*Koen Kerremans**

A Comparative Study of Terminological Variation in Specialised Translation

Abstract

This article presents an ongoing project in which terminological variation is studied in the context of specialised translation. The project aims to find out whether certain patterns or tendencies can be derived from a comparative analysis of terminological variation in source texts and their translations. New insights about terminological variation in specialised translations could contribute to the development of a new type of specialised translation dictionary which will better account for the different ways in which a given thought is expressed. In this article, I will present the methodology that was set up in this project to carry out the comparative analysis. Currently, the analysis is carried out on English, Dutch and French texts related to biodiversity issues (including climate change, air pollution, invasive species, etc.).

1. Introduction

The view that terms should be used unambiguously to refer to clearly defined concepts (see e.g. Wüster 1979 or Felber 1981) has dominated terminology research for quite some years. To a large extent, it has determined what information should be taken up in specialised dictionaries and how this information should ideally be presented to the dictionary users.

A general observation is that specialised dictionaries are mainly focused on providing users an understanding of the way a subject field is organised by presenting information about the concepts that belong to the subject field (see e.g. Pearson 1998, Collet 2004a). The concept – which is the starting-point of the terminological analysis – is identified by means of an original term and its meaning is explained in a definition. According to the traditional, prescriptive view in terminology, alternative ways of expressing the concept (for instance by means of lexical variants) should be restricted because it is believed that these alternative expressions may hamper the exchange of information in specialised communicative settings.

This article presents a research project which looks at specialised dictionaries from the perspective of the translator, one of the most important users of these dictionaries (Collet 2004a). Translators who need to translate a domain-specific text, consult specialised dictionaries to acquire a better understanding of particular concepts or the subject field, to familiarise themselves with the terminology and to look up possible translation equivalents of terms they encountered in their source text. Previous studies have pointed out that specialised dictionaries only partially meet the requirements of translators in this respect because they very often lack a specification of how a term really 'behaves' in specialised discourse (e.g. Collet 2004a, Gerzymisch-Arbogast 2008). Questions such as 'are there any collocational restrictions?' or 'what term fits best in the context of my translation?' remain very often unanswered. One explanation for this is that in specialised dictionaries it is very often neglected that besides their referential function (i.e. they point to domain-specific concepts), terms fulfill additional roles in specialised discourse. Seen from a textual perspective, for instance, Collet (2004b) pointed out to the fact that a term also contributes to text coherence. This explains why in specialised texts, authors sometimes deviate from the tra-

* *Koen Kerremans*
Centrum voor Vaktaal en Communicatie
Department of Applied Linguistics
Erasmushogeschool Brussel
Belgium
koen.kerremans@ehb.be

ditional view of using only one term consistently throughout the text to refer to a particular concept. From a communicative point of view, for instance, terms are used in specialised discourse to transfer thoughts from sender to receiver. “Each sender adapts his expression to the specific communication situation” (Cabré 1995: 8). This explains why in certain communicative contexts a lexical variant is sometimes preferred over a term that appears in a specialised dictionary. For a translator, who works within the reality of specialised discourse, it is important to know what source and target language terms are commonly used to address a concept in a specific communicative setting. Such information may be further complemented with a specification of how source language terms were actually translated in real discourse.

One of the aims of the project discussed in this article is to find out how a representation of the different ways in which a term is translated in actual discourse can be linked to the traditional way of structuring specialised dictionaries. This will be examined on the basis of a study of terminological variation in specialised translations. The project examines whether and how terminological variation occurring in source texts is reflected in specialised translations. It aims to find out whether certain patterns or tendencies can be derived from a comparative analysis of terminological variation between source and target texts. Currently, the analysis is carried out on English, Dutch and French texts related to biodiversity issues (e.g. climate change, air pollution, invasive species, etc.). New insights about terminological variation in specialised translations will hopefully contribute to the development of a new type of specialised translation dictionary which better reflects actual use of terminology in specialised source texts and translations.

In this article, I will present the methodology that is worked out in this project to compare the use of terminological variation in source and target texts. Before this methodology is presented, I will first explain how the object of study – i.e. terminological variation – is defined in the scope of this project. As terminological variation is studied in the context of translation, it is perceived from a textual and a contrastive perspective (see section 2). In section 3, I will provide a brief survey of related studies that have contributed to defining some of the research targets set forth in this project. Section 4 will be devoted to the presentation of the project’s methodology. Finally, in section 5, I conclude.

2. Object of study

This section looks at the notion of terminological variation. As was stated in the introduction, terminological variation is studied from textual and contrastive perspectives in this project.

Seen from a textual perspective, the unit of understanding is the starting-point in the analysis of terminological variation in a source text. The notion of unit of understanding (UoU) was introduced in sociocognitive terminology theory in order to clarify the inadequacy of classical concept theory for the conceptual structure of most specialised fields (Temmerman 2000). As in traditional terminology theory a ‘concept’ is by definition clearly delineated and preferably referred to by one term only, we will avoid the term ‘concept’ in the remainder of this article and instead refer to units of understanding (at the highest level of abstraction) which allow for prototypically structured understanding, i.e. fuzziness and multiple ways of lexicalisation.

