"Translation equivalents in translation corpora and bilingual dictionaries: the case of approximators in English and French"

Goossens, Diane

Document type: Communication à un colloque (Conference Paper)

Référence bibliographique


Available at:
http://hdl.handle.net/2078.1/111832
[Downloaded 2018/12/01 at 03:31:52]
Translation equivalents in translation corpora and bilingual dictionaries: the case of approximators in English and French

Diane Goossens

Keywords: translation equivalents, bilingual dictionaries, translation corpora, quantity approximation.

Abstract

This paper reports on an investigation of the translations of ‘approximator + number’ combinations (e.g. about 200) in English and French using a translation bidirectional parallel corpus of news reporting and examining the entries for the approximators selected in three bilingual dictionaries. This study analyses the major tendencies that are found in the corpus when translating six commonly used approximators occurring around numbers into French and into English: the approximators plus de, près de, environ, dépasser, quelque and approximators formed using a number and the suffix –aine are analysed for the French to English translation direction, and the approximators more than, about, up to, around, over and some are investigated for the English to French translation direction. The entries for these approximators are also scrutinized in three bilingual dictionaries: the Harrap’s Unabridged, Le Grand Robert et Collins électronique and the Grand Dictionnaire Hachette Oxford. The paper focuses more specifically on the types of examples given for each approximator, examining whether the quantity approximation meaning of these items is well represented. Several bidirectional translation equivalence issues are also discussed as some translation equivalents mentioned in one direction are not listed in the other direction in the same dictionary. Based on corpus evidence, the study suggests several ways of improving the treatment of the items under study in bilingual dictionaries. These include the introduction of labels that would inform the user about the context in which the approximator is preferred, for instance in informal contexts or with certain types of quantities. The variety of items identified in the corpus may also help lexicographers list translation equivalents from a wider variety of grammatical categories.

1. Introduction

Numbers are an integral part of our lives. Everyday texts such as newspaper articles are full of amounts of money or reports about quantifiable events. These quantities may be expressed in a precise way (e.g. 5,135 people) or in a more imprecise way, that is approximating quantities (e.g. about 10,000 jobs). A number of linguistic devices can be used to approximate quantities, including the use of approximators like around with numbers. The purpose of this study is (1) to analyse the various translations used in a translation corpus of news reporting when translating six commonly used approximators occurring around numbers in French and in English, and (2) based on corpus evidence, to evaluate the treatment of approximators in bilingual dictionaries and suggest ways of improving it.

2. Quantity approximation in English and French

Approximators used around numbers of the type ‘around 1000’ have so far received little attention. Most studies of quantity approximation investigate a limited set of items belonging to the same grammatical category, for example the adverbs about, around and approximately as in Channell (1994) and Ruzaite (2007). A recent onomasiological corpus-driven study of quantity approximation in English and French (De Cock and Goossens, in preparation) has shown that a wider variety of linguistic devices are used to express quantity approximation in business news
reporting. The study concentrates on combinations of approximators and numbers denoting quantities that were automatically retrieved using part-of-speech-tagged corpora. Numbers that did not denote quantities (e.g. 5 pm) were manually discarded. Concordances for the remaining numbers were scrutinized for potential approximators, defined as ‘any linguistic item occurring around a number denoting a quantity and signaling an approximation of this quantity’ (e.g. more than £3m).

This paper focuses on the translation equivalents of a selection of approximators that were identified in De Cock and Goossens (in preparation). The approximators were selected among the most frequently occurring approximators in the English and French business news corpora analysed in De Cock and Goossens (in preparation) and were chosen to be representative of the preferred grammatical categories identified in the study, for example adverbs, ‘MORE/LESS than’ approximators and prepositions. The French approximators selected for the present investigation include plus de, près de, environ, dépasser, quelque and approximators formed using a number and the suffix –aine. The following English approximators are examined: more than, about, up to, around, over and some.

