Enriching Bilingual Dictionaries with Corpus-Based Data: First Steps Towards an Improved Description of Verbs in General Bilingual Dictionaries Thanks to a Popular-Science Corpus

Cécile Frérot[1] and Amélie Josselin-Leray[2]

1. Introduction

1.1 Corpora, translation and dictionaries

Corpora have grown increasingly attractive for language professionals over the past decade and, in specialised translation, they have gained considerable interest as they help enhance the target language production by providing information missing from conventional resources such as dictionaries, more specifically regarding term selection and idiomatic expressions (Bowker, 1998). However, general bilingual dictionaries are still widely used by specialised translators as shown by a recent survey on the role and treatment of terms, i.e. words belonging to specialized languages, in general bilingual dictionaries (Josselin-Leray, 2005). Therefore, we investigate how a popular-science corpus can help improve the treatment of terms in this type of dictionary.

1.2 The inadequate treatment of verbs in terminology and in lexicography

We focus on the description of verbs as it is a part-of-speech, just like adjectives and adverbs, which has been long been discarded by terminologists, as shown repeatedly by L'Homme (1998, 2002, 2004) and Kübler (Kübler 2002, Kübler and Frérot 2003). L'Homme (1998: 70) gives four reasons why verbs should be described more thoroughly: (i) "le verbe dénote une activité propre à un domaine de spécialité" (the verb refers to an activity which is speficic to a specialised field), (ii) "l'emploi des formes verbales est soumis à des conventions" (the use of verbal forms obeys certain rules), (iii) "il existe des différences d'un domaine de spécialité à l'autre" (some differences are to be found depending on the field), (iv) "les définitions générales doivent être précisées" (general definitions must be refined).

Our emphasis on verbs is further justified by the problems raised by phraseology in specialised translation (Roberts, 1994) and the translation-oriented perspective peculiar to bilingual dictionaries. Some efforts have been made in recent specialized dictionaries: users can now turn to bilingual dictionaries of specialized cooccurents, like the one designed by Meynard (2000) in the field of the Internet. However, in general language dictionaries, verbs are often the least well-treated part-of-speech, especially regarding syntax. This is the conclusion reached by Cowie (1989) regarding general monolingual dictionaries, but the same can be said of general bilingual dictionaries.

[1] Université Stendhal Grenoble III, France
e-mail: Cecile.Frerot@u-grenoble3.fr
[2] CLLE-ERSS, Université Toulouse II le Mirail and CNRS, France
e-mail: josselin@univ-tlse2.fr

1.3 Overview of the study

The goal of the study is to see to what extent the use of a popular-science corpus can be a first step towards an improved description of verbs in bilingual dictionaries. Section 2 describes the comparable corpus and the translation corpus used for the study as well as the corpus-based tools, and the bilingual dictionaries the comparison is based on. Section 3 is fully dedicated to the linguistic analysis of the five verbs under study and it is based on the description of the verb entries in the dictionaries and the corpus-based data. Concluding remarks are given in section 4.

2. Description of tools used for the analysis

2.1 Presentation of the two corpora

For our study, we have used two types of corpora that were originally designed for a large-scale study of the role and treatment of terms in general language dictionaries (Josselin-Leray, 2005). One is a French-English comparable corpus and the other a French-English translation corpus. There is no need to prove any more that they complement each other (Altenberg and Granger, 2002).

2.2.1 Comparable corpus

We use the term "comparable corpus" in the same meaning as Altenberg and Granger (2002: 7-8): "Comparable corpora consist of original texts in each language, matched as far as possible in terms of text type, subject matter and communicative function". "Original texts" means that the texts were not translated.

Our comparable corpus is a French-English popular-science corpus dealing with volcanology. It comprises approximately 400,000 words per language and has been written by native speakers- whole texts originating from France and Quebec on the one hand, the USA, Canada and Great-Britain on the other hand. The time period covered by the corpus ranges from 1980 to 2002.

As far as the type of text is concerned, three main reasons account for *popular science* texts: (*i*) lexical items found in popular science texts can be considered as actual terms (Delavigne, 2001); (*ii*) since a popular science corpus is by definition aimed at non-specialists, lexical items used in such a corpus should logically be found in the entries of general dictionaries aimed at the general public (as opposed to specialized dictionaries); (*iii*) the typical explanatory style of popular science texts is very useful to lexicographers. These popular-science texts are classified according to two main criteria– *discourse* and *genre*. The corpus aims at representing the three following discourse levels (based on Pearson 1998; Meyer and Mackintosh 1996): (*i*) "semi-popularized" discourse, written by experts for initiates (*e.g. Scientific American*– US; *Pour la Science*– FR) ; (*ii*) popularized discourse, written by relative experts for uninitiated (*e.g. New Scientist*– GB; *Discover*– US; *Science et Vie*– FR) ; (*iii*) instructional discourse, written by teachers for pupils (*e.g. A Teacher's Guide to the Geology of Hawaii Volcanoes National Park*). It also contains texts from various genres: it can be divided into actual texts and glossaries. The texts are themselves subdivided into the following categories: press (newspapers and magazines) and non-

press (textbooks, books, exhibition texts, Web documents). Table 1 summarizes the structure of the corpus.