The textual perspective allows me to search for the different lexical expressions in the source text that refer to the same unit of understanding. These terminological variants are said to have co-referential status. Terms having co-referential status can be found on the basis of an analysis of lexical chains. Rogers (2007: 17) defines a lexical chain as consisting of cohesive ties sharing the same referent, lexically rather than grammatically expressed” (2007: 17). This definition indicates that not all expressions that refer to a particular unit of understanding are considered relevant. This is illustrated by means of 2 text samples in which only the relevant data have been

marked in bold. The first text sample is taken from a working document of the European Commission about invasive species[1]:

- "**Invasive Alien Species**" are alien species whose introduction and/or spread threaten biological diversity. […]"
- **Invasive species** (**IS**) negatively affect biodiversity. […]"
- The costs of control, although lower than the costs of continued damage by the **invader**, are often high. […]"

In the second text sample[2], the following items have been marked:

- In some instances these **new arrivals** are so successful that they are no longer a biological curiosity but a real threat, causing serious damage […]"
- Non-native species that have such a negative impact are known as **Invasive Species** or **IS**. […]"

All sentences in the samples refer to a unit of understanding labelled in English as 'invasive alien species'. Invasive alien species are alien species that enter a new habitat, manage to thrive there until they become a real threat for the endemic fauna and flora. In the first sentence of the first text sample, the term 'invasive alien species' refers to the unit of understanding and is therefore underlined. The longer string 'alien species whose introduction and/or spread threaten biological diversity' also refers to the unit of understanding but is disregarded because it is a description. For the same reason the longer string 'Non-native species that have such a negative impact' in the last sentence of the second text sample is not considered either. Note that the hypernyms in both these descriptions – i.e. 'alien species' and 'non-native species' – have not been underlined because they clearly refer to another unit of understanding in both these contexts. We may find contexts where the two hypernyms act as co-referents of 'invasive alien species'. Only in those cases would they be considered relevant data.

In the second sentence of the first text sample, both terms 'invasive species' and the abbreviated form 'IS' are considered relevant. The same holds for the term 'invader' in the third sentence.

In the first sentence of the second sample, the unit of understanding is addressed as 'new arrivals' and 'they'. Since the latter is not a lexical expression but a grammatical one, it is not considered part of the lexical chain (cf. supra).

The terminological variants highlighted in the text samples above, are assigned the unique label of the unit of understanding (i.e. 'NON_NATIVE_INVASIVE_SPECIES'). This will be further explained in section 4. The results of the analysis from the previous two text samples are shown in Table 1.

UNIT OF UNDERSTANDING	ENGLISH TERMS
NON_NATIVE_INVASIVE_SPECIES	• invader • invasive alien species • invasive species • IS • new arrivals

Table 1. Terminological variants linked to 'NON_NATIVE_INVASIVE_SPECIES'

1 The reference number of this document is: SEC(2008) 2886.

2 The second example is taken from a communication of the European Commission. The reference number of this document is: COM(2008) 789 final.

The cluster of terminological variants may eventually consist of various types of terminological variation (e.g. orthographic variation, lexical variation, morphological variation, reduction, permutation, etc.). For a classification of denominative variants based on formal criteria, see e.g. Freixa (2002). Some studies deal with several of these subtypes (e.g. Depierre 2007, Carreño 2008) while others focus on just one subtype (e.g. Bowker 1997, Bowker/Hawkins 2006).

In the textual approach, only terminological variation in the source text is considered. The contrastive perspective allows me to examine the way terminological variation in the source text is reflected in the translation.

The starting point in this contrastive analysis is the unit of translation, i.e. a text segment which is treated by the translator as one single cognitive unit. Units of translation have been studied at the level of the text, paragraph, sentence, phrase or word (for a survey, see e.g. Zhu 1999). As the current project aims to study how the terminological variant is translated, the unit of translation is defined as the combination of a source language term – of which examples were shown in the previous paragraph – and its equivalent in the target text. This is shown in Table 2. The examples are taken from an English text about biodiversity and its Dutch translation. Both documents were published by Greenfacts[3].

UNIT OF UNDERSTANDING	UNIT OF TRANSLATION (EN-NL)	SOURCE
INVASIVE_SPECIES	invasive alien species – invasieve soorten	Greenfacts
INVASIVE_SPECIES	invasive alien pests – invasieve schadelijke soorten	Greenfacts

Table 2. Examples of units of translation (English-Dutch)

It should be noted that no restrictions are placed on the types of utterances extracted from the target language texts. If a term in the source language is for instance translated as a pronoun or a longer description, the equivalents are regarded as relevant data. The fact that a translator has translated a terminological variant as a personal pronoun could be a way for him/her to deal with variation occurring in the source text.

3. Research framework

As was stated in the introduction, this project starts from a general observation that specialised dictionaries for translators do not pay sufficient attention to the way terms behave in specialised discourse and, for instance, which term fits best in the specific translation context (Collet 2004a). According to Durieux (1995), translators should always take into account the following 3 contexts during the translation process: the linguistic context, the situational context and the cognitive context.

The linguistic context allows translators to pay attention to the words that often co-occur with a given term. For instance, the term can be part of a collocation so that its use is constrained.

In the situational context, the translator takes into account the cultural and psychological backgrounds of sender and receiver and tries to find the most acceptable translation. Within this situational context, the translator should also be aware of the fact that there are different types of specialised discourse, depending on the different backgrounds of sender and receiver. For instance, when specialists communicate with one another about a specialised topic in their field of expertise, they will probably use a different set of terms as compared to situations when they communi-

3 Greenfacts was a Belgian non-profit organization, created in December 2001 by individuals from different backgrounds, with the aim to bring complex scientific consensus reports on health and the environment to the reach of non-specialists (source: http://about.greenfacts.org/organization/index.htm). The bilingual text sample was taken from the following documents: the English version Scientific Facts on Biodiversity: A Global Outlook" and the Dutch version 'Wetenschappelijke Feiten over Biodiversiteit: een Wereldwijd perspectief").

cate about this topic with non-specialists (cf. Cabré 1995, Freixa 2006). Translators have to take this fact into consideration when they look for suitable translation candidates.

