

# How reliable are online bilingual concordancers?

An investigation of *Linguee*, *TradooIT*, *WeBiText* and *ReversoContext* and their reliability through a contrastive analysis of complex prepositions from French to English



---

Marie Genette

Supervisors: Prof. Sylviane Granger & Hilde Hasselgård

Master's Thesis in Linguistics and its Applications for a Multilingual Society

Spring Term 2016

UNIVERSITÉ CATHOLIQUE DE LOUVAIN & UNIVERSITETET I OSLO



# TABLE OF CONTENT

<b>ABSTRACT</b> .....	<b>VI</b>
<b>ACKNOWLEDGMENT</b> .....	<b>VII</b>
<b>LIST OF TABLES</b> .....	<b>VIII</b>
<b>LIST OF FIGURES</b> .....	<b>IX</b>
<b>ABBREVIATIONS IN ALPHABETICAL ORDER</b> .....	<b>X</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1. AIM AND SCOPE.....	1
1.2. OBJECT OF THE RESEARCH.....	2
1.3. BACKGROUND .....	3
1.3.1. <i>Corpus linguistics</i> .....	3
1.3.1.1. Corpora used in contrastive research.....	4
1.3.1.2. Alignment systems .....	6
1.3.2. <i>Contrastive linguistics</i> .....	6
1.3.2.1. Definition and aim.....	6
1.3.2.2. Relevant terminology .....	7
1.3.2.3. Correspondence measurements .....	9
1.4. ANALYSIS ON THE BASIS OF FRENCH COMPLEX PREPOSITIONS .....	10
1.5. STRUCTURE OF THE THESIS.....	12
<b>2. DATA AND METHODOLOGY</b> .....	<b>13</b>
2.1. CORPORA .....	13
2.2. MATERIAL.....	14
2.3. METHODOLOGY.....	14
2.3.1. <i>Selection of the compared items</i> .....	15
2.3.2. <i>Juxtaposition</i> .....	16
2.3.3. <i>Comparison</i> .....	16
<b>3. SURVEY OF ONLINE TRANSLATION TOOLS</b> .....	<b>18</b>
3.1. ONLINE BILINGUAL DICTIONARIES .....	18
3.1.1. <i>Overview and previous work</i> .....	18
3.1.2. <i>Description of each OBD</i> .....	19
3.2. ONLINE BILINGUAL CONCORDANCERS.....	20
3.2.1. <i>Introduction</i> .....	20
3.2.2. <i>Terminological clarification</i> .....	20
3.2.3. <i>Description of each OBC</i> .....	23
3.2.3.1. Overview .....	23
3.2.3.2. Languages.....	24
3.2.3.3. Process: data collection and alignment system .....	25

3.2.3.4.	Corpus integration .....	26
3.2.3.5.	More and better data .....	29
3.2.3.6.	Users .....	29
3.2.3.7.	User input .....	30
3.2.3.8.	Hybridization .....	31
3.2.3.9.	Operating of the OBCs .....	33
	• Interface.....	33
	• Search system.....	34
	• Efficiency of access.....	35
	• Sentence pairs.....	35
	• Bi-texts .....	36
	• Grouped translations.....	36
3.2.4.	<i>Assessment of the OBCs</i> .....	38
3.2.4.1.	Linguee.....	39
3.2.4.2.	TradooIT.....	44
3.2.4.3.	WeBiText .....	45
3.2.4.4.	ReversoContext .....	47
<b>4.</b>	<b>ANALYSIS.....</b>	<b>48</b>
4.1.	INTRODUCTION.....	48
4.2.	MACRO-ANALYSIS.....	48
4.3.	MICRO-ANALYSIS OF THE TRANSLATIONS .....	51
4.3.1.	<i>En matière de</i> .....	51
4.3.1.1.	Frequent English translations suggested by the tools .....	51
4.3.1.2.	Contribution of each tool.....	53
4.3.1.3.	Translation control: back-translation and mutual correspondence.....	54
4.3.2.	<i>Au sein de</i> .....	55
4.3.2.1.	Frequent English translations suggested by the tools .....	55
4.3.2.2.	Contribution of each tool.....	55
4.3.2.3.	Translation control: back-translation and mutual correspondence.....	56
4.3.3.	<i>Par rapport à</i> .....	57
4.3.3.1.	Frequent English translations suggested by the tools .....	57
4.3.3.2.	Contribution of each tool.....	58
4.3.3.3.	Translation control: back-translation and mutual correspondence.....	59
4.3.4.	<i>Dans le cadre de</i> .....	59
4.3.4.1.	Frequent English translations suggested by the tools .....	59
4.3.4.2.	Contribution of each tool.....	61
4.3.4.3.	Translation control: back-translation and mutual correspondence.....	61
4.3.5.	<i>En faveur de</i> .....	62
4.3.5.1.	Frequent English translations suggested by the tools .....	62
4.3.5.2.	Contribution of each tool.....	63
4.3.5.3.	Translation control: back-translation and mutual correspondence.....	63
4.3.6.	<i>À la fin de</i> .....	64
4.3.6.1.	Frequent English translations suggested by the tools .....	64

4.3.6.2.	Contribution of each tool.....	65
4.3.6.3.	Translation control: back-translation and mutual correspondence.....	65
4.3.7.	<i>Dans le domaine de</i> .....	66
4.3.7.1.	Frequent English translations suggested by the tools .....	66
4.3.7.2.	Contribution of each tool.....	67
4.3.7.3.	Translation control: back-translation and mutual correspondence.....	68
4.3.8.	<i>En raison de</i> .....	68
4.3.8.1.	Frequent English translations suggested by the tools .....	68
4.3.8.2.	Contribution of each tool.....	69
4.3.8.3.	Translation control: back-translation and mutual correspondence.....	70
4.3.9.	<i>Aux côtés de</i> .....	70
4.3.9.1.	Frequent English translations suggested by the tools .....	70
4.3.9.2.	Contribution of each tool.....	72
4.3.9.3.	Translation control: back-translation and mutual correspondence.....	73
4.3.10.	<i>Au service de</i> .....	73
4.3.10.1.	Frequent English translations suggested by the tools .....	73
4.3.10.2.	Contribution of each tool.....	74
4.3.10.3.	Translation control: back-translation and mutual correspondence .....	75
<b>5.</b>	<b>DISCUSSION AND CONCLUSION.....</b>	<b>76</b>
5.1.	SUMMARY OF THE ANALYSIS.....	76
5.1.1.	<i>The OBCs and OBDs' degree of similarity to authentic language</i> .....	76
5.1.2.	<i>Translation process of CPs from French to English</i> .....	80
5.1.3.	<i>Contrastive insights</i> .....	81
5.2.	CONCLUSION AND FURTHER CONSIDERATIONS.....	81
	<b>REFERENCES.....</b>	<b>83</b>
	<b>APPENDICES.....</b>	<b>95</b>

## ABSTRACT

This contrastive study consists in evaluating the reliability of online bilingual concordancers (OBCs) on the basis of their translation quality from French to English. To this end, ten French complex prepositions (CPs) are first searched for in translation corpora (Label France and PLECI\_news corpora) so as to spot their translations in authentic language. This reveals that French CPs are not systematically but most commonly rendered in English by divergent correspondences (mostly simple prepositions but also verbs) and also that numerous possibilities are generally available, i.e. none of the French CPs has a strict equivalent in English, which implies that context is essential in order to select the correct translation.

In a second stage, the same French CPs are queried in four OBCs (*Linguee*, *TradooIT*, *WeBiText* and *ReversoContext*), where the first 30 translations are identified and compared, as well as in three online bilingual dictionaries (*Larousse*, *Reverso-Collins* and *Oxford*). All the translations suggested are subsequently tested against the corpora's findings to evaluate their degree of similarity to authentic language, which is the basis for the issue of reliability. The analysis reveals that the translations of the dictionaries poorly match those found in the corpora, implying a poor degree of reliability.

On the other hand, the translations provided by the first sentence pairs in the OBCs are not identical but similar to authentic language, although these tools do not contribute much as compared to the information found in corpora. However, two of the OBCs under investigation (*TradooIT* and *ReversoContext*) offer a unique feature, where the various translations to a queried term are indexed before the random pairs of sentences together with frequency information, i.e. the grouped translations option. As the word-alignment system is more accurate in *ReversoContext*, this particular OBC proves to be the most reliable. Because the amount of data available is much larger than in the corpora, it (a) brings out new information on the variety of possible translations, (b) better shows in which context a translation should be used and (c) better draws the line between extremely frequent translations, moderately frequent ones and infrequent ones. Several improvements should be effected in the future however, concerning the distinction between source and target languages and the possibility to filter the results according to the different corpora. Moreover, the OBC should display the exact size of each corpus as well as unequivocal information on frequency.

## ACKNOWLEDGMENT

First and foremost, I would like to thank my supervisors, professors Sylviane Granger and Hilde Hasselgård. I sincerely believe that being supervised by two linguists has been extremely helpful, as it allowed me to go forward with my research and writing with twice as much advice. I am also grateful for their enthusiasm (both as professors and supervisors), availability and support, as I have been guided through the whole process with useful tips on interesting articles, encouragements, ideas and ways of improving the thesis.

On a more personal note, I would like to seize this opportunity to thank professor Granger for lending me an attentive ear back in the third year of my bachelor, when everything seemed too overwhelming. Her deep understanding and useful suggestions helped me get my head above water again, take my exams and eventually get my degree.

I would also like to thank my parents for their unconditional support and their open mind. They really did everything they could for my blossoming, and I most certainly would not be here, handing out my MA thesis, if it was not for them.

Finally, special thanks to all my friends and especially to my partner, Jérémy, for his presence, his encouragements, his positive attitude as well as for his helpful expertise in computer sciences.

Oslo, 18<sup>th</sup> May, 2016,

Marie

## LIST OF TABLES

**Table 1.** Number of occurrences of the CPs in Label France and PLECI\_news corpora

**Table 2.** Most frequent translations of *en matière de* in the three types of tools.

**Table 3.** Most frequent translations of *au sein de* in the three types of tools.

**Table 4.** Most frequent translations of *par rapport à* in the three types of tools.

**Table 5.** Comparison of the English correspondences of *par rapport à* in OE and TE in the PLECI\_news corpus.

**Table 6.** Most frequent translations of *dans le cadre de* in the three types of tools.

**Table 7.** Most frequent translations of *en faveur de* in the three types of tools.

**Table 8.** Most frequent translations of *à la fin de* in the three types of tools.

**Table 9.** Most frequent translations of *dans le domaine de* in the three types of tools.

**Table 10.** Most frequent translations of *en raison de* in the three types of tools.

**Table 11.** Most frequent translations of *aux côtés de* in the three types of tools.

**Table 12.** Most frequent translations of *au service de* in the three types of tools.



## LIST OF FIGURES

- Figure 1.** Paradigm of correspondences (Johansson 2007: 25).
- Figure 2.** Classification of CPs within the adposition category.
- Figure 3.** Possible comparisons in the PLECI corpus.
- Figure 4.** Efficiency of the word-alignment and highlighting options in *ReversoContext*.
- Figure 5.** Corpus information available from French to English in *TradooIT*.
- Figure 6.** Frequency of *en matière de* according to the corpora in *TradooIT*.
- Figure 7.** List of corpora in *WeBiText*.
- Figure 8.** Corpus information in *ReversoContext*.
- Figure 9.** Community aspects in *ReversoContext*.
- Figure 10.** Phraseological issues in *Linguee*'s bilingual dictionary.
- Figure 11.** *Wikipedia* Section in *Linguee*.
- Figure 12.** Incremental function in *Linguee*.
- Figure 13.** Results information in *TradooIT*.
- Figure 14.** Direct access to the parallel web pages in *WeBiText*.
- Figure 15.** Grouped translations in *TradooIT*.
- Figure 16.** Grouped translations in *ReversoContext*.

## **ABBREVIATIONS IN ALPHABETICAL ORDER**

**CPs:** complex prepositions

**GTs:** grouped translations

**MWUs:** multi-word units

**OBCs:** online bilingual concordancers

**OBDs:** online bilingual dictionaries

**OE:** original English

**OF:** original French

**POS:** part-of-speech

**SL:** source language

**TE:** target English

**TF:** target French

**TL:** target language

**TMs:** translation memories

**TTs:** translation tools

# 1. INTRODUCTION

## 1.1. AIM AND SCOPE

From the beginning of the computer area in the 1950s, language-related activities and professions have witnessed an evolution of the tools at hand, but, as suggested by García Hernández (2014: 2) and Gracia (2015: 1), an even more radical modification has conceivably transpired with the advent of the Internet. A panel of revolutionary possibilities were then introduced to various types of language users from around the world at the click of a mouse, from more interactive, collaborative and hybrid resources to easier and quicker access to larger amount of data. We will concentrate on tools providing translations.

However, users may feel at a loss when they are faced with the proliferation of online translations tools (TTs) and wonder which one to turn to. Whatever the type of language enquiry, we believe that the most fundamental aspect to consider is quality. Because these tools are accessed on the Internet, this quality implies various factors, listed by Khawaja *et al.* (2010: 37) as the following: trust,<sup>1</sup> navigability, responsiveness, efficiency, functionality, ease of use, usefulness, information quality and web appearance. While most of these features will be taken into account,<sup>2</sup> this study first and foremost concentrates on the information quality, defined as “the concern that information provided is accurate, updated, and appropriate” (Loiacono *et al.* 2002: 19). The information provided in online TTs being related to language and, more precisely, translation, it needs to be evaluated by an empirical linguistic analysis, which is called for by Alonso Jiménez (2013: 20). To this end, a specific set of words and their translations will be analysed and tested against authentic language, accessed through corpora, namely French complex prepositions (CPs) in the process of grammaticalization (see Section 1.4.).

Despite the importance of the assessment of such quality, the scarcity of studies related to online TTs is patent, especially when we prune the concept of quality to a micro-perspective, construed as the inherent quality of the translations provided (e.g. Désilets *et al.* 2008a is interested in macro-reliability, such as the rate of alignment error). To our knowledge, the only kindred studies are the evaluation of *TransSearch*<sup>3</sup> by Danlos & Roze (2011). They have

---

<sup>1</sup> The concept of trust relates to the privacy security.

<sup>2</sup> Alonso Jiménez (2013) and Gelpí (2004) concentrate on these features for *Linguee* and online bilingual dictionaries respectively.

<sup>3</sup> *TransSearch* is an online bilingual concordancer (<http://tsrali3.com>. Accessed on 2 November 2014) that will not be part of the present research due to its charged access.

adopted a similar methodology to that of the present study, in the sense that they too have selected a type of words which give rise to translation problems, namely discourse connectors (*en effet* and *alors que*) and analysed them in contrast with corpora. In their results, Danlos & Roze (2011: 5) suggest a poor matching of the translations offered by *TransSearch* with the results observed in the corpora or the dictionaries, which is explained by the disregarding of the zero correspondence type (*ibid.*). They also believe that the tool cannot provide relevant results because it works at the sentence level and thus cannot take the discursive context into account. Other similar studies will be discussed in Section 3.2.4. since they directly address one of the OBCs under consideration here. These include the investigation of *WeBiText* by Simard (2013), Volk *et al.* (2014)'s research on online tools and Bourdaillet & Langlais (2012)'s article on *TransSearch*, *Linguee* and *TradooIT*. Despite the lack of scholarly research (mentioned in Abel 2012: 87, Grauer 2010: 3, Simard 2013: xix), doubts have already been voiced however, reflected by the indication to use further tools (dictionaries and corpora) to control the results found in the online ones (Alonso Jiménez 2013: 7, Kübler 2013 and Van Bolderen 2012).

## 1.2. OBJECT OF THE RESEARCH

Given their novelty, non-automatic online TTs (i.e. which provide previously translated texts) will constitute our object of study, with the aim of compensating for the shortage of studies regarding their reliability. We will more specifically focus on those which “return pairs of sentences where the query and one of its translations are identified” (Bardouillet & Langlais 2012: 1), an option commonly referred to as (online) bilingual concordance (OBC). On account of the large number of online TTs providing this service, special attention will be directed at those that are freely available and offer the language pair from French to English, namely *Linguee*,<sup>4</sup> *TradooIT*,<sup>5</sup> *WeBiText*,<sup>6</sup> and *ReversoContext*.<sup>7</sup> A brief presentation of the tools in addition to a review of previous work will be given before the examination of their reliability. Besides OBCs, we will also probe some online bilingual dictionaries (OBDs), namely *Larousse*,<sup>8</sup> *Reverso-Collins*,<sup>9</sup> and *Oxford*<sup>10</sup> to contrast the results offered by two different types of online TTs. This analysis will show whether users can work exclusively

---

<sup>4</sup> <http://www.linguee.fr>.

<sup>5</sup> <https://www.tradooit.com>.

<sup>6</sup> <http://www.webitext.com/bin/webitext.cgi>.

<sup>7</sup> <http://context.reverso.net/traduction/>.

<sup>8</sup> <http://www.larousse.fr/dictionnaires/francais-anglais>.

<sup>9</sup> <http://dictionnaire.reverso.net>.

<sup>10</sup> <http://www.oxforddictionaries.com.proxy.bib.ucl.ac.be:8888>.

with OBCs (or certain OBCs), be better off using other types of TTs (e.g. dictionaries, corpora) or better benefit from a combination of various tools, which is the hypothesis of this study.

### 1.3. BACKGROUND

As mentioned earlier, because the OBCs' reliability depends on the accuracy of their translations, we need to analyse and assess them, however not on the basis of intuition but against corpora. As a consequence, the fields of contrastive linguistics and corpus linguistics will be of major importance to this study.

#### 1.3.1. CORPUS LINGUISTICS

Corpus linguistics is the “study of language based on examples of ‘real life’ language use” (McEnery & Wilson 2001: 1) and is neither an independent field nor a theory but a methodology which can be adopted for almost all linguistic areas. We will therefore provide theoretical considerations on corpora (which are defined below by McEnery *et al.* 2006) in this section and focus on the methodology in Section 2.3.

A collection of machine-readable authentic texts (including transcripts of spoken data) which is sampled to be representative of a particular language or language variety. (McEnery *et al.* 2006: 4-5)

The use of corpora, especially in electronic format, has profoundly changed how linguistic research is conducted as well as how language is conceived, moving from the abstract study of language as a system to the concrete study of language in use (Johansson 1999: 3). Corpora also allow for a quantitative approach to language because of the large amount of data they contain, and researchers do not need to rely on intuitive and subjective descriptions anymore, which offers a scientific reliability to the field of linguistics.

Even though corpora should not be taken for granted, as they cannot possibly represent the whole language (Hoffmann 2005: 6) and may contain ungrammatical instances, the positive results brought out arguably outnumber the possible pitfalls. Due to their “abundant, more authentic and contextualized data” (Buyse *et al.* 2013: 509), corpora enable linguists “to perceive what may be invisible to the naked eye” (Johansson 1999: 21), for example, the existence of multi-word units (MWUs) (Cobb 2013: 79) and “not only traditional categories but also the phraseological patterns which tend to be semantically compositional and therefore less salient” (Granger and Lefer 2013: 1-2). Also, they prove useful for registers and genres

specificities, as well as specific terminology. Finally, while all linguistic areas can greatly benefit from corpus use, Krzeszowski (1990: 203) strongly argues that “systematic contrastive studies are incomplete and inadequate unless supported by quantitative data at all levels of linguistic analysis”. Because this is the type of study involved here, we will now describe the types of corpora that can best serve it (also see Granger 2010).

### **1.3.1.1. Corpora used in contrastive research**

As mentioned in Altenberg & Granger (2002: 7), “the terminology used to describe the different types [of corpora] is inconsistent and confusing”. In contrastive research, the first crucial distinction to make is that between bilingual and multilingual corpora, if the source language can be translated into various languages (the second term being a “general inclusive term”, *ibid.*). More importantly, the second distinction is that between translation and comparable corpora.

**Translation corpora** consist of original texts in one language with their translations in one or more other languages. These texts should therefore “express the same meanings and have the same discourse functions” (Johansson 1999: 5), which is helpful when one is looking to establish paradigms of correspondences, as in the present research. As Altenberg & Granger (2002: 8) explain, translation corpora can either be unidirectional (as the Label France corpus, see Section 2.1.), when one language is the source and the other is the target, or bidirectional, when both languages are the translations and sources of one another. In any case, Salkie (2008: 5) qualifies the distinction between the original language on the one hand and the target language on the other as the absolute bare minimum for a translation corpus to be considered as of quality. The compilation of such corpora can be fraught with pitfalls, purely and simply because there are fewer translated texts than there are original ones, especially when working with less dominant languages or even depending on the direction of translation (Johansson 1999: 6). Besides this practical problem which prevents translation corpora from giving a full and balanced representation of the languages compared (Altenberg & Granger 2002: 9), their main disadvantage is the potential presence of translators’ idiosyncrasies (also in Zanettin 1998: 618) and of incorrect translations (Granger and Lefer 2013: 12). Two additional problems must be accounted for: translated texts (a) may very well contain translationese, i.e. “deviance in translated texts induced by the source language” (Johansson & Hofland 1994: 26) and (b) reflect the typical patterns of translation itself rather than those of a particular language (see Baker 1996 and 2007 for a discussion on translation universals). For all these reasons, linguists have voiced their concern on restricting one’s study to translation

corpora, warning that they should be used as “complementary sources of cross-linguistic data” (Altenberg & Granger 2002: 9). Investigating the second type of corpora is arguably a good option to alleviate these problems and provide more trustworthy results.

**Comparable corpora** contain original texts in each language and can be matched according to various criteria such as time, domain, genre, target audience, subject, communicative function, etc. As they represent natural language, Johansson (1999: 5) indicates that they “should allow safe conclusions to be drawn on similarities and differences between the languages compared”. Altenberg & Granger (2002: 8) add that since they are “unaffected by translation effects”, they can also serve as control to support (or contradict) findings from translation corpora. Nevertheless, even comparable corpora should be analysed carefully, i.e. as information circulates worldwide, influences even pervade resources of original language.<sup>11</sup> Finally, Johansson (1999: 5) advocates that the greater issue when using comparable corpora is to know what to compare, which is corroborated in Altenberg & Granger (2002: 8). Also, **parallel corpus** is another frequently used label that will serve as an umbrella term for the two types of corpora described above but as referring to corpora combining both comparable and translation data (Hasselgård 2010: 100), such as the PLECI corpus (see Section 2.1.), the English-Norwegian Parallel Corpus and the English-Swedish Parallel corpus.

Finally and directly related to our research, it needs to be acknowledged that the use (and collection) of corpora can be “difficult, time-consuming, laborious (and) tedious [...]” (Buyse *et al.* 2013: 509). Therefore, the fact that corpora provide quantitative support combined with these negative aspects paradoxically render their use both necessary and limited among professionals and students in translation (Kübler 2013, Volanschi 2007: 30, Alonso Jiménez 2013: 7 and Simard 2013: 38). As a result of this difficulty,<sup>12</sup> Kübler (2013) explains that students and translators head towards the Internet instead and use, for example, *Linguee*. Furthermore, Volk *et al.* (2014: 3172) acknowledge that while “for the medium to advanced language learner or second-language user, the advantages of parallel corpora are apparent”, beginners should refer to bilingual dictionaries. The analysis will reveal whether bilingual dictionaries, at least the online kind, can be used as a primary and unique resource for this type of users.

---

<sup>11</sup> Personal communication (Granger 2015)

<sup>12</sup> Kübler lists (2013) a series of further issues that seems to be the reason why translators shy away from corpora, such as the scarcity of corpora for specialized domains, the lack of updates for rapidly evolving domains, the lack of corpora for certain languages, the unbalance of corpora in terms of genres, the excess of different interface, the copyright, cultural differences, etc.

### **1.3.1.2. Alignment systems**

Hasselgård (2010: 101) and Altenberg & Granger (2002 : 10) explain that one would not be able to exploit corpora to their full potential without proper alignment, whose system links each unit in the original text to its corresponding unit in the translated one according to a particular threshold, such as paragraph, sentence, phrase or word (Altenberg & Granger 2002 : 10). Alignment can be performed automatically through statistical principles (such as cognate words, sentence length, typography, etc.), but the results need to be verified manually.<sup>13</sup>

The authors explain that alignments at the word level prove more challenging, since “a given word in the source text may be rendered by many translation equivalents and structural paraphrases, and sometimes none at all”, hence sentence alignment is more often opted for (Altenberg & Granger 2002 : 10-11). Despite this difficulty, Volk *et al.* (2014: 3172) strongly argue that “automatic word alignment enables new search options that are interesting for translators and linguists alike” and that all systems should therefore “tap [...] the potential of this new technology” (2014: 3177).

## *1.3.2. CONTRASTIVE LINGUISTICS*

### **1.3.2.1. Definition and aim**

Contrastive analysis, or contrastive linguistics, which is “the systematic comparison of two or more languages with a view to describing their similarities and differences” (Hasselgård 2010: 98), has a dominant position in linguistic. Altenberg & Granger (2002: 5), however, insist that what we are dealing with is a revival of the field, which “had its glory back in the 1960s, before falling into disfavour” (see Ebeling & Ebeling 2013 for a thorough history of the field). At first, the field was aimed at translation studies and language teaching, with a major concern at measuring the differences between languages likely to “cause problems for foreign language learners with a particular mother-tongue background” (Johansson 1999: 3-4). The claim that difficulties in learning a language can be fully predicted progressively weakened to a hypothesis “based on the evidence of language interference” and the phenomenon of interlanguage (Hasselgård 2010: 99-100).

In addition to its practical applications, contrastive linguistics is now acknowledged as a descriptive field as well, as it can both offer valuable insights into the languages compared

---

<sup>13</sup> Problems arise, for example, when a sentence has been divided into two in the translation, or conversely (Altenberg & Granger 2002: 10).



and “formulat[e] accurate descriptions of individual languages” (Johansson 1975: 15), increasing our knowledge with features that would have most probably gone unnoticed with separate studies of the two languages.<sup>14</sup> In the arduous process of attesting the correspondence between items in different languages, James (1980: 168) maintains that the most decisive aspect is to avoid equating items at all levels and to compare like with like so as to offer a valuable “frame of reference” (Hasselgård 2010: 98), as “it is only against a background of sameness that differences are significant” (Connor & Moreno 2005: 5). Various types of comparison bases, which is referred to as *tertium comparationis*, exist,<sup>15</sup> but translation equivalence is identified as the most reliable one for contrastive analysis (as well as cross-linguistic studies in general) by James (1980: 178), Altenberg & Granger (2002: 15), Johansson (1999: 5) and Hasselgård (2010: 99).

### 1.3.2.2. Relevant terminology

To follow Johansson’s terminology (2007) and Chesterman’s advice (1998: 31), the term *equivalence* will not be used in the analysis, inter alia because the degree of equivalence between two items systematically depends on the context, but the major reason is that “[...] no two items in different languages are equivalent or identical” (Ebeling & Ebeling 2013: 23). Instead, the term *translation* will refer to the items used to render the French CPs in English while the more comprehensive term *correspondence*<sup>16</sup> (Hasselgård 2010: 101) will be used when the relations described between the compared items is “bidirectional so that a word or phrase in a source text has a correspondence in the translation and vice-versa” (Johansson 2007: 23), as in the sections dealing with back translation (see definition below).

The different types of correspondences are embraced in a concept referred to by Johansson (2007: 23) as the translation paradigm. We will follow the two classification parameters of Ebeling (2015),<sup>17</sup> namely direction of translation and expression. The second parameter (see Figure 1) opposes overt expressions, which are visible in the target text, from zero expressions (see example 1), which are either omitted or added in the target text. An important distinction applies for overt correspondences, in that the correspondences can be of

---

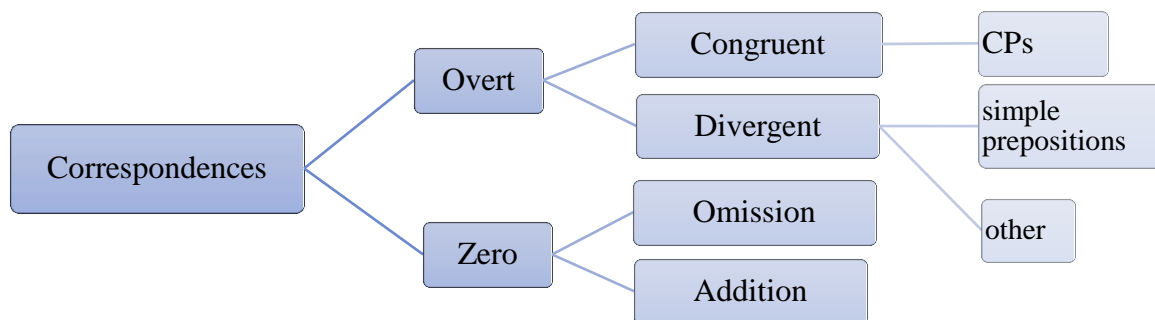
<sup>14</sup> As Hasselgård (2010: 98) points out, while contrastive analysis is not bound to any particular field of linguistics, it most generally focuses on a single pair of languages.

<sup>15</sup> Statistical equivalence, system equivalence, semantico-syntactic equivalence, rule equivalence, substantive equivalence, pragmatic equivalence (Krzyszowski 1990 : 23), but also grammatical categorization equivalence, genre equivalence, etc. (Hasselgård 2010 : 98)

<sup>16</sup> The term will not be equated to bad translations, as in Lederer (2014: 45), where she writes “[...] the poor quality of a translation by correspondence is immediately apparent”.

<sup>17</sup> Personal communication (Ebeling 2015)

the same formal category, namely congruent (see examples 2-3), or of a different one, namely divergent (which is also called transposition in Vinay and Darbelnet 1977), as in example (4). In our analysis, divergent correspondences will embrace simple prepositions on the one hand and other grammatical categories on the other (such as verbs, for example). Even though our primary concern is not to establish the translation paradigms of French CPs, the process is necessary in order to contrast the translations suggested in the OBCs, and ultimately evaluating their reliability as a TT.



