011. <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.2145&rep=rep1&type=pdf>>.
- MYKING, Johan (2009): (Yet another) taxonomy of motivations. *Terminology Science and Research*. 20. Visited 13 May, 2011, <<http://lipas.uwasa.fi/hut/svenska/iitf/tsr2005/vol20/>>.
- ROGERS, Margaret (2004): Multidimensionality in concepts systems: A bilingual textual perspective. *Terminology*. 10(2):215-240.
- ROGERS, Margaret (2007): Lexical chains in technical translation. A case study in indeterminacy. *In: Bassegy ANTIA, ed. Indeterminacy in Terminology and LSP*. Amsterdam/Philadelphia: John Benjamins.
- SAGER, John (1997): Term Formation. *In: Sue-Ellen WRIGHT and Gerhard BUDIN, eds. Handbook of Terminology Management*. Amsterdam/Philadelphia: John Benjamins, 25-41.

- SUÁREZ DE LA TORRE, Mercedes (2004): *Análisis contrastivo de la variación denominativa en textos especializados: del texto original al texto meta*. Barcelona: Institut Universitari de Lingüística Aplicada/Universitat Pompeu Fabra.
- TEMMERMAN, Rita. (2000): *Towards New Ways of Terminology Description: The Sociocognitive-Approach*. Amsterdam/Philadelphia: John Benjamins.
- TOURY, Gideon (1995): *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.
- WÜSTER, Eugen (1979/1991): Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie. In: Leopold BAUER, ed. *Schriftenreihe der technischen Universität Wien*. Vienna/New York: Springer.

## APPENDICES

### Appendix 1. Corpus references

#### Galician Texts:

- (SG) SURIS-REGUEIRO, Juan Carlos, GARZA-GIL, Dolores (2004): Avaliación de danos directos e indirectos. Metodoloxía e plan de traballo para o caso do Prestige. In: Albino Prada BLANCO and María Xosé Vázquez RODRÍGUEZ, eds. *Efectos económicos, sociais e ambientais da marea negra do Prestige*. Madrid: Consello da cultura galega, 311-351.
- (GN) GARCÍA NEGRO, María do Carme (2004): Consideración sobre o estudo dos danos económicos nos sectores productivos directamente vinculados ás actividades mariñas. O caso do Aegean Sea. In: Albino Prada BLANCO and María Xosé Vázquez RODRÍGUEZ, eds. *Efectos económicos, sociais e ambientais da marea negra do Prestige*. Madrid: Consello da cultura galega, 354-419.
- (DC) VILLASANTE, Carlos Sebastián, GARCÍA NEGRO, María do Carme, CARBALLO PENELA, Adolfo, et al. (2007): Sobre a valoración económica dos efectos do vertido do Prestige na pesca comercial na Costa da Morte. In: *X Foro dos Recursos Mariños e da Acuicultura das Rías Galegas e I Foro Iberoamericano dos Recursos Mariños e da Acuicultura*, 665-682.

#### English texts:

- SURIS-REGUEIRO, Juan Carlos, GARZA-GIL, Dolores (2004): Evaluation of direct and indirect damages. Methodology and work program for the Prestige case. In: Albino Prada BLANCO and María Xosé Vázquez RODRÍGUEZ, eds. *Economic, social and environmental effects of the "Prestige" spill*. Madrid: Consello da cultura galega, 311-351.
- GARCÍA NEGRO, María do Carme (2004): Considerations on the study of economic damage in productive sectors directly linked with marine activities. The case of the Aegean Sea. In: Albino Prada BLANCO and María Xosé Vázquez RODRÍGUEZ, eds. *Economic, social and environmental effects of the "Prestige" spill*. Madrid: Consello da cultura galega, 354-419.
- GARCÍA NEGRO, María do Carme, VILLASANTE, Carlos Sebastián, CARBALLO PENELA, Adolfo, et al. (2009): Estimating the economic impact of the Prestige oil spill on the Death Coast (NW Spain) fisheries. *Marine Policy. The International Journal of Oceans Affairs*. 33:8-23.