Finally, the cognitive context also determines the choice for a suitable translation. The cognitive context is defined by Durieux as the "stock mnésique qui se constitue au cours de l'assimilation du sens d' un discours ou d'un texte" (1995: 217). Before he starts translating, the translator needs to verify how a specific thought is constructed in the source text and what terms were introduced in the source text to support that process. Afterwards, the translator must examine how he can best present this thought in the translation and what translation equivalents he can use for this.

If developers of specialised dictionaries want to better meet the needs of specialised translators, they must start from the reality of specialised discourse (cf. section 1). This means looking at the way terms in source and target languages behave in different contexts: "Term descriptions on a systems level (as terminology in Wüster's sense) need to be complemented by contextual data of their possible and likely contaminations or frequency of 'constellations' and would then be of great practical help for LSP translators in all phases of the translation process […]"(Gerzymisch-Arbogast 2008: 25). Such information may be further complemented with a specification of how source language terms were actually translated in real discourse (cf. section 1).

This research is first and foremost motivated by the fact that terminological variation is a widespread phenomenon in specialised discourse (Freixa 2002). For that reason, it makes much sense to examine how this variation could ideally be presented in the specialised dictionary. Apart from that, it is also important to identify and categorise the different problems that it causes in specialised translation (Suárez de la Torre 2004). In this project, it is assumed that specialised translations tend to reflect the terminological variation appearing in source texts. The different translation options for a given source language term should therefore be presented in specialised translation dictionaries.

The claim that terminological variation appearing in source texts is reflected in specialised translations, is based on studies of universal features of translation (e.g. Blum-Kulka 1986, Baker 1993, Toury 1995; Laviosa 1998, Tirkkonen-Condit 2002; Mauranen 2004). For instance, 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 on the syntactic level but also on the lexical level: i.e. translators propose terminological equivalents in the target text based on term choices that were made in the source text. To understand how the variation is transferred to the target text, it is important to raise additional research questions such as: What parameters or dimensions tend to correlate with a higher degree of terminological variation encountered in source and target texts? What units of understanding tend to have more terminological variants? etc.

Furthermore, it is assumed that due to the linguistic, situational and cognitive contexts (cf. supra) that translators need to take into account when translating a text, it is not always possible for them to translate a source language term consistently, i.e. by means of the same translation equivalent. The different translation possibilities should therefore be displayed in specialised dictionaries. Given the close intertextual relation between a source text and its translation, it would seem reasonable to expect that a set of source language terms that designate a common object or conceptual structure is replaced by a set of conceptually equivalent target language terms. However, Rogers (2004) showed that this is not necessarily so. She observes that the translation of terms is linguistically more creative than often assumed and argues that this may be due to the polysemy and synonymy of terms which establish complex relations between forms and meanings. She also observes that phraseological contexts in which terms are embedded may vary cross-linguistically and that these contexts may have a perspectivising role for the unit of understanding. Based on these insights, Rogers concludes that prescriptive "measures which aim to establish one-to-one relations across languages may in some cases result in a loss of functionality in the textual medium (2004: 233).

A study of the different ways in which source language terms are represented in the target text will contribute to a better understanding of what translators consider to be 'equivalent'. Interlingual variation – i.e. variation resulting from the different ways in which a source language utterance was translated – can be studied in a framework of equivalence. Choices made by translators with respect to the translation of specific units can be motivated by the need to establish equivalence between source and target texts. This means that equivalence becomes an anchor to study terminological variation in the context of translation.

Equivalence in translation is much broader a notion than equivalence in terminography. The latter is restricted to equivalents referring to the same unit of understanding, whereas translators apply different techniques which allow them to search for equivalents that sometimes refer to other units of understanding as is shown in the sample below[4].

- (English) In the recent past, the rate and risk associated with alien species introductions have increased significantly, notably because of increased travel, trade and tourism. **[>ACTIVITY]**
- (French) Depuis peu, le rythme d'introduction d'espèces exotiques et les risques apparentés ont fortement augmenté, surtout à cause de l'augmentation des voyages, du commerce et du tourisme. **[>ACTIVITY]**
- (Dutch) In het recente verleden zijn de risico's die gepaard gaan met invasieve soorten sterk toegenomen, mede bepaald door de toename in reizen, handel en toerisme. **[>SPECIES]**

The English source talks about 'alien species introductions' which refers to a particular activity or event. This is also the case in the French translation where 'alien species introductions' has literally been translated as 'introduction d'espèces exotiques'. In the Dutch translation, however, the translator decided to change the perspective from alien species introductions to invasive species (which may be the result of an introduction) and translated alien species introductions as 'invasieve soorten' (i.e. 'invasive species').

What needs to be examined is how often translators make such a translation decision. The answer to this question will ultimately weigh on the decision whether or not to include these as valid translation options in the specialised translation dictionary. It can not be assumed that every translation option found in the final version of a translation – i.e. the version in which the translation gets published – is correct. Translating a source language term by means of a conceptually related term would probably not be the translator's first option. Nevertheless, a translator may have good reasons to do so. For instance, in an attempt to avoid too much repetition in the translation, he may decide to replace the most common translation equivalent of a source language term by means of an alternative expression. During the analysis of terminological variation in the specialised translation, it will be assumed that all translations are correct and should therefore be considered relevant. After all translations have been identified, a statistical analysis of the corpus may reveal which translations are the most common translation candidates of a given source language term and which ones are the isolated incidents.

