Good Contexts for Translators-A First Account of the Cristal Project

Amélie Josselin-Leray, Cécile Fabre, Josette Rebeyrolle, Aurélie Picton, Emmanuel Planas

Good Contexts for Translators–A First Account of the Cristal Project

Amélie Josselin-Leray*, Cécile Fabre*, Josette Rebeyrolle*, Aurélie Picton**, Emmanuel Planas***
*CLLE-ERSS, University of Toulouse & CNRS, France; **FTI, University of Geneva, Switzerland; ***LINA, University of Nantes, France
josselin@univ-tlse2.fr, cecile.fabre@univ-tlse2.fr, rebeyrol@univ-tlse2.fr, aurelie.picton@unige.ch, emmanuel.planas@univ-nantes.fr

Abstract

This paper questions the notion of good contexts for translators and describes an experiment which tests the usefulness of two specific kinds of contexts in a translation task, namely (1) contexts that provide conceptual information about a term, and (2) contexts that provide linguistic information about the collocational profile of this term. In the experiment, trainee translators are asked to use several types of resources, including a set of pre-annotated contexts of various types, and to identify the contexts that they consider to be the most relevant for their task. We present the first results of this experiment, which confirm our general assumption about the usefulness of such rich contexts and indicate some differences regarding the use of contexts in the source and target language. This study takes place in the CRISTAL project whose aim is to retrieve from bilingual comparable corpora the contexts that are the most relevant for translation and to provide them to users through a CAT tool.

Keywords: CAT tools; corpus resources for translators; Knowledge-Rich Contexts

1 Introduction

Even though it is widely acknowledged as being essential to the translator, the very idea of context in translation is hard to define (Baker 2006: 321) and it also "lacks a definition that can be applied in the everyday work of a professional translator" as stated by Melby & Foster (2010: 1). Therefore, when one wants to provide translators with tools that better meet their needs–such as improved CAT tools–, one should in the first place wonder about what makes a context relevant for them. In other words, what is a 'good context' for translators? This is one of the questions the CRISTAL project[1] tries to give an answer to. The main aim of the CRISTAL project, an acronym that stands for "Knowledge-Rich Contexts for Terminological Translation" ("Contextes Riches en Connaissances pour la Traduction Terminologique" in French), is to automatically retrieve from bilingual comparable corpora the contexts that are the most relevant for translation and to provide them to users through the CAT tool developed by Lingua & Machina, the Libellex Platform.

The first part of this paper reviews the translators' needs regarding context. It seems necessary to first identify which type of information is essential for translators, to see how this information is recorded in the tools most commonly used by translators, i.e. dictionaries, term banks and corpora, and how satisfied translators are about the way that information is recorded. In order to refine the notion of "good contexts" for translators, in part 3 we investigate what a "good example" in lexicography and what a "Knowledge-Rich Context" in terminology are, and introduce the distinction between conceptually-rich and linguistically-rich concepts. Part 4 then focuses on one aspect of the methodology of the CRISTAL project: an experimentation involving trainee translators in order to refine our idea of a "good context for translators". Finally, part 4 presents the very first results of the experiment.

2 Some Facts about the Needs of Translators Regarding Context

As stated by Rogers & Ahmad (1998: 195), "one of the translator's prime needs is for context-sensitive information". We may wonder what the notion of context-sensitive information encompasses and what sources of information translators can rely on–or not.

2.1 What do Translators need Contextual Information for and Where do they Find it?

2.1.1 Context in Translation: a Preliminary Definition

As thoroughly explained by Melby & Foster (2006), specialists in many fields (e.g. philosophy, psychology, pragmatics, and functional linguistic) have discussed the notion of context, and various definitions have been written. The three facets of context as defined by Halliday (1999), i.e. *context of situation*, *context of culture* and *co-text* are all particularly relevant in translation. However, in this paper, we will only focus on what Halliday calls *co-text*. While both *context of situation* and *context of culture* are outside of language itself, *co-text* specifically pertains to language in use. It can broadly be defined as the surrounding discourse of an utterance. Therefore, our definition of "context" in this paper will

[1]CRISTAL is a three-year project (2012-2015) funded by the French National Agency for Research (ANR-12-CORD-0020). It involves four partners: a computing research team at the University of Nantes, France (LINA), a linguistics research team at the University of Toulouse, France (CLLE-ERSS), the Translation Technologies team from the Faculty of Interpreting and Translation at the University of Geneva in Switzerland, and a firm specializing in multilingual text management (Lingua et Machina).

be limited to *co-text*, and will rely on the definition provided by Fuchs[2]:

What is called *context* is the linguistic environment of an element (phonetic unit, word or group of words) within an utterances; i.e. the units that precede and follow it. Thus, in the utterance "Marie est jolie comme un cœur", the element *comme* has as its immediate context "jolie…un cœur" and its wider context "Marie est jolie…un cœur". By extension, the word *context* is also applied to the utterance(s) which precede(s) and follow(s) a given utterance in discourse. [3]