Type of Document			**Ex. of French Titles**	**Number of Words**		**Number of Words**
Texts	**Non-Press**	**Academic Textbooks**	*Eléments de Géologie*	12,689	*Essentials of Geology*	10,590
		General Interest books	*Les Volcans et leurs Secrets*	81,151	General-Interest Publications of the USGS[3] on Volcanoes	91,208
		Web Documents	*Le Volcanisme de Djibouti*	26,732	*How Volcanoes Work*	21,603
		Exhibition	*Des Volcans et des Hommes*		∅	
	Press	**General Dailies**	*Le Monde, La Presse*	145,369	*The Guardian(GB) The New York Times (US) The National Post (CD)*	148,642
		Popular Science Magazines	*Science et Vie, Science et Avenir*	58,609	*New Scientist (GB), Science News (US)*	58,633
		Semi-Popular Science magazines	*Pour la Science*	15,689	*Scientific American*	16,094
		Discovery and Travel Magazines	*Géo, Terre Sauvage*	56,446	*Earth, National Geographic*	54,152
Glossaries				**3,227**		**4,686**
Total				**400,486**		**405,546**

Table 1.

2.2.2 Translation corpus

We use the term translation corpus in the same meaning as Altenberg and Granger (2002: 8): "Translation corpora consist of original texts in one language and their translations into one or several languages. If the translations go in one direction only (from language A to language B), they are unidirectional; if they go in both directions (from language A to language B and from language B to language A they are bi-directional". Our corpus is a bi-directional corpus as some translations are from French into English and some translations are from English into French. The corpus is much smaller than the comparable corpus because of how difficult it is to find translated texts in electronic format. It is also slightly more specialized than the comparable corpus. The information is summed up in table 2.

[3] USGS: United States Geological Survey.

Type	Title	Number of Words	Date	Direction of translation	Title	Number of Words	Date
Semi-Popular Science Magazine	Pour la Science	70 132	1979 - 1999	ENG ⇒ FR	Scientific American	67 762	1979 – 1999
Special issue of the letter of the Parc des Volcans d'Auvergne	Volcanisme et Volcans d'Auvergne	12 482	2002	FR ⇒ ENG	Volcanism and Volcanoes of Auvergne	11 804	1999
Popular-Science book	Les Feux de la Terre-Histoires de Volcans	28 092	1991	FR ⇒ ENG	Volcanoes-Fire from the Earth	26 268	1993
Total		**110 712**				**105 706**	

Table 2.

2.3 Corpus analysis tools

Exploring the comparable corpus and the translation corpus implied using corpus analysis tools. In order to take full advantage of the comparable corpus and focus on the grammatical category of verbs, we used a corpus-based terminology extraction tool, Syntex (Bourigault *et al.*, 2005) which was built for French and English noun extraction as well as verb and verb phrase extraction.

However, we had to use a concordancer to be able to make occasional and precise searches in the comparable corpus - a possibility that is ruled out with a terminology extraction tool. Using AntConc[4] proved very useful in that respect.

Finally, exploring a translation corpus made it necessary to use a tool that was both a bilingual aligner and a bilingual concordancer and we turned to LogiTerm[5].

2.4 Dictionaries

For our study, we have used the following bilingual (English-French, French-English) general dictionaries on Cd-Roms: *(i) Harrap's Shorter French and English Dictionary* (2000) (henceforth HAR), which has been chosen because of its relatively exhaustive treatment of scientific and technical words and *(ii) Oxford Hachette French-English, English-French Dictionary* (1996) (henceforth OXHA), given its corpus-based approach[6].

3. Analysis of five examples

3.1 Choice of the verbs and method of analysis

For this study, we have decided to have a closer look at five verbs in particular: the English verbs *erupt, ooze, extrude* and *spew*, and the French verb *se réveiller*. Several

[4] http://www.antlab.sci.waseda.ac.jp/
[5] Built in Ottawa (Canada) by Terminotix.
[6] There was no major change in the most recent editions of those dictionaries concerning the verbs we studied. This is the reason why we kept to the above-mentioned editions.

criteria were taken into accoung when we chose which verbs to study: language (English or French), terminological status (relevance to the field), frequency of occurrence in the corpora, relevance to our translation-oriented perspective.

First and foremost, we need to say that we decided to focus mainly on English verbs (four out of five verbs under study) because our native language is French, and we usually translate from English into French, and so do our student translations. The next thing that comes to mind is the terminological status of the verb. How to decide on the terminological status of a lexical item is a thorny issue, which has been studied in detail by Delavigne (1995). In fact, not all verbs mentioned in this study have the same status: two of the five verbs can be considered as actual terms belonging to the field of volcanology: *erupt* and *se réveiller*. To decide on their status, we partly relied on the criteria used by L'Homme (2002) to include verbs in specialized dictionaries, i.e. "Normally, verbs [...] are included in specialized dictionaries if they are not used in other contexts (e.g. the French term *configurer* '*configure*' appears in dictionaries of computing, since it exists only in that field), or if they have a meaning that cannot be described using definitions found elsewhere (e.g. *to surf*, in the field of the Internet, has a metaphorical meaning that cannot be captured using definitions found in other dictionaries, for instance, general language dictionaries)". We also used the morphological criterion (when verbs are derived from nouns or vice-versa, e.g. *erupt / eruption, se réveiller / réveil*). Another criterion was whether the verb combined with a term in noun form (e.g. *a volcano erupts / un volcan se réveille*). The verb *erupt* was also chosen because it was the most frequent verb pertaining to the field of volcanology in the English comparable corpus (554 occurrences), and because its most frequent equivalent in French (*entrer en éruption*) has a more complex syntactic pattern than the term in English. As for the French verb *se réveiller* (78 occurrences in the French comparable corpus), it especially caught our attention because it is a metaphorical term, something which does not translate easily.