By also considering those translation decisions that deviate from the ones that we would expect, this research project is also related to studies of translation shifts (e.g. Blum-Kulka 1986, Aubert 1997, Al-Zoubi/Al-Hassnawi 2001, Salkie 2001, Osimo 2008). Examples of such shifts are:

- semantic modification: i.e. a source language term is translated into a target language term which has a more specific or more general meaning.

4 The samples were taken from three documents published on the website of Greenfacts. The English document ("Scientific Facts on Biodiversity. A Global Outlook") is the original document. The Dutch ("Wetenschappelijke Feiten over Biodiversiteit. Een Wereldwijd Perspectief") and French ("Consensus Scientifique sur la Biodiversité. Perspectives mondiales") versions are the translations. The documents can be downloaded from the following website: http://www.greenfacts.org/en/global-biodiversity-outlook/

- addition: i.e. extra information is added to the term in the target language (e.g. a synonym, a definition, an explanatory note).
- deletion: i.e. certain information is left out in the translation.
- mutation: i.e. the source language term is replaced by a term with a different meaning.

The project intends to examine what translation shifts are most commonly applied by translators, taking into account several parameters such as subject field, text type, degree of specialisation, language or source. Based on Blum-Kulka (1986)'s explicitation hypothesis we would expect that addition frequently occurs in translations. Based on Tirkkonen-Condit (2004)'s proposed Unique Items Hypothesis (UIH), we could hypothesise that translators will not very often replace a source language term by a more specific term in the target language. The Unique Items Hypothesis states that target-language specific elements, which do not have equivalents in the source language, tend to be under-represented in translated texts, since "they do not readily suggest themselves as translation equivalents" (Tirkkonnen-Condit 2004: 177-178).

Summarising the first part of this article, it can be observed that the analysis of terminological variation starts from the unit of understanding. It is first examined how the unit of understanding is expressed in source texts based on the analysis of co-referents (see Section 2). Next, it is examined how the relevant source language expressions were translated into the target language, taking into account various parameters such as subject field, language, text source, etc. It is assumed that insights derived from this comparative analysis may contribute to the development of a new type of specialised translation dictionary that shows how source language terms were translated in actual translations. In the next section, the methodology will be presented that was set up to be able to perform the comparative study.

4. Research methodology

The research methodology combines both textual and contrastive perspectives in the analysis of terms, source language variants and translation equivalents. A comparative analysis of the patterns of variation is carried out on a multilingual corpus (English, Dutch and French) of specialised texts and their translations. The corpus currently consists of texts related to biodiversity issues (e.g. climate change, air pollution, invasive species, etc.). The search for terminology and variants started from a collection of glossaries and specialised texts related to these areas.

The aim of the workflow is to arrive at a stage where it becomes possible to compare the results of the analysis in source and target texts. In order to arrive at this stage, a number of intermediate steps need to be taken. These steps as well as the output data are visualised in Figure 1:

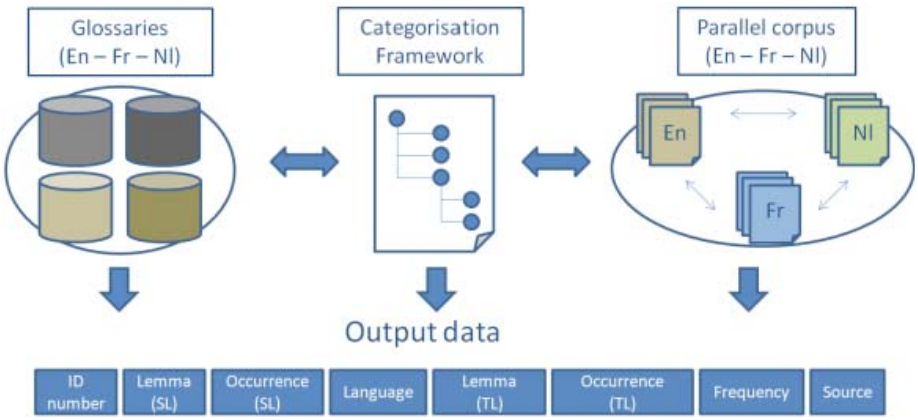

Figure 1. Methodological steps and output data

Existing glossaries are used as a starting-point for setting up a categorisation framework (cf. section 4.1). This framework is further extended with terms and variants that are manually extracted from the source texts in the parallel corpus (cf. sections 4.2 and 4.3). By means of the extended term list, all term occurrences in the source segments of the aligned corpus are automatically annotated. For each annotated unit in the source segments, a translation equivalent is manually highlighted in the corresponding target segment(s) (cf. section 4.4). The comparative analysis of terminological variation in source and target languages is based on the resulting annotated texts (cf. section 4.5). The methodological steps will be briefly discussed in the following subsections.

4.1. Analysis of glossaries

Before a terminological analysis was carried out on the texts in the corpus, a categorisation framework was first set up (Kerremans 2004). This framework lists the units of understanding that needed to be searched for in the corpus of source texts (cf. section 4.2). In order to set up such a framework for the biodiversity domain, several freely available biodiversity terminology lists were consulted. An important source was e.g. the terminology list taken from the website of the Belgian Clearing House Mechanism, which is maintained by the Royal Belgian Institute of Natural Sciences. This website provides information about the Convention on Biological Diversity (CBD)[5] and its implementation in Belgium. The glossary covers 576 English terms, 561 French terms and 88 Dutch terms related to biodiversity.

