Representing intra- and interlingual terminological variation in a new type of translation resource: a prototype proposal

Koen Kerremans
Vrije Universiteit Brussel

ABSTRACT

In this study, terminological variation pertains to the different ways in which specialised knowledge is expressed in written discourse by means of terminological designations. Choices regarding the use of term variants in source texts (i.e. intralingual variation) as well as the different translations of these variants in target texts (i.e. interlingual variation) are determined by a complex interplay of contextual factors of several kinds. For translators, it is therefore important to know the different language options (i.e. variants) that are available when translating terms and to know in which situational contexts certain options are more likely to be used.

To this end, translators often consult bi- or multilingual translation resources (e.g. terminological databases) to find solutions to certain translation problems. Different possibilities are offered in terminological databases to represent and visualise intra- and interlingual variants. In conventional terminology bases, terms in several languages usually appear on concept-oriented term records. This particular way of structuring and visualising terminological data has its roots in prescriptive terminology in which terms are merely viewed as 'labels' assigned to clearly delineated concepts (Picht and Draskau 1985). In ontologically-underpinned terminological knowledge bases or TKBs, terminological data tend to be represented in networks comprised of conceptual and semantic relations (Kerremans et al. 2008; Faber 2011; Durán Muñoz 2012; Peruzzo 2013). As opposed to traditional ways of representing terminological data (e.g. on the basis of alphabetically sorted lists, tables or matrices), such networks allow for a flexible and dynamic visualisation of data that may be connected to one another in several ways.

The aim of this article is to reflect on how visualisations of terms, variants and their translations in networks can be improved by taking into account the contextual constraints of the texts in which they appear. To this end, a novel type of translation resource has been developed, resulting from a semi-automatic method for identifying intralingual variants and their translations in texts.

A prototype visualisation of this resource will be presented in which terms, variants and their translations appear as a contextually-conditioned network of 'language options'. The proposed model derives from the Hallidayan premise that each language option or choice acquires its meaning against the background of other choices which could have been made. The choices are perceived as *functional*: i.e. they can be motivated against the backdrop of a complex set of contextual conditions (Eggins 2004). Changing these contextual conditions causes direct changes in the network of terminological options that are shown to the user.

1. Introduction

Choices regarding the use of term variants in source texts (i.e. intralingual variation) as well as the different translations of these variants in target texts (i.e. interlingual variation) are determined by a complex interplay of contextual factors of several kinds (Freixa 2006). For translators, it is therefore important to know the different language options (i.e. variants) that are available when translating terms and to know in which situational contexts certain options are more likely to be used.

To this end, translators often consult bi- or multilingual translation resources to find solutions to certain translation problems. However, such 'structured resources' never fully cover the wealth of options available in language. By separating terms from their 'natural environment' (i.e. the texts in which they appear), a lot of valuable information on which translation decisions should be based is lost. This is why translators also often resort to 'unstructured resources': i.e. texts originally written in the source and target languages or previously translated texts.

In a recently conducted study on terminological variation, it is argued why the representation of intra- and interlingual variation in existing multilingual termbases is too restrictive to account for the wealth of potential linguistic options to express units of specialised knowledge (or units of understanding) in source and target texts (Kerremans 2014). Based on this study, a new type of translation resource has been worked out in which intra- and interlingual variation retrieved from parallel texts (i.e. source texts and their translations) is structured according to semantic and contextual criteria.

The aim of this article is to discuss how intra- and interlingual terminological variants in this resource can be visualised in a dynamic and flexible graph to be used by translators. The idea for this type of visualisation further builds on recent initiatives in multilingual cognitive-oriented terminology studies to represent the conceptual organisation of a specialised field as a relational network comprised of units of understanding (denoted by terms in multiple languages) and different types of conceptual relations (see Section 2).

The graph representation in our approach differs from these initiatives in the sense that several contextual parameters will be taken into consideration when visualising intra- and interlingual variants for a given unit of understanding that were retrieved from a corpus of parallel texts (see Section 3). Terminological data in the envisaged graph representation will be dynamically structured as translators will have the possibility to zoom into specific occurrences of terms and translations in selected registers (see Section 4). Apart from summarising the basic principles underlying the prototype, we will also briefly reflect on our future work regarding its implementation (see Section 5).