The remaining three verbs (*extrude, spew, ooze*) are not terms *per se* of the field of volcanology. They happen to be frequent cooccurents of terms such as *volcano, lava, magma* etc (e.g. *Viscous dacitic-to-rhyolitic lavas generally **ooze** out of the volcano's central vent*). And as such they need to be studied because it is a well-known fact that collocations, and especially verbal collocations, are very difficult to translate and the translator often has no resource to turn to when confronted with them. Student translators in particular often have difficulty in translating verbal collocations. When asked to translate the following sentence into French *When basaltic lava is extruded, the dissolved gases escape quite freely* many of our students failed to translate the collocation *lava / extrude* and gave a word-for-word translation (*extruder*). Furthermore, *Ooze* and *spew* are particularly tricky because they often combine with a number of prepositions and adverbs: *ooze out, spew forth* etc.

3.2 Analysis of the treatment of the verbs in the macrostructure of the bilingual dictionaries

We started by extracting data from the comparable corpus concerning the five verbs under analysis. We skimmed through the concordance lines provided by AntConc and the contexts provided by Syntex and roughly established the frequencies of the verbs and their various constructions. These are given in the following table:

		Comparable corpus
erupt	**Total**	**554**
	Tr	42 (7.6%)
	Intr.	512 (92.4%)/
extrude	Total	**47**
	Tr	42 (89%)
	Intr	5 (11%)
se réveiller		**78**
ooze	Total	**42**
	Tr	5 (11.9%)
	Intr	37 (88.1%)
spew	Total	**106**
	Tr	77 (72.6%)
	Intr	29 (27.4%)

Table 3.

After analysing the data from the corpus, we turned to the bilingual dictionaries to see how these five verbs were treated in the macrostructure. At first sight it seemed that dictionaries did satisfy our needs since all English verbs were included as headwords in the English-French part of the dictionaries, and the verb verb was found in the French-English part of the dictionaries. However, a closer look at the data provided by the corpus showed that there were two kinds of problems: either the description was incomplete, or it was very difficult to decide whether the headword recorded in the dictionary did correspond to the field of volcanology.

3.2.1. Incomplete information

When we first skimmed through the data for *erupt* in the comparable corpus, we found out, surprisingly enough, that *erupt* could be used in a transitive manner, as shown in the following examples:

After periods of volcanic quiescence lasting as long as hundreds of thousands to even a million years, some Hawaiian **volcanoes erupt** alkalic volcanic **products**.
Pillow lavas are volumetrically the most abundant type because **they are erupted** at mid ocean ridges.

An in-depth analysis of the the 554 occurrences of the verb *erupt* data provided by the corpora led us to the conclusion that the transitive use was not a rare phenomenon: we found forty-two occurrences of this construction in the comparable corpus, and in different sources of the corpus (see figures in table 3). However, neither dictionary records this type of construction, which is a real problem for the translator who cannot use the equivalent given by the dictionary for the intransitive construction (*"entrer en éruption"*) to translate the transitive construction. The very same happens to the verb *extrude*: according to the corpus, it has both a transitive and an intransitive construction (see figures in table 3), but in OXHA, one only finds the transitive construction. The problem gets even more complex when verbs are used in combination with various prepositions, which is the case for *ooze* and *spew*, as will be seen later in 3.3.1.

3.2.2 Relevance to the field

When faced with a terminological verb which can also be used in general language, such as *erupt* and *se réveiller*, or with a "collocational" verb that can be used in a wide range of contexts, the translator needs to be reassured as to whether the sense division of the dictionary entry and the correspondant equivalent he uses are the correct ones. The means the dictionary should logically provide is a systematic field label. As a rule, neither dictionary uses the field label 'volcanology'; however, they do have a 'geology' label, which can act as a generic term for volcanology. The label 'geology' is in fact only used in one subdivision of one entry in one of the two dictionaries: for the intransitive construction of *extrude* in HAR. There is a field label for that construction in OXHA, but it pertains to industry, which is irrelevant in our case.

Another indirect means that can be used by lexicographers to show the user what field a particular sense division refers to is actants. By *actant*, we mean a word which indicates the context, when the choice of equivalent depends on the context in which the source language word is used. An actant may be a typical subject or typical object of a headword, which acts in the immediate structure in which the word is used and is able to fit into the structure of the sentence[7]. When the headword is a verb, the actant may be a noun serving as a subject, or as direct object of that verb. In our case, we expected to find the actant *volcano* as this is the most relevant key term in our field. We did find it in the entry *erupt* in both OXHA and HAR, as shown in table 4.

OXHA	HAR
erupt (...) / intransitive verb 1 **[volcano]** entrer en éruption;	**erupt** (...)vi (a) **(of volcano)** entrer en éruption

Table 4.

However, when we decided to look up the verb *se réveiller*, there was a discrepancy between the two dictionaries, as shown in the following table :

OXHA	HAR
II **se réveiller** verbe pronominal 3 (après période d'inertie) [personne, peuple] to wake up; [nature] to reawaken; **[volcan]** to become active again;	2 **se réveiller** (a) (d'un dormeur) to wake (up), to awakese réveiller d'un sommeil agité to wake from a troubled sleep (...) (b) (de sentiments) to be awakened or roused or stirred (up) (c) (**de la nature**, la végétation, l'économie) to revive

Table 5.

In the case of HAR, how can the (student) translator be sure that he should choose the equivalent given in sense division (c) ? It is only through a look at the corresponding noun, *réveil*, that he/she can find some helpful information: HAR gives *the renewed rumblings* or *the awakening of a volcano* as possible translations of *le réveil du volcan*. Finally, there can be no actants or indication of field whatsoever. This is the case of the verb *spew* in HAR, whose entry is to be found in table 6.

[7] Based on the *Bilingual Canadian Dictionary Methodology*, Roberts 1998.

Spew vti Br Familiar, colloquial (vomit) vomir, Slang dégueuler Slang it makes you want to spew! ça fait gerber!