The glossary of the Belgian Clearing House Mechanism was further extended with (English) terminology from other sources and structured in a categorisation framework. After duplicates are removed, a unique label is assigned to each remaining term. Terms that refer to the same unit of understanding are preceded by the same unique label. This unique label represents the unit of understanding and is therefore called the 'UoU label'. The UoU label is used to cluster all terminological variants. For instance, the UoU label 'DOMESTIC_BIODIVERSITY' is used to classify the following terms:

- DOMESTIC_BIODIVERSITY – domestic biodiversity
- DOMESTIC_BIODIVERSITY – agricultural biodiversity
- DOMESTIC_BIODIVERSITY – agrobiodiversity
- DOMESTIC_BIODIVERSITY – farmland biodiversity

The resulting framework is used to automatically find and annotate exact matches in the source language documents of the parallel corpus (see section 4.2). Afterwards, the source texts were manually analysed and variants were added to the categorisation framework (section 4.3).

4.2. Corpus compilation

The corpus consists of texts related to biodiversity issues (e.g. air pollution, climate change, invasive species, etc.). It features both source texts and translations in Dutch, English and French. The source texts and translations are aligned on the level of the sentence wherever this is possible. All texts in the corpus are freely downloadable from the web.

Texts in the corpus are classified according to multiple parameters or dimensions. Several parameters should be taken into consideration when studying terminological variation in texts: e.g. the level of specialisation of the texts in the corpus, the text types or the subject field (see e.g. Gallardo 2000). Freixa (2002), for instance, postulated that denominative variation occurs more often in less specialised texts. She also claimed that depending on the level of specialisation of

5 The CBD is an international agreement signed by more than 150 government leaders at the Earth Summit, in Rio de Janeiro, in 1992. It was negotiated under the guidance of the United Nations. The main objective of the CBD is to identify general goals and policies with respect to the conservation of biological diversity.

the texts, different types of denominative variation and degrees of conceptual equivalence are distinguished.

According to Gerzymisch-Arbogast (2008), it is important to measure the correlations between the degree of terminological variation in a text and text external parameters such as the ones previously mentioned. The author believes that this will allow us to identify the parameters that should be prioritised. The author notes that "[l]ittle is known about the actual interplay of parameters on different levels (micro- and macro-levels) which govern how sense as a whole is established" (Gerzymisch-Arbogast 2008: 9).

A typical approach to classification generally takes the shape of a taxonomy of text-types. A problem of this type of classification is that the notion of text-type is "of such a wide applicability that it can subsume a bewildering range of text-form variants" (Emery 1991: 567). A second common type of classification – i.e. based on text function – is also problematic as several functions may be discerned in a text. Text classification may also be based on different communicative contexts (cf. Pearson 1998) or 'situations' such as province or domain (see e.g. Mason 1982). What Emery (1991) proposes is a model that integrates different perspectives in text classification.

The dimensions according to which texts have been classified so far in this project are visualised in Table 3. The first dimension is the domain. For each domain (e.g. air pollution, climate change, etc.) a different table is created. Next, texts are classified according to the language (i.e. English, Dutch or French). It is also specified whether a text is the source text (ST) or a translation (TT). In case of translation, the language of the source text is placed between brackets. For instance, if a text is classified as 'English – TT(Nl)' it means that it is an English translation of a Dutch source text.

SOURCE	**DOMAIN**								
	ENGLISH			**FRENCH**			**DUTCH**		
	ST	**TT(Fr)**	**TT(Nl)**	**ST**	**TT(En)**	**TT(Nl)**	**ST**	**TT(En)**	**TT(Fr)**
BFIS, EC, KBIN, etc.									

Table 3. Text classification table

Apart from the domain and language parameters, texts are also classified according to the source from which they were taken. Examples of sources containing texts related to biodiversity are the European Environment Agency (EEA), BFIS (Belgian Forum on Invasive Species), the European Commission (EC), the Royal Belgian Institute of Natural Sciences (KBIN), etc. The classification shown in Table 3 needs to be further refined in such a way that it becomes useful for translators. This requires a further investigation of text typologies within specialised translation (see e.g. López Rodríguez 2000, Mayor Serrano 2007).

4.3. Analysis of source language terminology

Source language terms in the categorisation framework are automatically searched for in the source language sentences of the aligned corpus files. Next, the source text is read and new terms or variants are added to the existing categorisation framework. Term identification is based on an analysis of co-referents, leaving out grammatical expressions (cf. section 2).