**Figure 1.** Paradigm of correspondences (Johansson 2007: 25).

- |     |                  |  |
|-----|------------------|--|
| (1) | FR <sup>18</sup> | Ces mesures ont permis de confiner le virus de la poliomyélite en grande partie dans le sud du pays et de consolider les bases pour de futures campagnes nationales <b>en faveur de</b> la santé.      |
|     | ENG              | Thanks to these efforts, the polio virus has been largely contained within the southern region of the country, and the basis for future national [Ø] health campaigns has been strengthened (TradooIT) |
| (2) | OF               | [...] lectorat s'est prononcé à plus de 60 % <b>en faveur de</b> l'Union européenne, tandis que [...]  |
|     | TE               | [...] whose electorates voted more than 60 per cent <b>in favour of</b> European Union, while [...] (PLECI_news)   |
| (3) | OF               | [...] dans la Déclaration de Bamako (Mali) <b>en faveur de</b> la démocratie et des droits de l'Homme [...]  |
|     | TE               | [...] by the Bamako (Mali) Declaration <b>in support of</b> democracy and human rights [...] (Label France)  |
| (4) | OF               | [...] des zones qui ont pus [sic] que d'autres payé le prix du sang <b>en faveur de</b> l'ancrage européen qu [sic] point de le plébisciter (l 'Alsace et la Lorraine), [...]                          |
|     | TE               | Areas which have paid a heavier price in blood than others also <b>favoured</b> firm attachment to Europe, to the point of returning a resounding "yes " [...] (PLECI_news)                            |

<sup>18</sup> When the translation direction is known, we will specify whether an example comes from an original or a target text (with the acronyms OF, TF, OE and TE). On the other hand, the examples collected from tools which do not distinguish the original or source language from the target one will be simply accompanied with a notification of their language (FR for French and ENG for English).

In this analysis, the term *literal* will also be used when a correspondence is a, idiomatic word-for-word translation, such as example (2) above.

### 1.3.2.3. Correspondence measurements

Altenberg & Granger (2002: 17) explain that when comparing two items in different languages, various methods permit to circumvent the problem of “superficially corresponding structures” (Ebeling & Ebeling 2013: 24). First of all, **quantitative measurement** is used to bring out recurrent correspondences (Kzreszowski 1990: 27). Altenberg & Granger (*ibid.*) warn that this might lead to disregarding “valuable evidence and miss the cross-linguistic insights that ‘unexpected’ translations often provide”. However, our primary goal being the evaluation of OBCs in terms of the translations they suggest, we believe that this method is not only relevant but also simply necessary.

A second method, namely **back-translation** (see Ivir 1983, 1987) “restrict[s] the comparison to forms in L2 that can be translated back into the original forms in L1” (Altenberg & Granger 2002: 17). By offering deeper insights into the degree of correspondence, this method will control the corpora’s findings for the direction French to English and further support or deny the OBCs’ quality. Concretely, the data for the back-translation will be collected by searching for the French CPs starting from target French (referred to as TF) in the PLECI\_news corpus using *ParaConc* (Michael Barlow), so as to spot where they come from (see example 5).

- |     |    |  |
|-----|----|--|
| (5) | OE | THE Lisbon summit in March 2000 made it explicit that the principal aim of European education policy was the production of profitable human capital <b>for</b> economic competitiveness .  |
|     | TF | [...] le Conseil européen de Lisbonne avait fixé comme principal objectif à la politique de l'Union en matière d'éducation de produire un capital humain rentable <b>au service de</b> la compétitivité économique. (PLECI_news) |

Finally, we combine these two methods to measure the degree of correspondence between two items in the different languages by calculating the **mutual correspondence**, which is described by Altenberg & Granger (2002: 17-18) as “a valuable diagnostic of the degree of correspondence between items or categories in different languages”. The calculation, which presupposes a parallel corpus such as the PLECI corpus, takes into consideration the number of times that the items translate one another as well as the number of occurrences of the items in the source texts, with the following formula:

$$\frac{(A_t + B_t) \times 100}{A_s + B_s}$$

The authors explain that, according to this measure, two items in two different languages will have to systematically be the translation of one another to have a mutual correspondence of 100% (which is extremely rare) and, conversely, their mutual correspondence will be of 0% if they are never translated by each other. However, it does not allow us to define what equivalence is and what is not. Also, Altenberg (1999: 262) warns that low mutual correspondence does not necessarily equate to “a gap or better choice in one of the language” but can reflect the fact that a certain item is translatable by various items. The three methods are implemented in Section 4.

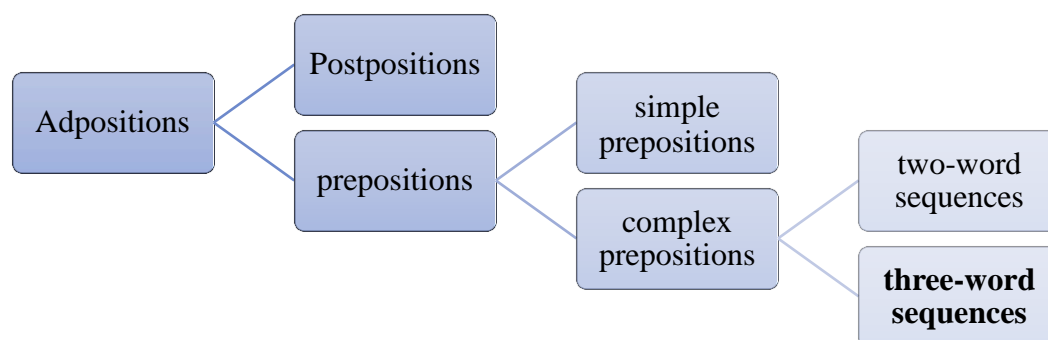
#### 1.4. ANALYSIS ON THE BASIS OF FRENCH COMPLEX PREPOSITIONS

Because the thesis concentrates on OBCs, the set of words selected to conduct the analysis required to assess their reliability will be presented here rather than in a later chapter. Together with their descriptions, we will suggest reasons for choosing to focus on French CPs in the process of grammaticalization.

In his book, Hoffman (2005: 5) disapproves of the absence of book-length investigations as well as of the lack of short studies on the subject of CPs (see Roy & Svenonius 2009 and Adler 2001). Mentioned in grammars as *group prepositions*, *phrasal prepositions* or *compound prepositions* (Hoffmann 2005: 26), CPs are one of the two possible types of prepositions, which, together with postpositions (e.g. *ago*) are part of the adposition group (see Hagège 2010). CPs are subdivided into two and three-word sequences (see Figure 2). The **two-word prepositions** are generally composed of an adverb, an adjective or a conjunctive followed by a preposition, simple and commonly used (e.g. *because of*) or two prepositions (e.g. *aside from*) (Brenda 2014: 63). **Three-word prepositions**, the focus of this thesis, consist of a preposition, a noun (accompanied by a determiner or not)<sup>19</sup> and another preposition, such as the examples *in view of*, *on the grounds of* or *with regard to* found in Quirk *et al.* (1985: 671).

---

<sup>19</sup> Campubri (1997: 188) concentrates on the presence or absence of determiner in his study, but this will not be the case in this thesis.



**Figure 2.** Classification of CPs within the adposition category.

The majority of CPs have a simple counterpart, e.g. *in spite of* vs. *despite* (Brenda 2014: 64 and Hoffmann 2005: 25) and function likewise (Gaatone 1976: 185),<sup>20</sup> i.e. they are followed by “a noun, pronoun or the ‘-ing’ form of a verb, and shows its relation to another part of the sentence” (Macmillan Dictionary).<sup>21</sup> For this reason and given that they are memorized as lexical units, CPs are to be considered as wholes (Hoffmann 2005: 2 and Pottier 1987). This is to be understood within the framework of the increased interest in recurrent word-combinations, referred to as the idiom principle (Sinclair 1991). It postulates that native speakers use institutionalized sentence stems retrieved from the long term memory, rather than entirely spontaneous ones, and in a greater extent than what had been suggested before.

In his research, Hoffmann (2005) analyses CPs from a functionalist viewpoint, studying their process of grammaticalization, already accounted for three centuries ago, as grammarians understood that grammatical words were derived from lexical ones (2005: 53). This vantage point allows for fuzzy boundaries and gradience between the different categories (Hoffmann 2005: 59) and accounts for a loss of identity (Campubri 1997: 186), a shift from concrete to abstract meaning or, more precisely, from spatial to non-spatial meaning in the case of CPs, (Marque-Pucheu 2001: 35), a loss of compositional meaning and a process of generalization (Hoffman 2005: 54). However, because some CPs allow great flexibility, their establishment as a grammatical category is still debated (Hoffmann 2005: 25), e.g. *in (hot) pursuit of*. Sinclair (1991: 109), however, argues that idiom expressions tolerate a considerable degree of variation. We will study the establishment of the CPs under consideration here, and to what extent it impacts their treatment in OBCs and OBDs.

<sup>20</sup> Earle (1892: 515) even argues that CPs are on the way of becoming simple ones.

<sup>21</sup> <http://www.macmillandictionary.com/dictionary/british/preposition>. Accessed on 17 March 2016.

## 1.5. STRUCTURE OF THE THESIS

The methodology of this analysis will be expounded in Section 2, along with a presentation of the corpora employed. We will then turn to a survey of online TTs, first with OBDs (Section 3.1.) and second with OBCs (Section 3.2.). These are presented last since, broadly speaking, they are a combination of dictionaries and corpora. The analysis will be presented in Section 4, before the conclusion in Section 5. Section 6 lists the bibliographic references as well as the online resources and is followed by appendices.

## 2. DATA AND METHODOLOGY

The corpora, material and the methodology will be described in this chapter. First, we described the different corpora used as a support of the correspondences found in the OBCs and OBDs. The final data used is then presented, followed by a description of the methods adopted.

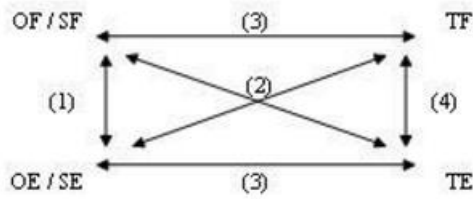
### 2.1. CORPORA

In the present research, the Label France corpus<sup>22</sup> and the PLECI corpus, both rather small by today's standards, are used. The first is a unidirectional translation corpus providing texts in original French (OF), matched in terms of genre (magazine articles on politics, economics, culture, technology and tourism as well as transcribed interviews) and time (between 1998 and 2008), together with their translations into English. The original texts contain exactly 823,996 words, while the translated ones consist of 834,790 words. These were automatically aligned at sentence level with *Alinea* (Olivier Kraif) and can be used with the concordancer *ParaConc*.

The second corpus, the PLECI (for Poitiers-Louvain Échange de Corpus Informatisés), is a parallel corpus, i.e. it is a combination of comparable and translation data for French and English. It was collected by the University of Louvain and the University of Poitiers and includes literary prose (which will not be used in this thesis, as a pilot study produced scarce occurrences of CPs) and newspaper articles. The latter contained 394,995 words in OF (original French), 353,985 in TE (target English), 470,936 in OE (original English) and 552,228 in TE (target English) when collecting the data for the present research, but the data is continually increasing. The corpus allows for different comparisons: between (a) original languages, (b) target languages, (c) French and English in both directions of translation, (d) French as a source and a target language, and (e) English a source and a target language (see Figure 3). It is both sentence and paragraph-aligned.

---

<sup>22</sup> <https://www.uclouvain.be/en-258636.html>. Accessed on 4 January 2016.



**Figure 3.** Possible comparisons in the PLECI corpus.

## 2.2. MATERIAL

As a random selection of French CPs would have biased the objectivity of this thesis, they were selected from a list of n-grams created by Granger and Lefer (2012). For their article on phrasal entries in bilingual dictionaries, these authors have extracted all 2-5 French n-grams with a minimum frequency of 20 from the Label France corpus. 6000 n-grams were collected with this corpus-driven method, but only the complete and relevant ones constitute the final list of 422 n-grams. From all the CPs included in that list, we pruned down our final material to those with the highest number of hits when combining the Label France and PLECI\_news corpora (see Table 1).

<b>Corpora</b> <b>French CPs</b>	<b>Label France</b>	<b>PLECI_news</b>	<b>Total</b>
<b>En matière de</b>	131	35	<b>166</b>
<b>Au sein de</b>	87	39	<b>126</b>
<b>Par rapport à</b>	55	19	<b>74</b>
<b>Dans le cadre de</b>	50	16	<b>66</b>
<b>En faveur de</b>	41	16	<b>57</b>
<b>À la fin de</b>	37	10	<b>47</b>
<b>Dans le domaine de</b>	34	7	<b>41</b>
<b>En raison de</b>	26	14	<b>40</b>
<b>Aux côtés de</b>	35	4	<b>39</b>
<b>Au service de</b>	35	4	<b>39</b>

**Table 1.** Number of occurrences of the CPs in Label France and PLECI\_news corpora.

## 2.3. METHODOLOGY

Because the analysis involves the support of corpora, corpus linguistics will play a primary role. Two approaches are possible when working with corpora, namely corpus-based and corpus-driven approaches, which are both in strict opposition with intuition-based methodologies. Altenberg & Granger (2002: 15) explain that the first is an umbrella term for both but, in a more restricted sense, it refers to research starting from a precise hypothesis and



using the data to either infirm or confirm their postulate. On the other hand, the second approach analyses corpora “with minimal theoretical presuppositions” (Hunston & Francis 2000: 318). Although the line between the two is rather thin, this thesis employs the first approach, as it uses corpora to support or reject the reliability of the translations suggested in OBCs. Finally, our analysis is also based on contrastive linguistics, and more precisely on the methodology described in Krzeszowki (1990: 35), since translations are at the core of the evaluation of the reliability of OBCs.

### 2.3.1. SELECTION OF THE COMPARED ITEMS

In Section 1.3., we discussed French CPs, a grammatical category which gives rise to translation problems and calques with “a preliminary characterisation of these in terms of some language-independent theoretical model” (Altenberg & Granger 2002: 14). The selection of this category of words lies on two assessments. First, the alarming observation on the poor phraseological coverage of dictionaries (Granger and Lefer 2012) and, more precisely, their ill-presentation of prepositions (Cosme & Gilquin 2008: 259, Brala 2002: 1), and second, the need for more examinations of prepositions through a contrastive lens (such as the study of Norwegian CPs and their translations into English and French by Egan & Graedler 2015). Cosme & Gilquin (2008: 271) note that such research would drastically improve the field of bilingual lexicography as well as second-language acquisition, as prepositions cause difficulties to learners at all levels (Désilets *et al.* 2008b: 1-2), for which Cosme & Gilquin suggest the following explanation:

While they [prepositions] are often regarded as having clear translation equivalents in most languages, a detailed analysis usually reveals a large number of language-specific uses. (Cosme & Gilquin: 2008: 261)

Désilets *et al.* (2008b: 2) even indicate that there is indeed no one-to-one correspondence for prepositions from one language to another, which proves problematic given their high frequency. These translation problems, also mentioned in Chuquet & Chuquet (2006: 189), could be due to the distinct utilisation of “the common cognitive endowment of prepositions” by the different languages (Zelinski-Wibbelt 1993: 20), which implies likely deviations with their metaphoric extensions (Lindstromberg 2001: 82). Other factors also play a role in the difficulty to translate prepositions, namely phraseology and polysemy (Cosme & Gilquin 2008: 266), along with the heavier and wider use of prepositions in English (Downing & Locke 1992: 580), which forces French to resort to alternatives (Cosme & Gilquin 2008: 263).

Lewis (1997: 64) also postulates that multi-word sequences are rarely rendered literally and the proportion of congruence is even lower when translating restricted combinations in comparison with freer ones (Nesselhauf 2003: 236). All these features combined explain why learners have difficulties in using prepositions correctly, and also greatly support the need for an extended analysis on their use and their translations.

### 2.3.2. JUXTAPOSITION

The second stage of Krzeszowki's methodology consists in identifying cross-linguistic correspondences, which is also referred to as translation spotting (Simard 2003: 65). The identification of the translations in the corpora will first take place. To this end, each French CP will be searched for in the OF section of the Label France and the PLECI\_news corpora using *ParaConc*. The output will then offer the pairs of sentences containing the queried term in OF, matched with the corresponding translated English sentences. We must identify the correspondence in each sentence pair manually, as there are not aligned at word level and hence not highlighted (see Appendix 1 for an illustration of *ParaConc*).

For *Linguee*, *TradooIT*, *WeBiText* and *ReversoContext*, the analysis will take into account the first 30 results provided by each tool.<sup>23</sup> In *TradooIT* and *ReversoContext*, spotting the translations is faster because of the highlighting option but still requires manual verification. In *Linguee* and *WeBiText* however, this manual work is inevitable, since highlighting completely lacks accuracy in the first and is simply absent in the second. Also, *TradooIT* and *ReversoContext* resort to a second method to show translations, i.e. by listing them before the pairs of sentences together with raw frequencies (see Section 3.2.3.9. for more details), which will also be examined. This will permit a consideration of the translations through more abundant data and also show whether the first examples provided are representative of the whole data available on those two OBCs. The translations will also be identified in the OBDs (*Larousse*, *Reverso-Collins* and *Oxford*).

### 2.3.3. COMPARISON

In this last stage, the translations suggested by the OBCs and OBDs will be contrasted with the ones found in the corpora, which will shed light on the pros and cons of each type of tools and indicate whether the OBCs (and the OBDs) reflect authentic usage, i.e. whether they are

---

<sup>23</sup> This implies that pairs of sentences containing more than one occurrence of the queried CP will count as several results. Generally, however, the first 30 results correspond to the first 30 pairs of sentences offered in the tools.

reliable. Also, the “degree and type of correspondence between the compared items” will be evaluated in this comparison stage (Altenberg & Granger 2002: 14), using the measurements covered in Section 1.3.2.3. This will hopefully give new insights into the similarities and divergences of French and English and improve our knowledge of both individual languages regarding CPs. Concretely, the data for back-translation will be collected by searching for the French CPs starting from TF in the PLECI\_news corpus using *ParaConc*. The parallel sentences will reveal what the French CPs have as their sources in OE.

The TTs will also be studied and compared from a macro-perspective, i.e. analyzing variety of context, target language correctness, highlighting accuracy, alignment errors, etc. The macro-quality will greatly influence the degree of trust users can put in the OBCs, and will determine whether one of the tools outperforms the others and/or if they seem to be equal regarding macro-quality or lack thereof.

### 3. SURVEY OF ONLINE TRANSLATION TOOLS

This chapter will provide information on online TTs, focusing on OBDs and OBCs. More emphasis will be placed on the second, as we will endeavour to delineate and define this type of tool in a first stage, and describe the tool reviewed here (*Linguee*, *TradooIT*, *WeBiText* and *ReversoContext*) in a second. The literature available on these OBCs will be provided at the end of the section.

#### 3.1. ONLINE BILINGUAL DICTIONARIES

##### 3.1.1. OVERVIEW AND PREVIOUS WORK

Due to their growing popularity, OBDs have been the focus of extended studies (see García Hernández 2014), as has electronic lexicography in general (see Granger & Paquot 2012). In fact, they have become so widely used that several publishing houses, such as Macmillan and Oxford, have declared that printed version of their dictionaries will no longer be published (Rundell 2015: 5). Helpful to the present investigation is Gelpí's (2004) evaluation of English-Spanish OBDs, where she endeavours to establish a list of factors determining their reliability.

In particular, an online bilingual dictionary must have a real and public author; it must be oriented to the main addressee or users; it must be designed according to its lexicographical function and main objectives; it must be adequately organized with regard to hyperstructure, access structure, macrostructure, microstructure and iconic structure. At the same time, a good online bilingual dictionary should be usable, updated, hypertextual and should offer some degree of satisfaction. (Gelpí: 2004: 10)

Gelpí (2004: 3) argues that the popularity of OBDs, while offering many advantages (they are “easy, quick and cheap”), also creates numerous disadvantages, among which the presence of low-quality products. These can prove difficult to discern for users, even more so if these users do not question their quality in the first place. While Gelpí's criteria are undoubtedly valuable when assessing the reliability of an OBD, there is no mention of the main object of such tools, namely the quality of the translations they provide.

Müller-Spitzer *et al.* (2011) also throw light on online dictionaries, through a user-survey (1,074 users) testing various useful criteria to distinguish high-quality dictionaries from low-quality ones (2011: 204). Whilst several of these are directly relevant to this thesis (e.g. clarity, links to other dictionaries, adaptability, speed, up-to-date content, corpus integration, etc.), one is precisely the object of this thesis, namely the reliability of the content. Müller-

Spitzer *et al.* (2011) pave the way for the current analysis, as they show in their study that this particular factor is rated as the most fundamental one by the users.

In their study on the phraseological needs of advanced learners, Granger and Lefer (in press) have, as in this research, selected a type of words which gives rise to translation problems, namely metadiscursive lexical bundles (such as *au bout du compte*, *en l'occurrence*, *sans parler de*, etc.), and analysed them in the same OBDs as in this study and in contrast with corpora. In their results, the authors conclude that the bilingual dictionaries often suggest calque translations<sup>24</sup> which differ from what can be found in parallel corpora. Their analysis also reveals that translations can be of a different order (i.e. *en fin de compte* translated by an adverb such as *ultimately*).

### 3.1.2. DESCRIPTION OF EACH OBD

*Larousse* offers several services on its online platform, amongst which linguistic tools, namely a French monolingual dictionary and a bilingual dictionary from French to English, Spanish, German, Italian, Chinese and Arabic. The French – English section offers 250,000 words and expressions as well as 400,000 translations. The entries display pronunciation and conjugated forms, and provide hyperlink for every word. Quite a lot of advertisements clutter the page, even though it remains intuitive and clear.

*ReversoDictionary* offers the languages French, Spanish, Dutch, Italian, Russian, Japanese, Chinese, English, German, Arabic, Portuguese, Hebrew, Polish and Korean, all combinable. In the first section of the entries, the OBD shows how the word or the phrase is translated, and then all the results from the community are presented separately, in a second section. The third section displays the entry of the *Collins* dictionary (the one reviewed here), followed by a very useful section of results from English to French. The last section provides translations in context (rooting from *ReversoContext*). Finally, it includes an incremental option (i.e. a search engine which predicts words and phrases according to the letters already typed in by the user)<sup>25</sup>,

The *Oxford Dictionary* is a charged platform of various linguistic resources, such as thesaurus, grammar, monolingual, bilingual dictionaries, etc. There is no incremental search,

---

<sup>24</sup> “a loan translation, especially one resulting from bilingual interference in which the internal structure of a borrowed word or phrase is maintained but its morphemes are replaced by those of the native language” (Dictionary.com). <http://www.dictionary.com/browse/calque>. Accessed on 23 March 2016.

<sup>25</sup> <http://www.thefreedictionary.com/Incremental+search>. Accessed on 2 January 2016.

but the tool provides the pronunciation and genre of each word, along with hyperlinks. There are no advertisements in the research pages.

## 3.2. ONLINE BILINGUAL CONCORDANCERS

### 3.2.1. INTRODUCTION

In the literature, *Linguee*, *TradooIT*, *WeBiText* and *ReversoContext*, which are not academic products, are arbitrarily referred to as (bilingual/parallel) concordancers (Goulet *et al.* 2012, Bowker 2012: 394, Delisle *et al.* 2013, Désilets *et al.* 2008a; 2010 and Simard 2013: 17), “website(s) for language learners” (Baisa *et al.* 2014: 63), online parallel corpora (Kübler 2013), web-based search tools (Volk *et al.* 2014), search engines (Portal 2011, Grauer 2010: 3, Simard 2013: xvii), online translation help tools (Taravella 2011: 7, Désilets *et al.* 2010: 1), translation memories (referred to as TMs) (Désilets *et al.* 2008a), “google of parallel texts” (*ibid.*), multilingual dictionaries (Bouchard 2012) or computer-assisted translations tools (or CAT tools) (McDuff 2011). The scientific articles and reviews also generally overlook the inherent dissimilarity between the different tools. For example, Goulet *et al.* (2012) grouped together *TradooIT*, *TransSearch*, *WeBiText*, *TextStat*<sup>26</sup> and *Wordsmith Tools*,<sup>27</sup> while setting *Linguee* apart in the category of translation and localization tools.

Faced with this terminological chaos and fuzzy boundaries, we need to delineate the category of the tools under consideration with respect to other existing TTs and to rectify the lack of steady label (see Section 3.2.2.). Once the subject of interest has been delimited, an assessment of the tools will be presented in Section 3.2.4. Before this academic discussion, a description of each website will be included, so as to enable a better visualisation of the tools (see Appendix 2 for a summary of their features).

### 3.2.2. TERMINOLOGICAL CLARIFICATION

None of the terms mentioned above is incorrect in essence, but there are all either too imprecise or restricted. A good start for situating the tools among the plethora of TTs is therefore to first define and grasp this category (see Appendix 3 for a representation of the tools).

Opinions greatly differ concerning their use, with several authors partially criticizing the fusion between translation and technologies (see Hutchins 2001, Kübler 2013, Precup-

---

<sup>26</sup> Software producing word frequency lists and concordances.

<sup>27</sup> This paid software is meant to search patterns in a language.

Stiegelbauer 2013 and Volanschi 2007) and others enthusiastically investigating it (see Désilets *et al.* 2008a, 2008b, 2009 and Lagoudaki 2006). Froeliger (2007) endeavors to distance himself from this alleged fight opposing humans to machines through a presentation of the potential of translation technologies.

Second, it is paramount to distinguish the two different types of TTs, namely **computer-assisted translation tools** and machine translations (often referred to as MTs). The latter are defined as “automated translation or translation carried out by a computer”,<sup>28</sup> such as *Systran*.<sup>29</sup> Machine translations have received massive attention in the literature (see Brown *et al.* 1990, Kay 1997, Somers 1999, Hutchins 2001 and Kübler 2007) and will not be further discussed here. The former, on the other hand, encompass “software tools built to help translators”,<sup>30</sup> and are therefore the umbrella term for the tools investigated here, along with electronic dictionaries, spell checker programs, TMs, concordancers, alignment software, terminology banks, etc.

As they contain authentic examples of previously translated texts, the tools reviewed here have no connection in relation to online automatic aid, i.e. which are performed by a computer program (e.g. *Google Translate*,<sup>31</sup> *Bab.la*,<sup>32</sup> *Free Translation*<sup>33</sup> or *Babelfish*).<sup>34</sup> This may imply a post-editing stage from the users so as to adapt the findings and reach the benefit result, as it is not possible to type full sentences in the search bar and thus no word-for-word translations of their query (Mansfield 2013: 5-6, Xhark 2010). For that reason, Alonso Jiménez suggests (2013: 9) a great resemblance to **TMs**, defined by Macklovitch & Russell as follows:

[...] particular type of translation support tool that maintains a database of source and target-language sentence pairs, and automatically retrieves the translations of those sentences in a new text which occur in the database. The broader definition regards TM simply as an archive of past translations, structured in such way as to promote reuse. (Macklovitch & Russell 2000: 137)

This description evokes the tools under investigation, since both systems “store and index previously translated content in an organised way, so that users can later retrieve from it as much information as possible when queried” (Lagoudaki 2006: 3). This analogy is also

---

<sup>28</sup> [http://kantanmt.com/documents/Machine\\_Translation.pdf](http://kantanmt.com/documents/Machine_Translation.pdf) . Accessed on 15 April 2015.

<sup>29</sup> <http://www.systransoft.com>.

<sup>30</sup> [https://www.transdraft.com/university/cat\\_tools](https://www.transdraft.com/university/cat_tools). Accessed on 15 April 2015.

<sup>31</sup> <https://translate.google.com>.

<sup>32</sup> <http://en.bab.la>.

<sup>33</sup> [www.freetranslation.com/fr/](http://www.freetranslation.com/fr/).

<sup>34</sup> <http://www.babelfish.fr>.

alluded to in the “about” page of *Linguee*, where the team writes that their tool offers “access to millions of texts translated by other people”.<sup>35</sup>

The system defined above is also referred to as a **concordancer**, which accounts for the wide use of the term to label *Linguee*, *TradoolIT*, *WeBiText* and *ReversoContext* in the above-mentioned studies. It appears to be a valid option, due to the possibility of searching terms and receiving results in the shape of a list of parallel sentences, although these tools provide additional services, such as bilingual dictionaries.<sup>36</sup> However, the term may give rise to ambiguity since it primarily refers to programs such as *ParaConc*.

Fifth, the term **corpus** cannot define the tools for the pure and simple reason that they are not proper corpora but only permit the investigation of (certain) corpora (Alonso Jiménez 2013: 6). In their case, the material is **harvested from the World Wide Web**,<sup>37</sup> which does not imply that the tools use the whole Web but rather have a more or less definite set of web-corpora, which further distinguishes our tools from other OBCs, such as *TransSearch*. Moreover, the four tools lack a crucial feature mentioned earlier, namely the distinction between source and target languages, as alluded to in Van Bolderen (2012).

The terms **website** or online services (Bowker 2012: 394), albeit noticeably too vague, imply another important tenet, i.e. the tools are not software to be bought or downloaded but directly accessible online. Also, the tools are virtually on a par with **search engines**, in the sense that they both search and return results according to specified keywords in a search bar.<sup>38</sup> The term, however, is not specific enough because it also embraces programs and systems such as *Google*<sup>39</sup> or *Startpage*.<sup>40</sup> The most glaring discrepancy lies in their primary goal to furnish segments of bi-texts, not full websites or PDFs (Van Bolderen 2012).

On the basis of all these attributes, the tools examined here can be described as non-automatic bilingual online tools which retrieve pre-existing and manual sentence pairs from the Web used as a corpus. This explanation is similar to Simard (2013: xvii), who describes them as “un outil pratique et convivial [...] à partir (duquel) les utilisateurs peuvent faire des recherches dans une collection de textes bilingues provenant du Web”. This definition is

---

<sup>35</sup> <http://www.linguee.com/english-french/page/about.php>. Accessed on 4 January 2016.