Table 6.

The presentation is particularly confusing because there are several labels before the first equivalent: a geographic one, and a register one. How can the translator be sure that he is dealing with the same verb in general language and volcanology, and that the only possible equivalent is *vomir* and not *dégueuler* or *gerber*? In fact, the entry *spew* especially in HAR seems to be oversimplified, something we will now see in more detail with the analysis of the microstructure, and in particular the analysis of the presentation of grammatical information.

3.3 Analysis of the treatment of the verbs in the microstructure of the bilingual dictionaries

3.3.1 Presentation of grammatical information : The case of phrasal verbs and verbs followed by prepositional phrases (*spew* and *ooze*)

The analysis of all the occurrences of the verb *spew* (109)[8] in the comparable corpus yielded a wealth of information concerning the syntactic patterns it can be used in, as shown in table 7.

Use		**Frequency in the comparable corpus**	**Example from the comparable corpus**	**HAR**	**OXHA**
Transitive		**77 (72,6%)**			
1. Spew		59	The volcano has been **spewing** gas and ash for the past month	-	+
Phrasal Verbs	2. Spew out	16	the 732-metre-high volcano **has spewed out** ash and rocks	+	+
	3. Spew forth	2	Scientists have long puzzled over why certain volcanoes, (...) **spew forth** lava	+	-
Intransitive		**29 (27,4%)**			
4. Spew + ∅		3	when it suddenly began to shake and swell and **spew**	-	+

[8] Out of the 109 occurrences extracted with AntConc, 106 only were included in our analysis. Indeed, as the concordancer uses raw text - discarding the possibility of restricting the query to verbs only - some occurrences may be irrelevant. This was the case for *spewing* which can be a verb (gerondive) or an adjective. For instance, the following occurrence was not analysed as *spewing* is not used as a verb :
`Clifford Farrell, a fireman, gazed at the` **`spewing`** `vents of the Soufriere Hills`

5. Spew + PP (out of)		5	molten rock and ash **spewed out of** the ground	-	-
6. Spew + PP (from)		14	Increasing sulfur dioxide **spewing from** volcanic vents	-	-
7. Spew + PP (onto)		1	the effects of ash that **spewed onto** Plymouth and nearby villages	-	-
8. Spew + PP (into)		2	high-level nuclear waste **spewing into** the atmosphere	-	-
Phrasal verbs	9. Spew out	2	a huge fountain of lava and ash **spewed out** to form a perfect volcanic cone	+	+
	10. Spew out + PP (onto)	1	before they [magmas, molten rock] **are spewed out** onto the surface	-	-
	11. Spew forth	1	the pent-up waters **spewed forth** with startling speed	+	+

Table 7.

The verb *spew* can be simply transitive, or intransitive, as in (1) and (4), or it can be part of a phrasal verb, such as *spew out* (2, 9) or *spew forth* (3, 11), which, themselves can be transitive or intransitive. Moreover, all these patterns can be followed by additional prepositional phrases (PP) such as PPs introduced by *from* (6). How do the two bilingual dictionaries deal with this type of information? We have summed up the presence or absence of the various constructions in table 7, but let us now have a look at the entries in more detail:

OXHA	HAR
spew **I transitive verb** 1 **(also spew out)** vomir [smoke, lava, propaganda]; cracher [insults, coins, paper]; 2 [!](also spew up) dégobiller[!][food, drink]. **II intransitive verb** 1 **(also spew out, spew forth)** [lava, smoke, insults] jaillir;	**spew** **vti** Br Familiar, colloquial (vomit) vomir, Slang dégueuler Slang it makes you want to spew! ça fait gerber! **spew forth, spew out** **1 vi** (of lava, flames etc) jaillir, fuser (from, of de) Figurative use (of propaganda, lies, etc) fuser **2 vt sep** (lava, Fig propaganda) déverser **spew up vi, vt sep** Br Familiar, colloquial (vomit) vomir, Slang dégueuler
2 [!](also spew up) dégobiller[!].	

Table 8.

First, even though the corpus shows that the simple transitive pattern (1) is most frequently used in the corpus (59/109), this does not strike the reader when looking at the entries, especially in the HAR, which only has one entry for both the transitive and the intransitive construction. It is a space-saving technique used by lexicographers when the equivalent which is given is the same (i.e. *vomir*), but it can sometimes be confusing. According to the OXHA, anyway, there are in fact two different equivalents: *vomir* for the transitive verb, and *jaillir* for the intransitive one.

Another difference between the two dictionaries is that the hierarchy of information is presented in a different way: the OXHA has decided to include phrasal verbs within the entry *spew*, given as synonyms of the headword *spew* for some meanings only, whereas the HAR has decided to have a separate entry for the two phrasal verbs *spew forth* and *spew out*. The two policies are acceptable if two conditions are fulfilled. First, the presentation has to be consistent within a dictionary, but this is not the case: for the verb *ooze*, OXHA has a subentry for the phrasal verb *ooze out* (it should be given as a synonym of the intransitive construction of *ooze* if the dictionary were consistent), and HAR includes the same phrasal verb in the middle of an example, without any distinction from the verb *ooze*:

1 vi suinter (from de) *the wound was oozing with pus/blood* du pus/sang suintait de la plaie *water that* ***oozes out*** *from the rock* eau qui sourd du rocher *the walls were oozing with water* les murs suintaient, l'eau suintait des murs *the mud oozed up between her toes* la boue sourdait lentement entre ses orteils

Table 9.

Second, the information should be complete enough. Table 7 shows that, for instance, OXHA does not list the transitive use of *spew forth*, something that might be explained, however, by the fact that it is less commonly found (*cf.* the low frequency in the corpus).