3. English and French approximators in a translation corpus

3.1. Data and method.

Two journalese subcorpora of the PLECI corpus (Poitiers-Louvain Échange de Corpus Informatisés), a bidirectional translation parallel corpus of fiction and journalese, are used to analyse how a selection of approximators are translated into English and French. The first subcorpus (French to English translation direction) is made up of 277,556 words of French newspaper articles (from Le Monde Diplomatique and Le Monde) translated into English. The second subcorpus (English to French translation direction) contains 255,142 words of newspaper articles (from various sources including Le Monde Diplomatique, Time Magazine and The Economist) translated into French.

Six approximators were selected for each translation direction. The occurrences of the approximators selected and their translation equivalents were retrieved from the PLECI corpus using ParaConc (Barlow 2002). Only combinations of numbers denoting quantities (e.g. 500 jobs and not 5 pm) and approximators (e.g. nearly 80%, plus de 30,000 enfants) are examined.

3.2. Results of the corpus analysis

3.2.1. Main tendencies. Three main tendencies are observed when scrutinizing the concordances in PLECI: (1) the combination ‘approximator + number’ is translated by an equivalent combination using an approximator and a number (e.g. environ 12% translated by about 12%); (2) the combination ‘approximator + number’ is translated without using an approximator (e.g. over 120m is translated by 120 millions); (3) the quantity expressed by the combination ‘approximator + number’ in the original text does not appear in the translated text, and, as a result, the approximator is not translated (e.g. (...) une fraction modeste (environ 10 %) du budget américain is translated by a tiny fraction of the US equivalent).

Table 1 shows the number of tokens for each of these categories. Cases that do not fit into any of these categories are labeled as ‘other’ (4). An example of item included in this category is the translation of the approximated
quantity using a phrase denoting an imprecise quantity without using a number: *quelques centaines de milliers de francs* translated by *small fortune*.

Table 1. Translation of the approximators studied in the PLECI corpus (journalese).

<table>
<thead>
<tr>
<th>Approximator</th>
<th>(1) Translation = approximator + number</th>
<th>(2) Translation = number</th>
<th>(3) No translation</th>
<th>(4) Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source language: French</td>
<td>% total n° of tokens</td>
<td>% total n° of tokens</td>
<td>n° of tokens</td>
<td>n° of tokens</td>
</tr>
<tr>
<td>plus de</td>
<td>82.47% (127)</td>
<td>16.88% (26)</td>
<td>0.65% (1)</td>
<td>0</td>
</tr>
<tr>
<td>près de</td>
<td>78.79% (26)</td>
<td>18.18% (6)</td>
<td>3.03% (1)</td>
<td>0</td>
</tr>
<tr>
<td>environ</td>
<td>68.75% (22)</td>
<td>15.63% (5)</td>
<td>12.5% (4)</td>
<td>3.13% (1)</td>
</tr>
<tr>
<td>suffix –aine (e.g. douzaine)</td>
<td>78.95% (60)</td>
<td>11.84% (9)</td>
<td>6.58% (5)</td>
<td>2.63% (2)</td>
</tr>
<tr>
<td>quelque</td>
<td>57.14% (16)</td>
<td>42.82% (12)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dépasser</td>
<td>88.89% (16)</td>
<td>11.11% (2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Source language: English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>more than</td>
<td>83.75% (67)</td>
<td>12.5% (10)</td>
<td>2.5% (2)</td>
<td>1.25% (1)</td>
</tr>
<tr>
<td>about</td>
<td>62.22% (28)</td>
<td>35.56% (17)</td>
<td>2.22% (1)</td>
<td>0</td>
</tr>
<tr>
<td>up to</td>
<td>58.82% (10)</td>
<td>41.18% (7)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>around</td>
<td>86.96% (20)</td>
<td>13.04% (3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>over</td>
<td>61.54% (16)</td>
<td>38.46% (10)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>some</td>
<td>57.14% (24)</td>
<td>33.33% (14)</td>
<td>4.76% (2)</td>
<td>4.76% (2)</td>
</tr>
</tbody>
</table>