Once the text is read, the categorisation framework is used to automatically find and highlight terms in the English sentences in the aligned corpus file. Each matching term is placed between identification tags. The identification tag provides two types of information: it links the term to a unit of understanding by showing the UoU label (cf. section 4.1). It also shows a unique number for each term per sentence. This is illustrated by means of the following annotated sample:

2	(En) If so and you live in central or northern Europe, the '||1-KILLER_SLUG||killer slug||1-KILLER_SLUG||' is probably one of your personal enemies.
2	(Nl) Als dat zo is en u woont in het midden of noorden van Europa, dan is de 'killerslak' waarschijnlijk een van uw persoonlijke vijanden.
2	(Fr) Si c'est le cas et que vous vivez en Europe centrale ou septentrionale, la « limace tueuse » est probablement l'un de vos ennemis personnels.
3	(En) The ||1-KILLER_SLUG||slug||1-KILLER_SLUG||, which attacks your herbs and vegetables relentlessly, seems immune to control measures.
3	(Nl) Deze slak, die het genadeloos gemunt heeft op uw kruiden en groenten, lijkt immuun voor beheersingsmaatregelen.
3	(Fr) Ce mollusque, qui s'attaque inlassablement à vos légumes et fines herbes, semble résister aux mesures d'éradication.
4	(En) The ||1-KILLER_SLUG||killer slug||1-KILLER_SLUG||, known scientifically as ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||, is also called the '||3-KILLER_SLUG||Spanish slug||3-KILLER_SLUG||' because it is native to the Iberian peninsula.
4	(Nl) De killerslak, met de wetenschappelijke naam Arion lusitanicus, wordt ook wel 'Spaanse wegslak' genoemd, omdat hij een inheemse soort op het Iberisch schiereiland is.
4	(Fr) La limace tueuse, dont le nom scientifique est Arion lusitanicus, est également appelée « limace ibérique » étant donné qu'elle est originaire de la péninsule ibérique.

This sample is taken from the text 'Killer slugs and other aliens' published by the European Environment Agency[6]. The English source text is aligned with its Dutch and French translations on sentence level (wherever possible). In this sample, a line number is followed by the sentence in either English, Dutch or French. In the English sentence preceded by line number 2, the term killer slug is automatically marked as '||1-KILLER_SLUG||'. Number 1 indicates that this is the first term identified in this sentence. KILLER_SLUG is the UoU label. During the automatic annotation, only source sentences are considered.

In the English sentence preceded by line number 4, we find apart from 'killer slug' also the terms 'Arion lusitanicus' and 'Spanish slug'. As these two terms refer to the same unit of understanding 'KILLER_SLUG' they are identified by the same UoU label. Note that the identification tags also show that 'Arion lusitanicus' is identified as the second term found in this sentence and 'Spanish slug' as the third. In section 4.4, it will become clear why this first part of the identification tag is so important.

It should be noted that although the annotation of the source language terminology was automatically performed, a manual analysis is currently still required to verify whether all source language terms were correctly annotated. Errors occur in the case of terms that are used in different clusters, such as hypernyms. An example of this is given in the English sentence preceded by line number 3 where the word 'slug' refers to the killer slug and therefore should be assigned the UoU label 'KILLER_SLUG'.

4.4. Analysis of translation equivalents

After the source language terms are automatically annotated (and manually verified), the corpus file is used to manually identify the translation equivalents. A translation receives the same identification tag as its corresponding term in the source language. In the sample taken from the text

6 The full reference of this sample is: Killer slugs and other aliens – EEA Signals 2009: Eight environmental stories for Europe. The document was published in January 2009 by the European Environment Agency and can be viewed here: http://www.eea.europa.eu/pressroom/newsreleases/killer-slugs-and-other-aliens (accessed on the 9th of November 2009).

discussed in the previous section, it can be derived that ‘killer slug’ in the English sentence 2 is translated as ‘killerslak’ in Dutch and ‘limace tueuse’ in French.

2	(En) If so and you live in central or northern Europe, the '||1-KILLER_SLUG||killer slug||1-KILLER_SLUG||' is probably one of your personal enemies.
2	(Nl) Als dat zo is en u woont in het midden of noorden van Europa, dan is de '||1-KILLER_SLUG||killerslak||1-KILLER_SLUG||' waarschijnlijk een van uw persoonlijke vijanden.
2	(Fr) Si c’est le cas et que vous vivez en Europe centrale ou septentrionale, la « ||1-KILLER_SLUG||limace tueuse||1-KILLER_SLUG|| » est probablement l’un de vos ennemis personnels.
3	(En) The ||1-KILLER_SLUG||slug||1-KILLER_SLUG||, which attacks your herbs and vegetables relentlessly, seems immune to control measures.
3	(Nl) Deze ||1-KILLER_SLUG||slak||1-KILLER_SLUG||, die het genadeloos gemunt heeft op uw kruiden en groenten, lijkt immuun voor beheersingsmaatregelen.
3	(Fr) Ce ||1-KILLER_SLUG||mollusque||1-KILLER_SLUG||, qui s'attaque inlassablement à vos légumes et fines herbes, semble résister aux mesures d’éradication.
4	(En) The ||1-KILLER_SLUG||killer slug||1-KILLER_SLUG||, known scientifically as ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||, is also called the '||3-KILLER_SLUG||Spanish slug||3-KILLER_SLUG||' because it is native to the Iberian peninsula.
4	(Nl) De ||1-KILLER_SLUG||killerslak||1-KILLER_SLUG||, met de wetenschappelijke naam ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||, wordt ook wel '||3-KILLER_SLUG||Spaanse wegslak||3-KILLER_SLUG||' genoemd, omdat hij een inheemse soort op het Iberisch schiereiland is.
4	(Fr) La ||1-KILLER_SLUG||limace tueuse||1-KILLER_SLUG||, dont le nom scientifique est ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||, est également appelée « ||3-KILLER_SLUG||limace ibérique||3-KILLER_SLUG|| » étant donné qu’elle est originaire de la péninsule ibérique.