After giving an overview of the problem of the treatment of grammatical information related to verbs, let us now deal with the core of our analysis: the problem of the translation of verbs.

3.3.2 Translation

Our translation-oriented study led us to focus on the equivalents provided by the bilingual dictionaries. This section is divided into sub-sections, each title addressing a particular issue or problem regarding the description of equivalents. We put the emphasis on the information related to the equivalents - actants in particular - as these are of special relevance to translators who want to pinpoint the right translation. The last two sub-sections aim at showing in greater detail how the corpus can help enhance the equivalents in the dictionaries under study.

3.3.2.1 Discrepancies in the French equivalents provided by the dictionaries

To begin with, the HAR gives *expulser* and *faire jaillir* as equivalents of the verb *extrude*, mentionning that in the field of geology the equivalent is *s'épancher*. However, none of these verbs are mentionned in the OXHA whose equivalent is *faire sortir*.

Another example is the information related to the verb *ooze* : the equivalent *suinter* is provided by both dictionaries but then OXHA uses *s'écouler* as equivalent of *ooze out* while HAR does not.

The most telling example we wish to mention is the verb *vomir* : it is given as equivalent of *spew* by OXHA in the field of volcanology (the actant is *lava*) whereas

HAR does not mention that it can be used in this specialised sense. However, to our surprise, we found that *vomir* is said of a volcano in the Fr-En section of the HAR (see table 10):

VOMIR
1 vt
(b) (d'une cheminée, **d'un volcan** etc) (fumée, flammes etc) to vomit, to belch forth, to spew out

Table 10.

3.3.2.2 Confusing data or lack of data associated with the equivalents

Some of the data associated with the equivalents can be regarded as confusing – sometimes misleading. For instance, the HAR doesn't provide any equivalent in a lemmatized form for the transitive use of the verb *ooze* - while it does for the intransitive use - but gives only one example (sentence) followed by a French equivalent : *the wound was oozing pus/blood du pus/sang suintait de la plaie.* This information is confusing as translators are very likely to wonder whether the verb *suinter* is said only of a *wound/plaie* or whether it can be applied to some other nouns. Ideally, translators expect to find in a dictionary one equivalent in the lemmatized form followed by a mention on actants (subjects and objects in the case of verbs) as well as an example showing how it is used in language[9].

Another example showing some confusing data associated with the equivalents is the verb *se réveiller*, the problem of which was mentionned earlier (see section 3.2.2, table 5). Indeed, the sense division (c) in the HAR entry of *se réveiller* [*(c) (de la nature, la végétation, l'économie) to revive*] - more specifically the information on the actant *nature* - is tricky : even though *nature* is the most approaching sense of *volcano*, the information is not enough accurate to enable (student) translators choose the equivalent. In section 3.2.2, we have shown that the helpful information had to be found in the corresponding noun entry where *réveil* is said of a volcano.

The confusing data can also be exemplified with *s'épancher* given as equivalent of the verb *extrude* in the HAR ; *extrude* is said of a rock in the field of geology and the abbreviation *etc [Geology (of rock **etc**) s'épancher]* suggests that *s'épancher* can apply to other semantically-related nouns. Which are they exactly ? The question remains unanswered and the information does not seem enough accurate for translators.

Finally, the information provided by the bilingual dictionaries is sometimes incomplete as is the case for OXHA which gives *s'écouler* as equivalent of *ooze out* without any indication on the actant (subject) to be used with the verb. Therefore, the question arises whether *s'écouler* can apply to the field of geology or volcanology but it remains unanswered.

[9] The type of information displayed in this entry is all the more confusing as two lexical combinations are mentioned to be used in the figurative sense of *ooze* (*to ooze confidence déborder d'assurance / to ooze charm exsuder un charme mielleux*) followed by a sentence (*Figurative use this place just oozes wealth cet endroit sue l'opulence*).

3.3.2.3 Differences in the actants provided by the dictionaries and the corpus

A thorough analysis of the nouns serving as direct object of the verb *spew* in the comparable corpus shows that the most common nouns associated with that verb (as well as *spew out* and *spew forth*) are *ash* and *lava*, which account for about 50% of the total occurrences, followed by *gas* and *rock* – 15% each. When we compared the data found in the corpus and the data from the dictionaries, we found that the HAR only provides *lava* as direct object of *spew* while the OXHA gives *lava* and *smoke*. However, based on the information found in the corpus, the noun *smoke* accounts for less than 2% of the total occurrences. As we will see in more detail in section 3.3.2.6, using the corpus-based data in the dictionaries can be a useful source of linguistic knowledge to enrich the current information and make it more accurate.

The same applies to the intransitive use of *ooze*. We have conducted a similar corpus analysis in the comparable corpus and found that the word combinations *lava*[10] [subject] + *ooze* and *magma*[subject] + *ooze* are the most frequent in the corpus as shown in table 11. However, they are found neither in OXHA, nor in HAR. Besides, the number of occurrences with the nouns *water* and *mud* - given in the HAR - accounts for only 10% of the total occurrences.

Subjects of the verb *ooze*	**Frequencies in the comparable corpus**
Magma	6 (16,2%)
lava + compounds (*basalt lava, dacitic-to-rhyolitic lavas, lava dome,...*)	27 (73%)
volcano / water / mud / basalt	1 (10,8%)

Table 11.

3.3.2.4 Several equivalents provided by the bilingual dictionaries

One of the thorniest issues in bilingual dictionaries is the number of equivalents provided to the users. Indeed, translators are very often faced with several equivalents and just do not know which one should be used.