<sup>36</sup> According to Buyse *et al.* (2013: 509), the integration of the dictionary section allows “dictionaries to play a more prominent role again”.

<sup>37</sup> For a debate on whether the web is a corpus, see Kilgarriff & Grefenstette (2003).

<sup>38</sup> [http://www.webopedia.com/TERM/S/search\\_engine.html](http://www.webopedia.com/TERM/S/search_engine.html). Accessed on 22 January 2016.

<sup>39</sup> <https://www.google.be>.

<sup>40</sup> <https://startpage.com>.



applicable to more tools than those studied here, such as *LEO*,<sup>41</sup> *Globse*,<sup>42</sup> *TAUS*<sup>43</sup> or *MyMemory*.<sup>44</sup> These will not be reviewed however, because they do not meet all the criteria described above.

### 3.2.3. DESCRIPTION OF EACH OBC

This section is devoted to a presentation of the different OBCs, with a description of their operating<sup>45</sup> and specificities, but mostly focusing on their linguistic features, notably those described by Granger as “the profound changes brought about by the electronic medium” (2012: 2), namely corpus integration, more and better data, efficiency of access, hybridization, user input and customization. The latter is a key difference with respect to paper dictionaries (Granger 2012: 4), but the four OBCs are unfortunately neither adaptable nor adaptive.

#### 3.2.3.1. Overview

Both in the literature and more informal documentations, *Linguee* is often referred to as innovative (García Hernández 2014: 56), the first of its genre (Xhark 2010) and unique (Portal 2011, Schoppman 2014). We will try to fathom the reasons for the predominance of this German<sup>46</sup> tool, created by two computer engineers and first released online in 2009 for the German-English language pair in a beta version. Green (2011) advertises that one year after its official release in May 2010, the website<sup>47</sup> rapidly became the largest and most frequented bilingual dictionary in the world. On the “about” page, the creators describe the website as being fast, user-friendly and reliable, but there are no empirical tests of this alleged quality to date. They also emphasize the variety of contextual registers at play, a statement that cannot be supported here.

Less information is to be found for the three remaining tools, but Simon McDuff, computer engineer and creator of *TradooIT*, informs us that the Canadian tool was first released in a reduced version available in French and English in 2011, and has now evolved into a more

---

<sup>41</sup> [http://dict.leo.org/frde/index\\_de.html](http://dict.leo.org/frde/index_de.html).

<sup>42</sup> <https://fr.glosbe.com>.

<sup>43</sup> <https://www.taus.net>.

<sup>44</sup> <https://mymemory.translated.net/français/>.

<sup>45</sup> For more information, it is possible to consult the Help section in the websites, which is very complete and thorough in *TradooIT* (<http://wiki.tradooit.com>, Accessed on 11 January 2016.). There are also tutorials available on *Linguistech* for *Linguee*, *TradooIT* and *WeBiText* ([http://linguistech.ca/WeBiText\\_Individual\\_E\\_TUTCERTT\\_I\\_PartI](http://linguistech.ca/WeBiText_Individual_E_TUTCERTT_I_PartI), Accessed on 15 June 2015). The creators also provide instructions for the latter in Désilets *et al.* (2008a).

<sup>46</sup> As Geyken notices (2015: 16), Germany is quite prolific in the field of electronic lexicography, with, in addition to *Linguee*, *LEO*, *dict.cc* and *PONS*, which has been online since 2001 and contains 10 million words and phrases.

<sup>47</sup> In addition to the online version, a mobile version and a brand-new application are now available.

hybrid and multilingual tool.<sup>48</sup> The idea of *WeBiText*, on the other hand, dates back to 2005 (Franz 2011: 18), and was created in Canada in 2008. Its creators are published linguists, who promote it as relevant for both general and specialized language (Désilets *et al.* 2010: 6).<sup>49</sup> No information is to be found regarding the beginning or the creators of *Reverso*.

### 3.2.3.2. Languages

- *Linguee*

Shortly after the launch of the tool, French, Spanish and Portuguese were added to the German and English languages. Portal (2011) claims that there were no similar tools for the French language, but *WeBiText* was already available. In 2013, Italian, Russian, Chinese, Japanese, Dutch, Mandarin and Polish became available as well, “cover(ing) the mother tongues of more than 78% of all Internet users worldwide”<sup>50</sup> (Green 2013) and “targeting one quarter of non-native English-speaking Internet users” (Portal 2011). In 2014, the website introduced new official EU languages so as to guarantee its internationalization (Schoppman 2014). As of 2016, however, there are still no signs of Norwegian, Irish or Croatian. It is also important to note that the combinations are restricted, i.e. English can be combined with any language but Japanese, Chinese and Russian are only combinable with English, which amounts to a total of 234 language combinations. An interface is available for every language, which facilitates its access and use to non-English speakers.

- *TradooIT*

*TradooIT* offers a rather small possibility of language pairs, namely English, French and Spanish, but they are all-combinable. Paradoxically, there is no interface in Spanish, which might encumber the use of the website for users not proficient in French and English. Also, it seems that more information is displayed in the French interface.

- *WeBiText*

30 languages are provided in *WeBiText*, namely Arabic, Bulgarian, Chinese, Czech, Danish, Dutch, English, Estonian, Finish, French, German, Greek, Haitian creole, Hungarian, Inuktitut, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Rumantsch, Russian, Slovak, Slovenian, Spanish, Swedish and Welsh. The creators

---

<sup>48</sup> <http://wiki.tradooit.com/home/dou-vient-tradooit>. Accessed on 11 January 2016.

<sup>49</sup> 18% of a sample of 100 queries was judged to pertain to specialized language in one of their study. General language problems still seem to dominate, however (Désilets *et al.* 2010: 5).

<sup>50</sup> They used this source: <http://www.internetworldstats.com/stats7.htm>. Accessed on 8 January 2016.

acknowledge that the languages, which are not all-combinable, are not served equally, the best supported pair being English – French (Désilets *et al.* 2010: 3, Gallimore 2011). Désilets *et al.* (2010: 4) have observed that users use the pivot technique, i.e. they use an intermediate language between a less-documented language pair, which “implies that the tool can be useful, even for languages for which there is not a lot of parallel data on the web”. We believe, however, that (a) counting on this method is hazardous given the transfer errors that will most probably be included in the process and that (b) this task should be offered to users directly, not delegated to them.

- *ReversoContext*

The bilingual concordancer is available in English, Spanish, Italian, German, Russian, Portuguese, Hebrew and Japanese.<sup>51</sup> The interface can be set in French, English, Spanish, German, Chinese, Portuguese, Romanian, Czech, Italian and Polish.

### 3.2.3.3. Process: data collection and alignment system

The qualification of OBCs as non-automatic mentioned earlier is in fact specific to the type of data provided in their platform, i.e. manual sentences from previously translated texts, not to the process employed to collect these resources. This process is based on an automatic web crawler extracting external examples from bilingual websites.<sup>52</sup> The pairs of sentences are then checked, either manually as in *WeBiText* (Désilets *et al.* 2010: 7), or automatically as in *Linguee* and *TradooIT*. *Linguee*'s main objective is to eliminate unreliable translations, such as those rooting from automatic translation. According to the website, the data of *ReversoContext* is not revised.

The OBCs are aligned at the level of the sentence but *TradooIT* and *ReversoContext* offer evidence of a certain degree of sophistication regarding word-alignment. More accurately, they highlight the queried term as well as its translation, which is impossible in the corpora, albeit with a certain degree of imprecision (Volk *et al.* 2014: 3174). *ReversoContext* is however able to distinguish the English items which appear in the target sentence as translation of the French queried term from those which are not<sup>53</sup> (see Figure 4).

---

<sup>51</sup> [http://www.ReversoContext.net/text\\_translation.aspx?lang=FR](http://www.ReversoContext.net/text_translation.aspx?lang=FR). Accessed on 14 January 2016.

<sup>52</sup> In addition to this automatic process, *WeBiText* also uses an on-demand process for a small proportion of its data, which is slower but facilitates the incorporation of new domains (Désilets *et al.* 2008: 5)

<sup>53</sup> The tool also makes the distinction in the total of results.

Des pratiques optimales en matière de prévention ont également été recensées dans divers domaines.

Best practices in prevention have also been identified in various areas.

**Figure 4.** Efficiency of the word-alignment and highlighting options in *ReversoContext*.

#### 3.2.3.4. Corpus integration

While it has become unthinkable to conceive a (monolingual) dictionary without corpora at hand (De Schryver 2003: 167), the corpus integration now also works “downstream, as an integral part of the electronic dictionary to which users have direct access and which they can mine for themselves” (Granger 2012: 3).

- *Linguee*

Direct corpus access is extremely restricted in *Linguee*, if not non-existent, as users are limited to the first 30 pairs of sentences,<sup>54</sup> with no possibilities to consult further pages of results anymore. In addition to this major downside, the website fails to provide users with any information with respect to its corpora, in terms of size or description. The *Linguee* team explains that its data stems from companies’ websites, certificates, government and parliament documents, scientific publications (Portal 2011), but also “PDF, patent specifications, multilingual commercial websites, marketing papers and [...] highly specialized online shops” (Green 2011, Schoppman 2014). However, as Gallimore (2011) complains, even a clear list is nowhere to be found. As a consequence, users miss out on the very point of using corpora, i.e. assessing the frequency of one possible translation against the others, comparing frequencies across corpora, selecting and working with a specific register, etc. In view of the material, we can presume a certain limitation concerning registers in any case, as there is no data from news, fiction or speech, which may lower the number of variants and therefore reduce the perspective users can have on the translation paradigm of a particular word or MWU. However, Frahling, one of the two creators of the tool, argues that *Linguee* is useful for domain-specific structures because of the large amount of genres it contains that need be expressed in English (Portal 2011). This is corroborated by Arcan *et al.* (2013) and Xhark (2010).

- *TradooIT*

*TradooIT* is by far the most accurate OBC with regard to corpus information, as it provides a clear list of the 62 corpora employed, together with their code, title, description, last update,

---

<sup>54</sup> This number can even lower in some cases, varying between 26 and 30.

number of documents, of segments and of words<sup>55</sup> (see Figure 5). This information adapts to the language pair selected and its direction, attesting the precision of the website. Because of this precision, the website is able to inform users of the exact frequency of each item when queried. Unfortunately, the frequency of the item's translations is only available for those which are listed in the grouped translations (see section 3.2.3.9 for more details). It is possible to select a specific corpus in the left column of the interface (see Figure 6), or from the search bar. The first option might be more useful in some cases, as the column only provides the corpora containing occurrences of the queried term. The second, on the other hand, is a quick way of knowing whether a queried term exists in a specific corpus. In both cases, the frequency is provided, as we can see in the illustration.

Code	Titre	Description	Dernière mise à jour	Documents	Segments	Mots
NEWSGCCA	Centre des nouvelles du Canada	Toutes les nouvelles Internationales, nationales et régionales depuis 1997.	Il y a 15 heures	55 866	4 001 811	31 175 467
CINEUROPA	CINEUROPA	Cineuropa est le premier portail européen consacré au cinéma et à l'audiovisuel publié en quatre langues.	04 avril 2016	20 537	258 157	6 284 370
TED	Ideas worth spreading	Ideas worth spreading	04 avril 2016	1 880	187 138	3 904 286

**Figure 5.** Corpus information available from French to English in *TradooIT*.

#### Sources

HANSARD [4781]  
 CORDIS [4687]  
 EUROPARL [4619]  
 NEWSGCCA [4549]  
 OECD [2620]  
 UNESCO [2398]  
 LOICAN [2363]  
 NEWSONCA [2090]  
 STATCAN [1854]  
 ICGCCA [1666]

**Figure 6.** Frequency of *en matière de* according to the corpora in *TradooIT*.

- *WeBiText*

89 corpora are integrated to *WeBiText*, 63 more than 6 years ago (Désilets *et al.* 2010: 3). They notably stem from the Canadian government and the European Union websites, but no exact figures are provided. The only information is to be found in Désilets *et al.* (2010: 3),

<sup>55</sup> [https://www.tradooit.com/corpus\\_stats](https://www.tradooit.com/corpus_stats). Accessed on 3 February 2016.

who indicate the size of the main domains, which are between 181 million and 106,000 pages. However, *WeBiText* affords the opportunity to choose among the different registers and corpora, directly from the homepage and arguably more readily than in *TradooIT* (see Figure 7). The possibility to load one's own corpus and analyse it is a unique feature of the OBC that is nowhere to be found in other similar tools.

And a predefined corpus from our list:

<b>All sites in category General Corpora</b>	[-]
click here to open or close the list of corpora and then click on corpus name to select it	
No predefined corpus	
<b>.ca domain</b>	[+]
<b>Finance</b>	[-]
All sites in category Finance	
Bank of Canada ( <i>bankofcanada.ca</i> )	
Canada Mortgage and Housing Corporation ( <i>cmhc-schl.gc.ca</i> )	
Desjardins ( <i>www.desjardins.com</i> )	
European Central Bank ( <i>www.ecb.int</i> )	
L'Autorité des marchés financiers ( <i>www.lautorite.qc.ca</i> )	
TMX ( <i>tmx.com</i> )	
<b>General Corpora</b>	[+]
<b>Government of Canada</b>	[+]
<b>Health</b>	[+]
<b>Legal</b>	[+]

Figure 7. List of corpora in *WeBiText*.

- *ReversoContext*

The data contains millions of translated texts from dialogues, official organisations and other multilingual websites (such as [www.opensubtitles.org](http://www.opensubtitles.org)), thus providing users with formal as well as informal language. A list of the 19 corpora used is provided with a reference but no information about frequency<sup>56</sup> (see Figure 8). The creators plan to both create a filter for users to select a specific corpus and to enlarge the domains they cover.<sup>57</sup>

<sup>56</sup> <http://context.reverso.net/traduction/legal>. Accessed on 14 January 2016.

<sup>57</sup> <http://context.ReversoContext.net/traduction/about>. Accessed on 14 January 2016.

## Mentions légales pour les corpora du dictionnaire contextuel

✓ **Corpus:** opensubtitles2012. **Licence:** non spécifiée. **Références:** <http://www.opensubtitles.org/>, Jörg Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012).

✓ **Corpus:** multiUN. **Licence:** non spécifiée. **Références:** <http://www.euromatrixplus.net/multi-un/>, MultiUN: A Multilingual corpus from United Nation Documents, Andreas Eisele and Yu Chen, LREC 2010.

**Figure 8.** Corpus information in *ReversoContext*.

### 3.2.3.5. More and better data

One of the most significant improvements brought about by online TTs is the ability to overcome the previous space limitation imposed on paper dictionaries. Therefore, more data, examples and multimedia resources can be included, while collocations can eventually receive the attention and space they deserve. Great care must be taken, however, not to overwhelm users with information (Granger 2012: 3). In the four OBCs studied here, the most obvious addition of data as compared to printed language tools is the inclusion of examples. *WeBiText* does not seem to exploit this possibility of having more space and resources within the same interface.

### 3.2.3.6. Users

Users are not the focus of this research, but we need to consider the user-friendliness and the targeted type of users in our evaluation. The creators of *WeBiText* indicate that their tool might best serve freelance translators or those working in small or medium organisations, as they may lack access to large parallel corpora (Désilets *et al.* 2008a: 1). Franz (2011: 18) believes that *WeBiText* falls short as a public service, its main target being researchers in any case. Likewise, *TradooIT* first aims at language experts rather than language learners. However, McDuff argues that the bilingual concordancer can also prove useful to any type of user.<sup>58</sup> On the other hand, the aim of the *Linguee* team is to offer a polyvalent tool, which is not an advantage according to Granger & Lefer (in press).

Users who need to write texts in a foreign language on a daily basis turn to *Linguee* for help; professional translators seek inspiration for the best-matching terms and phrases with *Linguee*, too. (Schoppman 2014)

---

<sup>58</sup> <http://wiki.tradooit.com/home/-qui-sadresse-tradooit>. Accessed on 5 October 2015.

### 3.2.3.7. User input

Grauer (2010: 3) views online TTs as a combination of “the Wiki-system of user-generated content (and) traditional approaches”. The crowdsourcing system is nevertheless to be put into perspective, as users can only rate translation results, with a reduced, even non-existent authority as far as editing is concerned. Désilets *et al.* (2007: 11) note that “when dealing with collaboration at that kind of scale, our intuitions about what can and cannot happen are often wrong”. Paradoxically, *WeBiText* seems to be devoid of any sort of collaborative work. The only possible interaction is through a feedback section appearing below the search bar after a search where the users can assess the results as useful or not. In *TradooIT*, the community aspect is limited to a thumb-up/thumb-down rating system, whose use verges on 0%.

- *Linguee*

Users used to have the opportunity to rate each sentence pairs through a thumb-up/thumb-down system, which put the best-rated results in the first position.<sup>59</sup> The remnants of this system, i.e. the sign ⚠ warning that the translation could be wrong, still exist but there is no longer any ✓ next to a sentence pair to inform users that it has been checked. According to Volk *et al.* (2014: 3174), these “[were] a welcome functionality since many automatic alignments on the sentence level and subsequently on the word level are incorrect”. As opposed to what the website advertises, the section *Contribute* has disappeared as well, along with the possibility for users to suggest entries for the dictionary. The feedback Section is still available however. The disappearance of the collaborative options, highly valued by users, is quite surprising since Fink, the second developer of *Linguee*, had earlier stated that users’ recommendations were at the root of their success (Portal 2011).

- *ReversoContext*

*ReversoContext* is undoubtedly the most collaborative OBC. After creating a personal account, users can contribute to the website, but also create their own list of vocabulary, a unique feature which increases the user-friendliness of the tool. The creators indicate that the collaborative dimension is crucial to the proper functioning of the website as well as its improvement, as registered users can (a) give their opinions on the translations, (b) add words and expressions in the collaborative dictionary, (c) edit other users’ entries, (d) comment and

---

<sup>59</sup> <http://www.commentcamarche.net/news/5852780-linguee-com-nouvel-outil-de-traduction-contextuelle-en-ligne>. Accessed on 28 October 2014.



(e) report problems. These options (see Figure 9) are easily accessible from each sentence pair, after hovering over it.

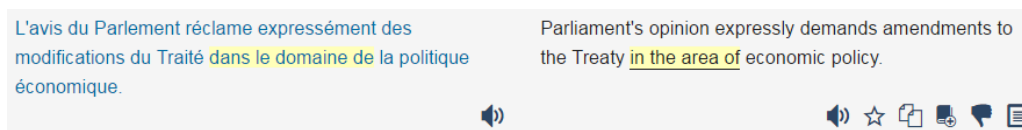


Figure 9. Community aspects in *ReversoContext*.

### 3.2.3.8. Hybridization

As Bowker explains (2012: 395), users need tools that are adaptable, and there is a “noticeable trend towards offering hybrid resources, where a dictionary is one part of a larger, integrated collection of language resources”, as in the OBCs. Back in the late 90s, Stig Johansson already had the idea of merging bilingual dictionary, parallel corpus and contrastive grammar together so as to compensate for their respective shortcomings and provide learners with better pedagogical resources (see Ebeling 1999). Ebeling (1999: 32-33) explains the difficulty of measuring up to the challenge as regards more complex grammatical research and polysemous words however, and strongly stresses the need for manual editing.

- *Bilingual dictionaries section*

With the exception of *WeBiText*,<sup>60</sup> the OBCs under investigation also provide a bilingual dictionary. In *Linguee*, it appears in a section above the pairs of sentences and contains information about gender in French, pronunciation, inflection<sup>61</sup> and examples of common word combinations. Alonso Jiménez (2013: 13) argues that it would be better described as a list of vocabulary, however. These two sections (namely the dictionary and the sentence pairs ones) can easily be reduced, which is useful with larger entries or when the user wants to focus on one of the two sections. Also, only part of the dictionary entry appears at a time, and it is the user who will have the choice to display more information. This responsiveness is characteristic of *Linguee* as compared to the other tools, which Lew (2010: 299) refers to as presentation space, i.e. the amount of data that can be displayed on screen to the user at a given time.

<sup>60</sup> The OBC only provides a link to *TERTIUM Plus*, a terminology and linguistic data bank ([http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&index=frt&\\_index=frt&srchtxt=dans+le+domaine+de](http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&index=frt&_index=frt&srchtxt=dans+le+domaine+de). Accessed on 1 April 2016).

<sup>61</sup> It is displayed for English verbs as well as English and French plurals (Green 2011).

On a more critical note, *Linguee*'s dictionary seems to leave room for improvement, especially regarding its phraseological relevance. Although a key objective of the creators, the results for MWUs often lack accuracy (see Figure 10). Note that this search does not suggest any results in the *TradooIT*'s dictionary, while *Reverso* provides an entire entry in its dictionary section.



**Figure 10.** Phraseological issues in *Linguee*'s bilingual dictionary.

The bilingual dictionary in *TradooIT* is more limited, and simply disregards phraseology altogether. Searches such as *en matière de* will be useless as far as the dictionary is concerned, for example, since it only offers translations of the noun *matière*. Indisputably, *ReversoContext* is the best resource regarding lexicography, as it is directly linked to the dictionary section of the *Reverso* website with a book icon next to the search bar. Users will then access the various sections of the dictionary as described in Section 3.1.2

- *Other services*

*TradooIT* and *ReversoContext* are certainly the most hybrid tools. The former offers supplementary services accessible from its homepage, arguably directed at professionals rather than language learners, namely a TM, a pre-translation system, an extension module for Microsoft Word and a system of data sharing.<sup>62</sup> The latter, on the other hand, comprises services for both everyday users (a spelling corrector, a conjugator, a grammar section and an automatic translator) and experts (a pro translator and a localizer, which are both new options). In contrast, *Linguee* only offers one recent additional option, namely a *Wikipedia* section (see Figure 11). This section only seems to appear with simple words, and not systematically.

<sup>62</sup> For more information, see <http://wiki.tradooit.com/guide-dutilisation>. Accessed on 17 October 2015.



Figure 11. Wikipedia Section in Linguee.

### 3.2.3.9. Operating of the OBCs

- *Interface*

Each OBC offers a help section and a feedback section in its interface, along with a possibility for changing the language pair and the interface language. *ReversoContext*<sup>63</sup> and *Linguee* have greatly improved their respective interfaces so as to become sober and intuitive. In comparison, *TradooIT* is perhaps less responsive, notably because of the numerous options and information displayed, but its interface remains clear and well-harmonized, as in *WeBiText*.

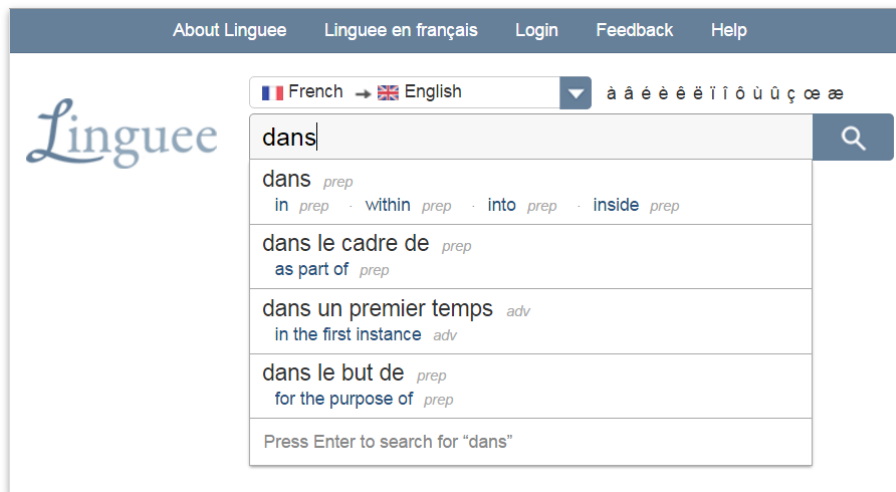
Overall, users are neither overwhelmed with an excess of data nor disturbed by unrelated or useless details. In fact, *TradooIT* and *WeBiText* are completely free from advertisements. In *Linguee*, users will receive 100 ad-free searches with the creation of their free account (Gallimore 2011), but the interface is not cluttered in any case. The excess of advertisements is most noticeable in *Reverso*, despite the creators' claim to restrain their presence for the good of the users.<sup>64</sup>

<sup>63</sup> While the design has become lighter in *ReversoContext*, the main homepage of the *Reverso* website might seem disordered because of its numerous tools and the excess of advertisements.

<sup>64</sup> <http://www.ReversoContext.net/more.aspx?lang=FR>. Accessed on 14 January 2016.

- *Search system*

The four tools function in a similar way, i.e. users type in their request (single words, MWUs, “rare expressions and specific technical terms”)<sup>65</sup> in the search bar and pairs of sentences are displayed accordingly. Unfortunately, *Linguee* and *ReversoContext* are the sole tools to really take advantage of the possible alternatives offered by online technology, such as incremental search bar. This accelerates the searching process since it presents a useful overview of the possible words or MWUs available (see Figure 12). As illustrated in the Figure, *Linguee* is more sophisticated in terms of information, providing the part-of-speech (referred to as POS) of the searched term as well as its possible translations together with their POS. If users type the whole query, more specific suggestions will be displayed, however without further information.



**Figure 12.** Incremental function in *Linguee*.

Except for *WeBiText*, queried terms can be refined using a series of metacharacters, which are generally combinable (see Appendix 4 for a list of the metacharacters). It is explained on *ReversoContext* that capital letters, accent and hyphens will modify the results (*peut-être* or *peut être*), even though the program is tolerant. Users are advised to search for conjugated forms rather than the bare form of the verbs.

As acknowledged by the creators, *WeBiText* only searches for the exact expression entered by the users (Désilets *et al.* 2010: 7), whereas the other OBCs are more flexible. The auto-correction is either inexistent or not optimal, even in *Linguee*, contrary to Gallimore’s statement (2011). In fact, *ReversoContext* seems to be the only one displaying relevant results

<sup>65</sup> <http://www.linguee.com/english-french/page/about.php>. Accessed on 4 January 2016.

when a mistake or a typo has been introduced in the search. Finally, the language detection only works when writing in one of the two languages of the language pair selected, i.e. none of the tools will recognize a query in Spanish if working with the French – English pair.

- *Efficiency of access*

Granger mentions (2012: 4) hyperlinks as a feature contributing to the efficiency of access, as they facilitate navigation inside and outside the tool, but this option is only available on *Linguee* and *ReversoContext*. More precisely, users can click on any word in the dictionary and in the sentence pairs to access its result page. In *ReversoContext*, this possibility is restricted to the highlighted words in the target language and the suggested translations underneath the search bar. *TradooIT* only offers the option for the words listed in the terminology section.

- *Sentence pairs*

The sentence pairs are displayed immediately after pressing *Enter*, except for *WeBiText* which requires a few seconds to provide the results. *TradooIT* and *ReversoContext* furnish the exact amount of time needed to collect the results as well as their total number (see Figure 13). As previously mentioned, *Linguee* is limited to around 30 examples, but the other three OBCs are virtually unlimited in terms of examples. For that reason, they display their respective sentence pairs separately, with ten per page in *TradooIT*,<sup>66</sup> 20 in *ReversoContext*<sup>67</sup> and a dozen in *WeBiText*, with the possibility to consult the remaining results in the other pages. Each example is accompanied by a URL, which enables the user to check the website where the example comes from and enquire the register.



**Figure 13.** Results information in *TradooIT*.

*Linguee* regularly informs users that the sentence pairs have not been reviewed, which contradicts the team’s claim that their learning-machine algorithm checks all the data. The *Linguee* team argues that checking each of the billion sentences would not be manageable.<sup>68</sup> Also, the team acknowledges that although the URL might “give a hint to the source and

<sup>66</sup> Users can change the numbers of results displayed in their settings.

<sup>67</sup> It is now necessary to have an account in *ReversoContext* so as to be able to consult further examples.

<sup>68</sup> <http://www.linguee.com/english-french/page/faq.php#sourceandtargetlanguage>. Accessed on 14 July 2015.

target language”, it is often impossible to tell the country or the original language of the website. The other tools’ results change when changing the direction of translation, but they do not precisely tell which one is the source either. Finally, *Linguee* proposes two distinct sections for each language when the queried term exists in both French and English.

- *Bi-texts*

*Linguee* is the only one to lack this option, but more contextual information is accessible when the sentences start with square brackets. The bi-texts are highly useful, since they do not run the risk of disappearing, as opposed to websites, because they are stored within the OBC’s data. Furthermore, *WeBiText* comes with a web icon allowing see both pages side by side (see Figure 14). This option seems to be one of the features most frequently praised by users, as it allows them to immediately get a sense of the type of document at hand and the importance of the word within the document (Désilets *et al.* 2010: 2).



**Figure 14.** Direct access to the parallel web pages in *WeBiText*.

- *Grouped translations*

This extremely valuable feature with respect to both reliability and user-friendliness, only available on *TradooIT* and *ReversoContext*, enables users to have a clear and rapid understanding of the correspondences of a particular query together with their frequencies (see Figures 15 and 16).<sup>69</sup> *TradooIT* outperforms *ReversoContext* by providing a precise number of its overall data, but the second seems more accurate. This option allows users to refine their search to one translation variant, whose exact frequency is provided, and access results where the queried term has been translated with this variant only. This can help users insofar as each translation is illustrated by examples separately, showing contextual differences more clearly, hence helping users choose the most appropriate translation. A similar but definitely not as sophisticated option is available on *Linguee*, as users can click on an information icon to prune the results to a specific translation. However, the bilingual

<sup>69</sup> Users need to click on the grouped translations to access the frequency information in *ReversoContext*.

dictionary being so poor at phraseological coverage, it does not prove useful in the case of an analysis of CPs. Moreover, there is no information on frequency and the number of examples is severely limited.