2.1.2 What do Translators Need Context for?

Following Roberts & Bosse-Andrieu (2006: 203), let us remind here that the translation problems translators have to face can be considered as source text-related (for comprehension of the source-text) or target-text related (for transfer into the target text) and classified into three main categories: *encyclopedic*, *linguistic* or *textual*. *Encyclopedic* problems encompass "general subject-related problems as well as more specific problems dealing with proper nouns–that is, a lack of familiarity with the topic of the text or with specific places or people mentioned in the text"; *linguistic problems* are defined as "those attached to specific words and phrases–that is, problems related to the comprehension or translation of a given word or phrase"; finally, *textual* problems are those concerned with text types and the internal organization or reproduction of a given text type.

Bowker (2011, 2012) draws a list of those items of contextual information that can prove "useful" for the translator to solve his source-text and target-text-related problems. They can be summed up as follows: (*i*) information about usage; this of course includes collocations, in particular which general-language words collocate with terms (see also Roberts 1994: 56), (*ii*) information about the frequency of use of a particular word or term, (*iii*) information about lexical and conceptual relations (such as synonymy, meronymy, hyperonymy etc.) (see also Marshman, Gariépy & Harms 2012, Rogers & Ahmad 1998), (iv) pragmatic information about style, register and genre (see also Varantola 1998), (*v*) information about usages to avoid. It thus seems to us that the items that are not situation-linked fall into the following categories: *conceptual* information and *linguistic* information.

To solve those problems and to make decisions, translators need external help, which they typically get by consulting other human experts and conventional resources such as dictionaries and term banks (Rogers & Ahmad 1998: 198).

2.1.3 Where do Translators Find Contextual Information?

As mentioned by Varantola (2006: 216), "the translator's problem-solving techniques have changed dramatically over the past decade or so". In addition to the above-mentioned conventional resources (monolingual and bilingual dictionaries–which have undergone radical changes–; term banks), translators now also partly rely on the information provided by corpora.

- Dictionaries

It is mostly through examples that dictionaries provide contextual information. The empirical study on scientific and technical words (i.e. terms) in general bilingual and monolingual dictionaries carried out by Josselin-Leray (2005) has shown that up to 80.3% of users turned to dictionaries to find information about how to use the term in a sentence and that bilingual dictionaries always ranked higher in that respect. Among the respondents who chose that answer, it was the « language professionals » user group (which includes translators) that was mostly represented.

- Term Banks

Term banks typically provide contextual information through the "context" section of the terminological record. The importance given to contextual information in terminological resources by translators is confirmed by the findings of the survey by Duran-Muñoz (2010): examples were considered to be "essential data" by the respondents, and among "desirable data", one found "a greater variety of examples" and "semantic information (semantic relations, frames)".

- Corpora

Although corpus data is obviously intrinsically made of contextual information, it is a resource which seems still quite scarcely used by translators, as shown by the results of the survey by Duran-Muñoz (2010): only 5.09% of the participants quoted (parallel) corpora as being a terminological resource they "used *more*[4] when translating". However, 41.8% of the respondents to the Mellange Survey[5], which was carried out in 2005-2006 among trainee translators and professional translators, do claim they use corpora in their translation practice (the most frequent type being the corpora in the target language).

Although there seems to be a wide array of resources translators can turn to when they need contextual information, these resources do not necessarily meet the translators' needs.

[2] http://www.universalis.fr/encyclopedie/concept/. Consulted on April 2, 2014.
[3] We translated the quotation.
[4] We added the italics.
[5] http://mellange.eila.jussieu.fr/Mellange-Results-1.pdf. Consulted on April 10, 2014.

2.2 The Shortcomings of Existing Resources regarding Context

2.2.1 The Dissatisfaction of Translators regarding Contextual Information in Existing Resources: a Hard Fact

We found it relevant to first look at the findings of various empirical surveys on the use of conventional resources by translators.

- Dictionary Use

Before starting to compile the *Bilingual Canadian Dictionary*, Roberts (1994: 56) carried out a survey among its potential users in order to clearly identify their needs and reached the following conclusion: "Between one-third and one-half the members in each user group [of sophisticated second language users] appreciated, to varying degrees, the number of examples presented in their present most frequently used dictionaries. But between one quarter and one half of each group felt that improvement was needed in that respect". The study by Josselin-Leray (2005) reached the same conclusion: although users were overall satisfied by the examples provided by their dictionaries (between 41.3% and 67.5% of users said they were satisfied), the level of satisfaction was lower for bilingual dictionaries, and lower among the "language professionals" user group.