Table 1 reveals that the combinations of approximators and numbers are translated using a similar combination (i.e. approximator + number, (1)) in at least 50% of the cases in both translation directions, with percentages ranging from 57.24% (*quelque* and *some*) to 88.89% (*dépasser*). Translations using a number only without any approximator (2) account for 11.11% to 42.82% of the cases. Cases where the combination ‘approximator + number’ is not translated (3) and ‘other’ cases (4) only account for a fairly low percentage (from 0.65% to 12.5%) of the translations in PLECI. The rest of this ‘results’ section will therefore focus on cases (1) and (2).

The combination ‘approximator + number’ is translated using a similar combination slightly more often in the French to English direction (in 75.83% of the cases on average) than in the English to French direction (68.03% of the cases on average). It is noticeable that the approximator is more often left out in the translations in the English to French direction (29.42% of the translations on average vs. 19.41% for the French to English direction). As a result, the approximated quantity is translated using a number only, which, when it is a round number, can be interpreted as a quantity approximation too as ‘a round number on its own may be understood as an approximation’ (Channell 1990: 100). For instance, *Rolls-Royce has slashed its workforce (…) to just over 3,000* is translated by *Rolls-Royce a réduit ses effectifs à 3 000 personnes*. However, the absence of an explicit approximator in the translation may arguably introduce ambiguity because a round number such as 3,000 can potentially refer both to a precise number (i.e. exactly 3,000) or an approximation (e.g. 3,013).

3.2.2. Translation equivalents of approximators in the PLECI corpus. When ‘approximator + number’ combinations are translated using a similar combination, various items are used as translation equivalents for the approximators under study. Table 2 lists the different translation equivalents used in the PLECI corpus.
Table 2. Translation equivalents of the approximators studied in the PLECI corpus (journalese subcorpora). French and English.

<table>
<thead>
<tr>
<th>Approximator (total number of tokens ‘approximator + number’ combinations)</th>
<th>Translation equivalents used in PLECI (number of tokens)</th>
</tr>
</thead>
</table>
| French source language:  
plus de (127)  
près de (26)  
environ (22)  
suffix –aine (60) (e.g. douzaine)  
dépasser (16)  
English source language:  
more than (67)  
about (28)  
up to (10)  
around (20)  
over (16)  
some (24) | more than (88), over (31)  
to exceed (2), about (1), almost (1), and over (1), less than (1), some (1), upwards of (1)  
almost (12), nearly (12), about (1), upwards of (1)  
about (9), some (7), around (4), in the order to (1), more than (1)  
dozen (4), or so (3), several (3), around (2), some (2), more than (2), almost (1), a score of (1), less than (1)  
about (5), some (5), around (2), or so (2), up to (1), several (1)  
in excess of (6), more than (6), over (2), above (1), to exceed (1)  
plus de (55), supérieur à (3)  
environ (17), suffix –aine (4), plus de (2), quelque (2), à peu près (1), près de (1), avoisiner (1)  
jusqu’à (8), plus (1), ou moins (1)  
environ (10), près de (3), autour de (2), avoisiner (2), jusqu’à près de (1), quelque (1), bonne (1)  
plus de (13), dépasser (2), environ (1)  
quelque (10), environ (9), suffix –aine (1), suffix –ier (1), plus de (1), près de (1), un peu moins de (1) |

The analysis of the translation equivalents used in PLECI shows that a variety of translation equivalents is used for each approximator, ranging from 3 (up to and over) to 11 (English translations of approximators containing the suffix –aine) different translation equivalents. The most frequently used translations are often similar to their equivalent in the original text in terms of word class (e.g. the adverb around is mostly translated by the adverb environ) and meaning (e.g. over mostly translated by plus de, both meaning ‘more than’).