2. Graph representations of multilingual terminological knowledge

In ontologically-underpinned terminological knowledge bases or *TKBs*, terminological data tend to be represented in network representations comprised of conceptual and semantic relations (Kerremans et al. 2008). As opposed to traditional ways of representing terminological data (e.g. on the basis of alphabetically sorted lists, tables or matrices), such networks allow for a flexible and dynamic visualisation of terminological data.

Particularly relevant for the present study is the fact that the methodological principles underlying TKBs are increasingly applied to the creation of bi- or multilingual special language resources for translators (Durán Muñoz 2012; Peruzzo 2013).

An example of an advanced implementation of a multilingual TKB is the *EcoLexicon*[1] database (León Araúz et al. 2011). *Ecolexicon* is targeted towards "different user groups, such as translators, technical writers, environmental experts, etc., who wish to expand their knowledge of the environment for the purpose of text comprehension or generation.". Given this objective, the database is allegedly primarily concerned with the conceptual organisation of the environmental domain. Descriptions of possible uses or preferences of terms and variants in certain communicative contexts is not provided, which seems to us an important limitation of a multilingual TKB for translators.

We have therefore defined a model for a new type of translation resource that specifically covers the choices that were made by translators when confronted with multiple terminological variants for a given unit of understanding. This translation resource is comprised of semantically and contextually-structured, term-based translation units that were extracted from a multilingual parallel corpus (Kerremans 2014).

3. Structure of the translation resource

Term-based translation units are the primary building blocks of the resource presented in this article. Each translation unit (TU) is further classified according to text-related and semantic categories:

- Text-related categories are properties originally assigned to the bitext (i.e. the combination of a source text and its translation) from which the TU is extracted. Examples of such categories are text type, text source, language, text topic, etc.

- Semantic categorisation involves classifying the English term in the TU according to the 'concept' to which it refers in the source text. This means that each term in the source texts is marked with a unique identification code – i.e. a so-called 'cluster label'. Terms extracted from the source texts that carry this label appear in the same 'cluster' of terminological variants (Kerremans 2011).

[1] http://ecolexicon.ugr.es/en/aboutecolexicon.htm

4. A prototype proposal for visualising intra- and interlingual terminological variants

The proposed graph representation of the translation resource derives from the Hallidayan premise that each language option or choice acquires its meaning against the background of other choices which could have been made. The choices are perceived as *functional*: i.e. they can be motivated against the backdrop of a complex set of contextual conditions (Eggins 2004). Changing the contextual conditions causes direct changes in the network of terminological options that are shown to the user. This idea is illustrated by means of Figure 1.

In a bilingual view, the user will first specify a search term in the source language. This will activate in the contextual (i.e. semantic, situational and linguistic) filters different fields that are associated with the search query. In case the search term is connected to multiple conceptual clusters (see Section 3), the user will be able to select the proper cluster label. The result of the search query is visualised in a graph representation. Situational (see *Text options*) and linguistic (see *Lemma* and *POS options*) criteria or filters are used in this example to highlight or deactivate certain connections or nodes in the source and target languages, causing constant changes or shifts in the structure of variants.

Bilingual view					
	[add SL search term or ClusterID]		**Search**		
Text options	**Framework**	**Legal status**	**Source**	**Intertextual ref.**	**Texts**
	✓ EU ✓ Non-EU	✓ Leg ✓ Non-Leg	✓ CoR ✓ EC ✓ EEA ✓ EESC ✓ GRE	✓ 1 (e.g. PreLex ref.) ✓ 2 ✓ …	✓ TextID 1 ✓ TextID 2 ✓ …
Cluster option(s)	✓ GHG_EMISSION_REDUCTION				
Source language options (English)			**Target language options**		
			Dutch		
Lemma option(s)	✓ emission reduction		✓ afname van emissie ✓ broeikasgasemissie verminderen ✓ …		
POS option(s)	✓ noun ✓ noun noun ✓ …		✓ adjective ✓ noun ✓ noun verb ✓ noun prep noun ✓ …		