The information provided by the HAR for the translation of the verb *ooze* is very telling in that respect; the French equivalent *suinter* is given followed by a number of examples (sentences) in which the verb *sourdre* is used (see table 12):

1 vi
suinter (from de)
the wound was oozing with pus/blood du pus/sang suintait de la plaie
water that oozes out from the rock eau qui sourd du rocher
the walls were oozing with water les murs suintaient, l'eau suintait des murs
the mud oozed up between her toes la boue sourdait lentement entre ses orteils
Figurative use to ooze with confidence déborder d'assurance

Table 12.

[10] As well as some compound nouns.

A closer look at the sentences shows that *sourdre* is the translation provided for the phrasal verbs *ooze out* and *ooze up*, which raises the question of the synonymy between:

- those phrasal verbs and the verb entry *ooze;*
- the French equivalents *suinter* and *sourdre.*

Unfortunately, no other information is provided and the entry displayed as such can hardly be helpful to translators. Actually, the translation may depend on the actant[11] and on the type of construction (verb or phrasal verb), but the HAR fails to provide any accurate information.

3.3.2.5 Are the equivalents attested in the corpus ?

One of the most straightforward uses of a corpus is to check whether a given word is used in language. Therefore, we conducted an analysis aimed at observing whether the equivalents provided by the bilingual dictionaries were attested in both the comparable corpus and the translation corpus. We have limited our search to the equivalents pertaining to the field of geology/volcanology using the field label provided by the dictionaries (example : *(Geology) s'épancher*) as well as the information on the actants (example : *lava* for *spew*) whenever possible. We also included the equivalents for which no information was provided (example : *s'écouler*) in order to see whether they were used in the field of geology/volcanology. The results[12] are summed up in table 13.

Verb entry	Equivalent	Comparable corpus	Translation corpus	Specialised sense[13] in HAR and OXHA
Ooze	Suinter	3	1	-
	Sourdre	1	2	-
	S'écouler	44	8	-
Extrude	**S'épancher**	**33**	**10**	**+**
	Expulser	19	13	-
	Faire jaillir	2	0	-
Spew (forth /out / up)	**Jaillir**	**78**	**12**	**+**
	Fuser	**1**	**0**	**+**
	Déverser	**5**	**3**	**+**
Spew	**Vomir**	**27**	**6**	**+**
Se réveiller	**Revive**	**1**	**0**	**-**

Table 13.

This analysis is only at a very preliminary stage and a lot of work remains to be done to thoroughly analyse the corpus-based data in order to be able to draw definite conclusions. However, some of the results displayed in table 13 can show some general trends.

Generally speaking, not all the equivalents are well attested in the corpus. Nevertheless, it should be noticed that among the five equivalents for which the specialised sense is mentioned in the dictionaries, three of them are widely attested in the corpus (*s'épancher*: 33/11 occurrences; *vomir*: 27/6 occurrences; *jaillir*: 78/12

[11] Are the following nouns *wound, water, wall* and *mud* the only nouns to be used with the verb *ooze* ? Are they the most commonly used or do they represent the generic word of a semantic class ?

[12] We skimmed through all the occurrences and only took into acount those belonging to the field of geology/volcanology.

[13] Field label or information associated with actants.

respectively in the comparable and the translation corpus). However, the number of occurrences for *fuser* and *déverser* is rather low suggesting that other equivalents may be used.

Another significant result is the number of occurrences for *s'écouler* and *expulser* (respectively 44 and 19) for which no information pertaining to the field was provided in the dictionaries. The figures show that both equivalents are widely used in the field of volcanology. However, further work is necessary and implies skimming through all the occurrences to clearly identify the actants associated with each verb (subject/object). The information regarding the actants could then be included in the dictionaries as none is actually used.

Regarding the English equivalent of the French verb *se réveiller*, it is worth noticing that *revive* is said of a volcano only once in the comparable corpus *(Professor Steven Sparks, the chief scientist at the Montserrat Volcano Observatory, said Montserrat is also likely to* ***revive*** *as soon as the eruptions stop)* while no occurrence was found in the translation corpus. This figure suggests that other potential equivalents may be in use and could therefore be found in the corpus. In the next section, we investigate such a possibility and show how the corpus could be used to help enhance the content of the dictionaries.

3.3.2.6 How can the corpus help improve the dictionaries ?

In the previous sections (3.3.2.2 in particular), we have shown that the information pertaining to actants in the verb entries analysed can be confusing and incomplete. Therefore, we first illustrate how the corpus-based data can be of special relevance in that respect (3.3.2.6.1). Then, we turn to translation of the verb, strictly speaking, and focus on the potential equivalents provided by both the comparable and the translation corpus (3.3.2.6.2).

3.3.2.6.1 Actants

As mentioned earlier (3.3.2.5), the verb *s'écouler* is widely attested in the comparable corpus and the OXHA does provide it as equivalent of *ooze out.* However, the dictionary fails to mention its actants (the typical subjects of the verb in this case). Our corpus analysis shows that *lava* is one of the most frequent nouns associated with the verb *s'écouler* (table 14). Therefore, including this noun in the dictionary would make the equivalent much more accurate and useful to translators.

le stade ultime est la **lave** qui **s'écoule** calmement à la surface
les **laves** les plus visqueuses ne peuvent pas **s'écouler**
la **lave** continue de **s'écouler**
tandis que les **laves** basaltiques **s'écoulent** avec fluidité
De la **lave** fraîche **s'écoule** continuellement de ces volcans
La **lave s'écoule** en grande partie dans ce réseau complexe
La **lave**, visqueuse, ne peut **s'écouler** et dresse

Table 14.