The sample also shows why it is important to provide in each sentence a separate number for each identification tag. If left out, it would not be possible to see how the different variants for the unit of understanding ‘KILLER_SLUG’ in the sentences preceded by line number 4 would be translated. By taking into account the separate number of each identification tag, it is possible to derive the following translation correspondences for the sentences preceded by line number 4:

ENGLISH (source)	DUTCH (translation)	FRENCH (translation)
• ||1-KILLER_SLUG||killer slug||1-KILLER_SLUG||	• ||1-KILLER_SLUG||killerslak||1-KILLER_SLUG||	• ||1-KILLER_SLUG||limace tueuse||1-KILLER_SLUG||
• ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||	• ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||	• ||2-KILLER_SLUG||Arion lusitanicus||2-KILLER_SLUG||
• ||3-KILLER_SLUG||Spanish slug||3-KILLER_SLUG||	• ||3-KILLER_SLUG||Spaanse wegslak||3-KILLER_SLUG||	• ||3-KILLER_SLUG||limace ibérique||3-KILLER_SLUG||

Assigning the correct tags to the translation equivalents is currently manually performed. As this work is to a large extent repetitive, an annotation tool is currently developed in order to automate the annotation of both source and target language terms. This annotation tool will make use of a translation dictionary that increases in size as more parallel texts are analysed in the corpus.

4.5. Analysis of results

Once a parallel text is annotated, the source language terms and their translations are automatically extracted and structured in a table format (cf. Table 4). Apart from source and target language occurrences, the table also includes the UoU label (retrieved from the annotated corpus file), the

target language, the number of times a source language term was translated as a specific target language term (i.e. ‘Frequency’) and a reference to the corpus file (i.e. ‘Source’). For each occurrence, the corresponding lemma is manually added. The results for KILLER_SLUG are shown in the following table.

Unit of understanding	**SL occurrence**	**SL term**	**TL**	**TL lemma**	**TL occurrence**	**Frequency**	**Source**
KILLER_SLUG	arion lusitanicus	arion lusitanicus	Nl	arion lusitanicus	arion lusitanicus	1	EEA_030409
KILLER_SLUG	arion lusitanicus	arion lusitanicus	Fr	arion lusitanicus	arion lusitanicus	1	EEA_030409
KILLER_SLUG	killer slug	killer slug	Nl	killerslak	killerslak	4	EEA_030409
KILLER_SLUG	killer slug	killer slug	Fr	limace tueuse	limace tueuse	4	EEA_030409
KILLER_SLUG	slug	slug	Nl	slak	slak	2	EEA_030409
KILLER_SLUG	slug	slug	Fr	limace	limace	1	EEA_030409
KILLER_SLUG	slug	slug	Fr	mollusque	mollusque	1	EEA_030409

Table 4. Example of an output format

This output format makes it possible to derive conclusions with respect to both intra- and interlingual terminological variation (cf. section 2). With respect to intralingual variation it can be derived that KILLER_SLUG has 3 different denominations in English, 3 in Dutch and 4 in French. It can also be derived that in English, the term ‘killer slug’ is the most frequent term used in this text to refer to KILLER_SLUG. The same holds for the translations of this term in Dutch (‘killerslak’) and French (‘limace tueuse’). The interlingual variation shows the different translations of a given source language term. It can be derived from this table that the terms in the English text – i.e. ‘arion lusitanicus’ and ‘killer slug’ – have been translated consistently whereas the English hypernym ‘slug’ was translated in French as ‘limace’ and ‘mollusque’. At this point only references are mentioned at the end of each line in the table. Texts will need to be classified according to certain dimensions to be able to detect correlations between text dimensions and tendencies in intra- and interlingual terminological variation (cf. section 4.2).

On the basis of this Table 4, it becomes possible to automatically compare the number of variants in the three languages and to present the results of the comparison for each unit of understanding. Results of such a comparison can be classified according to the following categories:

- E=N: the number of terms in English and Dutch is the same
- E=F: the number of terms in English and French is the same
- E<N: there are more terms in Dutch than in English
- E<F: there are more terms in French than in English
- E>N: there are more terms in English than in Dutch
- E>F: there are more terms in English than in French

For KILLER_SLUG, the results are presented in Table 5:

Unit of understanding	EN	NL	FR	E<N	E=N	E>N	E<F	E=F	E>F
KILLER_SLUG	3	3	4	0	1	0	1	0	0
…	…	…	…	…	…	…	…	…	…
TOTAL									
…	…	…	…	…	…	…	…	…	…

Table 5. A comparison of clusters

The results for this unit of understanding show that terminological variation is reflected in the specialised translations (cf. section 3). In the next step, it needs to be examined how the terminological variation in the source text was actually translated into the target languages. This requires a qualitative analysis of the results in Table 4.

5. Conclusion

In this article, I have presented an ongoing project in which terminological variation is studied in the context of specialised translation. It was shown how I intend to carry out a comparative analysis of terms and variants on a parallel corpus (English, French and Dutch) consisting of texts and translations related to biodiversity issues. It is hoped that new insights about terminological variation in specialised translations could contribute to the development of a new type of specialised translation dictionary which better reflects inter- as well as intralingual terminological variation. Results from a small sample of the corpus seem to confirm the working hypothesis that specialised translations will contain at least as many terminological variants for a given unit of understanding as in the source texts or even more (cf. Table 3). Following this comparison, I intend to study how the different source language terms and variants were actually translated into the target languages. This requires a qualitative analysis based on existing models of translation shifts (see e.g. Al-Zoubi/Al-Hassnawi 2001). Such a qualitative analysis will start from the results presented in section 5.4 (cf. Table 4).