The screenshot shows the TradoIT interface for the query 'en raison de'. On the left, a sidebar lists 'Traductions groupées' with various English equivalents and their frequencies: 'due to [6461]', 'because of [5007]', 'As a result of [2028]', 'for reasons of [628]', 'because [334]', 'based on [295]', 'Credit by [266]', 'owing to [257]', ', given [256]', and 'increasing [165]'. The main area displays a table of translations from French to English, including 'in consequence of', 'due to', 'in view of', 'on the grounds that', 'at a rate of', 'thought-out', 'rightly', 'cause', and 'reason'. Below the table, there are sections for 'Français' and 'Anglais' with example sentences and source information.

Figure 15. Grouped translations in *TradoIT*.

The screenshot shows the ReversoContext interface for the query 'en raison de'. The search bar shows the query in French, with 'French' and 'English' selected. Below the search bar, a list of translations is displayed in blue boxes: 'because of [prep./adv.]', 'due to [adv.]', 'owing to [prep.]', 'as a result of', 'on account of [prep.]', 'by reason of [prep./adv.]', 'in view of [adv.]', 'for [adv.]', 'by virtue of [adv.]', 'on the basis of [adv.]', 'as a consequence of [adv.]', 'on the grounds of [adv.]', 'on grounds of', 'in the light of', 'resulting from', 'in accordance with [adv.]', and 'in consequence of [adv.]'. A filter icon (a funnel) is circled in red at the bottom right of the page.

Figure 16. Grouped translations in *ReversoContext*.

*ReversoContext* surpasses *TradoIT* in the area of grouped translations owing to an extremely useful feature: the filter option, circled in Figure 16 above. This option allows users to test a translation of their choice and also partially compensates for cases where a recurrent or valuable translation is absent from the list of grouped translations. In the second case, this option can only be as good as the users nevertheless, in the sense that they will have to be



aware of the possible translations in order to search for it. In any case, it does show the cutting-edge degree of sophistication of *ReversoContext*.

#### 3.2.4. ASSESSMENT OF THE OBCs

As said earlier, several authors have already addressed the case of OBCs. Few of them directly discuss the issue of reliability however, and this number continues diminishing when pruning the concept of reliability to translation quality. We will also witness the imbalance in the number of studies according to the OBCs, with *Linguee* being the most scrutinized.

Combined, the scientific studies suggest that, more often than not, authors are torn between the pros and cons of using OBCs. Among the disapproving arguments put forward is the ambivalence as to the dichotomy between the SL and the TL (Kübler 2013), a major weak point already mentioned earlier. Furthermore, Precup-Stiegelbauer (2013: 1776) urges the supervision of a critical user, which hints at their inappropriateness for beginners. This caveat is also corroborated in other articles (Buyse *et al.* 2013: 509) and non-academics reviews ([www.commentcamarche.com](http://www.commentcamarche.com) and Altissa). Users indeed need have a certain proficiency in the language, as they have to select the most appropriate translation themselves. While it is true that *Linguee*'s PDG Frahling has deliberately created its tool as an alternative to machine translations, where the translation is imposed, *Linguee* and the other OBCs should definitely warn beginners not to rely blindly on the results.

These adverse observations are offset in Désilets *et al.* (2008a: 1), Simard (2013: 37) and Zanettin (1998: 617), who echo one another saying that the possibility to consult authentic translations is a particularly effective way to resolve translation problems. More specifically, *WeBiText* is promoted as offering better coverage in comparison with *TransSearch* (Désilets *et al.* 2008a: 8) or more traditional tools (Simard 2013: 40). Simard (*ibid.*) also argues that its translations are both acceptable and varied (*ibid.*). *Linguee* is also valued in Alonso Jiménez (2013: 11-12) on the grounds of its incremental search bar, the URL, the context and the crowdsourcing resort, as Désilets *et al.* (2008a: 11) were arguing for. We have seen earlier that this last feature is no longer available on *Linguee* however.

Finally and paramount for this study is the new approach progressively entering the literature, namely the objective of more user-driven applications, as “research in Translation Technology is often carried out by people and teams that have little knowledge of how translators actually work” (Désilets *et al.* 2009: 1). So as to avoid gaps between translation technologies and the needs of users, several authors endeavour to understand users' practice



(even though most of them focus on translators, see Taravella 2011, Désilets *et al.* 2009, Lagoudaki 2006 and Macklovitch *et al.* 2008), a process which has ultimately resulted in the creation of *WebiText* by Alain Désilets and his colleagues. These latter conducted a Contextual Inquiry of 8 professional translators,<sup>70</sup> for whom they assert (a) the ability to tell noisy results from reliable ones and to rapidly find what they want from a list, (b) the satisfaction when finding one acceptable solution in the first 10 or 20 results, (c) the willingness to be given the choice of various relevant solutions, which indicates a preference for coverage over precision (Désilets *et al.* 2009: 7-9). Also, they argue that although translators are worried about the trustworthiness of the data and its origin, it will not prevent them from employing the tool<sup>71</sup> (Désilets *et al.* 2008a: 4). This study finally shows that translators use public and multi-domain resources, without any reluctance to work with the less controlled ones or those containing translated texts. Consequently, Désilets *et al.* (2009: 6) further assess that developers, when creating TMs (i.e. translation memories) or OBCs, should not “[worry] about domain, quality control, nor direction of translation [...] because of [professional’s] critical judgement”.<sup>72</sup> Taravella (2011) corroborates this viewpoint in her investigation of linguistic technology:

Les répondants ne craignent pas, en majorité, que les outils de technologie langagière introduisent des erreurs de traduction, nuisent à la fluidité ou à la cohérence de la traduction, ralentissent leur pensée ou leur créativité, ni qu’ils soient trop lents ou que l’environnement de travail soit mal adapté à leurs besoins. (Taravella 2011: 10)

This is antonymic to Kübler (2007: 11), since she indicates that OBCs are not necessarily used by professional translators, but mostly by individuals who need to write in a second language, i.e. individuals that should not be asked so much critical judgement.

#### **3.2.4.1. Linguee**

In a press release, Fink announced that *Linguee* had received “several positive reviews, notably from linguistic experts” (Portal 2011). Comments and non-academic articles are enthusiastic, and linguists are generally positive as well, except Kübler (2013) as we have

---

<sup>70</sup> The authors argue that their inquiry is relevant because, unlike previous ones, they observed the subjects in their natural environment with a non-artificial task, focusing on one homogenous group and on their use of tools, namely translators with a moderate level of automation. However, they warn that while the number of subjects is enough to obtain insights and results, readers should be careful not to generalize the results, as the study might not be varied enough in terms of country and language coverage (Désilets *et al.* 2009: 2-3).

<sup>71</sup> Désilets *et al.* (2009: 11) suggest incorporating a voting system, but it still not available.

<sup>72</sup> For example, they continue searching even after finding a valuable solution from a reliable tool or source (Désilets *et al.* 2009: 7).

seen above. Linguists allude more or less extensively to some of the features, pros and cons of *Linguee* within the framework of another subject but, to date, Buyse *et al.* (2013) and Alonso Jiménez (2013) are the only ones concentrating exclusively and thoroughly on the website. Generally, the website is only mentioned, as in Mann (2015: 234) or Geyken (2015: 16). Finally, *Linguee* is not analysed but presented in García Hernández (2014: 59) and Huet & Langlais (2013: 203-205), where they also present *TradooIT* (2013: 205-208) and *Linear B* in a survey of computer-assisted translation tools comparable to *TransSearch*, whose effectiveness is the main purpose of the study.

In less academic reviews, *Linguee* is positively rated by Xhark (2010), Altissa, commentçamarche.com and Gallimore (2011), mainly because of the contextual provision -a major flaw of automatic translation-, the clear interface, the large data, the incremental search bar,<sup>73</sup> the phraseological aspect, the domain-specific provision and its relevance in a professional context (also emphasized by Portal 2011). In Linguistech<sup>74</sup>, Gallimore (2011) compares *Linguee* and *WeBiText*, and she mentions the impossibility neither of entering one's own corpus nor of choosing a specific corpus-based domain. Racicot (2014) suggests in his blog that *Linguee* (and *TradooIT*) does not provide sufficient or convincing examples and contains errors, but there is no empirical evidence.

- Buyse *et al.* (2013)

The objective of this study, which is to empirically test “free online data-driven lexicographic instruments”, shares similarities with this thesis, inasmuch as the authors evaluate the tools' positive effects on user-friendliness and test scores as compared to more traditional lexicographic tools (Buyse *et al.* 2013: 507). The difference lies in the language pair studied, namely English-Spanish, and in their perspective. The authors centre more on pedagogical outcomes, i.e. the purpose is to advise students on how to use these various tools appropriately in their learning process (*ibid.* 508), while we are more interested in translation trustworthiness as such (even though the two issues are evidently related). The reason for focusing on such tools roots from (a) the authors' criticism that dictionaries are time-consuming, limited and deprived of contextual information and (b) their students' heavy use of *Linguee* (*ibid.* 508-509). Therefore, they have tested the efficiency of this OBC through a small experiment with three different groups of Dutch students with a B2 level in Spanish in a 2-hour writing test: (a) an experimental group using *Linguee*, (b) a first control group using

---

<sup>73</sup> We have seen that this function is not at all optimal.

<sup>74</sup> <http://linguistech.ca>.

corpora and traditional online dictionaries, and (c) a second control group with all the aforementioned tools at their disposal (*ibid.* 509). Against their hypothesis (viz. an outperformance of the third group), it turns out that students using exclusively *Linguee* scored better,<sup>75</sup> especially when considering their vocabulary use (*ibid.* 510). The second group of students, the one using corpora and dictionaries, showed a higher degree of satisfaction with respect to the quality of the tools, but complained the most at the time constraint. This suggests that these tools are too time-consuming (*ibid.* 510). As a solution, the authors suggest the creation of a more hybrid tool.

- *Alonso Jiménez (2013)*

The purpose of Alonso Jiménez (2013) is also much related to ours, the quality of *Linguee* being at the centre of her attention. Although she discusses (*ibid.* 5) various characteristics directly related and applicable to this study, such as “peer-review, human review and history of editions in dictionary entries, traceability of content, as well as collaborative edition [...]”,<sup>76</sup> her perspective is principally dissimilar, as she analyses the quality of *Linguee* from a macrostructural viewpoint, i.e. the value of *Linguee* as a website, while our viewpoint is microstructural, i.e. the value of *Linguee* as an TT. More accurately, she tests various criteria essential for a website to be valuable according to Pinto,<sup>77</sup> and obtains positive results. As mentioned earlier, she herself strongly stresses the need for a microstructural investigation, but suggests that the quality of the translation will unlikely be homogenous given the diversity of websites used in the data (*ibid.* 19).

- *Mansfield (2013)*

In his investigation into freely available tools from *Google*, Mansfield (2013: 5) aims at “wean(ing) students off the machine translator”, on which they rely without any post-editing and “without taking into account the paradigmatic and syntagmatic conditions imposed by the context” (Buyse *et al.* 2013: 508). To this end, he tests 40 Anglophone students on the use of a French corpus so as to immerse them in the authentic, non-automatic language, through *Linguee*. His study suggests a more intense activity from learners when using *Linguee* rather than automatic TTs, which is undoubtedly an advantage in the process of language learning.

---

<sup>75</sup> Except for spelling mistakes.

<sup>76</sup> As explained in Section 3.2.3.7., the function of peer-reviewing, collaborative aspect are not as strong as before, and the history of editions in dictionary entries is not applicable anymore.

<sup>77</sup> These features are the following: authorship, updates, content, accessibility, functionality, navigation and design.

- Kübler (2013)

*Linguee* is discussed in this study showing the internal and external limitations of corpora but stressing their importance in the process of translation. Kübler (*ibid.*) believes that the translations in the OBCs are not always reliable because of the difficulty in deciding whether a translation is correct or not, even with the warning sign. Although we are inclined to follow her hypothesis, there is no empirical test in her study. She disapproves of *Linguee* because (a) it only provides certain registers, (b) does not allow users to study a textual genre and (c) is deprived of terminological context. In conclusion, she acknowledges the usefulness of the tool, but warns that users should verify its results with monolingual corpora, as well as dictionaries and terminological databases. Finally, she addresses an issue that none of these tools will solve, namely the importance of the translators' general knowledge.

- Volk *et al.* (2014)

In their article on the innovations of parallel corpora, Volk *et al.* (2014: 3172) focus on an evaluation of the Canadian Hansard (which comprises proceedings of the Canadian parliament from 1986 to 2007 and amount to 8.3 million sentence pairs), but also discuss *Linguee*, *Globse* and *TradooIT*, arguing that they are well-designed (with the exception of the second), valuable and useful, even though the sophistication of the search bar should be refined. They employ a similar methodology as ours to evaluate the various tools, but with different language pairs, namely English – German and English – Spanish. They test the translations of three verb-particles (*fool around*, *knock down* and *speed up*) as well as three noun-compounds (*oil tanker*, *lung cancer* and *board meeting*). The difference is that they evaluate those translations intuitively, i.e. without the support of corpora. Their results, however, reveal that *Linguee*'s highlighting can be erroneous, that the information in the dictionary is not supported by the results in the examples and that the word-alignment is incorrect (*ibid.* 3174). On the other hand, its large coverage of translations is second to best, behind *Globse* (*ibid.* 3177). This study allows us to realise how “online parallel corpus query systems have become popular recently”, both in public and researcher systems, which have not employed automatic word alignment yet (*ibid.* 3177).

- Arcan *et al.* (2013)

*Linguee* is also discussed in Arcan *et al.* (2013). In their attempt to compensate for the lack of translations for domain-specific vocabulary (financial reporting taxonomy in their case), the authors constructed their own parallel corpus on the basis of two other corpora (Europarl and

JRC-acquis), *Wikipedia*, *DBpedia* as well as *Linguee*, with a total of 200,000 aligned sentences. *Wikipedia* and *Linguee* were selected for their “extensive multilingual data” (*ibid.* 204). Their study indicates that creating and using a new, domain-specific corpus improves translation quality, and that the web proves useful for mitigating the lack of such data (*ibid.* 205), which is also put forward by Désilets *et al.* (2010: 1). Also, while they stress (Arcan *et al.* 2013: 205) the importance and usefulness of statistical machine translations as well as the need for new resources in the field, they believe that manual editing of the translations should become a major concern, albeit time-consuming, as it permits filtering among correct and incorrect translations while eliminating mistakes. Although not an examination of *Linguee* as such, this article is relevant for our reflection, inasmuch as the authors employ the tool in their research, which might imply that it has started to acquire a certain value in the academic world.

- *Grauer (2010)*

Concerned with the difficulty to translate collocations, Grauer (2009) conducts an empirical evaluation on 40 students and their use of Internet-based tools. His results show that students, despite their web competence and knowledge, restrict their online resources to few tools (such as *LEO*, *dict.cc*<sup>78</sup> and *PONS*<sup>79</sup>) and therefore do not benefit from the tremendous help available. According to its results, students are not satisfied with these tools, either because there were no results or because they were difficult to find (Grauer 2009: 19). Because *Linguee* did not exist at the time, Grauer (2010) decided to compare the results of the aforementioned research in a small-scale study<sup>80</sup> which yielded positive results. He chose to investigate the tool to measure the extent to which it could help translators with respect to collocations (*ibid.* 2). His conclusion (*ibid.* 19-20) is that “diversified and intelligent use of the Internet produces quality translation results”. Since this use requires experience, he also believes that it should therefore be included in education programs (*ibid.* 19-20)

- *Bourdaillet & Langlais (2012)*

Bourdaillet & Langlais (2012) endeavour to identify and solve some of the problems occurring with three OBCs, namely *Linguee*, *TradooIT* and *TransSearch*, mainly focusing on the issue of infrequent translations on the basis of several sequences of words (e.g. *meanwhile*, *de façon répétée*, *sur ces entrefaits*, etc.), using the Canadian Hansard bilingual

---

<sup>78</sup> <http://www.dict.cc>

<sup>79</sup> <http://fr.pons.com/traduction>

<sup>80</sup> He only tested one difficult German collocation, namely *eine Strafe verhängen* (Eng. *to impose a penalty*)

corpus as a control tool. The results yielded by this study show that the translations are only partial, often wrong or altogether absent. This is all the more appalling considering that translations are indeed present in the bi-texts of the OBCs but are simply not found by the tool and, as a consequence, not provided. According to the authors, this is mainly due to the lack of precision of the word alignment at play, which, once improved, would greatly diversify the set of infrequent and idiomatic translations presented to the users. For now, because they co-occur only a few times with the queried term in the data, their lexical associations tend to be poorly estimated by statistical translation models.

#### 3.2.4.2. TradooIT

While we have observed a general consensus on the positive aspects of *Linguee* despite the numerous weak points mentioned, *TradooIT* definitely achieves more favourable assessments. The scarcity of scholarly resources, however, probably bears some connection with this approval. Additionally, the treatment of the tool is generally not at the centre of the few studies available, hindering the authors from conducting a thorough analysis.

We can find the equivalent of the *Linguee* press releases, namely the newscasts, available on the website<sup>81</sup> and the blog<sup>82</sup> of its creator, which consists in describing the updates and functions of the tool. New posts are directly accessible from the homepage in News section, but the latest entry dates back to 2014. Finally, a tutorial is available on Linguistech, in which the author (Ouellet 2012) promotes the scarcity of noise (i.e. irrelevant or incorrect results) and the sophistication of the tool.

- *Volk et al. (2014)*

The previously-mentioned study from Volk *et al.* (2014: 3174) also discusses *TradooIT*, of which they test the English – Spanish pair. Multifarious features of the tool are promoted: (a) the presence of a definite list of the corpora, (b) the precise information about sizes (c) the results sorting according to the frequency of the translation variants, (d) the unambiguity of the highlighting, (e) the filter option, described as a main advantage for translators, (f) the collaborative voting and finally, (g) the possibility to consult the bi-texts. They do have one criticism, however, namely alignment errors.

---

<sup>81</sup> <http://www.tradooit.com/nouvelles.php>. Accessed on 13 January 2016

<sup>82</sup> <http://blog.tradooit.com>. Accessed on 13 January 2016.

- *Gallimore (2011)*

In her review, Gallimore (2011) clearly states her preference for the tool as compared with *Linguee* and *WeBiText*. Her appreciation rests on (a) the warrant for the OBC to remain free, (b) the presence of *TERTIUM* within the website, (c) the speed, (d) the access of slang through the corpus of subtitles, (e) the improvement of the algorithm through the users rating and, in accordance with the previous studies, (f) the accurate highlighting of the terms as well as (g) the various filters.<sup>83</sup> Finally, and on a more subjective level, she believes that the layout is more appealing than in *Linguee* and *WeBiText*. Her claim on the distinction between SL and TL is not supported here and, irrespective of her caveat on the impending presence of advertisements, the website is still free from them.

#### 3.2.4.3. WeBiText<sup>84</sup>

The tool is mentioned in McBride (2011: 6) and in a blog entry (Anonym: 2012) where users recommend to use *TradooIT* instead. The platform Linguistech also provides tutorials (Van Bolderen 2011, Franz 2011) and a comparison between the tool and *Linguee*, as mentioned earlier (Gallimore 2011). Besides these non-scholar reviews, *WeBiText* is the best-documented tool, thanks to several studies written by its creator, Alain Désilets, who presents it as a help to find solutions to translation problems (2010: 1). Moreover, it is also the focus of Simard's thesis (2013).

- *Désilets et al. (2007, 2008a and 2009)*

Désilets and his colleagues started their project on the creation of *WeBiText* after several assessments based on inquiries: (a) a one-million-page TM is sufficient to resolve 76% of translation problems<sup>85</sup> (Désilets *et al.* 2007: 10), (b) web-based, heterogeneous texts are a relevant resource for building general corpora and would therefore add value to existing TMs (Désilets *et al.* 2008a: 1) and, more importantly, (c) translators were already using the web to manually search for parallel texts, as it offers large and varied data (Désilets *et al.* 2009). They warn, however, that, “while the concept of building TMs from Web corpora holds great promise, more research may be needed to evaluate its actual potential for language pairs other than English-French” (Désilets *et al.* 2008a: 10). The authors argue, however, that *WeBiText*

---

<sup>83</sup> She also mentions the auto completion search bar, but, as we have seen, this feature is not available on *TradooIT*.

<sup>84</sup> The OBC is still free of charge, contrary to the claim of Van Bolderen (2011) and Franz (2011: 18).

<sup>85</sup> The list of translation problems was created on the basis of a contextual inquiry with professional translators that they conducted, published Désilets *et al.* (2009).

is different from other OBCs such as *MyMemory* and *TAUS*, which work with donated corpora and do not provide enough data per domain to be relevant (Désilets *et al.* 2008a: 2).

- *Désilets et al. (2010)*

According to Désilets *et al.* (2010: 2), several features allow the assessment of *WeBiText* in terms of reliability, namely the URL, the parallel web-pages access and the context. We understand that trustworthiness refers here to the origin of the data, not to translation quality. Their plan for the future is to design the website to keep up with heavy traffic, add highlighting, support advanced and more flexible query syntax, add more specifications in the selection of corpora, assist users who want to use the pivot technique and automate the process (Désilets *et al.* 2010: 7). They believe that “the rapid growth of the traffic on *WeBiText* is a clear indication that it is meeting a need in the translation industry” (*ibid.*) and that the popularity should remain high and steady (Désilets *et al.* 2010: 6) for the following reasons: (a) the tool can serve all kinds of translators, (b) it is easy to use and does not require any training, and (c) there is no need for approval from the top of the organisation. The tool, however, is by no means as popular as the other OBCs.

- *Simard (2013)*

In her Ph.D. thesis, Simard (2013) investigates the usefulness of three different TTs for medical vocabulary, namely the *Dictionnaire anglais-français des sciences médicales et paramédicales*, *TERTIUM* and *WeBiText*, whose corpora, process and interface are described. The evaluation of their usefulness is based on a comparison of the coverage (quantitative evaluation) and acceptability of their solutions (qualitative evaluation), which is highly similar to our purpose. Simard (2013) conducts her research on the basis of 203 translation problems revealed by 8 article summaries. The results suggest that *WeBiText* outperforms the others, with 62% of coverage against 37 and 30 for the two other tools respectively. The OBC also provides twice as many solutions, all considered acceptable according to 88 language specialists. She believes that it is overall a larger and richer resource, especially considering that the health domain is not the biggest corpus available on the website. Simard (2013: 93) suggests that an acceptable solution is more likely to appear in *WeBiText* given the abundance of its solutions. On the downside, bad solutions are also more likely to be present for the exact same reason. In her conclusion, she expects that parallel corpora and OBCs will be increasingly used in the professional translation field, for they allow users to find immediate answers to concrete questions. She calls for further research implementing an even more



practical methodology, where the researcher starts by detecting the translation problems together with the users.

#### 3.2.4.4. ReversoContext

Whilst the only resources available used to be confined to a blog<sup>86</sup> and the Facebook page of the tool, a number of studies directly and indirectly related to *Reverso* have recently appeared. *ReversoDictionary* is discussed in Miller for example, whilst *ReversoLocalize* is discussed in Le Ny (2014) and thoroughly deliberated in Peron & Morado Vázquez (2015) and Peron (2013). There is no account of the reliability of the *ReversoContext* from our standpoint

- *Esplà-Gomis et al. (2015)*

The acceptance of the alignment system of the tool is discussed in Esplà-Gomis *et al.* (2015), who study the language pairs English – Spanish and English – German. In their quest to evaluate machine translation quality estimation, the authors account for the richness of *ReversoContext* as a source of bilingual information and the reliability of its resources owing to its manually-translated text data (*ibid.* 5). They mention two weaknesses notwithstanding, namely its low coverage and the fact that results are only displayed when there are a minimum number of occurrences (*ibid.*). This last drawback is now circumscribed however, as users can enter a specific translation in the filters.

- *Le Ny (2014)*

*ReversoContext* is mentioned in Le Ny (2014: 11), who indicates that errors in the highlighted segment might occur, for which he suggests assigning a warning sign. The main purpose of the article is to identify key indicators to assess quality in human translation and TMs, but he focuses on the overlooked issue of alignment quality (*ibid.* 2-3). The creators of the tool also warn that while the alignments are correct at the sentence level, there might be mismatches between the query and the translation highlighted.<sup>87</sup>

- *Granger & Lefer (in press)*

These authors indicate that *ReversoContext* may be seen as an improvement as compared to OBDs, as it offers genuine examples. However, because of the confusion between SL and TL as well as the unbalanced genre representation, such a tool might pose difficulties, especially to learners.

---

<sup>86</sup> <http://blog.ReversoContext.net>. Accessed on 14 January 2016.

<sup>87</sup> <http://context.ReversoContext.net/traduction/about>. Accessed on 14 January 2016.

## 4. ANALYSIS

### 4.1. INTRODUCTION

With a clearer picture of the OBCs in mind, we can now move on to the main purpose of this thesis, i.e. a thorough and empirical analysis of their reliability. The general macro-analysis will be presented first, considering the quality of the tools in displaying translations. In a second stage, we will discuss the English translations of the ten French CPs<sup>88</sup>. Those will be presented in tables and in decreasing order of frequency according to the Label France corpus, along with those of the PLECI\_news corpus, the OBCs<sup>89</sup> (both those from the first 30 sentence pairs and those from the grouped translations) and the OBDs. Our main focus being the reliability of OBCs,<sup>90</sup> we will mainly discuss relative frequencies to highlight the differences and similarities of the tools as compared to the corpora as well as to show the specific contributions of each type of tool. Also, the translations will be annotated in the tables as ✓ if they are found in the tool or as ✗ if they are not. Because of the plethora of translations suggested for each French CP, a threshold has been established to bring out the most recurrent ones, i.e. either those representing more than 5% of the Label France corpus or those provided by a majority of the tools. The lists of the most common translations will therefore be somewhat biased in favour of the Label France corpus, since it is used as an inclusion criterion. As explained in Section 1.3.2.3., the corpora's findings will be tested by back-translations and the mutual correspondence calculation.

### 4.2. MACRO-ANALYSIS

- *Translation corpora*

The analysis does not reveal any major problem within the translation corpora, other than few missing translations which do not bias the results.

- *OBCs (first random examples)*

Several types of problems occur in the OBCs, such as lack of source variety, which can lead to a situation where the same sentence is repeated throughout the results or, more generally, to a lack of contextual variety, where a low number of topics covers the majority of the sentence

---

<sup>88</sup> The structure of this section follows the same order as in Table 1.

<sup>89</sup> Since the tools are constantly updated, it is important to note that the data of the first 30 sentence pairs was collected in September 2015 while the analysis of the grouped translations was carried out in March 2016.

<sup>90</sup> Moreover, raw frequencies are difficult to compare in this analysis, as each French CP has a different number of occurrences according to each OBC and corpus.

pairs. Repeated sentences occur in all the OBCs, except in *Linguee*, which systematically provides a series of unique pairs of sentences, but *WeBiText* is undeniably the most problematic tool. For example, it produces 25 cases of zero correspondences for *en matière de* because a few sentences are repeated several times. This results in a misrepresentation of the remaining translations and an overrepresentation of this type of correspondence.

A second problem, which only appears in *WeBiText*, is an error of target language, i.e. the French sentences are translated into another language than English. In seven out of the ten collections of the French CPs' translations, 2 to 15 French sentences are translated in German, seriously jeopardizing the usefulness of the tool. *WeBiText* is also the only one to include error messages.

Alignment at the sentence level is defective in all the OBCs, however it is not systematic and occurs with varying degrees of intensity. Most of the cases appear in *WeBiText* with generally 2 to 4 problematic sentence pairs for each search, whereas there are only few cases in the other OBCs. Word-alignment, on the other hand, is non-existent in *WeBiText* and seems problematic in *Linguee*, as the items highlighted as translations often lack accuracy or are simply wrong. The system is more reliable in the other two OBCs, especially in *ReversoContext*, as it differentiates cases where an item is the translation of the queried term from cases where it is not.

- *OBCs (grouped translations)*

*TradooIT* frequently provides erroneous translations in its grouped translations due to alignment errors at the word level (see example 6)<sup>91</sup>, which also occurs in *ReversoContext* but at a considerably lower rate. On the other hand, *ReversoContext* frequently fails to index frequent and valuable translation. Also, this analysis shows that the whole data of these OBCs is usually ill-represented by the first 30 random sentence pairs. Taken to its extreme, this results in (a) cases where the first 30 pairs of sentences do not include a frequent translation indexed in the grouped translations or (b) conversely, cases where a translation recurrently appears in the first sentence pairs but is not presented in the grouped translations (especially high-frequency simple prepositions such as *in*). In *ReversoContext*, however, it is possible to access the translations absent from the grouped translations via the filter option.

- (6) FR « Il s'agit d'établir et d'évaluer si la PME possède les capacités nécessaires pour mener ou participer au projet de développement proposé **en matière**

---

<sup>91</sup> In a majority of these cases, the translations are incorrectly indexed because they involve a zero correspondence.

**de** recherche et de technologie.  
ENG ‘This consists of assessing and evaluating whether the SME possesses the necessary capability to lead or participate in the proposed **research** and technology development project. (TradooIT)<sup>92</sup>

In addition, one key issue is the frequency information, which should be better displayed, and its possible inaccuracy and ambiguity (for example, differences of frequency depending on where we look for the information).<sup>93</sup> Besides the fact that the translations do not appear according to their respective frequency (e.g. *in terms of* is in fourth position for the queried term *en matière de* despite its higher number of occurrences), it is arduous to pin down the exact number of occurrences of a particular translation through the list of grouped translations. For instance, *support of* (1555 hits) is indexed in *TradooIT* as a translation of *en faveur de*. However, it corresponds to sentences involving both *in support of* and *to support*, which impedes us from having a clear idea of the frequency of each item.