Another example we wish to mention is the French verb *déborder*, which is given by the HAR as equivalent of *ooze* only in a figurative use *(to ooze with confidence déborder d'assurance)* and used with *butter* in the OXHA. This verb is actually used in the field of volcanology as shown by the occurrences displayed in table 15) and co-occurs in the comparable corpus with the noun *lave* as well as some compounds (*remontée de lave, coulée de lave*).

la **lave débordant** du cratère à intervalles des **remontées de lave** visqueuse **débordent** de son cratère ses **coulées de lave débordent** rarement de son enclos

Table 15.

In section 3.3.2.2, we highlighted the use of the abbreviation *etc* in the verb entry of *extrude* (*[Geology (of rock **etc**) s'épancher]* which suggests that, apart from *rock, s'épancher* can also be said of some other semantically-related nouns. A closer look at the fifty-three occurrences of *s'épancher* in the comparable corpus provides an answer as it shows that *lave* and *magma* frequently co-occur with *s'épancher*: *la lave s'épanche tranquillement sur le fond de la mer* ; *la lave provient d'un magma profond qui s'épanche en surface.*

3.3.2.6.2 Equivalents

One of the most telling examples highlighting how helpful the use of corpus in translation can be is the French verb under study. Generally speaking, translating metaphorical terms, as is the case of *se réveiller* (78 occurrences in the comparable corpus), is a thorny issue as it is very difficult to know whether a given metaphor in the source language can be translated as such into the target language.

When looking up a translation of the French collocation *volcan / se réveiller* in both bilingual dictionarires, users are very likely to be dubious as OXHA provides the equivalent *to become active again* but HAR, under the entry *réveil*, gives *the renewed rumblings* or *the awakening of a volcano*. The latter may be questionable as it looks like a word-for-word translation, suggesting some kind of "translationese". An in-depth analysis of the translation corpus (Frérot, Josselin, to be published) shows that *awakening*, as well as the verbs *awake/re-awaken*, are used to translate *se réveiller* in the Fr-En translations (*i.e.* where French is the source language) as illustrated by the examples in table 16 :

[FR] Dès leur découverte, la question a été posée de savoir si les volcans d'Auvergne pouvaient **se réveiller**.	[ENG] Since their discovery it has commonly been asked whether the volcanoes of the Auvergne could **re-awaken.**
[FR] La vengeance des dieux est terrible, le volcan **se réveille**. 1184 personnes périssent dans les nuées ardentes et les coulées de boue.	[ENG] The vengeance of the gods was terrible, and the volcano - which had been dormant for six centuries - **awoke**: 1184 people died in the ash and mud.
[FR] Hekla est l'un des volcans les plus actifs d'Islande. Depuis son paroxysme de 1104, qui fut, avec celui de l'Oraefajokull en 1362, la plus forte éruption explosive des temps historiques dans ce pays, il a connu 167 **réveils**, le dernier en janvier 1991.	[ENG] Hekla (above, in a 16th-century print) is one of Iceland's most active volcanoes. Since its eruption in 1104 it **has awakened** 167 times, most recently in January 1991.

Table 16.

However, *reawaken* and *reawakening* are also used in the En-Fr section of the translation corpus (table 17) :

[ENG] The rapid scientific response to the **reawakening** of Mount St. Helens in 1980 drew on earthquake and ground movement-monitoring techniques developed at Kilauea.	[FR] Les techniques de surveillance des séismes et des mouvements du sol mises au point au Kilauea ont permis aux géologues de réagir rapidement face au **réveil** du mont Saint-Helens en 1980.
[ENG] Thus any small-scale activity that might have preceded its **reawakening** went unnoticed.	[FR] En conséquence, toute marque d'activité même faible, qui avait peut-être précédé son **réveil**, passa inaperçue.

Table 17.

The analysis of the data in the English part of the comparable corpus shows that *reawaken* is used by native speakers. Besides, it provides some other potential equivalents such as : *to be again in eruption, to be about to erupt, to be about to blow, to rumble back to life, to threaten to erupt, to begin to erupt, to erupt again, to erupt into life*.

The diversity of potential equivalents is also exemplified with the translations of *extrude* (table 18) and *spew* (table 19) found in the translation corpus and the next step implies examining in greater detail this corpus-based data together with the data in the comparable corpus.

[ENG] Basalt is the commonest rock formed by the solidification of magma **extruded to** the surface of the earth, the moon and perhaps other bodies in the solar system.	[FR] La solidification du magma qui **fait éruption** à la surface de la terre donne presque toujours naissance à de la roche basaltique.
[ENG] Domes of viscous lava **were**	[FR] Des dômes de lave visqueuse **se**

extruded inside the crater on June 13-20	**mirent en place** à l'intérieur du cratère, du 13 au 20 juin
[ENG] Some lavas **extruded** along a branch of the North Atlantic rift west of Greenland (but still within the compass of the Iceland plume) contain up to 20 percent magnesium oxide.	[FR] les laves **émises** dans le rift de l'Atlantique Nord, à l'Ouest du Groenland (associées au panache islandais), en contiennent jusqu'à 20 pour cent
[ENG] Like the young North Atlantic, this rift **extruded** vast amounts of magma onto the continental margins	[FR] Comme le rift de l'Atlantique Nord, ce rift **a déversé** de grandes quantités de magma sur les marges continentales

Table 18.