3.2.3. Types of quantities approximated and preferred translation equivalents. A closer inspection of the concordances in PLECI reveals that various types of quantities are approximated in the corpus. The following types of quantities were identified in quantity approximations: number of people (e.g. more than 200 prisoners), amounts of money (e.g. environ 90 000 francs), time durations (e.g. over a thousand years), measures (e.g. près de 250,000 km²), percentages (e.g. quelque 20%) and ‘other’ types of quantities (e.g. some 60 countries).

An exploratory investigation of the possible effect of the types of quantities that are approximated on the way in which the approximator is translated was carried out on a selection of approximators. Were selected the approximators whose top two translation equivalents occur with very similar frequencies (see Table 2): près de (equally often translated by almost and nearly – 12 times), environ (9 times translated by about, 7 by some), dépasser (translated 6 times as in excess of and more than), and some (10 quelque and 9 environ).

This exploratory study shows that there seems to be a tendency to prefer one translation equivalent over the other with certain types of quantities. This tendency was especially noticed around percentages, for example environ is more often translated by about than some, dépasser is translated by in excess of (instead of more than) and some is more often translated by environ than quelque when approximating a percentage. On the other hand, the translation of dépasser by more than exhibits a tendency to occur around amounts of money. The approximators under scrutiny also seem to display preferred translation equivalents around numbers of people, for example some is more often translated by quelque than environ when referring to an approximate
number of people. The only exception is the approximator près de for which no similar tendency was observed as its translation equivalents almost and nearly are equally used around all types of quantities. However, a larger translation corpus would need to be investigated to check whether these tendencies are confirmed and to explore similar tendencies with additional approximators. The analysis of a comparable corpus in English and French may also provide clues as to the preferred approximators occurring around the types of quantities most often approximated in the corpus.

4. Translation equivalents of English and French approximators in bilingual dictionaries

In order to evaluate how bilingual dictionaries meet the needs of users when dealing with approximators, the entries of the twelve approximators under study are scrutinized in three bilingual dictionaries: Harrap’s Unabridged (2001, paper version⁸), Le Grand Robert et Collins électronique (2009, electronic version) and Grand Dictionnaire Hachette Oxford (2004, electronic version).

Table 3 lists the various translation equivalents found in the PLECI corpus and those recorded in the three bilingual dictionaries consulted. A distinction is made between full translation equivalents and additional translations given in the examples in each entry (in italics in Table 3).

The discussion of the findings focuses on (1) the examples given in bilingual dictionaries, (2) the issue of bidirectional translation equivalence in bilingual dictionaries, and (3) the way corpus findings may contribute to improving the treatment of approximators in bilingual dictionaries.

**Table 3. Translation equivalents of the French and English approximators studied in the PLECI corpus (journalese) and in bilingual dictionaries.⁹**

<table>
<thead>
<tr>
<th>Approximatos</th>
<th>PLECI translation corpus</th>
<th>Harrap’s Unabridged (paper version)</th>
<th>Le Grand Robert et Collins électronique</th>
<th>Grand Dictionnaire Hachette Oxford (electronic version)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corpus only</td>
<td>Corpus + at least 1 bilingual dictionary or all dictionaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Source language = French</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plus de</td>
<td>to exceed, about, almost, and over, less than, some, upwards of</td>
<td><em>more than</em>, over</td>
<td>more than, over</td>
<td>more than, over</td>
</tr>
<tr>
<td>près de</td>
<td>about, upwards of</td>
<td><em>almost</em>, <em>nearly</em></td>
<td>nearly, <em>almost</em></td>
<td>nearly, <em>almost close on</em></td>
</tr>
<tr>
<td>environ</td>
<td>some, in the order to, more than</td>
<td><em>about</em>, <em>around</em></td>
<td><em>about</em>, <em>around</em></td>
<td><em>about</em>, or thereabouts, or so</td>
</tr>
<tr>
<td>number + suffix –aine</td>
<td>almost, less than, some</td>
<td><em>about</em>, suffix –s, dozen, or so, several, around, more than, a score of</td>
<td><em>about</em>, <em>around</em>, or so, <em>roughly</em></td>
<td>about, or so, a score of several, around, dozens</td>
</tr>
<tr>
<td>quelque</td>
<td>up to, several</td>
<td><em>about</em>, or so, some, around</td>
<td>around, about, or so</td>
<td><em>some</em>, <em>about</em> or so</td>
</tr>
<tr>
<td>dépasser</td>
<td>above</td>
<td><em>more than</em>, over, in excess of,</td>
<td><em>to exceed</em>, to go beyond, in excess</td>
<td><em>to exceed</em>, in excess of</td>
</tr>
</tbody>
</table>
4.1. Examples