A second phenomenon inclines us to remain critical as regards frequency information. For *en faveur de*, for example, the list of grouped translations does not seem exhaustive in *TradooIT*, as the number of occurrences of the indexed translations combined only account for 35.3% of the total occurrences of the French CP. The most frequent translations indexed in *ReversoContext* have the opposite problem, as they exceed 100%, which either implies that the total occurrences of *en faveur de* is incorrect or that sentence pairs are attached to several grouped translations. In either case, this shows a lack of precision and will confuse users. That being said, one must remember that *TradooIT* and *ReversoContext* are the only tools to display raw frequency.

- *OBDs*

The entries vary considerably from one OBD to the others, both in terms of the translations suggested and the information included, with the *Reverso-Collins* and *Oxford* at each extreme (see Appendix 5 for a list of the entries). *Oxford* is generally more extensive on additional information such as examples in use. More importantly, the ten French CPs are presented in sub-entries or highly contextualized examples in *Reverso-Collins*, while two of them are presented separately with the status of headword in *Oxford*. *Larousse* outperforms the others in this area, as seven French CPs are headwords.

---

<sup>92</sup> The highlighting reflects the one offered in *TradooIT*.

<sup>93</sup> e.g. The link for *for* as a translation of *en faveur de* indicates a total of 9116 occurrences but, at the bottom of the page, the number 16,378 is given. While small differences are harmless given the total data of *ReversoContext*, this almost doubles what had been previously indicated.

A consequence of their secondary status is their difficult accessibility, already raised by Granger & Lefer (in press). *Larousse* does not systematically give direct access to the French CPs and their respective translations, which are almost never highlighted. A first possibility is where users are directed to the relevant information (in two cases only) but most of the time, they will need to click on the relevant link in the left column of the website to access the translations (see Appendix 6 for an example). Generally, users will have to peruse the whole page in *Reverso-Collins* and *Oxford*. However, while the relevant information is neither highlighted nor indexed in the first, it is generally highlighted in the second.

Another problem in *Oxford* is that manual search sometimes brings out further information, left hidden by the tool. For example, typing *dans le cadre de* in the search bar offers scant results. However, when typing the noun *cadre* in the search bar and manually scrolling down the page, one can find a much richer solution (see Appendices 7 and 8). This is a real problem in terms of accessibility and even reliability, as it impedes users to directly access all the information available on the dictionary, either because they do not have the time to perform manual searches or because they are simply unaware that there might be other possible solutions to look for.

### 4.3. MICRO-ANALYSIS OF THE TRANSLATIONS

#### 4.3.1. EN MATIÈRE DE

##### 4.3.1.1. Frequent English translations suggested by the tools

When combining the results of the two translation corpora, the OBCs and the OBDs, the list of translations of *en matière de* amounts to 40 items, mostly of the congruent type. Several of these are extremely rare when translating the French CP (e.g. *pertaining to*, *on the subject of*, *related to* etc.), but a comparison between the tools brings out the most recurrent ones (see table 2). This list reveals that divergent correspondences are in fact more frequent (*in*, *for*, *on*), and that zero correspondences are a common option. Literal translations are rather infrequent.

Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
TTs	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
In	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗
In the area of	✓	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗
For	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
on	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗
In terms of	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗
With regard to	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✗
Zero corr.	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
When it comes to	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗

Table 2. Most frequent translations of *en matière de* in the three types of tools.

The table above shows one strong tendency, i.e. *Linguee*, *TradooIT* and *ReversoContext* suggest in their first 30 sentence pairs the most recurrent translations found in the corpora, while the other tools differ on those, except for the simple preposition *for* (see example 7) and the zero correspondences (see examples 8-9).<sup>94</sup> A closer look at the frequencies reveals that the respective proportion of each translation varies from one tool to the other. The translations *in*, *in the area of*, *on*, *in terms of*, *with regard to* and *when it comes to* (see examples 10-14), albeit not very frequent (2.3% in Label France), are provided by almost all the tools, and therefore included in the table. *WeBiText* does not provide most of the recurrent translations because of sentences repeated throughout its first random pairs of sentences. Few of these frequent translations are indexed in the grouped translations, while they are all absent from the OBDs.

- (7) OF Elargir la responsabilité **en matière de** protection des droits humains signifie [...]  
TE Widening responsibility **for** the protection of human rights means [...]  
(Label France)
- (8) FR Le médecin qui délivre un certificat d'incapacité ou un certificat de prorogation en informe promptement le malade et en avise promptement un conseiller **en matière de** droits.  
ENG A physician who issues a certificate of incapacity or a certificate of continuance shall promptly advise the patient of the fact and shall also promptly notify a [Ø] rights adviser (TradooIT)
- (9) FR Puisqu'elle n'a pas *d'incidence* **en matière de** concurrence, cette décision devrait relever exclusivement de la compétence des États, décidant en fonction de leurs priorités propres.  
ENG Since it does not *affect* [Ø] competition, the decision on this matter should fall exclusively within the competence of the Member States and be made

<sup>94</sup> In this case, the zero correspondences are due to a shift of grammatical category, either from the use of a CP in French to a compound noun in English (as in example 7) or from a noun + CP structure in French (e.g. *incidence en matière de qqch*) to a verb in English (e.g. *to affect sth*), as in example (8).

- on the basis of their own priorities (TradooIT)
- (10) FR Les parties contractantes (États membres) s'engagent à atteindre et à maintenir un niveau de sûreté constamment élevé **en matière de** gestion du combustible usé et des déchets radioactifs.
- ENG It represents a commitment by Contracting Parties (member countries) to achieve and maintain a consistently high level of safety **in** the management of spent fuel and radioactive waste (TradooIT)
- (11) OF Quoi, en France, a marqué la décennie 90 **en matière de** consommation ?
- TE What, in France, was the distinctive feature of the nineties' decade **in the area of** consumption? (Label France)
- (12) FR Il a été souligné que la législation nationale **en matière de** migrations internationales devait s'aligner sur les normes internationalement reconnues en matière de droits de l'homme.
- ENG It was stressed that national legislation **on** international migration should reflect internationally agreed human rights standards (ReversoContext)
- (12) OF ... sion historique d'une brutalité inouïe **en matière de** niveau de salaires, de couverture sociale [...]
- TE [...], it represents a severe step backwards **in terms of** wages, social welfare, protection of minors, public safety, working hours, paid leave and unemployment benefit (PLECI\_news)
- (13) FR En second lieu, je souhaite insister sur le respect des règles s'imposant aux institutions **en matière de** classification des crédits entre dépenses obligatoires et dépenses non obligatoires.
- ENG Secondly, I should like to insist on the respect for the rules that is incumbent on the institutions **with regard to** classifying appropriations as compulsory or non-compulsory expenditure (TradooIT)
- (14) OF **En matière de** métissage culturel, Sandira est un hybride [...]
- TE **When it comes to** cultural cross - fertilisation, Sandira is an amazing hybrid (Label France)

#### 4.3.1.2. Contribution of each tool

The corpora bring out that both CPs and simple prepositions can be used to translate *en matière de* in English, and the OBCs show these two possibilities, however with extremely dissimilar proportions. *WeBiText* is particularly deceitful due to its poor macro-quality and should be disregarded.

The grouped translations listed in *TradooIT* and *ReversoContext*<sup>95</sup> are very dissimilar to the list of translations found in the corpora, which does not mean that the frequent translations are absent from the tools but that they may have not been indexed. Another problem is that the first 30 examples do not necessarily match the grouped translations. The main contribution of the grouped translations is that they indicate further items to translate *en matière de* than what had been suggested not only by the first 30 sentence pairs but also by the corpora (e.g. *over, in the field of, regarding, concerning, as regards, relating to, with respect to, in relation to, in respect of*, etc.). These translations are not necessarily absent from the corpora, but their

<sup>95</sup> 40,399 and 138,280 results respectively.

respective frequencies were low enough to be a criterion of exclusion from our table in the beginning of the section.

The OBDs are absolutely deviating from the results found in the corpora, as none of the most frequent translations of *en matière de* is suggested, while infrequent, even non-existent ones in the corpora are provided, such as *as far as sth is concerned* (see example 15).

- (15) FR **en matière de** cuisine/d'art/d'emploi  
 ENG **as far as** cooking/art/employment **is concerned** (Larousse)

#### 4.3.1.3. Translation control: back-translation and mutual correspondence

With a back-translation analysis (producing 47 hits), we observe that the translation *en matière de* is triggered by seven different items in OE (*in, for, about, on, in terms of, in all matters related to and of*), while *en matière de* in OF is translated by a total of 40 items in TE according to our data. This observation suggests that the lexical concept can be expressed in English by a broader range of items compared to French, but this needs to be further studied with larger bidirectional corpora. In one third of the cases, the French CP is an addition in TF, also called zero correspondence. The simple prepositions *for* and *in* are the second most common triggers (15% of the occurrences).

For *en matière de* and *in terms of*, the measurement of the mutual correspondence provides a result of 8.1%, while the other pairs of correspondences reach an even smaller degree of mutual translatability. However, hasty conclusions should not be drawn as this could simply reflect a situation where there is no prototypical correspondence but a vast array of them, as explained by Altenberg (1999: 262). While we can be certain that *en matière de* and *in, for* or *on* are not equivalents, they should still be considered as valuable correspondences.

The majority of the literal translations, namely *in the matter of* (see example 16), *in matters of* and *on the matter of*, are primarily found in the Label France corpus. The grouped translations in *ReversoContext* suggest that they are extremely infrequent, which could suggest that they have permeated from French in the corpora. However, the PLECI\_news corpus is too small to further inquiry this problematic, as there are virtually no occurrences of these English CPs, neither in TE nor in OE.

- (16) OF [...] le rôle fondamental de l'Etat, dont la responsabilité, **en matière de** parité, est de créer un environnement [...]  
 TE In its progress report, UNESCO stressed the fundamental role of the state, whose responsibility **in the matter of** parity is to create an environment [...] (Label France)



### 4.3.2. AU SEIN DE

#### 4.3.2.1. Frequent English translations suggested by the tools

Fewer possibilities are available to translate the French CP *au sein de*, with approximately 20 items retrieved from the three types of tools. From this list, we observe that the French CP does not have a literal correspondence, and is most commonly translated by simple prepositions rather than complex ones (e.g. *in the midst of*, *in the heart of*, etc.). Finally, there is only one instance of zero correspondence. Only two of the translations are listed in Table 3, as the others do not meet the criteria described in Section 4.1. (i.e. the Tables present the translations representing at least 5% of the corpora or those provided by most of the OBCs).

Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
Within	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
In	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗

Table 3. Most frequent translations of *au sein de* in the three types of tools.

*Within* (see examples 17) is found in all the tools, while *in* (see example 18) is absent from the grouped translations of *ReversoContext* and from the OBDs. The ranking and proportion of these translations are different from one type of tool to the other, which shows how frequency must be read carefully. For example, *within* translates *au sein de* in 43.6% of the cases in the corpora, in 20% of the cases in *WeBiText* and in 86% of the cases in *ReversoContext*.

- (17) OF [...] a même pris soin de signaler sa présence **au sein de** la foule par "une croix sur la gauche [...]"  
 TE The ever attentive Joe even took the trouble to indicate his presence **in** the crowd with "a cross to the left of the gibbet" (PLECI\_news)
- (18) FR Ce but a été atteint au terme de discussions approfondies **au sein de** la communauté internationales aux conférences de Jomtien (1990), de Salamanque (1994) et plus récemment de Dakar (2000).  
 ENG This goal has been achieved following extensive discussions **within** the international community following conferences in Jomtien (1990) in Salamanca (1994) and most recently in Dakar (2000) (TradooIT)

#### 4.3.2.2. Contribution of each tool

Both corpora clearly show that simple prepositions are more frequently used in order to translate the French CP *au sein de*, and the first examples in the OBCs are in accordance with this result. This is further corroborated in the grouped translations of *TradooIT* and

*ReversoContext*<sup>96</sup> which, because of their larger data, provide an even clearer quantitative separation between the prepositions *within* and *in*<sup>97</sup> on the one hand, and other possible translations on the other (e.g. *among*, *across*, *throughout*, *into*, etc., which represent less than 1.5% of the total occurrences of *au sein de* in both tools).

Despite the recurrent problems at the macro-level, the usefulness of the grouped translations and the filter option is clear, as they display the different possible translations in more detail, through the examples. In doing so, *ReversoContext* provides a wider variety of translations than the corpora and better shows their relevance. For example, we can see that *across*, although extremely infrequent (0.62% in *ReversoContext*), seems to be a valuable translation of *au sein de* when referring to a geographical zone (see example 19).

- (19) FR Les fluctuations dans le développement économique **au sein de** la région euroméditerranéenne [sic] ont toujours provoqué d'importants flux migratoires.
- ENG Variations in economic development **across** the Euro-Mediterranean region have always been an important cause of migration (ReversoContext)

All the OBDs provide one possible translation for *au sein de*, namely *within*. While this translation is supported by the other tools and is in fact the most recurrent possibility according to them, the simple preposition *in* is missing from the three dictionaries. Accentuating this discrepancy as compared to the other tools is the *Larousse*'s entry, where *in the midst of* and *in the bosom of* are suggested whereas they are virtually absent from both the OBCs and the corpora.

#### 4.3.2.3. Translation control: back-translation and mutual correspondence

The back-translation corroborates the previous results. First, it shows that the French CP in TF mostly comes from *within* and *in* in OE. Second, the majority of the triggers are simple prepositions, as there is only one occurrence of a CP, namely *at the heart of*. Third, there is also only one case of zero correspondence, where the French CP is an addition as compared to OE (see example 20). According to the mutual correspondence calculation, *au sein de* and *within* are equivalent at 13.02%, which is still substantial given that the simple preposition is polysemous in English.

- (20) OE [...], and an [Ø] E.U. average of 48.9 % - employee social charges and

---

<sup>96</sup> 30,384 and 55,801 results respectively.

<sup>97</sup> The simple preposition was absent from the grouped translations and searched for in the filter search bar in order to get its frequency.

labor market rigidities have provoked a considerable flight of small and medium - sized business, primarily to Britain.

TF [...], et 48,9 % en moyenne **au sein de** l'UE -des cotisations sociales et les rigidités du marché du travail ont provoqué une fuite considérable des petites et moyennes entreprises, principalement vers la Grande-Bretagne [...] (PLECI\_news)

### 4.3.3. PAR RAPPORT À

#### 4.3.3.1. Frequent English translations suggested by the tools

The results show that *par rapport à* has 38 possible translations in English, either congruent or divergent, but not zero or literal. Table 4 indicates the most recurrent items.

Types of TT TTs	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
Compared with	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗	✓
Compared to	✓	✗	✓	✗	✓	✓	✗	✓	✗	✗	✗
In relation to	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
From	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✓
Than	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗
Over	✓	✓	✗	✓	✗	✓	✓	✓	✗	✗	✗
On	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗

Table 4. Most frequent translations of *par rapport à* in the three types of tools.

As opposed to the case of *au sein de*, where *within* and *in* clearly stand out from the rest of the possible translations, the line between extremely frequent translations and less frequent ones is more blurry. This is exacerbated by the fact that the frequency and the proportion of each translation are slightly different from one tool to the other. For example, the most recurrent one in the corpora, namely *compared with* (see example 21), only occurs once in two of the OBCs. Finally, the tools provide the most recurrent translations found in the corpora (see examples 22-27), except for *WeBiText* and the OBDs.

- (21) OF Avec une croissance de 3 % **par rapport à** 2005 et 15,9 millions d'arrivées en 2006, [...]  
 TE With a growth of 3 % **compared with** 2005 and 15.9 millions [sic] visitors in 2006, [...] (Label France)
- (22) FR Le BAIIA du secteur Produits d'épicerie au deuxième trimestre est demeuré relativement stable **par rapport à** la même période l'an dernier.  
 ENG Our Grocery Products Sector EBITDA for the second quarter remained

- relatively stable **compared to** the same period last year (Linguee)
- (23) FR La valeur du contrat se déplace **par rapport à** l'instrument ou la devise sous-jacente.  
 ENG The contract value moves **in relation to** the underlying instrument or currency (ReversoContext)
- (24) FR Le Airbus et le Cessna se sont croisés à un endroit où le débattement angulaire est d'environ 150 à 160 degrés **par rapport à** l'orientation visuelle normale du contrôleur d'aéroport.  
 ENG The position at which the Airbus and the Cessna crossed is at an angular displacement of approximately 150 to 160 degrees **from** the normal visual orientation of the airport controller (TradooIT)
- (25) OF En fait, leur engagement réel a peu diminué **par rapport à** celui des générations précédentes ; [...]  
 TE In fact, their real commitment is very little less **than** it has been for earlier generations ; [...] (Label France)
- (26) OF [...], en augmentation de 30 % **par rapport à** 1995.  
 TE [...], a 30 per cent increase **over** 1995 (PLECI\_news)
- (27) FR Les élections législatives du 30 mai 1999 représentent une amélioration substantielle **par rapport à** celles de 1995.  
 ENG The parliamentary elections of 30 May 1999 were a substantial improvement **on** those held in 1995 (ReversoContext)

#### 4.3.3.2. Contribution of each tool

The Label France corpus shows that CPs are the most common translations of *par rapport à* in English, but that simple prepositions are highly valuable. This is also indicated in the OBCs, however less clearly in *WeBiText* and *ReversoContext*. The majority of the most recurrent translations of the French CP are also indexed in the grouped translations of *TradooIT* and *ReversoContext*.<sup>98</sup> Their advantage is to provide additional recurrent translations that are not found in the first sentence pairs of the OBCs (except *ReversoContext*) or the corpora, namely *relative to*, *with respect to*, *about* (see example 28), as well as *against* and *according to*, which are only indexed in *ReversoContext*.

- (28) FR Mais c'est plus **par rapport à** la quantité qu'à la qualité.  
 ENG But it's more **about** quantity than quality (ReversoContext)

Finally, the translations suggested in the OBDs differ from those found in the OBCs and the corpora. The most obvious observation is that the currency of divergent correspondences is largely underestimated, as only four simple prepositions (namely *regarding*, *against*, *towards*, which are not supported by the corpora, and *from*) are suggested, and only by a minority of the dictionaries.

<sup>98</sup> 34,289 and 89,278 results respectively.

#### 4.3.3.3. Translation control: back-translation and mutual correspondence

It is difficult to establish a pattern from the 25 occurrences of *par rapport à* in TF, as they come from a variety of items in OE, each of which occurring between one and three times. What is suggested is that the French CP is not an addition compared with the source texts. Secondly, it seems that *par rapport à* is mostly divergent as compared to its correspondences in OE, as it mainly has simple prepositions as its source. Third, the list of recurrent correspondences in OE looks perceptibly different from the one in TE (see Table 5), which indicates that there may be a difference between English as SL and English as TL. This contrast needs to be interpreted with caution, however, (a) because the English translations listed in Table 4 are those compiled from all the OBCs, the OBDs and the corpora whilst OE could only be analysed through the PLECI\_news corpus, and (b) because it is likely that a larger corpus would produce dissimilar outcomes.

<b>Language English corr.</b>	<b>OE</b>	<b>TE</b>
<b>Against</b>	✓	✗
<b>From</b>	✓	✓
<b>Over</b>	✓	✓
<b>Compared with</b>	✓	✓
<b>Compared to</b>	✗	✓
<b>In relation to</b>	✗	✓
<b>Than</b>	✗	✓
<b>On</b>	✗	✓

**Table 5.** Comparison of the English correspondences of *par rapport à* in OE and TE in the PLECI\_news corpus.

Owing to the data used and the fact that simple prepositions are polysemous, the only degree of correspondence that can be measured in this case is between *par rapport à* and *compared with*, which is of 16.66%. This does not imply that the other translations should not be included in bilingual dictionaries or pedagogical however, but they should definitely be illustrated with authentic examples.

#### 4.3.4. DANS LE CADRE DE

##### 4.3.4.1. Frequent English translations suggested by the tools

Among the 37 translations of *dans le cadre de* identified in the three types of tools, congruence is the most recurrent type of correspondence (21 CPs in total), while divergence is scarce (10 simple prepositions) and zero correspondence extremely rare. The most frequent translations are the six simple prepositions and CPs shown in Table 6.

Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
TTs											
English corr.											
As part of	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
In the context of	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	✗
Within the framework of	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Under	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	✗
Within	✓	✗	✓	✗	✓	✓	✓	✓	✗	✗	✓
In	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗

Table 6. Most frequent translations of *dans le cadre de* in the three types of tools.

In the Label France corpus, *as part of* (see example 29) is the most common translation of the French CP, followed by *in the context of* and *within the framework of* (see examples 30-31). On the other hand, the simple prepositions *under*, *within* and *in* (see examples 32-34) are more infrequent. While almost all the translations are provided by the OBCs (both in the first random examples and the grouped translations), none of them follow the same quantitative pattern, which differs from the one found in the Label France corpus in each case. Except for the *Oxford* dictionary, the OBDs are confined to congruent correspondences, but they all seem ill-representative of the corpora.

- (29) OF « En liaison avec les entreprises et **dans le cadre de** groupements de projets, nous participons [...]  
TE « Together with industry and **as part of** project associations, we participate in [...] (Label France)
- (30) OF **Dans le cadre de** l'Afrique occidentale française (AOF) [...]  
TE Colonisation, **in the context of** French West Africa, had led to fresh movements of populations on a different scale, [...] (PLECI\_news)
- (31) FR **Dans le cadre de** cette leçon, l'idée est qu'un jeune participant offre le soutien technique aux participants âgés.  
ENG **Within the framework of** this lesson, an idea is to have one of the younger participants provide the technical support for the older participants (Linguee)
- (32) FR Ce type de campagnes peut être financé **dans le cadre de** notre programme de santé publique.  
ENG Such campaigns can be supported **under** our public health programme (Reverso Context)
- (33) FR Demander à l'OMPI, **dans le cadre de** son mandat, d'étendre la portée de ses activités visant à réduire la fracture numérique, [...]  
ENG To request WIPO, **within** its mandate, to expand the scope of its activities aimed at bridging the digital divide [...] (WeBiText)
- (34) FR Nous entendons un autre argument **dans le cadre de** ce débat, à savoir que cela ne nous coûtera rien, que les Américains veulent tant notre approbation

qu'ils sont prêts à tout absorber.

ENG The other argument **in** this whole debate is that it is a free ride; that the Americans are so interested in us rubber stamping their program that they are willing to give us a free ride (TradooIT)

#### 4.3.4.2. Contribution of each tool

The Label France corpus indicates that while congruent correspondences are more frequent (*as part of, in the context of, within the framework of*), simple prepositions can be used to translate the French CP *dans le cadre de* as well. The OBCs provide both types of correspondence. More interestingly and despite their macro-issues, the grouped translations of *TradooIT* and *ReversoContext*<sup>99</sup> suggest more possibilities to translate the French CP than the corpora, e.g. *in connection with, on the occasion of, for the purpose of, within the context of*, etc.

Finally, the OBDs are once more confined to congruent correspondences, except for the *Oxford* dictionary which suggests *within* and *on*. *Reverso-Collins* suggests only one translation while *Larousse* is richer, both regarding translation variety and examples, however not as much as *Oxford*. *Larousse* also suggests a translation that is absent from the corpora or the OBCs (*within the scope of*), but *Oxford* seems to have useful propositions which are mentioned in the grouped translations of *ReversoContext* and *TradooIT* (*on the occasion of*) and should therefore be examined in larger corpora.

#### 4.3.4.3. Translation control: back-translation and mutual correspondence

No strong tendencies can be brought out by a back-translation analysis, since there are only 17 occurrences of *dans le cadre de* in TF in the PLECI\_news corpus. It seems to corroborate the results described above however, as the French CP mainly has *as part of, within, in* and *under* as its source. There are no instances of *in the context of* and *within the framework of*, which would suggest that the relatively high frequency of this CP in TE is a feature of translation rather than of the English language, but this needs to be further investigated in larger corpora. Out of the various correspondence pairs, *dans le cadre de* and *as part of* have the highest mutual correspondence, with a percentage of 11.43%.

---

<sup>99</sup> 40,026 and 88,424 results respectively.

#### 4.3.5. EN FAVEUR DE

##### 4.3.5.1. Frequent English translations suggested by the tools

According to the three types of tools, there are 18 translation alternatives for the French CP *en faveur de* in English, both congruent and divergent (simple prepositions and verbs). Most translations only appear once or twice in one or two of the tools (e.g. zero correspondences only appear once in *Linguee* and *TradooIT*), but four of them stand out from the rest (see Table 7).

Types of TT TTs	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
Verb	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓
For	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓
In favour of	✓	✓	✓	✓	×	✓	✓	✓ <sup>100</sup>	✓	✓	✓
In support of	✓	×	✓	✓	×	✓	✓	✓	×	×	×

Table 7. Most frequent translations of *en faveur de* in the three types of tools.

In the Label France corpus, *en faveur de* is most commonly translated by verbs such as *to promote* (see example 35), *to support*, *to encourage*, *to favour*, *to boost*, etc. Each of these verbs can be interpreted as infrequent but taken as a whole, this type of correspondence amounts to 31.7% in the translation corpus. This analysis shows that OBCs can also provide correspondences of a totally different order, both within their first sentence pairs and their grouped translations. Once more, we observe that, except for *WeBiText*, the OBCs are somewhat similar to the corpora in terms of the translations they provide (see examples 36-38), but that the translations are almost systematically ranked differently and are of various quantitative sizes according to the tool. For example, *in favour of* accounts for 50% of the first 30 sentence pairs in *ReversoContext* and only 20% of the Label France corpus' occurrences of the French CP. It must be noted that at least a third of the sentence pairs in the OBC belong to a similar context, namely politics, which probably exaggerates the frequency of this particular translation. Finally, *Oxford* better accounts for the various types of translations available to translate the French CP, as *Larousse* and *Reverso-Collins* only suggest *in favour of*.

- (35) OF [...] renforcé en 1999 par de nouvelles mesures **en faveur de** la qualité agricole et du développement [...]

<sup>100</sup> One of the grouped translations is labelled *in favour* in *ReversoContext* but all the examples provided contain the CP in favour of and have therefore been added together with the grouped translation *in favour of*.



- TE [...] reinforced in 1999 by new measures **promoting** agricultural quality and rural development , the number of organic farms has increased on average by 26 % per year in the European Union (Label France)
- (36) FR Enumérer individuellement toutes les sources séparées de financement **en faveur de** ce projet.  
 ENG Individually list all separate sources of funding **for** this project (Linguee)
- (37) FR Je plaiderai **en faveur de** trois domaines d'amendement.  
 ENG There are three areas for amendment which I would plead **in favour of** (ReversoContext)
- (38) FR M. John McKay: Monsieur le Président, je crois avoir parlé pendant les 20 dernières minutes **en faveur de** la motion en général.  
 ENG Mr. John McKay: Mr. Speaker, I think I have spoken for the last 20 minutes **in support of** the generalized motion (TradooIT)

#### 4.3.5.2. Contribution of each tool

The corpora clearly show that different types of correspondence are possible for the translation of the French CP *en faveur de*. In that sense, the OBCs prove useful, as they also suggest a great variety of verbs in their first 30 pairs of sentences, which highlights their ability to show divergent correspondences. The grouped translations in *TradooIT* and *ReversoContext*<sup>101</sup>, which are quite well-represented by the first random examples in this case, do not suggest additional valuable translations as compared to the corpora. This could imply that the list of translations spotted in the corpora is rather exhaustive despite the small amount of data they contain.

Regarding the OBDs, *Larousse* and *Reverso-Collins* only suggest one of the most frequent translations, namely *in favour of*. *Oxford* proves to be more informative, as it both provides a larger variety of translations and includes more examples. This particular OBD does not bring out additional valuable information as compared to the corpora and the OBCs, however. Finally, the representativeness of authentic language in dictionaries is further questioned, since infrequent or non-existent items in the corpora or the grouped translations are presented as translations of *en faveur de*, e.g. *on account of* and *to the benefit of*.

#### 4.3.5.3. Translation control: back-translation and mutual correspondence

The French CP *en faveur de* (26 occurrences in TF) has verbs as its source, such as *to support*, *to promote*, *to foster*, *to advocate*, *to sustain*, etc., as well as the preposition *for*, which corroborates the results described above. However, since there are no occurrences of *in favour*

<sup>101</sup> 19,482 and 37,323 results respectively.

of, the presence of this particular CP in TE could be due to an influence of the SL, but we would need to test this hypothesis against larger corpora.

The mutual correspondence cannot be measured between the French CP and the different verbs in English since they are polysemous and that a few of them can be both verbs and nouns, which would bias the number of occurrences found in the corpus. Moreover, it is paramount to raise awareness of this possible type of correspondence, rather than to establish an exact percentage of the degree of correspondence. For the other items listed in Table 7, the mutual correspondence produces results below 9%, but this is arguably due to the polysemy of the items as well as the small data available on the PLECI\_news corpus.

#### 4.3.6. À LA FIN DE

##### 4.3.6.1. Frequent English translations suggested by the tools

Fewer items seem possible to translate the French CP *à la fin de* in English (14 in total). All these possibilities are congruent (*around the end of, with the end of, to the end of, at the end of and by the end of*), except for one divergent correspondence, namely *in late*, which is part of the most frequent translation (see Table 8). Zero correspondences are non-existent in the tools and the most frequent congruent correspondences are literal.

Types of TT TTs	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso- Collins	Oxford
English corr.											
At the end of	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
By the end of	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗
In late	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗

Table 8. Most frequent translations of *à la fin de* in the three types of tools.

There is a general consensus regarding the CP *at the end of* (see example 39), which is presented as significantly more frequent in all the tools as compared to *by the end of* and *in late* (see examples 40-41). *ReversoContext* is the only OBC where the last two translations are absent, as the totality of its 30 first sentence pairs uses *at the end of* to translate *à la fin de* (although the sentences come from different contexts and origin). However, *by the end of* represents 15.1% of its results when considering the whole data. A closer look at the frequency shows great variation from one tool to the other except for *in late*, which is rather

infrequent, but seems worth mentioning since it is provided by most tools. The OBDs only suggest one out of the three recurrent translations.