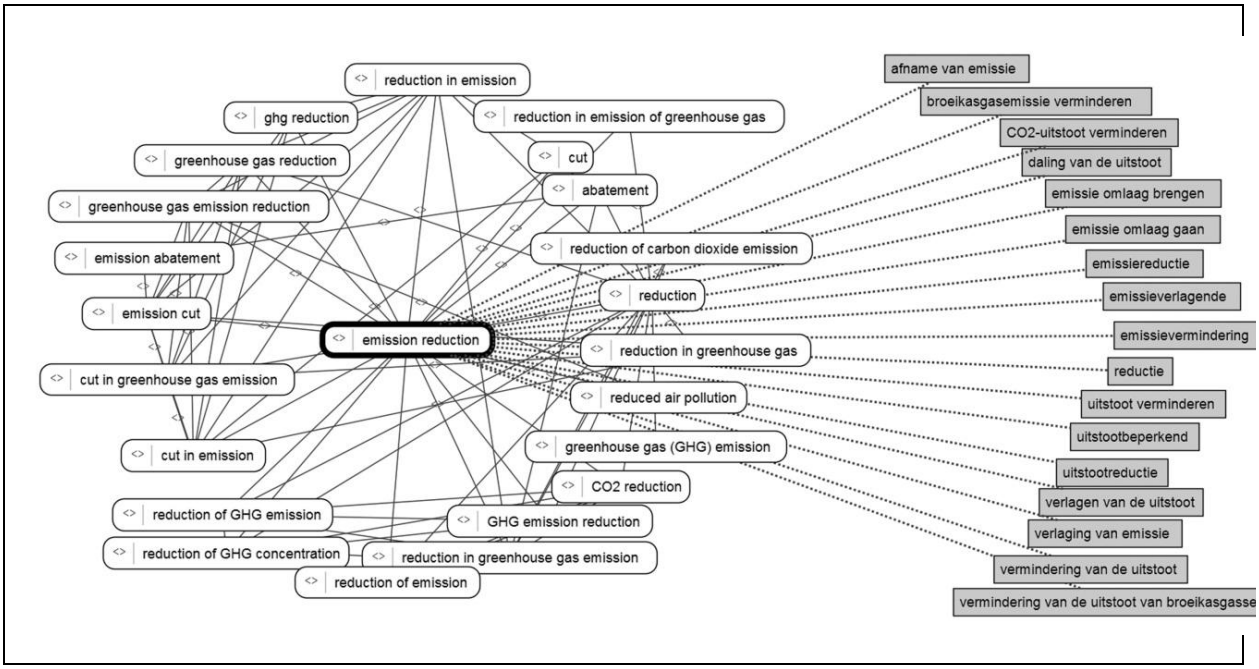

Figure 1: Example of visualising intra- and interlingual terminological variants

5. Conclusion

In this article we discussed how intra- and interlingual variants, extracted from parallel texts, can be used to populate a new type of translation resource. We discussed how this resource can be represented as an innovative graph. The resource can be perceived as an additional 'tool' that can be integrated in a computer-assisted translation (CAT) environment or workflow together with bi- or multilingual termbases and translation memories.

The next step in our research is to turn the ideas and requirements concerning the visualisation part into an actual implementation that could initially be tested in an experimental setting by means of a group of translators.

References

DURÁN MUÑOZ, ISABEL. 2012. *La ontoterminografía aplicada a la traducción*. Vol. 80. Studien Zur Romanischen Sprachwissenschaft Und Interkulturellen Kommunikation. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.

EGGINS, SUZANNE. 2004. *Introduction to Systemic Functional Linguistics: 2nd Edition*. London/New York: Continuum International Publishing Group.

FREIXA, JUDIT. 2006. Causes of denominative variation in terminology: A typology proposal. *Terminology* 12: 51–77.

KERREMANS, KOEN. 2011. Translating terminological variation: the case of biodiversity terminology. In *Translationswissenschaft: alte und neue Arten der Translation in Theorie und Praxis*, ed. Lew Zybatow, Alena Petrova, and Michael Ustaszewski, 71–77. Forum Translationswissenschaft 16. Frankfurt am Main/Berlin/Bern/Bruxelles/New York/Oxford/Wien: Peter Lang Verlag.

KERREMANS, KOEN. 2014. Terminological variation in multilingual Europe. The case of English environmental terminology translated into Dutch and French. PhD thesis, Brussel: Vrije Universiteit Brussel.

KERREMANS, KOEN, RITA TEMMERMAN, AND PETER DE BAER. 2008. Construing domain knowledge via terminological understanding. *Linguistica antverpiensia* 7: 177–191.

LEÓN ARAÚZ, PILAR, ARIANNE REIMERINK, AND PAMELA FABER. 2011. Environmental knowledge in EcoLexicon. In *Proceedings of the computational Linguistics-Applications Conference*, 9–16. Jachranka.

PERUZZO, KATIA. 2013. Terminological equivalence and variation in the EU multi-level jurisdiction: a case study on victims of crime. PhD Thesis, Trieste: Università Degli Studi Di Trieste.