The examples (in italics in Table 3) given in the bilingual dictionaries under scrutiny differ from one dictionary to another. Additional translations are often presented in the examples given for an approximator. For instance, in the Grand Dictionnaire Hachette Oxford no fewer than 10 additional English translations are provided in the examples recorded for the approximators formed with the suffix –aine. It is noteworthy that the other two dictionaries list fewer (three and four) additional translation equivalents among the examples given for the same items.

Furthermore, the ‘quantity approximation’ usage of an approximator is sometimes poorly or even not at all illustrated in the examples. In Le Robert et Collins électronique for instance, up to, which is frequently used to approximate quantities in the corpus, is translated by jusqu’à. That said, no instance of up to used as an approximator with a number denoting a quantity is given among the seven examples in the dictionary, the dictionary favouring translations of uses such as up to now or up to here (i.e. up to as part of expressions of time and place). On the other hand, the Grand Dictionnaire Hachette Oxford makes a distinction between these different uses of up to and includes up to now and up to here in separate subentries, thereby highlighting the different translations that are used depending on whether up to is used as a quantity approximator or not: quantity approximation (up to 50 dollars translated as jusqu’à 50 dollars), time expression (up to now translated as jusqu’à maintenant) and place expression (up to here translated as jusqu’ici).

4.2. Bidirectional translation equivalence issues.

This study also sheds light on some bidirectional translation equivalence issues. Table 3 shows that some translations given in one direction are not mentioned as translation equivalents in the other direction: for example, quelque is translated by about in all three dictionaries, whereas about is never translated by quelque in the dictionaries (it is translated as quelqu in the PLECI translation corpus). Similarly, around is translated by environ in two bilingual dictionaries.
while it is not mentioned as a translation equivalent of *environ* in the other direction in these dictionaries. Approximators formed using a number and the suffix –*aine* in French are provided as translation equivalents for the approximator *some* whereas *some* is never listed as a translation equivalent for ‘–*aine* approximators’.

4.3. *Corpus findings and the treatment of quantity approximation in bilingual dictionaries.*

Finally, an examination of the translation equivalents of approximators that are present only in the translation corpus shows that some approximators in PLECI may be good candidates for inclusion in dictionaries. Some of the most frequently occurring translations in PLECI are not listed as translation equivalents in the bilingual dictionaries under scrutiny (or at least in some of them): *some* as an equivalent of *environ*, approximate quantities formed by using the derivational suffix –*aine* as translations of *about* and *quelque* as an equivalent of *some*. The translations used only in PLECI also highlight the wider grammatical diversity of approximators which can be used as translation equivalents, illustrating the fact that translation corpora may help lexicographers free themselves from what Granger and Lefer (in preparation) refer to as the ‘categorial bias’ (i.e. ‘the tendency to translate a given source item into a word of the same grammatical category in the target language’) and start to translate adverbs such as *about* by verbs such as *avoisiner*. Nevertheless, as has been shown by Granger and Lefer (in preparation), translation corpora may contain errors. An example of incorrect translation found in the present study is the translation of *plus de* (meaning ‘more than’) as *less than* in PLECI. This underlines the need for highly skilled bilingual lexicographers to select good translation equivalent candidates from corpora for inclusion in dictionaries.