- (39) FR Supprime les redémarrages même s'ils sont nécessaires **à la fin de** l'installation.  
ENG Suppress reboots even if they were necessary **at the end of** the installation (ReversoContext)
- (40) FR Canal Satellite prévoit d'atteindre 600 000 abonnés **à la fin de** 1997.  
ENG Canal Satellite expects to achieve 600,000 subscribers **by the end of** 1997 (Label France)
- (41) FR La flotte de sous-marins de la classe Victoria continue de progresser vers un état opérationnel stable, où trois des quatre sous-marins seront disponibles pour les opérations **à la fin de** 2014.  
ENG The Victoria-class submarine fleet continues to progress towards a steady state, in which three of four submarines will be available for operations, which is anticipated to occur **in late** 2014 (TradooIT)

#### 4.3.6.2. Contribution of each tool

The corpora highlight the translation of *à la fin de* by English CPs (*at the end of* and *by the end of*) and one divergent, more infrequent correspondence (*in late*). These three translations are provided by all the OBCs except *ReversoContext*. The translation suggested in *TradooIT* and *ReversoContext*<sup>102</sup>, which are similar but not identical to the Label France corpus, are useful regarding more infrequent translations. Among these, *at the bottom of* seems noteworthy since it is systematically used when *à la fin de* refers to the concrete ending of a document (see example 42).

- (42) FR Vous trouverez **à la fin de** ce document quelques liens vers des sites web susceptibles de pouvoir répondre à votre demande.  
ENG **At the bottom of** this document however you can find a few links to websites that possibly can (ReversoContext)

The three dictionaries are in line with the corpora and the OBCs by suggesting *at the end of*, but do not include the other possibilities. Moreover, *Larousse* provides the translation *close of*, which is absent from the other tools. As this item is not illustrated by any contextual environment, it is rather demanding for users to understand its use.

#### 4.3.6.3. Translation control: back-translation and mutual correspondence

16 occurrences of *à la fin de* are found in TF in the PLECI\_news corpus, where they come from various CPs in English, including *to the end of*, *by the ending of*, *at the sth's end*, *by the end of* and *late in/in late*. Therefore, the back-translation agrees with the previous results, as

---

<sup>102</sup> 18,679 and 26,992 results respectively.

the French CP is not an addition compared to the English sentences and that most of its triggers are CPs. However, there are no instances of *at the end of*, which could suggest that the use of this CP as a translation of *à la fin de* is influenced by French, but only larger corpora could allow safe conclusions on that matter. The mutual correspondence for *à la fin de* and *at the end of* provides a result of 7.5%, so arguably, it should not be the only translation included in dictionaries or by the first random sentence pairs of the OBCs, as *ReversoContext* does.

#### 4.3.7. DANS LE DOMAINE DE

##### 4.3.7.1. Frequent English translations suggested by the tools

According to the tools, 19 items can translate the French CP *dans le domaine de*. 13 of these are congruent (e.g. *in the area of*, *in the sphere of*, *with respect to*, *with regard to*, *in relation to*, etc.), whereas the rest include divergent simple prepositions. Zero correspondences appear to be more frequent; at least in most OBCs. Table 9 includes the most recurrent translations, of which two are congruent and two are divergent.

Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
In the field of	✓	✗	✓	✓	✗	✓	✓	✓	✗	✓	✓
In the area of	✓	✗	✓	✓	✓	✗	✓	✓	✗	✗	✗
In	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✓
Zero corr.	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
On	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗

Table 9. Most frequent translations of *dans le domaine de* in the three types of tools.

The CP *in the field of* (see example 43), which is literal, is the most frequent translation of *dans le domaine de* in the Label France corpus (35.3% of the cases), which is reflected in the OBCs except for *WeBiText*. However, it is overrepresented in *ReversoContext* (86.6%), which consequently does not provide any of the other recurrent translations, namely *on*, *in* and *in the area of* (see examples 44-46). Zero correspondences (see example 47) translate the French CP in 13.3% of the cases in *Linguee* and *WeBiText*, but the proportion is lower in the Label France corpus, with a percentage of 5.9. Additionally, the data suggests a certain degree of flexibility of the phrase in TE (see example 48), as it does not follow the structure preposition-noun-preposition. The OBDs and the grouped translations of *TradooIT* and

*ReversoContext* poorly match the results of the corpora, which is due in the second case to translations that are actually present in the data but not indexed.

- (43) FR Les parties contractantes coopèrent **dans le domaine de** la recherche et du développement technologique sur la base des instruments existants.  
 ENG The Contracting Parties shall cooperate **in the field of** research and technological development in accordance with the existing instruments (ReversoContext)
- (44) FR Cette réunion a examiné les principales conclusions et recommandations issues des travaux de l'OCDE **dans le domaine de** l'eau réalisés en 2007 et 2008.  
 ENG This meeting discussed the key policy conclusions and recommendations emerging from the OECD work **on** water in 2007 and 2008 (WeBiText)
- (45) OF [...] de certains gouvernements, notamment **dans le domaine de** la lutte contre la discrimination.  
 TE But it does not explain everything and certainly not the political abdication of certain governments, particularly **in** the fight against discrimination (PLECI\_news)
- (46) FR Aujourd'hui, à un mois à peine des élections européennes, il advient donc que le traité d'Amsterdam restreint considérablement nos droits **dans le domaine de** la politique sociale.  
 ENG We are therefore now seeing, barely one month before the European elections, that the Amsterdam Treaty in fact limits our rights quite considerably **in the area of** social policy (TradooIT)
- (47) OF [...] apporte aussi à Alcatel ses technologies **dans le domaine de** l'ADSL  
 TE [...] the Canadian firm also brings to Alcatel its **[Ø]** ADSL (Asymmetric Digital Subscriber Line) technologies, fast Internet access via the telephone (Label France)
- (48) FR Les hommes diplômés sont surtout représentés **dans le domaine de** l'ingénierie, de la fabrication et de la construction, où ils occupent en moyenne de 62 à 87 pour cent des postes.  
 ENG **The field in** which men represent the highest proportion of graduates is 'engineering, manufacturing and construction', occupying on average between 62 and 87 per cent of places (TradooIT)

#### 4.3.7.2. Contribution of each tool

The Label France corpus brings out the CP *in the field of* as the most common translation of *dans le domaine de*, followed by *in the area of* and *in*. The first examples in *TradooIT* best represent these three translations, while *WeBiText* and *ReversoContext* seem out of line. This analysis shows that zero correspondences cannot be indexed in the grouped translations of *TradooIT* and *ReversoContext*.<sup>103</sup> The rest of the propositions suggested in the grouped translations are quite infrequent and do not contribute much as compared to the corpora's findings.

<sup>103</sup> Out of 25,066 and 42,278 results respectively.

The OBDs do not seem relevant for the translation of the French CP. *Larousse* is the least useful, as it does not include any of the most frequent translations and only provides one that is absent from the other tools, namely the phrase *as far as something is concerned* (see example 49). *Reverso-Collins* includes an example illustrating the flexibility of English (see example 50), but only *Oxford* informs us that the simple preposition *in* is another possibility. However, this translation is only presented in an illustrative example, which arguably encumbers its access. In other words, OBDs fall short on representing the variety of translations found in the corpora and the OBCs.

- (49) FR **dans le domaine de** la prévention, il y a encore beaucoup à faire  
 ENG **as far as** preventive action **is concerned**, there's still a lot to do (Larousse)
- (50) FR **dans le domaine de** l'environnement  
 ENG **in the** environmental **field** (Reverso-Collins)

#### 4.3.7.3. Translation control: back-translation and mutual correspondence

The back-translation would suggest that the French CP does not come from *in the field of* and *in the area of*, but from *in regard to*, *in* and *with respect to*. However, there are only 9 hits of *dans le domaine de* in TF, which clearly impedes us from establishing clear patterns and calculating the mutual correspondence of these pairs of items. Naturally, the mutual correspondence of *dans le domaine* and *in* is extremely low (0.04%), which reflects the polysemy of the English preposition more than anything else.

#### 4.3.8. EN RAISON DE

##### 4.3.8.1. Frequent English translations suggested by the tools

The juxtaposition of the OBCs, the OBDs and the corpora's results provides a list of 25 translations of the French CP *en raison de*. 15 of these are congruent (e.g. *in relation to*, *by virtue of*, *in light of*, *according to*, etc.), while the rest are divergent (simple prepositions and conjunctions). Zero correspondences and literal translations (*by reason of*) are virtually non-existent. Table 10 highlights the four most recurrent translations, which are mostly CPs.

Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
TTs	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
Because of	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Due to	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗
Because	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	✗
As a result of	✓	✓	✗	✓	✗	✓	✓	✓	✗	✗	✗

Table 10. Most frequent translations of *en raison de* in the three types of tools.

The CP *because of* (see example 51) stands out from the rest of the possible translations in the Label France corpus (53.8% of its findings), whilst its predominance is less marked in the PLECI\_news corpus and *WeBiText*. The translations *due to*, *because* and *as a result of* (see examples 52-54), clearly less frequent in the Label France corpus, are provided by most tools. Finally, the grouped translations, which indicate similar proportions in both tools, corroborate the results of the corpora, whereas the OBDs are at odds with the other tools.

- (51) FR Le terme intégration sociale échappe à toute définition **en raison de** son caractère en évolution constante.  
ENG The term social integration resists definition **because of** its constantly evolving nature (ReversoContext)
- (52) FR L'Institut Paul Scherrer (PSI) n'a pas épuisé non plus les microcrédits **en raison de** retards dus à des oppositions.  
ENG The Paul Scherrer Institute (PSI) did not fully use monies granted **due to** legal delays (Linguee)
- (53) FR Il convient de bien avoir à l'esprit que la commission du contrôle budgétaire se saisit de ces questions **en raison de** la multiplication des rapports spéciaux de la Cour des comptes.  
ENG It should be realised that the Committee on Budgetary Control addresses such issues **because** there are special reports by the Court of Auditors and these are becoming more numerous (TradooIT)
- (54) OF Qui plus est, ces îles où, **en raison de** l'esclavage et de la colonisation, [...]  
TE What is more, these islands where, **as a result of** slavery and colonisation, [...] (Label France)

#### 4.3.8.2. Contribution of each tool

The translation corpora indicate that CPs are more common to translate *en raison de*, and that one of these is clearly a better option, namely *because of*. In this case, the OBCs fail to provide the totality of the most frequent translations (except *TradooIT*) but they do suggest *because of* as more recurrent in their first sentence pairs. The grouped translations in

*TradooIT* and *ReversoContext*<sup>104</sup> contradict the predominance of *because of* and do not suggest any additional valuable translations as compared to the corpora.

The *Oxford* dictionary does not provide any results for the French CP *en raison de*. On the other hand, *Oxford-Collins* and *Larousse* both include *because of* together with examples. The first also suggests *on account of*, which is absent from the corpora and extremely rare in the OBCs.

#### 4.3.8.3. Translation control: back-translation and mutual correspondence

In the PLECI\_news corpus, the French CP *en raison de* mainly comes from *because of* (24% of the 29 occurrences), which confirms the corpora's findings described in the previous section. Other triggers are less frequent, e.g. *as* (see example 55), *because*, *after* (see example 56), *as a result of* or *by virtue of*. The mutual correspondence of *en raison de* and *because of* is of 9.7%, which shows that while it should be mentioned in the TTs, other translations need be included.

- |      |    |  |
|------|----|--|
| (55) | OE | That is why the extension of reservation to the OBCs causes controversy ; no political party dares oppose the move <b>as</b> the OBCs have considerable political clout in votes.  |
|      | TF | C'est pourquoi l'augmentation des quotas en faveur des OBC provoque une controverse. Aucun parti politique n'ose toutefois s'y opposer, <b>en raison de</b> leur poids politique dans la balance électorale (PLECI_news) |
| (56) | OE | [...] had to be shut down in 1995 <b>after</b> a sodium leak and fire, followed by evidence of negligence and a cover - up.  |
|      | TF | [...], a dû être fermé en 1995 à la suite de la tentative de camoufler un incendie par négligence survenu <b>en raison de</b> fuites de sodium (PLECI_news)  |

#### 4.3.9. AUX CÔTÉS DE

##### 4.3.9.1. Frequent English translations suggested by the tools

23 translations in total are found in the corpora, the OBCs and the OBDs. Half of these are congruent (e.g. *on the side of*, *together with*, *along with*, *close to*, etc.) whereas the other half are divergent (simple prepositions such as *among*, *beside*, *alongside*, *with*, etc.). The four main translations pertain to both groups (see Table 11). There are no cases of zero correspondences and the existing literal translations (*at/on/to side of*) are extremely infrequent.

---

<sup>104</sup> 27,573 and 39,031 results respectively.



Types of TT	Corpora		OBCs (first 30 examples)				OBCs (GT) <sup>105</sup>		OBDs		
TTs	Label France	PLECI	Linguee	TradooIT	WeBiText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
English corr.											
Alongside	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓
With	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
Along with	✓	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗
Beside	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✓

Table 11. Most frequent translations of *aux côtés de* in the translation corpora and the OBCs.

There is a frequency pattern between the different tools, with *alongside* (see example 57) as the most frequent translation of *aux côtés de*, except in *WeBiText*, where it is on a par with the simple preposition *with* (see example 58), and *with* as the second most frequent, except in *TradooIT*. Frequency-wise, the Label France corpus clearly highlights *alongside* as the most common translation (65.7% of its occurrences), but this predominance is only shown in *ReversoContext* and in the grouped translations of *TradooIT*. Despite their infrequency, *along with* (see example 59) and *beside* (see example 60) are provided by the majority of the tools or represent more than 5% of the Label France corpus and have therefore been included in the table. None of these translations are included in the OBDs, except for two of the simple prepositions in *Oxford*. Finally, *ReversoContext* fails to provide the majority of the frequent translations in its grouped translations.

- (57) OF Aussi, le LRMH participe-t-il, **aux côtés de** partenaires anglais, allemands et grecs, [...]  
TE The LRMH therefore participated, **alongside** British, German and Greek colleagues, [...] (Label France)
- (58) FR M. Freedman a exercé le droit **aux côtés de** Warren Winkler (à l'époque), Roy Filion et David Wakely avant son entrée à La Commission en 1979.  
ENG Mr. Freedman practised law **with** Warren Winkler (as he then was), Roy Filion and David Wakely before joining the Board in 1979 (WeBiText)
- (59) FR Le Canada, **aux côtés de** ses alliés internationaux, intervient en Afghanistan pour aider le peuple afghan à reconstruire son pays et l'empêcher de jamais redevenir un havre pour les terroristes.  
ENG Canada, **along with** its international allies, is engaged in Afghanistan to help the Afghan people rebuild their country and to prevent Afghanistan from ever again becoming a haven for terrorists (TradooIT)
- (60) FR Parfois, quelques personnes prennent place **aux côtés de** la personne qui présente la députation.  
ENG Sometimes, a few people will go up to the table to sit **beside** or stand behind the person presenting the deputation (Linguee)

<sup>105</sup> Out of 2,190 and 1,571 results respectively.

#### 4.3.9.2. Contribution of each tool

The main translation of *aux côtés de* found in the Label France corpus is the divergent *alongside*, but three further possibilities are brought out (*along with*, *with* and *beside*). Although the OBCs do not show the predominance of the first as compared to the three others in their first 30 examples, they provide these four translations and therefore seem to be representative of authentic usage, especially *ReversoContext*. The main advantage of the grouped translations in *TradooIT* and *ReversoContext*<sup>106</sup> is to directly access various translations and their authentic examples, as it shows that even more infrequent translations prove highly relevant in certain contexts. It is extremely valuable to have our attention drawn to, for example, *stand by/with*, as it is often used when translating *rester/se tenir aux côtés de* (see example 61). Also, *next to* seems to be a good option when translating *aux côtés de* in a concrete, geographical way (see example 62). These options were not necessarily absent from the translation corpora but, given their small data, these valuable translations might have been misinterpreted as irrelevant.

- (61) FR « Notre gouvernement reste inébranlable dans sa volonté de **se tenir aux côtés de** la population ukrainienne face à l’agression militaire persistante du régime Poutine, qui a déjà coûté la vie à plus de 5 300 personnes.  
ENG “Our Government remains steadfast in its commitment to **stand with** the people of Ukraine in the face of the Putin regime’s ongoing military aggression, which has already cost the lives of more than 5,300 people (TradooIT)
- (62) FR Tous les soirs, assis **aux côtés de** l’empereur d’Autriche.... Je jouais des duos avec lui... et corrigeais sa lecture de la musique.  
ENG Night after night I sat right **next to** the emperor of Austria playing duets with him, correcting the royal sight-reading (ReversoContext)

Out of the four most frequent translations of *aux côtés de* found in the Label France corpus, only two are provided in one of the three OBDs (i.e. *beside* and *alongside* in *Oxford*). The rest of the translations suggested in the dictionaries are literal correspondences, which wrongly suggests that this type of correspondence is the most common whilst it is extremely infrequent in the translation corpus. In addition, *Larousse* and *Reverso-Collins* could mislead users into believing that there are only one or two possibilities to translate the French CP. Interestingly, *Larousse* and *Oxford* suggest the translation *next to*, which could be interesting if illustrated, but it is not the case.

---

<sup>106</sup> 2190 and 1571 occurrences respectively.

### 4.3.9.3. Translation control: back-translation and mutual correspondence

The 9 occurrences of *aux côtés de* in TF found in the PLECI\_news corpus are never zero correspondences but come from *with* or *alongside*. The mutual correspondence of the French CP and *alongside* provides a result of 19.04%, which supports the presence of this English preposition in the tools.

### 4.3.10. AU SERVICE DE

#### 4.3.10.1. Frequent English translations suggested by the tools

The combination of the three types of tools' results provides a list of 20 alternative translations of *au service de*. The main correspondence type of these translations is congruent, as CPs, amongst which literal, are more numerous (15 in total, e.g. *in the service of*, *in view of*, *in the interest of*, etc.) as compared to the divergent ones, which only contain four simple prepositions (*for*, *with*, *at* and *to*). However, Table 12 shows that congruent correspondences are not systematically the most recurrent, since divergent ones (simple prepositions and verbs) are also part of it. Zero correspondences are extremely rare in the OBCs and are absent from the corpora.

Types of TT TTs	Corpora		OBCs (first 30 examples)				OBCs (GT)		OBDs		
	Label France	PLECI	Linguee	TradooIT	WeBText	Reverso Context	TradooIT	Reverso Context	Larousse	Reverso-Collins	Oxford
Verb	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓
In the service of	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗
At the service of	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✗
For	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✓

Table 12. Most frequent translations of *au service de* in the three types of tools.

The Label France corpus suggests that verbs (see examples 63) are the most frequent translation of the French CP *au service de* (such as *to serve*, *to work for*, *to aim at*, *to contribute*, *to supply*, *to benefit*, etc.). The verbs provided in the OBCs are similar to those, although their proportion is not as large, except in *Linguee* (see example 64). The frequency of the congruent *in the service of* (see example 65) is almost identical as the number of verbs in the corpus. All the tools rank these four recurrent translations differently and also with dissimilar proportions (see examples 66-67 for the remaining translations). Finally, the grouped translations agree with the corpus findings, as opposed to the OBDs. *Oxford*, however, shows its ability to suggest divergent correspondences.

- (63) OF [...] économique bien conduite peut être **au service de** la création culturelle.  
 TE A well - run economy can **serve** cultural creative activity (Label France)
- (64) FR Il souhaite mettre les compétences ainsi acquises **au service de** son pays.  
 ENG He plans to use his expertise **to help** his country (Linguee)
- (65) FR Mais il est temps que tu mettes tes enseignements **au service de** ton pays.  
 ENG But it is time that you put your accomplishments to use **in the service of** your country (ReversoContext)
- (66) FR Nous désirons accompagner ce développement par quelques orientations éthiques ciblées, soulignant que toute croissance doit toujours être **au service de** l'être humain et du bien commun. »  
 ENG We wish to accompany this development with some appropriate ethical guidelines which stress the fact that all growth must always be **at the service of** the human being and the common good." (TradooIT)
- (67) FR Ces programmes représentent un instrument indispensable **au service de** la croissance et de l'emploi.  
 ENG Those programmes are an indispensable instrument **for** growth and jobs (Linguee)

#### 4.3.10.2. Contribution of each tool

The translation corpora show that two types of correspondences are valuable options to translate the French CP *au service de*, namely congruent (English CPs) and divergent (one simple preposition and verbs), which again suggests that it is crucial to free other TTs from any part-of-speech constraint, i.e. where one item is translated by an item of the same category.

Interestingly, the OBCs generally include the most frequent translations found in the Label France corpus in their first random examples, including the divergent correspondences. *WeBiText* is undoubtedly less accurate in this area, since three out of the four translations presented in Table 12 are either rare or absent. The grouped translations are quantitatively different in *TradooIT* and *ReversoContext*,<sup>107</sup> but both tools contradict the quantitative information of the corpus. The verbs, for instance, represent less than 10% of their respective results.

Finally, *Reverso-Collins* has no results for this French CP, and the other two OBDs are in complete contradiction with both the corpora and the OBCs. *Larousse* does not provide any of the translations presented, and, while disregarding the translations *in the service of* and *at the service of*, *Oxford* is the only one to include the divergent correspondences.

<sup>107</sup> 9,156 and 6,630 results respectively.

#### **4.3.10.3. Translation control: back-translation and mutual correspondence**

Given the low frequency of the French CP in TF (13 occurrences), it is arduous to contrast the results but they still show that *au service de* can have as its source a verb, the simple preposition *for* as well as literal correspondences (such as *at/in/to the service of* and *as service for*). The mutual correspondence for *au service de* and *in the service of* is of 40%, but this high percentage may be influenced by the low number of both CPs in OE as well as TE and should be further investigated in larger data. It strongly supports, however, that the English CP should be provided in the bilingual dictionaries.

## 5. DISCUSSION AND CONCLUSION

This last chapter will first provide a summary of the analysis conducted in the previous sections. The results of this analysis concern three areas: (a) the OBCs and OBDs' degree of similarity to authentic language, (b) the translation process of CPs from French to English and (c) the relation between the two languages in this area. Regarding the first issue, the results enable us not only to provide an answer to our research question, i.e. how reliable OBCs are, but also to gain insights into the pros and cons of each type of TT as well as into their respective contributions. In the second part of the chapter, we will present concluding remarks and highlight avenues that could be explored in the future.

### 5.1. SUMMARY OF THE ANALYSIS

#### 5.1.1. *THE OBCS AND OBDs' DEGREE OF SIMILARITY TO AUTHENTIC LANGUAGE*

The study has shown that the agreement between the corpora and the online TTs varies from one French CP's analysis to the other. However, one outcome can be generalized: OBDs fundamentally differ from the corpora's findings while the OBCs partially reflect them, i.e. a majority of the OBCs provide a great deal but not all of the most recurrent translations spotted in the corpora, albeit with different quantitative supports. More precisely, the major difference between the corpora and the OBCs is not the list of translations suggested but the ranking and proportions thereof. Future research should look upon the issue of OBCs' reliability on the basis of other types of words and with larger corpora in order to corroborate this finding. As mentioned above, the contributions of each type of tool has been highlighted through the analysis as well. In view of those, we can confirm the hypothesis formulated in Section 1.2., as the study strongly supports Johansson (2007: 21) and his conviction that a combination of tools (within a same interface) would compensate for their respective shortcomings and therefore give access to a more thorough and coherent approach to language. The advantages, disadvantages and contributions of the TTs are the following:

- *Translation corpora*

Translation corpora have been used in this study to offer quantitative support to our analysis and as a control to evaluate the representativeness and authenticity of the results provided by the OBCs and the OBDs. Their main advantages are that they both increase our knowledge on the differences and similarities between languages and provide examples of language in use, rather than "citations of lexicographers and the usually context-free examples thought up by

grammarians” (Johansson 1999: 21). As Johansson (1999: 20) points out, corpora also highlight the degree of correspondence between two items and attest that this degree is dependent on the context. Finally, this type of TT makes a clear distinction between SL and TL and does not have part-of-speech constraints.

The main disadvantage in this case was their small size, so that they may not have contained valuable translations or may have misled us into disregarding potentially interesting ones due to their low frequency. The second drawback agrees with the caveat of Altenberg & Granger (2002: 17) cited earlier in Section 1.3.2.3., i.e. by resorting to quantitative measurement, we run the risk of missing out on valuable translations and hence on contrastive insights. This could be alleviated by relying on larger data, which is substantiated by the very fact that additional valuable translations were brought out by *ReversoContext*, which has a larger database than the Label France and the PLECI\_news corpora. Furthermore, we agree that referring to corpora might be time-consuming and cumbersome for beginners.

- *OBCs*

The main issues concerning the OBCs are that the SL and the TL are not distinguished and that the variety of registers is limited (perhaps to a lesser extent in *ReversoContext*). On the other hand, the analysis revealed two major positive aspects. First, the OBCs have a great capacity to show a variety of translations in their first random sentence pairs, which is all the more useful given that each possible translation is presented within a context. Second, they are able to provide translations of a different category. Whereas the OBCs all proved efficient vis-à-vis the second aspect, they are not to be lumped together regarding the first, as their reflecting authentic language considerably varies from one another and from one French CP to the others, and so is their respective usefulness. In short:

- *Linguee* is fast, responsive, excellent in terms of quality at the macro-level and it partially matches the corpora’s findings. These positive outcomes are stained by severe weak points however, as the tool only provides random sentence pairs (i.e. there are no grouped translations) without any possibility to enquire further examples than the first 30. Additionally, it is not word-aligned and consequently does not highlight the translations, hampering ease of access. In other words, this analysis suggests that the OBC’s popularity is not entirely justified.
- *WeBiText* is slow and particularly problematic regarding macro-quality, leading to extremely poor micro-quality. Its reliability as a TT has therefore been questioned

throughout the analysis to the point where its use is altogether strongly discouraged. This is at odds with Simard (2013: 40), who describes the tool as offering more acceptable and varied translations in comparison with more traditional tools. Also, all the improvements intended for the future by the creators (see p. 46) have not been effected.

- *TradooIT* is fast, moderate as far as its macro-quality is concerned, hybrid and offers comparable results in relation to the corpora. It stands out because the whole data can be investigated and because frequency as well as data size information are provided, which is more comparable to what a corpus would do. Moreover, the OBC offers the grouped translations option, the possibility to filter results according to corpora, a word-alignment system and a highlighting option, however with an amount of errors detrimental to its reliability. These findings agree with the study of Volk *et al.* (2014: 3174) explained in Section 3.2.4. (p. 44).
- *ReversoContext* is fast, hybrid, sophisticated, collaborative and accurate at the macro-level as well as regarding word-alignment and highlighting. Similarly to the previous OBC, it provides grouped translations, however much more precisely, and frequency information. In addition, the tool offers an extremely useful filter option,<sup>108</sup> but should provide information on its corpora size and offer the possibility to filter across the corpora, as in *TradooIT*. Besides the fact that the tool seems to be representative of authentic language, the most important outcome of our analysis is that, in this case, it brings out new information compared to the corpora. These two findings contribute to validating the use of this OBC because of its reliability, relevance and usefulness.

The grouped translations are arguably the best feature *TradooIT* and *ReversoContext* and are at the source of the second tool's usefulness. They allow users to compare the items both in terms of recurrence and, because of the authentic examples provided, in terms of contextual environment as well. Also, they mitigate all the drawbacks of the first random examples, which can overwhelm users, and particularly beginners, and be poorly representative of the actual translations contained in the whole data. Moreover, as the whole data can be investigated, these OBCs prove more useful than those only providing random pairs of sentences. The amount of data, being larger than in the translation corpora, also generates more translation possibilities, which contradicts Esplà-Gomis *et al.* (2015: 5) and their remark

---

<sup>108</sup> As we have seen, this option gives users access to a translation, its frequency and examples. Though extremely useful and sophisticated, it is not the perfect solution to compensate for the absence of recurrent correspondences in the grouped translations, because it implies that users should know the translation beforehand in order to search for it.



on the low coverage of *ReversoContext*. Also, the large data better shows the discrepancy between an extremely common translation, a relatively common one and an infrequent one. Again, because of the large data available, the tools facilitate decisions on which item suits best one particular context.

On the other hand, several problems in the grouped translations have induced that there is still room for improvement and, while it is more pertinent to rely on them rather than on the first examples, users should remain critical. The main issue concerns the poor matching between the grouped translations and the first random sentence pairs provided. It seems unfortunate to provide random examples that ill-represent translations statistically proved to be recurrent; however failing to provide a frequent translation in the grouped translations is arguably even more misleading for users, as they will miss out on a valuable possibility (e.g. *in* is not indexed as a translation for *en matière de* although it is its most frequent translation). The real question perhaps revolves around the relevance of presenting random examples when the website actually has the exact frequency of each variant at its disposal and allows users to filter the results according to one of them. Finally, the grouped translations do not bring out cases of zero correspondences correctly, as these are more commonly indexed within an erroneous grouped translation, especially in *TradooIT*, but this would require a highly sophisticated algorithm system. All things being equal, this is not the most problematic aspect to consider, as opposed to what Danlos & Roze (2011: 5) suggest in their conclusion.

To conclude on OBCs, we believe that their future lies in an improved version of the system adopted by *ReversoContext* combined with additional features offered by the other OBCs, i.e. the possibility to investigate the whole data, the frequency information, the hyperlinks, a strong word-alignment system (which shows the importance of macro-quality), a highlighting option, the collaborative community, a sophisticated dictionary section and especially the grouped translations option. On the other hand, the variety of registers must be expanded, while a solution must be found to separate SL from TL, which completely contradicts Désilets *et al.* (2009: 6), who consider these aspects as subsidiary.