References

Al-Zoubi, M.Q.R./Al-Hassnawi, A.R. 2001: Constructing a Model for Shift Analysis in Translation. In *Translation Journal* 5 (4) [online]. http://accurapid.com/journal/18theory.htm (accessed 15 November 2009).

Aubert, F.H. 1997: Translation Modalities – A Descriptive Model for Quantitative Studies in Translatology. In *Romansk Forum* 6, 1-25.

Baker, M. 1993: Corpus Linguistics and translation studies: Implications and applications. In Baker, M./Francis, G./Tognini-Bonelli, E. (eds.), *Text and technology: In honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 233-250.

Blum-Kulka, S. 1986: Shifts of cohesion and coherence in translation. In House, J./Blum-Kulka, S. (eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunther Narr, 17-35.

Bowker, L. 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. In *Terminology* 4 (2), 275-302.

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

Cabré, M. T. 1995: On diversity and terminology. In *Terminology* 2(1), 1-16.

Carreño, S.I. 2008: Characterizing Term Variation on an English-Spanish Parallel Corpus. *Proceedings of Multilingual and Comparative Perspectives in Specialised Language Resources*, Marrakech (CD-ROM).

Collet, T. 2004a: Esquisse d'une nouvelle microstructure de dictionnaire spécialisé reflétant la variation en discours du terme syntagmatique. In *Méta* 49 (2), 247-263.

Collet, T. 2004b: What's a term? An attempt to define the term within the theoretical framework of text linguistics. In *Linguistica Antverpiensia* (3), 99-111.

Depierre, A. 2007: Souvent HAEMA varie …: Les dérivés du grec HAEMA en anglais: Étude de cas de variation. In *Terminology* 13 (2), 155-176.

Durieux, C. 1995: Texte, Contexte, Hypertexte. *Cahier du CIEL* (1994–1995), 215-228.

Emery, P.G. 1991: Text classification and text analysis in advanced translation training. In *Meta* 36 (4), 567-577.

Felber, H. 1981: The Vienna school of terminology, fundamentals and its theory. In *Infoterm Series 7*, 69-86.

Freixa, J. 2002: *La Variació terminològica: Anàlisi de la variació denomitiva en textos de different grau d'especialització de l'àrea de medi ambient*. Barcelona: University of Barcelona.

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

Gallardo, S. 2000: Clon y clonación: Variabilidad conceptual en ciencias biológicas. In *Terminology* 6 (1), 59-78.

Gerzymisch-Arbogast, H. 2008: Fundamentals of LSP Translation. In *MUTRA* 2, 7-64.

Kerremans, K. 2004: Categorisation Frameworks in Termontography. In *Linguistica Antverpiensia New Series* 3/2004, 263-277.

Laviosa, S. 1998: Universals of translation. In Baker, M./Saldanha, G. (eds.), *Routledge Encyclopaedia of Translation Studies*. London/New York: Routledge, 288-291.

López Rodríguez, C.I. 2000: *Tipología textual y cohesión en la traducción biomédica inglés-español: un estudio de corpus*. Granada: Universidad de Granada.

Mason, I. 1982: The Role of Translation Theory in the Translation Class. In *Quinquerreme* 5 (1), 18-33.

Mayor Serrano, M.B. 2007: La importancia de la tipología textual pragmática para la formación de traductores médicos. In *Panace@. Boletín de Medicina y Traducción* 8 (26), 124-137.

Mauranen, A. 2004: Corpora, universals and interference. In Mauranen, A./Kujamäki, P. (eds.), *Translation Universals: Do They Exist?* Amsterdam/Philadelphia: John Benjamins, 65-82.

Osimo, B. 2008: Meaning in translation: A model based on translation shifts. In *Linguistica Antverpiensia New Series* 7/2008, 209-226.

Pearson, J. 1998: *Terms in Context*. Amsterdam/Philadelphia: John Benjamins Publishing.

Rogers, M. 1004: Multidimensionality in concepts systems: A bilingual textual perspective. In *Terminology* 10 (2), 215-240.

Rogers, M. 2007: Lexical chains in technical translation. A case study in indeterminacy. In Bassey, E.A. (ed.), *Indeterminacy in Terminology and LSP*. Amsterdam/Philadelphia: John Benjamins.

Salkie, R. 2001: A new look at modulation. In *Translation and Meaning* 5, 433-441.

Suárez de la Torre, M. M. 2004: *Análisis Contrastivo De La Variación Denominativa En Textos Especializados: Del Texto Original Al Texto Meta*. Barcelona: University of Barcelona.

Temmerman, R. 2000: *Towards New Ways of Terminology Description: The Sociocognitive-Approach*. Amsterdam/ Philadelphia: John Benjamins.

Tirkkonen-Condit, S. 2002: Translationese – a myth or an empirical fact?: A study into the linguistic identifiability of translated language. In *Target* 14(2), 207-220.

Tirkkonen-Condit, S. 2004: Unique items: over- or under-represented in translated language? In Mauranen, A./Kujamäki, P. (eds.), *Translation Universals: Do They Exist?* Amsterdam/Philadelphia: John Benjamins, 177-184.

Toury, G. 1995: *Descriptive Translation Studies and Beyond*. Amsterdam/ Philadelphia: John Benjamins.

Wüster, E. 1979/1991: Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie. In Bauer, L. (ed.). *Schriftenreihe der technischen Universität Wien*. Wien/New York: Springer.

Zhu, C. 1999: Ut Once More: The Sentence as the Key Functional Unit of Translation. In *Meta* 44 (3), 429-447.