[ENG] Historically, the formation of immense volcanic provinces-regions of intense eruptions possibly caused by enormous buoyant plumes of magma within the underlying mantle-**may have spewed** large amounts of gases and led to periods of global warming.	[FR] Ainsi, lors des éruptions volcaniques massives qui ont produit les immenses provinces volcaniques du Globe (des régions couvertes de basaltes probablement issus d'énormes panaches de magma au sein du manteau), **l'émission** rapide de grandes quantités de gaz dans l'atmosphère a peut-être perturbé temporairement le climat global, qui se serait réchauffé.
[ENG] Historical records show just such a. pattern: only six rift eruptions took place between 18-1840 and 1950, whereas 17 rift eruptions have occurred since 1950, when Kilauea **has spewed** lava at a much greater rate.	[FR] Les enregistrements historiques du volcan confirment cette hypothèse : seules six éruptions de crêtes ont eu lieu entre 1840 et 1950, alors que 17 éruptions ont été enregistrées depuis 1950, lorsque le **débit** du Kilauea était élevée.
[FR] L'Etna est la colonne du ciel **vomissant** les eaux que la mer déverse dans le gouffre de Charybde.	[ENG] " In ancient lore Etna is the pillar of the sky **spewing out** the waters that the sea pours into the whirlpool of the monster Charybdis.
[FR] L'énorme effusion de basalte du Laki, la plus volumineuse des temps historiques avec ses 12 kilomètres cubes, **vomit** plus de 500 millions de tonnes de gaz nocifs.	[ENG] The enormous outpouring of basalt at Laki, in southern Iceland-almost 3 cubic miles-was the biggest on record. It also **spewed out** more than 500 million tons of noxious gases.
[ENG] He stood raging at the summit, fire **spewing from** his mouth, and the mountain shook and rumbled.	[FR] Debout, écumant de rage sur le sommet de la montagne, le feu **jaillissait de** sa bouche et le sol tremblait et grondait.

Table 19.

4. Conclusion

In this paper, we have investigated the possibility of using a French-English comparable corpus as well as a translation corpus as a source of linguistic knowledge in order to improve the content of bilingual dictionaries. We have analysed the description of five verb entries (*extrude, erupt, ooze, spew* and *se réveiller*) in two bilingual dictionaries, focusing on both the macrostructure and the microstructure, and have explored the corpus-based information pertaining to those verbs. We have tried to show how this

linguistic information is relevant for translators as it enhances the quality of the description of the verb entries by providing accurate and new data – this is particulary true of the information related to the actants and the equivalents. There is no denying that the information found in the comparable and translation corpus is worth being extracted and above all thoroughly examined. Therefore, further work implies refining the current analysis, especially regarding the potential equivalents found in the translation corpus.

References

Altenberg, B. and S. Granger (2002) Recent Trends in Cross-Linguistic Lexical Studies, in Lexis in Contrast. Corpus-Based Approaches, B. Altenberg and S. Granger (eds.), Amsterdam, Philadelphie, John Benjamins, 3–48.

Bourigault, D., C. Fabre, C. Frérot, M.-P. Jacques and S. Ozdowska (2005) Syntex, analyseur syntaxique de corpus, Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles, Dourdan, France, 17–20.

Bowker, L. (1998) 'Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study', *Meta, 43, 4*, 631–51.

Cowie, A. P. (1989) Information on Syntactic Constructions in the General Monolingual Dictionary. Dictionaries: An International Encyclopaedia of Lexicography, eds. F. J. Hausmann et al., vol. 1, Berlin, New York, W. de Gruyter, 588–92.

Delavigne, V. (1995) 'Approche socioterminologique des discours du nucléaire', *Meta*, 40:2, 308–319.

Delavigne, V. (2001) Les Mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation, thèse de doctorat non publiée, Université de Rouen, Rouen

Frérot C. and A. Josselin-Leray (to be published) Contribution des corpus à l'enrichissement des dictionnaires bilingues généraux. Application au domaine de la volcanologie. Colloque des Jeunes Chercheurs en Didactique des Langues et en Linguistique, Presses Universitaires de Grenoble.

L'Homme, M.-C. (1998) 'Le statut du verbe en langue de spécialité et sa description lexicographique'. *Cahiers de lexicologie*, 73–2, 61–84.

L'Homme, M.-C. (2002) What can Verbs and Adjectives Tell us about Terms?, Sixth International Conference on Terminology and Knowledge Engineering, Nancy, 28–30 août 2002, 65–70.

L'Homme, M.-C. (2004) La Terminologie : principes et techniques, Montréal, Presses de l'Université de Montréal.

Josselin-Leray, A. (2005) Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Etude d'un domaine de spécialité : volcanologie, thèse de doctorat, Université Lyon II.

Kübler, N. (2002) Prédicats verbaux et arguments en anglais informatique, 22ème colloque du GERAS, Lyon.

Kübler, N. and C. Frérot (2003) Verbs in Specialized Corpora: from Manual Corpus-Based Description to Automatic Extraction in an English-French Parallel Corpus, Corpus Linguistics 2003 Proceedings, Lancaster University, 429–38.

Meyer, I. and K. Mackintosh (1996) 'The Corpus from a Terminographer's Viewpoint', *International Journal of Corpus Linguistics*, 1:2, 257–85.

Meynard, I. (2000) Internet. Répertoire bilingue de combinaisons lexicales spécialisées français-anglais, Brossard (Québec): Linguatech.

Pearson, J. (1998) Terms in Context, Amsterdam, John Benjamins.

Roberts, R.-P. (1994) Identifying the Phraseology of Languages for Special Purposes (LSPs), in Terminologie et linguistique de spécialité : études de vocabulaires et textes spécialisés, Edwards, H. P. and Lapierre, L. (Eds.), Actes de Langue Française et de Linguistique 1994–1995. Halifax : Universitas Dalhousiana, 61–74.

Dictionaries :

Dictionnaire Hachette-Oxford (1996). Hachette Livre, Oxford University Press (Cd-Rom).

Harrap's Shorter. Dictionnaire anglais-français (2000). Chambers, Harrap (Cd-Rom).