- *OBDs*

An extremely misleading tendency has been highlighted when comparing the results of OBCs and corpora with those of the dictionaries, which would be best qualified as inconsistent: the translations are different from one OBD to the other as well as from the corpora. Generally, the most frequent translations are not provided, while uncommon ones are. For instance, the

OBDs do not provide the translations *in*, *in the area of* and *for*, which are the most common ways to translate *en matière de*, but instead all present translations that are not supported by the corpora, such as *as regards* and *as far as something is concerned*, which coincides with Granger & Lefer (in press). The *Oxford* dictionary seems to be slightly more efficient in that matter, as in six cases out of ten, it is the only one to include an important translation. However, even in cases where the dictionary provides valuable translations, it never brings out new information in comparison with the other two types of tools.

In addition to the discrepancy as compared to authentic usage, the OBDs (a) are generally more inclined to suggest congruent correspondences, (b) restrict the list of possible translations,<sup>109</sup> which corroborates Granger & Lefer (in press)'s caveat that this type of tool is nowhere near to provide an equivalent richness of units as that revealed by corpora and (c) provide scarce contextual information. The last two problems show that OBDs do not exploit the space at their disposal. It is important to note that *Oxford* is richer in terms of the numbers of translations it suggests, but illustrative examples do not systematically accompany these different translations. As a consequence, the richness of information may backfire on users, as the lack of illustrative examples is not only terribly inconvenient but also simply misleading for beginners. In other words, these OBDs are deceitful regarding the translations of the French CPs and should not be used uncritically and without the support of corpora.

### 5.1.2. TRANSLATION PROCESS OF CPs FROM FRENCH TO ENGLISH

The analysis of the Label France and the PLECI\_news corpora revealed interesting aspects to consider when translating French CPs into English. First, there is usually a vast array of possible translations, which confirms Cosme & Gilquin (2008: 261). This finding is corroborated by the low percentage of mutual correspondence, i.e. there is no one strict equivalent for each French CP but different items that can be used depending on the context. Second, congruent correspondences are definitely not the only option, as divergent ones are most commonly used (simple prepositions, but also conjunctions and verbs). Also, the corpora brought out different patterns: (a) the group of the most recurrent translations (which we presented in Tables 2 to 12) varies in size according to the French CP (e.g. *au sein de* is most commonly translated by *within* or *in*, while there are seven frequent ways to translate *en matière de*) and (b) the line between highly and less frequent translations also depends on the French CP (e.g. *because of* clearly stands out from the rest of the recurrent translations of *en*

---

<sup>109</sup> In seven cases out of ten, *Reverso-Collins* only provides one possible translation.

*raison de*, as opposed to the case of *par rapport à*, where the different translations are more on a par).

### 5.1.3. CONTRASTIVE INSIGHTS

As mentioned in the introduction, contrastive linguistics gives insights into the languages compared, and this study has yielded interesting results regarding the area of CPs. While both languages have this type of items at their disposal, we have observed that a meaning expressed by a CP in French is not systematically rendered by one in English. The phrases between the two languages can share a similar number of words and be literal translations of each other (e.g. *à la fin de* vs. *at the end of*), be composed of a different number of words that are not literal translation of each other (e.g. *en raison de* vs. *due to*) or be of different category altogether (e.g. *en faveur de* vs. *for* or *to favour*). In other words, we can put forward that CPs are not used identically in French and English. Naturally, this would need to be cross-analysed with a study going from English to French.

In addition, contrastive linguistics also improves our knowledge of the individual languages compared. Again, this section would require a similar study working from English to French in order to be complete, but this study gives interesting insights into the status and stability of CPs in French. According to our analysis, these are still not fully recognized as recurrent MWU (multi-word units), as they are not all included in the OBDs reviewed<sup>110</sup> and most typically not granted the status of headword. However, this lack of establishment varies from one CP to the other, as they are presented differently both in terms of their status and the size of their entries.<sup>111</sup> The criteria of the CP's inclusion in bilingual dictionaries and of the amount of attention they deserve are unclear.

## 5.2. CONCLUSION AND FURTHER CONSIDERATIONS

The main purpose of this MA thesis has been to investigate the reliability of OBCs, which we believe to be of the utmost importance given their growing popularity. As we have seen in the assessment of the tools (Section 3.2.4.), OBCs are still rarely the focus of linguistic studies but even lesser attention has been paid to the quality of the translations they suggest. In order to compensate for this shortage of evaluation, we have conducted an empirical analysis on *Linguee*, *TradooIT*, *WeBiText* and *ReversoContext* on the basis of the translations provided

---

<sup>110</sup> Similarly, only five French CPs are provided as such in the *Linguee*'s dictionary.

<sup>111</sup> As can be seen in Appendix 7, *en matière de* is presented in a short sub-entry in *Oxford* whereas *par rapport à* or *en faveur de* are presented in separate and longer entries.

for ten French CPs. After a presentation of each tool and a general examination of their macro-structure, the translations have been compared to authentic language through two translation corpora (the Label France and the PLECI\_news corpora), as an evaluation based on intuition would have been detrimental to the objectivity of the results. Concurrently, the French CPs have also been searched for in three OBDs (*Larousse*, *Reverso-Collins* and *Oxford*), allowing us to spot differences and similarities as compared to the OBCs and to better identify their advantages and disadvantages.

Through the survey of the online TTs (Section 3), we have discovered that the OBCs are not identical, while the main outcome of the analysis is that their first 30 sentence pairs partially match the corpora's findings but do not contribute much. Most importantly, the analysis revealed that the option of the grouped translations displayed in *TradooIT* and *ReversoContext* is arguably the best feature of this type of tool and, although improvements should be effected in the future, it tends to confer a higher degree of reliability upon these two tools and more particularly upon the second one. This divergence between the two OBCs is explained by the accuracy of *ReversoContext*'s word-alignment system, which substantiates Bourdaillet & Langlais (2012) and their statement that improved word-alignment is the best way to increase the efficiency of concordancers.

In a broader perspective, this study has shown that both coverage and precision matter when translating, and that it requires tools offering not only the support of authentic language but also large amount of it, i.e. *ReversoContext* proved richer than the small translation corpora used mostly because of the size of its data. Second, we have seen how contextual information is crucial in order to determine which translation best suits a particular use of, in this case, a French CP. These two findings both incriminate the use of OBDs and encourage the use of OBCs. However, because they do not clearly indicate the direction of the translation and cannot be investigated as corpora (except for *TradooIT* and *ReversoContext*), they should only be used by a certain type of users, i.e. those with a proficiency as regards both languages so as to be able to spot downright mistakes and also question the translations suggested, which agrees with Precup-Stiegelbauer (2013: 1776). The belief that “[b]eginners in a new language might still be served better by a dictionary” (Volk *et al.* 2014: 3172) is nevertheless frowned upon, at least in their current state. In light of our results, we therefore conclude that users would better benefit from a combination of different TTs or, better even, from an interface providing a user-friendly version of the three types of tools reviewed here.

## REFERENCES

### 1. TRANSLATIONS TOOLS: OBCs, OBDs AND AUTOMATIC TRANSLATORS

**Babelfish.** Accessed on 2 November 2014. <http://www.babelfish.fr>.

**Bab.la.** Accessed on 2 November 2014. <http://en.bab.la>.

**FreeTranslation.** Accessed on 2 November 2014. <http://www.freetranslation.com/fr/>.

**Glosbe.** Accessed on 2 November 2014. <https://fr.glosbe.com>.

**Google Translate.** Accessed on 2 November 2014. <https://translate.google.com>.

**Larousse.** Accessed on 16 January 2016. <http://www.larousse.fr/dictionnaires/francais-anglais>.

**LEO.** Accessed on 16 October 2014. [http://dict.leo.org/frde/index\\_de.html](http://dict.leo.org/frde/index_de.html).

**Linguee.** Last accessed on 28 April 2016. <http://www.linguee.com/english-french>.

**MyMemory.** Accessed on 2 November 2014. <http://mymemory.translated.net>.

**ReversoContext.** Last accessed on 28 April 2016. <http://context.reverso.net/traduction>.

**Reverso-Collins.** Accessed on 16 January 2016. <http://dictionnaire.reverso.net/francais-anglais/>.

**TAUS.** Accessed on 16 October 2014. <https://www.tausdata.org/index.php>.

**TradooIT.** Last accessed on 28 April 2016. <https://www.tradooit.com>.

**TransSearch.** Accessed on 2 November 2014. <http://tsrali3.com>.

**WebiText.** Last accessed on 28 April 2016. <http://www.webitext.com/bin/webitext.cgi>.

### 2. ACADEMIC LITERATURE

ABEL, A. 2012. 'Dictionary Writing Systems and Beyond.' In Granger, S. & Paquot, M (eds.), *Electronic lexicography*. Oxford, UK: Oxford University Press, 83-106.

ADLER, S. 2001. *Les locutions prépositives : questions de méthodologie et de définition*. Travaux de linguistique 42-43. 157-170. Accessed on 22 December 2015. <http://www.cairn.info/revue-travaux-de-linguistique-2001-1-page-157.htm?iframe=true&width=100%&height=100%>.

- ALONSO JIMÉNEZ, E. A. 2013. “‘Linguee’ y las nuevas formas de traducir.’ *Skopos: revista internacional de traducción e interpretación* 2: 5-28. Accessed on 7 November 2014. <https://rio.upo.es/xmlui/handle/10433/851>.
- ALTENBERG, B. 1999. ‘Adverbial Connectors in English and Swedish: Semantic and Lexical Correspondences.’ In Ebeling, S. O. & Hasselgård, H. (eds.), *Out of Corpora. Studies in honour of Stig Johansson*. Amsterdam & Atlanta: Rodopi, 249-268.
- ALTENBERG, B. & GRANGER, S. 2002. ‘Recent Trends in Cross-Linguistic Lexical Studies.’ In Altenberg, B. & Granger, S. (eds.), *Lexis in Contrast. Corpus-Based Approaches*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 3-47.
- ARCAN, M., THOMAS, S-M., DE BRANDT, D. & BUITELLAR, P. 2013. ‘Translating the FINREP taxonomy using a domain-specific corpus.’ In Way, A., Simaan, K., Forcada, M. L., Grasmick, D. & Depaetere, H. (eds.), *Proceedings of the XIV Machine Translation Summit*. Nice, France, 2-6 September 2013, 199–206. Accessed on 2 December 2015. <https://www.insight-centre.org/sites/default/files/publications/mt-summit-2013-arcan-et-al.pdf>.
- BAISA, V. & SUCHOMEL, V. 2014. ‘SKELL: Web Interface for English Language Learning.’ In Horák, A. & Rychlý, P. (eds.), *Proceedings of the 8<sup>th</sup> Workshop on Recent Advances in Slavonic Natural Language Processing*. Karlova Studánka, Czech Republic, 5-7 December 2014, 63-70. Accessed on 27 December 2015. <https://nlp.fi.muni.cz/raslan/raslan14.pdf#page=71>.
- BAKER, M. 1996. ‘Corpus-based translation studies: The challenges that lie ahead.’ In Somers, H. (ed.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 175-186.
- BAKER, M. 2007. ‘Patterns of Idiomaticity in Translated vs. Non-Translated Text.’ *Belgian Journal of Linguistics* 21: 11-21.
- BOUCHARD, A. 2014. Recherche d’Information sur Internet (Perfectionnement): Méthodologie et Outils Disponibles. Paris, France: Urfist Paris. Accessed on 4 January 2016. [http://www.fouad-web.com/FouadWeb/Download/FWC/Veille\\_Technologique\\_1A/Recherche-internet.pdf](http://www.fouad-web.com/FouadWeb/Download/FWC/Veille_Technologique_1A/Recherche-internet.pdf).
- BOURDAILLET, J. & LANGLAIS, P. 2012. ‘Identifying Infrequent Translations by Aligning Non Parallel Sentences.’ In Simard, M., Liu, Q. & Quirk, C. (eds.), *Proceedings of the 10<sup>th</sup> Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*. San Diego, USA, 28 October-1 November 2012. Accessed on 14 April 2014. <http://rali.iro.umontreal.ca/rali/sites/default/files/publis/AMTA-2012-Bourdaillet.pdf>.

- BOWKER, L. 2012. 'Meeting the Needs of Translators in the Age of E-Lexicography: Exploring the Possibilities.' In Granger, S. & Paquot, M. (eds). *Electronic Lexicography*. Oxford, UK: Oxford University Press, 379 – 399.
- BRALA, M. 2002. 'Understanding and Translating (Spatial) Prepositions: An Exercise in Cognitive Semantics for Lexicographic Purposes.' *Working paper of the Research for English and Applied Linguistics*. Cambridge, Cambridge University. Accessed on 18 December 2015. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.8062&rep=rep1&type=pdf>.
- BRENDA, M. 2014. *The Cognitive Perspective on the Polysemy of the English Spatial Preposition "Over"*. Newcastle, UK: Cambridge Scholar Publishing.
- BROWN P. F., COCKE, J., DELLA PIETRA, S. A., DELLA PIETRA, V. J., JELINEK, F., LAFFERTY, J. D., MERCER, R. L. & ROUSSIN, P. S. 1990. 'A statistical Approach to Machine Translation.' *Journal Computational Linguistics archive* 16. 2: 79-85. Accessed on 9 December 2014. <http://dl.acm.org/citation.cfm?id=92860>.
- BUYSE, K. & VERLINDE, S. 2013. 'Possible Effects of Free on Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of Linguee and the Interactive Language Toolbox.' *Procedia - Social and Behavioral Sciences* 95: 507–512. Accessed on 17 December 2015. <http://www.sciencedirect.com/science/article/pii/S1877042813041955>.
- CAMPRUBI, M. 1997. 'Les locutions prépositives dans une approche contrastive.' *Faits de langues* 5.9: 185-192. Accessed 22 December 2015. [http://www.persee.fr/doc/flang\\_1244-5460\\_1997\\_num\\_5\\_9\\_1154](http://www.persee.fr/doc/flang_1244-5460_1997_num_5_9_1154).
- CHESTERMAN, A. 1998. *Contrastive Functional Analysis*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- CHUQUET, H. & CHUQUET, J. 2006. 'De la "Circonstance" à la "Cause": les Prédications en "with" et leur Traductions en Français.' In Chuquet, H. & Paillard, M. (eds.), *Causalité et Contrastivité, Études de Corpus*. Rennes, France: Presses Universitaires de Rennes, 189-210.
- COBB, T. 2013. 'Frequency 2.0: Incorporating homofoms and multiword units in pedagogical frequency lists.' In Bardel C., Lindqvist C. & Laufer B. (eds.), *L2 vocabulary acquisition, knowledge and use: new perspectives on assessment and corpus analysis*. Montréal, Canada: Eurosla, 79–107.
- CONNOR, U. M. & MORENO, A. I. 2005. 'Tertium Comparationis: A vital component in contrastive research methodology.' In Bruthiaux, P., Atkinson, D., Eggington, W. G., Grabe, W. & Ramanathan, V. (eds.), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevedon, England: Multilingual Matters, 153-164.

- COSME, C. & GILQUIN, G. 2008. 'Free and Bound Prepositions in a Contrastive Perspective. The case of "with" and "avec".' In Granger, S. & Meunier, F. (eds.), *Phraseology: An Interdisciplinary Perspective*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 259-274.
- DANLOS, L. & ROZE, C. 2011. 'Traduction (automatique) des connecteurs de discours.' *Actes 18e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*. Montpellier, France, 27 June-1 July 2011. Accessed on 23 October 2015. <http://utilisateurs.linguist.univ-paris-diderot.fr/~danlos/Dossier%20publis/TradConn-TALN'11.pdf>.
- DELISTE, J. & FIOLA, M. A. 2013 [2003]. *La Traduction Raisonnée – Manuel d'Initiation à la Traduction Professionnelle de l'Anglais vers le Français (3<sup>rd</sup> Ed.)*. Ottawa, Canada: University of Ottawa Press.
- DÉSILETS, A. 2007. 'Translation Wikified: How will Massive Online Collaboration Impact the World of Translation?' *Proceedings of the 29<sup>th</sup> International Conference on Translating and the Computer*. London, UK, 29-30 November 2007. Accessed on 16 January 2016. [www.mt-archive.info/Aslib-2007-Desilets.pdf](http://www.mt-archive.info/Aslib-2007-Desilets.pdf).
- DÉSILETS, A., FARLEY, B., PATENAUDE, G. & STOJANOVIC, M. 2008a. 'WebiText: building large heterogeneous translation memories from parallel web content.' *Proceedings of the 30<sup>th</sup> International Conference on Translating and the Computer*. London: UK, 23-28 November 2008. Accessed on 11 April 2015. <http://www.mt-archive.info/Aslib-2008-Desilets.pdf>.
- DÉSILETS, A., HERMET M. & SZPAKOWICZ S. 2008b. 'Using the Web as a Linguistic Resource to Automatically Correct Lexico-Syntactic Errors.' In Calzolari, N. (ed.), *Proceedings of the 6<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*. Marrakech, Morocco, 26 May-1 June. Accessed on 3 March 2015. [http://www.lrec-conf.org/proceedings/lrec2008/pdf/220\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/220_paper.pdf).
- DÉSILETS, A., MELANÇON, C., PATENAUDE, G. & BRUNETTE, L. 2009. 'How translators use tools and resources to resolve translation problems: an ethnographic study.' In Gerber, L. (ed.), *Proceedings of Beyond Translation Memories: New Tools for Translators Workshop MT summit XII*. Ottawa, Canada, 26-30 August 2009. Accessed on 11 April 2015. <http://mt-archive.info/MTS-2009-Desilets-2.pdf>.
- DÉSILETS, A., FARLEY, B. & STOJANOVIC, M. 2010. 'Using WeBiText to Search Multilingual Web Sites.' In Lavie, A., Dillinger, M. & Gerber, L. (eds.), *Proceedings of the 9<sup>th</sup> Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*. Denver, USA, 31 October–4 November 2010. Accessed on 19 February 2015. [https://www.researchgate.net/publication/50238287\\_Using\\_WeBiText\\_to\\_Search\\_Multilingual\\_Web\\_Sites](https://www.researchgate.net/publication/50238287_Using_WeBiText_to_Search_Multilingual_Web_Sites).



- DOWNING, A. & LOCKE, P. 1992. *A University Course in English Grammar*. Hertfordshire, UK: Prentice Hall.
- EBELING, J. 1999. 'Linking Dictionary and Corpus'. In Ebeling, S.O. & Hasselgård, H. (eds.), *Out of Corpora. Studies in honour of Stig Johansson*. Amsterdam & Atlanta: Rodopi, 31-45.
- EBELING, S. O. & EBELING, J. 2013. *Patterns in Contrast*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 13.33.
- EGAN, T. & GRAEDLER, A-L. 2015. 'Motion into and out of in English, French and Norwegian.' *Nordic Journal of English Studies* 14.1: 9-33. Accessed on 9 May 2016. <http://ojs.uib.no/ojs/index.php/njes/article/view/3087/2623>.
- ESPLÁ GOMIS, M., SÁNCHEZ MÁRTINEZ, F. & FORCADA, M. L. 2015. 'Using on-line Available Sources of Bilingual Information for Word-Level Machine Translation Quality Estimation.' In Forcada, M. L. (ed.), *Proceedings of the 18<sup>th</sup> Annual Conference of the European Association for Machine Translation (EAMT)*. Antalya, Turkey, 11-13 May 2015, 19-26. Accessed on 17 January 2016. <http://www.dlsi.ua.es/~mlf/docum/espla15p.pdf>.
- FRANZ, D. 2011. 'Aperçu des technologies langagières sur le marché.' Paper presented at Paper presented at the *Centre de recherche en technologies langagières (CRTL)*, Québec, Canada, 31 mai 2011. Accessed on 16 January 2016. <http://crtl.ca/display297>.
- FROELIGER, N. 2007. 'Pourquoi Avoir Peur de l'Informatisation en Traduction ?' *Tribune Internationale des Langues Vivantes* 43: 40-54. Accessed on 11 March 2015. <https://hal-univ-diderot.archives-ouvertes.fr/hal-01222590/>.
- GAATONE, D. 1976. 'Locutions prépositives et groupes prépositionnels.' *Linguistics* 14.167: 15-34. Accessed on 22 December 2015. [http://www.degruyter.com/dg/viewarticle/j\\$002fling.1976.14.issue-167\\$002fling.1976.14.167.15\\$002fling.1976.14.167.15.xml](http://www.degruyter.com/dg/viewarticle/j$002fling.1976.14.issue-167$002fling.1976.14.167.15$002fling.1976.14.167.15.xml).
- GARCÍA HERNÁNDEZ, A. 2014. 'Una investigación práctica para el avance desde el diccionario electrónico hacia el buscador 2.0.' MA thesis. Iceland: Universitatis Islandiae Sigillum. Accessed on 26 December 2015. [http://skemman.is/is/stream/get/1946/19235/44871/1/MFM\\_Almudena.pdf](http://skemman.is/is/stream/get/1946/19235/44871/1/MFM_Almudena.pdf).
- GELPÍ, C. 2004. 'Chapter 1: Reliability of Online Bilingual Dictionaries.' In Gottlieb, H. & Mogensen, J. E. (eds.), *Dictionary Vision, Research and Practice. Selected Papers from the 12<sup>th</sup> International Symposium of Lexicography, Copenhagen 2004*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 3-12.

- GEYKEN, A. 2015. 'Recent Developments in German Lexicography.' *Kernerman: Dictionary News* 23. 16-18. Accessed on 6 January 2016. <http://kdictionaries.com/kdn/kdn23.pdf>.
- GOULET, M. J. & DUPLESSIS, A. 2012. 'Focus Group on Computer Tools Used for Professional Writing and Preliminary Evaluation of LinguisTech'. In Piotrowski, M. (ed.), *Proceedings of the EACL 2012 Workshop on Computational Linguistics and Writing: Linguistic and Cognitive Aspects of Document Creation and Document Engineering*. Avignon, France, 23 April 2012, 39–47. Accessed 26 December 2015. [https://www.academia.edu/19046299/Focus\\_Group\\_on\\_Computer\\_Tools\\_Used\\_for\\_Professional\\_Writing\\_and\\_Preliminary\\_Evaluation\\_of\\_LinguisTech](https://www.academia.edu/19046299/Focus_Group_on_Computer_Tools_Used_for_Professional_Writing_and_Preliminary_Evaluation_of_LinguisTech).
- GRACIA, J. 2015. Multilingual dictionaries and the Web of Data. *Kernerman: Dictionary News* 23. 1-4. Accessed on 6 January 2016. <http://kdictionaries.com/kdn/kdn23.pdf>.
- GRANGER, S. 2010. 'Comparable and Translation Corpora in Cross-Linguistic Research. Design, Analysis and Applications.' *Journal of Shanghai Jiaotong University* 2: 14-21. Accessed on 3 March 2015. [sites-test.uclouvain.be/cecl/archives/Granger\\_Crosslinguistic\\_research.pdf](http://sites-test.uclouvain.be/cecl/archives/Granger_Crosslinguistic_research.pdf).
- GRANGER, S. 2012. 'Introduction: Electronic Lexicography – From Challenge to Opportunity.' In Granger, S. & Paquot, M. (eds). *Electronic Lexicography*. Oxford, UK: Oxford University Press, 1-11.
- GRANGER, S., & LEFER, M-A. 2012. 'Towards More and Better Phrasal Entries in Bilingual Dictionaries.' In Vatvedt Fjeld, R. & Torjusen, J. M. (eds.), *Proceedings of the 15<sup>th</sup> EURALEX International Congress*. Oslo, University of Oslo, 7-11 August 2012, 682-292.
- GRANGER, S. & LEFER, M-A. 2013. 'Enriching the Phraseological Coverage of High-frequency Adverbs in English-French Bilingual Dictionaries.' In Aijmer, K. & Altenberg, B. (eds.), *Advances in Corpus-Based Contrastive Linguistics*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 157-176.
- GRANGER, S. & LEFER, M-A. (in press). 'From General to Learners' Bilingual Dictionaries: Towards a more Effective Fulfilment of Advanced Learners' Phraseological Needs'.
- GRAUER, C. 2009. *Lesen, Verstehen und Übersetzen: Kollokationen als Handlungseinheiten der Übersetzungspraxis*. Trier, Germany: Wissenschaftlicher Verlag Trier.
- GRAUER, C. 2010. 'Using the Knowledge of the Crowd: Internet-Based Experiments on Collocation Translation'. In Bilić, V., Holderbaum, A., Kimmes, A., Kornelius, J., Stoll, C., & Trier, W. V. (eds.), *Translation in Translation 14*. Trier, Germany: Wissenschaftlicher Verlag Trier. Accessed on 1 December 2015. <http://www.t21n.com/homepage/articles/T21N-2010-14-Grauer.pdf>.
- HAGÈGE, C. 2010. *Adpositions*. Oxford Studies in Typology and Linguistic Theory. 208-214.

- HASSELGÅRD, H. 2010. 'Contrastive Analysis/Contrastive Linguistics.' In Malmkjær, K. (ed.), *The Routledge Linguistics Encyclopedia*. London, UK: Routledge, 98-101.
- HOFFMANN, S. 2005. *Grammaticalization and English Complex Prepositions: A Corpus-based Study*. Abingdon, UK: Routledge.
- HUET, S. & LANGLAIS, P. 2013. 'Translation of Idiomatic Expressions Across Different Languages: A Study of the Effectiveness of TransSearch.' In Neustein, A. & Markowitz, J. A. (eds.), *Where Humans Meet Machines: Innovative Solutions for Knotty Natural-Language Problems*. New York, USA: Springer, 185-209.
- HUNSTON, S. & FRANCIS, G. 2000. 'Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English.' *Computational Linguistics* 27.2: 318-320. Accessed on 27 April 2016. [www.mitpressjournals.org/doi/pdf/10.1162/coli.2000.27.2.318](http://www.mitpressjournals.org/doi/pdf/10.1162/coli.2000.27.2.318).
- HUTCHINS, W. J. 2001. 'Machine Translation over fifty years'. *Histoire Épistémologie Langage* 23.1: 7-31. Accessed on 16 March 2015. [http://www.persee.fr/web/revues/home/prescript/article/hel\\_075008069\\_2001\\_num\\_23\\_1\\_2815](http://www.persee.fr/web/revues/home/prescript/article/hel_075008069_2001_num_23_1_2815).
- IVIR, V. 1983. 'A Translation-Based Model of Contrastive Analysis.' *Jyväskylä Cross-Language Studies* 9: 171-178.
- IVIR, V. 1987. 'Functionalism in Contrastive Analysis and Translation Studies.' In Dirven, R. & Fried, V. (eds.), *Functionalism in Linguistics*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 471-481.
- JAMES, C. 1980. *Contrastive Analysis*. London, UK: Longman.
- JOHANSSON, S. 1975. *Papers in Contrastive Linguistics and Language Testing*. Lund, Sweden: CWL Gleerup.
- JOHANSSON, S. 1999. 'On the Role of Corpora in Cross-Linguistic Research'. In Johansson, S. & Ebeling, S. O. *Corpora and cross-linguistic research: Theory, method, and case studies*. Amsterdam & Atlanta: Rodopi, 3-24.
- JOHANSSON, S. 2007. *Seeing through Multilingual Corpora. On the use of corpora in contrastive studies*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- JOHANSSON, S. & HOFLAND, K. 1994. 'Towards an English-Norwegian Parallel Corpus.' In Fries, U., Tottie, G. & Schneifer, P. (eds.), *Creating and using English language corpora*. Amsterdam & Atlanta: Rodopi, 25-37.
- KAY, M. 1997. 'The Proper Place of Men and Machines in Language Translation.' *Machine Translation* 12.1-2, 3-23. Accessed on 12 December 2014. <http://dl.acm.org/citation.cfm?id=593157>.