Corpus findings may also inform lexicographers about the preferred use of some approximators in some contexts, for example in formal or informal contexts or with certain types of quantities. Some of the translation equivalents included in the bilingual dictionaries were not found in the corpus, for example *or so* as a translation of *environ* and *quelque*. A possible explanation is that approximators may exhibit different levels of formality, as has been demonstrated in Goossens and De Cock (2012). The treatment of the approximator *or so* in dictionaries seems to be an archetype of this phenomenon. It is labeled informal in several monolingual learners’ dictionaries whereas no register label is provided when *or so* is recorded as a translation equivalent in the bilingual dictionaries consulted. Results from corpus-driven investigations of quantity approximation in various business genres including business news reporting, academic business English and spoken business meetings (e.g. Goossens and De Cock 2012) offer further evidence for the preferred use of *or so* in more informal contexts as it never occurs in the more formal corpus of academic business English, while there are several occurrences of *or so* in the other more informal business genres examined.

Another related feature highlighted by corpus evidence which is lacking in the bilingual dictionaries under scrutiny is the lack of information about the type of items approximated. As exemplified in Section 3.2.3, some translation equivalents seem to be preferred with certain types of quantities in the translation corpus. A larger-scale corpus-based study might shed light on more general tendencies regarding the types of items or quantities associated with specific approximators, for instance to investigate whether time durations or percentages tend to combine with some approximators. The results of such an investigation might complement the information given in translation dictionaries, thereby guiding the user towards a potentially preferred translation equivalent.
5. Conclusion

The examination of a selection of ‘approximator + number’ combinations in an English-French translation parallel corpus shows that such expressions are not always translated using a similar combination, as the approximator is sometimes omitted in the translation and may be replaced by the use of a round number as an approximation for instance. When the approximator is translated, one or two translation equivalents can be regarded as generally preferred as they represent the majority of tokens. The analysis of the approximators in three bilingual dictionaries shed light on the unequal treatment of the ‘quantity approximation’ use of some approximators in the examples provided in the dictionaries and on some bidirectional translation equivalence issues.

Finally, the results of the corpus study carried out in this paper highlight several areas which may be improved in the treatment of quantity approximation in bilingual dictionaries: (1) the translation equivalents identified in the corpus exhibit a wider grammatical diversity than those recorded in bilingual dictionaries, (2) corpus findings suggest a possible link between the translation equivalent of an approximator and the type of quantity approximated and (3) register labels may be needed for some of the translation equivalents listed in bilingual dictionaries.

Notes

1 I acknowledge the support of the Fonds Spécial de Recherche (FSR, Université catholique de Louvain), which funds the research project within which I am currently completing my PhD thesis.
2 The PLECI corpus is paragraph- and sentence-aligned and contains translations of literary prose and newspaper articles into English and French.
3 Le Monde Diplomatique is an international newspaper published in 26 languages.
4 Example from the PLECI translation corpus (journalese), French to English direction.
5 Example from the PLECI translation corpus (journalese), English to French direction.
6 Example from the PLECI translation corpus (journalese), French to English direction.
7 Example from the PLECI translation corpus (journalese), French to English direction.
8 The paper version was used instead of the electronic version due to problems with the cd-rom.
9 approximator: translation found in the corpus + in all three bilingual dictionaries under study.
approximator: translation equivalent listed in all three bilingual dictionaries under study, on the same level (i.e. full translation equivalent or example).
approximator: additional translation equivalent only listed in the examples given in the bilingual dictionary.
10 The entries for the following approximators combining a number and the suffix –aine were scrutinized: dizaine, douzaine, quinzaine, vingtaine, cinquantaine, centaine.

References

A. Dictionaries


B. Other literature