- KHAWAJA, K. F. & BOKHARI, R. H. 2010. 'Exploring the Factors Associated With Quality of Website.' *Global journal of computer science and technology* 10.14: 37-45. Accessed on 16 April 2016. [http://globaljournals.org/GJCST\\_Volume10/7-Exploring-the-Factors-Associated-With-Quality-ofWebsite.pdf](http://globaljournals.org/GJCST_Volume10/7-Exploring-the-Factors-Associated-With-Quality-ofWebsite.pdf).
- KILGARRIFF A. & GREFENSTETTE G. 2003. Introduction to the Special Issue on Web as Corpus. *Computational Linguistics* 29.3: 333-347. Accessed on 22 April 2015. <http://www.mitpressjournals.org/doi/pdfplus/10.1162/089120103322711569>.
- KÜBLER, N. 2007. 'La traduction automatique : traduction machine?' *Tribune Internationale des Langues Vivantes*. 45. Accessed on 18 December 2014. [http://www.eila.univ-parisdiderot.fr/media/user/natalie\\_kuebler/recherche/traduction\\_machine\\_kubler\\_2007.pdf](http://www.eila.univ-parisdiderot.fr/media/user/natalie_kuebler/recherche/traduction_machine_kubler_2007.pdf).
- KÜBLER, N. 2013. 'Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ?' In Froeliger, N. (ed.), *Proceedings of the 2<sup>nd</sup> Colloquium Tralogy: Anticiper les technologies pour la traduction*, CNRS Auditorium, Paris, 17-18 January 2013. Accessed on 6 October 2014. <http://odel.irevues.inist.fr/tralogy/index.php?id=288>.
- KRZESZOWSKI, T. P. 1990. *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin & New York: Mouton de Gruyter.
- LAGOUDAKI, E. 2006. 'Translation Memories Survey 2006: Users' perceptions around TM use.' *Translating & the Computer* 28.1: 1-29. Accessed on 8 April 2015. <http://mt-archive.info/Aslib-2006-Lagoudaki.pdf>.
- LEW, R. 2010. 'Multimodal lexicography: The representation of meaning in electronic dictionaries.' *Lexikos* 20.1, 290-306. Accessed on 3 February 2015. <http://lexikos.journals.ac.za/pub/article/viewFile/144/85>.
- LEWIS, M. 1997. *Implementing the Lexical Approach. Putting Theory into Practice*. Melbourne, Australia: Thomson Heinle.
- LINDSTROMBERG, S. 2001. 'Preposition Entries in UK Monolingual Learners' Dictionaries : Problems and Possible Solutions.' *Applied Linguistics* 22.1: 79-103.
- LOIACONO, E.T., WATSON, R.T. & GOODHUE, D.L. 2002. 'WebQual<sup>TM</sup>: A Measure of Web Site Quality.' *Marketing theory and applications* 13.3: 432-438. Accessed on 16 April 2016. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.4410&rep=rep1&type=pdf>.
- MACKLOVITCH, E., & RUSSELL, G. 2000. 'What's Been Forgotten in Translation Memory.' In White, J. S. (ed.), *Proceedings of the 4<sup>th</sup> Biennial Conference of the Association for Machine Translation in the Americas (AMTA): Envisioning Machine Translation in*

- the Information Future*. Cuernavaca. Cuernavaca, Mexico, 10-14 October 2000, 137-145. Accessed on 4 June 2015. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.4422&rep=rep1&type=pdf>.
- MACKLOVITCH, E., LAPALME, G. & FABRIZIO, G. 2008. 'TransSearch: What are translators looking for?' In Dillinger, M. (ed.), *Proceedings of the 8<sup>th</sup> Biennial Conference of the Associations for Machine Translation in the Americas (AMTA)*. Waikiki, Hawaii, 21-25 October 2008, 412-419. Accessed on 29 November 2014. [www.rali.iro.umontreal.ca/rali/sites/default/files/publis/amta08-TransSearc-final.pfd](http://www.rali.iro.umontreal.ca/rali/sites/default/files/publis/amta08-TransSearc-final.pfd).
- MANN, M. 2014. 'Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen.' *Germanistische Linguistik* 4.223-224: 529-535. Accessed on 6 January 2016. <http://www.ajol.info/index.php/lex/article/view/126271/115794>.
- MANSFIELD, C. 2013. 'Studying French with Shared Web Documents on Google Drive.' Paper presented at Plymouth University. Accessed on 26 December 2015. <http://goo.gl/u2Gd8t>.
- MARQUE-PUCHEU, C. 2001. 'Les locutions prépositives: du spatial au non-spatial.' *Langue française* 129: 35-53. Accessed on 22 December 2015. [http://www.jstor.org/stable/41558982?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/41558982?seq=1#page_scan_tab_contents).
- MCBRIDE, C. 2012. 'Industry Insights: LinguisTech: Language technologies at your fingertips /L'industrie en marche : LinguisTech : les technologies langagières au bout des doigts.' *L'Actualité langagière/Language Update* 8.4: 5-7. Accessed on 16 January 2016. [http://publications.gc.ca/collections/collection\\_2012/tpsgc-pwgsc/S52-4-8-4.pdf](http://publications.gc.ca/collections/collection_2012/tpsgc-pwgsc/S52-4-8-4.pdf).
- MCENERY, T. & WILSON, A. 2001. *Corpus linguistics: An introduction*. Edinburgh, UK: Edinburgh University Press.
- MCENERY, T., XIAO, R., & TONO, Y. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. London, UK: Routledge.
- MILLER, T. L. 2012. 'Practical Implementation of Technology into Foreign Language Curricula.' In Sildus, T. (ed.), *Touch the World: Report of the Central States Conference on the Teaching of Foreign Languages*. Richmond, UK: Robert M. Terry. Accessed on 22 January 2016. <http://www.csctfl.org/documents/2012Report.pdf>.
- MÜLLER-SPITZER, C., KOPLINIG, A. & TÖPEL, A. 2011. 'What Makes a Good Dictionary? Empirical Insights from an Interdisciplinary Research Project.' In Kosem, I. (ed.), *Proceedings of Electronic lexicography (eLex) in the 21st century: New Applications for New Users*. Bled, Slovenia, 10-12 November 2011, 203-208. Accessed on 6 January 2016. <http://elex2011.trojina.si/Vsebine/proceedings/eLex2011-27.pdf>.

- NESSELHAUF, N. 2003. 'The Use of Collocations by Advanced Learners of English and some Implications for Teaching.' *Applied Linguistics* 24.2: 223-242.
- PERON, C. 2013. 'Évaluation d'une plate-forme de localisation : le cas de Reverso Localize.' MA Thesis. Genève: University of Genève. Accessed on 14 January 2016. <http://archive-ouverte.unige.ch/unige:33070>.
- PERON, C. & MORADO VÁZQUEZ, L. 2015. 'In-Context Online Localisation Tools: the Case Study of Reverso Localize.' *Localisation Focus-The International Journal of Localisation* 13.1: 4-15. Accessed on 14 January 2016. <http://irc.csisdmsz.ul.ie/sites/default/files/publications/LFV13-1-1.pdf>.
- POTTIER, B. 1987. 'La Subduction, la Métaphore et les Lexies.' *Cahiers de lexicologie*: 209-218.
- PRECUP-STIEGELBAUER, L. R. 2013. 'Automatic Translations Versus Human Translations in Nowadays World.' *Procedia – Social and Behavioral Sciences* 70, 1768-1777. Accessed on 27 May 2015. <http://www.sciencedirect.com/science/article/pii/S187704281300253X?np=y>.
- QUIRK, R., GREENBAUM, S., LEECH, G. & SVARTVIK, J. 1985. *A Comprehensive Grammar of the English Language*. London, UK: Longman.
- ROY, I. & SVENONIUS, P. 2009. *Complex Prepositions*. Colloque international de Caen: Autour de la Préposition, Sep 2009, Caen, France. 105-116. Accessed on 22 December 2015. [http://www.academia.edu/2679619/Complex\\_prepositions](http://www.academia.edu/2679619/Complex_prepositions).
- RUNDELL, M. 2012. 'The Road to Automated Lexicography: An Editor's Viewpoint.' In Granger, S. & Paquot, M. (eds.), *Electronic Lexicography*. Oxford, UK: Oxford University Press, 15-30.
- SALKIE, R. 2008. 'How can lexicographers use a translation corpus?' In Xiao, R, He, L. & Ye, M. (eds.), *Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies*. Zhejiang University, Hangzhou, China, 25-27 September 2008.
- DE SCHRYVER, G. M. 2003. 'Lexicographers' Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16.2, 143-199. Accessed on 18 March 2015. <http://ijl.oxfordjournals.org/content/16/2/143.full.pdf+html>.
- SIMARD, M. 2003. 'Translation spotting for translation memories.' In Mihalcea, R. & Pedersen, T. (eds.), *Proceedings of the Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and beyond*. Edmonton, Canada, 27 May-1 June 2003, 65-72. Accessed on 14 July 2015. <http://transsearch.iro.umontreal.ca/rali/sites/default/files/publis/SimardTranslationSpotting.pdf>.



- SIMARD, T. 2013. ‘Analyse comparative de la couverture et de l’acceptabilité des solutions d’un dictionnaire bilingue spécialisé, d’une banque de données terminologiques et d’un concordancier en ligne : application au domaine de la traduction médicale.’ Ph.D. Thesis. Ottawa: University of Ottawa. Accessed on 6 March 2015. <http://www.ruor.uottawa.ca/handle/10393/24929>.
- SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press.
- SOMERS, H. 1999. ‘Example-based Machine Translation.’ *Machine Translation* 14.2, 113-157. <http://link.springer.com/article/10.1023/A:1008109312730>.
- TARAVELLA, A.-M. 2011. ‘Rapport sommaire et préliminaire sur les résultats de l’enquête menée auprès des utilisateurs de technologies langagières en avril-mai 2011.’ Paper presented at the *Centre de recherche en technologies langagières (CRTL)*, Québec, Canada, 3 October 2011. Accessed on 6 March, 2015. <http://linguistech.ca/display958>.
- VINAY, J.-P. & DARBELNET, J. 1977. *Stylistique comparée du français et de l’anglais: méthode de traduction*. Paris, France : Didier.
- VOLANSCHI, A. 2007. ‘Outils informatiques et ressources électroniques pour les traducteurs.’ *La Tribune Internationale des Langues Vivantes* 43 : 24-39.
- VOLK, M., GRÄEN, J. & CALLEGARO, E. 2014. ‘Innovations in parallel corpus search tools.’ In Calzolari, N. (ed.), *Proceedings of the 9<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*. Reykjavik, Iceland, 26-31 May 2014. Accessed on 14 April 2015. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/504\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/504_Paper.pdf).
- ZANETTIN, F. 1998. ‘Bilingual Comparable Corpora and the Training of Translators.’ *Meta: Journal des traducteurs /Translators' Journal* 43.4: 616-630.

### 3. NON-ACADEMIC LITERATURE

- Altissia. 2014. ‘Quel est le meilleur traducteur en ligne gratuit ? – Test 2014.’ In *Altissia Blog*. Blog Article. Accessed on 28 October 2014. <http://www.altissia.com/blog/fr/quel-est-le-meilleur-traducteur-en-ligne-gratuit/>.
- Anonym. 2012. ‘Quand même, ils sont bons.’ In *Les Piles Intermédiaires*. Blog Article. Accessed on 8 January 2016. <http://lespilesintermediaires.blogspot.no/2012/11/quand-meme-ils-sont-bons.html>.
- Commentçamarche. 2010. ‘Linguee.com : nouvel outil de traduction contextuelle en ligne.’ In *News*. Blog Article. Accessed on 28 October 2014. <http://www.commentcamarche.net/news/5852780-linguee-com-nouvel-outil-de-traduction-contextuelle-en-ligne>.

- GALLIMORE, L. 2011. 'The adventures of WebiText and Linguee.' In *Linguistech Blog*. Blog Article. Accessed on 3 March 2015. <http://linguistech.ca/blogpost153> .
- GREEN, S. 2011. 'The largest bilingual web dictionary now greatly enhanced.' Press Release. Accessed on 1 December 2014. [http://www.linguee.com/press/EN/Linguee\\_Press\\_Release\\_06092011.pdf](http://www.linguee.com/press/EN/Linguee_Press_Release_06092011.pdf) .
- GREEN, S. 2013. 'Linguee.com Reaches 2 Billion Searches And Adds New Languages.' Press Release. Accessed on 1 December 2014. [http://www.linguee.com/press/EN/2013-12-04\\_LingueePressRelease.pdf](http://www.linguee.com/press/EN/2013-12-04_LingueePressRelease.pdf) .
- MCDUFF, S. 2012. 'Qu'est-ce que TradooIT?' Official Tutorial. Accessed on 4 December 2014. <http://wiki.tradooit.com/home/quest-ce-tradooit>.
- OUELLET, T. 2012. 'Tutoriel sur TradooIT, niveau I.' Non-official Tutorial. Accessed on 6 January 2016. [http://linguistech.ca/TradooIT\\_F\\_TUTCERTT\\_I](http://linguistech.ca/TradooIT_F_TUTCERTT_I).
- PINTO, M. 2004. Electronic Content Management Skills. Website. Accessed on 13 December 2014. <http://www.mariapinto.es/e-coms>.
- PORTAL, H. 2011. 'Le plus grand dictionnaire en ligne du monde fait peau neuve.' Press Release. Accessed on 28 October 2014. [http://www.linguee.fr/press/FR/Linguee\\_Press\\_Release\\_06092011.pdf](http://www.linguee.fr/press/FR/Linguee_Press_Release_06092011.pdf).
- RACICOT, A. 2014. 'Discriminer.' In *Au coeur du Français*. Blog Article. Accessed on 10 January 2016. <http://andrericot.ca/discriminer/>.
- SCHOPPMANN, C. 2014. 'Online Dictionary Linguee introduces 218 new language pairs.' Press Release. Accessed on 1 December 2014. [http://www.linguee.com/press/EN/2014-02-25\\_LingueePressRelease\\_218NewLanguagePairs.pdf](http://www.linguee.com/press/EN/2014-02-25_LingueePressRelease_218NewLanguagePairs.pdf).
- XHARK, M. 2010. 'Linguee: un traducteur contextuel et communautaire.' Blog Article. Accessed on 28 October 2014. <http://blogmotion.fr/internet/traducteur-contextuel-6126>.
- VAN BOLDEREN, T. 2012. 'Linguee Tutorial, Level I.' Non-official Tutorial. Accessed on 12 May 2015. [http://linguistech.ca/Linguee\\_E\\_TUTCERTT\\_I\\_PartI](http://linguistech.ca/Linguee_E_TUTCERTT_I_PartI).



# APPENDICES



Appendix 1. Searches of *en matière de* in PLECI\_news corpus using *ParaConc*.

	Linguee	TradooIT	WebiText	Reverso <sup>112</sup>
<b>Non-automatic</b>	✓	✓	✓	✓
<b>Web-based</b>	✓	✓	✓	✓
<b>Data</b>				
<b>Available</b>	✓	✓	✓	✓
<b>online</b>				
<b>Free of charge</b>	✓	✓	✓	✓
<b>Contextual</b>	✓	✓	✓	✓
<b>examples</b>				
<b>Grouped</b>	✗	✗	✓	✓
<b>translations</b>				
<b>Word-</b>	Problematic	✓	✗	✓
<b>alignment</b>				
<b>Bilingual</b>	✓	✓	✗	✓ (in another section)
<b>dictionary</b>				
<b>Other sections</b>	Wikipedia	Terminological bank, translation memory, pre-translation system, extension module for Microsoft Word, system data sharing	✗	Monolingual dictionary, Speller, Automatic translator, Grammar and conjugator help, Localize
<b>Beginning</b>	2010	2011	2009	No information
<b>visitors/day</b>	38,000 user	419 users	200 users	5,000,000 users

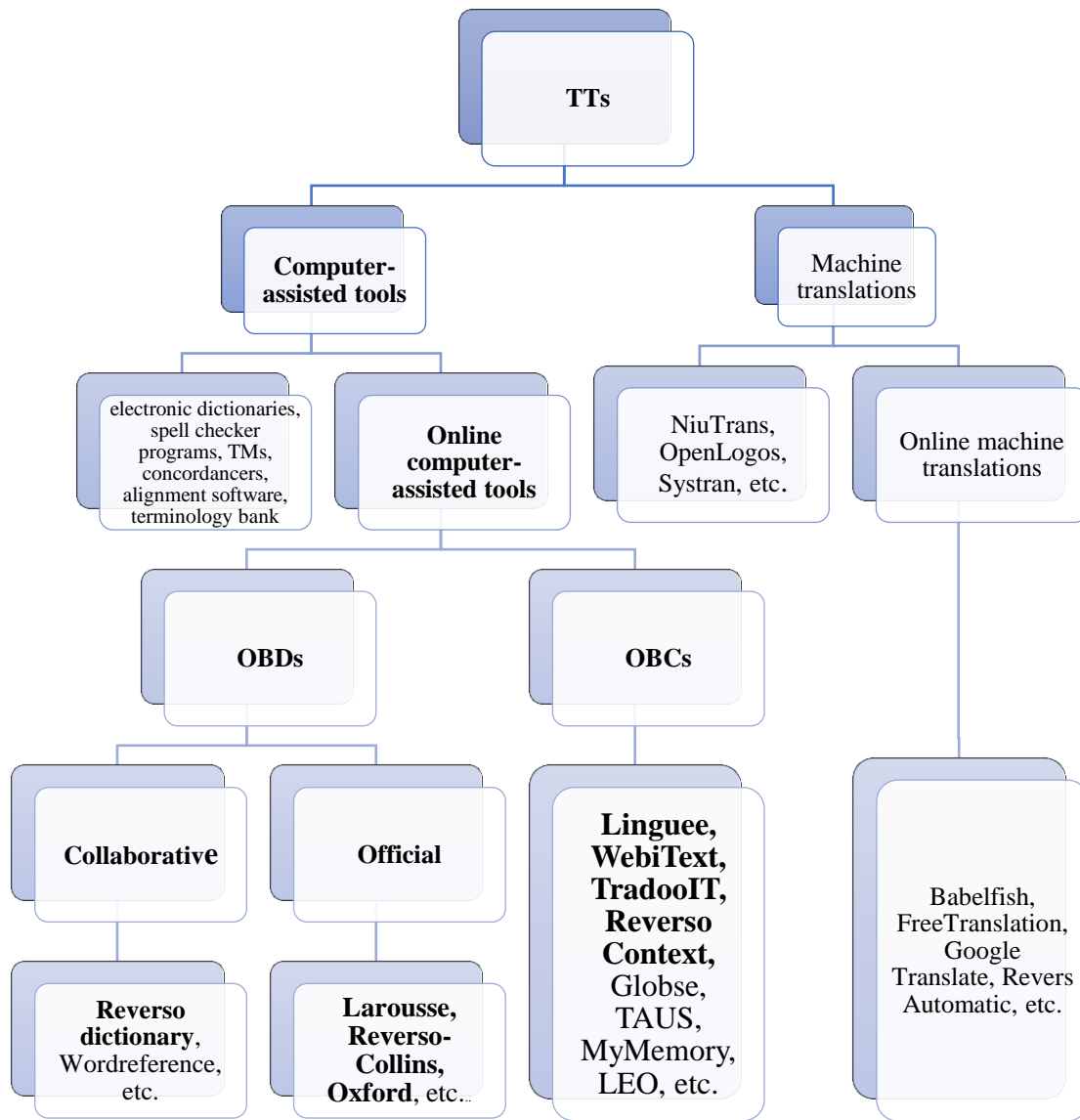
<sup>112</sup> In this table, we consider the entire *Reverso* website, even though we focus on *ReversoContext* in this thesis.

<b>Number of queries/days</b>	1,5 million queries (in 2011)	No information	3,500 queries (in 2010)	No information
<b>Creation of a free account</b>	✓	✓	✗	✓
<b>Interface language</b>	For every language available	French & English	English, French, Portuguese & Spanish	For 10 languages
<b>Language combination</b>	25	French, English and Spanish	19	8 (more for the other tools)
<b>Size</b>	1,000,000,000	464,450,884 words	285,000,000 pages	No information
<b>Number of corpora</b>	A dozen	62	89	19
<b>Corpora selection</b>	✗	✓	✓	✗
<b>Frequency information</b>	✗	✓	✓ (not for all corpora)	✓
<b>SL and TL distinction</b>	✗	✗	✗	✗
<b>Sources revision</b>	Not systematic	manual	manual	Not manual
<b>Incremental search bar</b>	✓	✗	✗	✓
<b>Search type</b>	Exact matched when using quotation mark, otherwise it brings up inflected forms and sentences with partial matches. <sup>113</sup>	Idem <i>Linguee</i>	Only searches simple terms and shows exact matches, no partial or inflected matches	Idem <i>Linguee</i>
<b>Results displaying</b>	30 at once	10 per page	A dozen at a time	20 per page
<b>Target word highlighted</b>	✓ (not accurate)	✓	✗	✓
<b>Bitexts</b>	✗	✓	✓	✓
<b>URL available</b>	✓	✓	✓	✓
<b>Speed</b>	Real-time display	Real-time display	Slow	Real-time display
<b>Customization</b>	✗	✓	✗	✓
<b>Hybridization</b>	✓	✓	✗	✓
<b>User input</b>	✗	✓ <sup>114</sup>	✗	✓
<b>Interface</b>	Clear and responsive	Clear	Clear	Clear and responsive
<b>Updates</b>	Regular	Regular	Latest in 2010	Regular
<b>Unlimited examples</b>	Up to 30	✓	✓	✓
<b>Mobile version</b>	✓	✗	✗	✓
<b>Language detection</b>	If on the correct pair	If on the correct pair	If on the correct pair	If on the correct pair
<b>Social media</b>	Facebook and Twitter	Facebook and blog	Facebook	Facebook, Twitter and blog
<b>Advertising</b>	✓	✗	✗	✓

**Appendix 2.** Summary of the tools' features.

<sup>113</sup> Gallimore (2011)

<sup>114</sup> Very limited



Appendix 3. Translations tools.

	Linguee	TradooIT	Reverso Context
Minus sign	Exclude words		
Plus sign	Include words	All the lemmatized forms	
Question mark		See the facultative words	
Asterisk		Indicate the obligatory presence of a word	
Quotation marks	to “find only sentences where the query words occur in the exact form and		

Punctuation	<i>order</i> <sup>115</sup>	Exclude some	results <sup>116</sup>
Hourglass			Include words

Appendix 4. Metacharacters of the OBCs's search bar.

Types of OBDs	Free OBDs		Charged OBDs
OBDs	Larousse	Reverso	Oxford
<b>French CPs</b>			
<b>En matière de</b>	en matière de locution prépositionnelle <b><u>as regards</u></b>  [en matière de cuisine <b><u>as far as cooking is concerned, as regards cooking</u></b> ]	en matière de <b><u>as regards</u></b>	[en matière de cuisine/d'art/d'emploi <b><u>as far as cooking/art/employment is concerned</u></b> ]
<b>Au sein de</b>	au sein de (soutenu) locution prépositionnelle <b><u>within</u></b> [au sein de la famille <b><u>in the bosom</u></b> OU <b><u>midst of the family</u></b> ]	au sein de [+équipe, institution] <b><u>within</u></b>	au sein de <b><u>within</u></b>
<b>Par rapport à</b>	par rapport à locution prépositionnelle [en ce qui concerne] <b><u>regarding</u></b> [comparativement à] <b><u>compared with, in comparison to</u></b>  <i>on constate un retrait de l'euro par rapport aux autres monnaies européennes the euro has dropped sharply against other European currencies</i>	par rapport à (=comparé à) <b><u>in relation to</u></b> (=à propos de) <b><u>with regard to</u></b>	also: par rapport à locution prépositive  1 (comparé à) <b><u>compared with, in comparison with</u></b>  <i>le chômage a augmenté par rapport à l'an dernier unemployment increased compared with last year il est généreux/petit par rapport à son frère he's generous/small compared with his brother</i> <b><u>[par rapport au dollar against the dollar]</u></b>  2 (en fonction de)

<sup>115</sup> <http://www.linguee.com/english-french/page/help.php?help=extendedsearch>. Accessed on 10 January 2016. For our analysis, quotation marks were not necessary, although the external examples might have been ordered differently if we had included them.

<sup>116</sup> *prise*, (with the comma) will exclude results where the words is employed in an MWU such as *prise en compte*.

			<p>[le nombre de voitures par rapport au nombre d'habitants the number of cars <u>in relation to</u> the number of inhabitants un angle de 40° par rapport à la verticale an angle of 40° <u>to</u> the vertical un changement par rapport à la position habituelle du parti a change <u>from</u> the usual party line]</p> <p>3 (vis-à-vis de) <b><u>with regard to, toward(s)</u></b> [notre position par rapport à ce problème our position <u>with regard to</u> this problem l'attitude de la population par rapport à l'immigration people's attitude toward(s) immigration]</p>
<b>Dans le cadre de</b>	<p>dans le cadre de locution prépositionnelle <b><u>within the framework</u></b> OU <b><u>scope of</u></b> [dans le cadre de mes fonctions <u>as part of</u> my job]</p> <p><i>cela n'entre pas dans le cadre de mes fonctions it falls outside <u>the scope of</u> my responsibilities</i></p>	<p>dans le cadre de prep. <b><u>within the framework of</u></b></p>	<p>2 (faire partie de) [s'inscrire dans le cadre de de to be <u>in line with</u>]</p> <p>also: dans le cadre de locution prépositive</p> <p>1 (à l'occasion de) <b><u>on the occasion of</u></b> (voyage, fête, rencontre) [dans le cadre de cette journée particulière <u>on this special occasion</u>]</p> <p>2 (dans le contexte de) <b><u>within the framework of</u></b> (lutte, politique, négociations, organisation) <b><u>as part of</u></b> (enquête, campagne, plan) [les manifestations]</p>

			<p>organisées dans le cadre du festival events organized <u>as part of</u> the festival]</p> <p><i>les négociations doivent avoir lieu dans le cadre de la UE</i> <i>negotiations must take <u>place within the framework of the EU</u></i></p> <p>[recevoir une formation dans le cadre d'une entreprise/d'une association to undergo training <u>within</u> a company/an association]</p>
<b>En faveur de</b>	<p>en faveur de locution prépositionnelle [à cause de] <b><u>on account of</u></b> [au profit de] <b><u>to the benefit of, in favour of</u></b> [en ma/votre faveur <b><u>in</u></b> my/your <b><u>favour</u></b>] [favorable à] <b><u>in favour of</u></b></p>	<p>en faveur de <b><u>in favour of</u></b> (Grande-Bretagne), <b><u>in favor of</u></b></p>	<p>also: en faveur de locution prépositive</p> <p>1 (à l'avantage de) <i>le jugement a été rendu en sa faveur</i> <i>the court decided <u>in his/her favour</u> (anglais britannique)</i> <i>la caissière s'est trompée en ma faveur</i> <i>the cashier gave me too much change</i> [les votes en faveur du candidat de l'opposition the votes <u>for</u> the opposition candidate]</p> <p>2 (pour aider) [des mesures en faveur des handicapés measures <u>to help</u> the disabled les mesures en faveur de l'emploi measures <u>to promote</u> employment intervenir en faveur de quelqu'un to intervene <u>on</u> somebody's behalf]</p> <p>3 (partisan de)</p>

			<p>[être en faveur de quelque chose to be <u>in favour</u> (anglais britannique) <u>of</u> something être en faveur de quelqu'un to be <u>for</u> somebody]</p> <p>4 (en considération de) <b>on account of</b> <i>ses torts ont été oubliés en faveur de sa compétence</i> <i>his/her failings were overlooked <u>on account of</u> his/her efficiency</i></p>
<b>À la fin de</b>	à la fin de locution prépositionnelle <b>at the end</b> OU <b>close of</b>	[à la fin de la guerre <b>at the end of the war</b> ]	à la fin de <b>at the end of</b>
<b>Dans le domaine de</b>	<i>dans le domaine de la prévention, il y a encore beaucoup à faire</i> <u>as far as preventive action is concerned, there's still a lot to do</u>	<p>dans le domaine de prep. <b>in the field of</b></p> <p>dans le domaine de qch <b>in the field of</b> sth [dans le domaine de l'environnement <u>in the environmental field</u>]</p>	<p><i>j'ai eu de la chance dans le domaine or sur le plan professionnel</i> <i>I've been lucky <u>in</u> my professional life</i></p> <p>[dans le domaine financier/philosophique <b>in the field of</b> finance/philosophy]</p>
<b>En raison de</b>	<p>en raison de locution prépositionnelle [à cause de] <b>on account of, because of</b></p> <p><i>le vol est annulé en raison du mauvais temps</i> <i>the flight has been cancelled <u>because of</u> bad weather</i></p> <p>[en proportion de] <b>according to</b></p>	<p>en raison de prep. <b>because of</b></p> <p>en raison de (=à cause de) <b>because of</b> [en raison d'une grève <b>because of</b> a strike en raison du mauvais temps <b>because of</b> the bad weather, <b>due to</b> the bad weather]</p>	No results
<b>Aux côtés de</b>	<p>[être aux côtés de quelqu'un to be <u>by somebody's side</u>]</p> <p>à côté de locution prépositionnelle [pas loin] <b>next to</b></p>	[être aux côtés de qn to be <u>by sb's side</u> ]	[être du côté de quelqu'un/quelque chose to be <u>on somebody's side/on the side of</u> something du côté britannique/français <u>on the</u> British/French <u>side</u> ]


			<p>also: à côté de locution prépositive</p> <p>1 (à proximité de) <b>next to</b></p> <p>also: aux côtés de locution prépositive</p> <p>(près de) [[littéral] [figuré] aux côtés de quelqu'un (être, rester) <u>beside</u> somebody, <u>at</u> <u>somebody's side</u> aux côtés de quelqu'un/quelque chose (se retrouver) <u>beside</u> ou <u>alongside</u> somebody/something] (siéger, s'engager, travailler) <b>alongside</b> somebody/something</p>
<b>Au service de</b>	<p>[entrer au service de quelqu'un to enter somebody's service]</p> <p><i>il a mis son savoir-faire au service de la société he put his expertise <u>at</u> <u>the disposal of the</u> company</i></p>	No results	<p>[être au service de son pays <u>to serve</u> one's country travailler au service de la paix to work <u>for</u> peace]</p>

Appendix 5. Results for the French CPs in the OBDs.

The image shows a digital dictionary interface for the word 'rapport'. At the top, there are language selection buttons for 'Français' and 'Anglais', along with a refresh icon. Below this, there are several small circular icons representing different dictionary sources. The main content area displays the word 'rapport' with its phonetic transcription [rapɔʁ] and the note 'nom masculin'. A numbered list of definitions is provided, including 'compte rendu - généralement' and 'report'. A 'MILITAIRE' section lists 'briefing' and its usage. An 'EXPR' section lists 'rapport annuel', 'annual report', 'rapport d'expert', 'audit report', and 'rapport financier', 'annual (financial) report OU statement'. A search bar at the bottom shows the query 'par rapport à' highlighted in a red box.

Appendix 6. Par rapport à in Larousse.




- 2 (faire partie de)  
*s'inscrire dans le cadre de*   
to be in line with

**Appendix 7.** *Dans le cadre de* in *Oxford* with the queried term *dans le cadre de*.

also: *dans le cadre de*  
locution prépositive

- 1 (à l'occasion de)  
**on the occasion of** (voyage, fête, rencontre)

*dans le cadre de cette journée particulière* 

on this special occasion

---

IN OTHER DICTIONARIES


---

- 2 (dans le contexte de)  
**within the framework of** (lutte, politique, négociations, organisation)


**as part of** (enquête, campagne, plan)

*les manifestations organisées dans le cadre du festival* 

events organized as part of the festival

*les négociations doivent avoir lieu dans le cadre de la UE* 

negotiations must take place within the framework of the EU

*recevoir une formation dans le cadre d'une entreprise/d'une association* 

to undergo training within a company/an association

**Appendix 8.** *Dans le cadre de* in *Oxford* with the queried term *cadre*.