- Term Banks

In the survey by Duran-Muñoz (2010), translators also had the opportunity to give their opinions regarding their needs; the second most repeated argument was "include more pragmatic information about usage and tricky translations", and the fifth one was "provide examples taken from real texts". The conclusion of her findings, found in Duran-Muñoz (2012: 82) is straightforward: "we can confirm that most of the terminological resources that are currently available (especially in electronic format) do not fulfill their requirements". Why is that so?

2.2.2 Why are Translators Dissatisfied?

Varantola (e.g. 1994) has written at great length about the *context-free* vs. *context-bound* dilemma faced by translators: dictionaries and term banks only provide context-free examples, i.e. examples that are perceived as prototypical and frequent, while what the translators need to find the suitable equivalent(s) is typically context-bound. Moreover, the examples/contexts provided are not varied enough. This is especially true of term banks, as underlined by Bowker (2011: 214-215) who explains the information found on those records is rather limited and usually consists in definitions and terms presented out of context, or in only a single context. She pinpoints a paradoxical situation in which the advances of research on terminology (especially the work on Knowledge-Rich Contexts, which we will introduce in 3.2) have not been integrated into the tools translators most commonly use, i.e. term banks.
However, Varantola (2006: 217) says the context-free/context-bound dilemma should now be qualified since "context-free definitions of concepts within a particular domain [which] were for a long time the theoretical ideal in terminological theory […] are now replaced by less rigid, contextually relevant definitions". She ascribes it to the availability of large corpora, whose role is also now central in dictionary-compiling. Some dictionaries now even "provide access to more examples in the form of concordances from the corpus data that lie behind the dictionary". Corpora are no panacea, though. One of arguments against corpora is that they are "tools of shallow intelligence" (Varantola, 2006: 223) when they are raw and non-tagged or POS-tagged, since the user "is left to handle the manipulation, dissection and interpretation of results". In other words, compiling the corpus and analysing the corpus can be too tedious a task for translators who often work under tight constraints[6].

2.3 The Translator's Ideal Workstation?

In 1996, Atkins already suggested (p. 526) that the dictionary of the future should "give its users the opportunity to make their own decisions about equivalences" : the users "should be able to consult as many examples as they need of words used in their various senses, each in a variety of contexts with a variety of collocate partners". More recently, Bowker (2011: 215) suggested: "it would be more helpful for translators to have access not simply to term records that provide a single 'best' term with a solitary context, but rather to information that would allow them to see all possible terms in a range of contexts and thus find the solution that works best in the target text at hand". She insists on the fact that looking at a wide range of contexts should not be considered as a waste of time, and that this has been made easier thanks to corpus-analysis tools that present information in an easy-to-read format. She goes even further by suggesting (Bowker 2012: 391) that translators have access to the whole of the information that lexicographers usually rely on when devising a dictionary entry:

> In order to arrive to that entry, lexicographers have gone through a number of intermediary steps, where they learn about the various characteristics of the words and concepts being described, such as their grammatical and collocational behaviours, the different relationships that hold between words and their underlying concepts, and the characteristics that are necessary and sufficient for distinguishing one concept in an intensional definition.

However relevant that objective might be, it seems rather ambitious and difficult to achieve in the very near future, all the more so as "lexicographers'needs are very different from translators' corpus needs" (Varantola 2006: 217). Narrowing

[6] Another point worth mentioning is that, in the survey carried out for the Mellange, even though 94.4% of the respondents said they used Google to research terminology, 10.2% found that Google was limited for finding information on language use because the "search results [did] not provide enough context to be useful".

down that objective to providing more corpus-based context data in a way that is more in keeping with the actual working conditions of translators seems more feasible, which is why the main aim of the CRISTAL project is to help design a CAT tool that provides translators with customized contexts automatically retrieved from comparable corpora.
In order to reach that goal–which can also sound ambitious, we first decided to refine the notion of 'good contexts for translators' by doing two things: (*i*) we first looked at the way lexicographers deal with examples in dictionaries and terminographers deal with "Knowledge-Rich Contexts" (part 3), (*ii*) we devised an experiment with trainee translators focusing on what we thought to be "good contexts" (part 4).

3 Dictionary Examples and Knowledge-Rich Contexts

The thoughts of lexicographers and terminologists on "good examples" or "Knowledge Rich Contexts" provide some valuable insight into what a good context for translators might be. After examining those two aspects, we give our own definition of Conceptually Rich Contexts and Linguistically Rich Contexts.

3.1 Good Dictionary Examples

Many studies have underlined the importance of the illustrative component in dictionaries as a means to provide typical contexts about a word's meaning and usage (Atkins & Rundell 2008). In monolingual as well as in bilingual dictionaries, examples are meant to help the dictionary user both in the production and the comprehension process. They have therefore diverse functions (Rey-Debove 2005, Roberts 1994, Siepmann 2005): they can provide syntagmatic information about word patterns and collocations, together with paradigmatic information about words that are semantically-related (synonyms, hyperonyms, etc.). They may also give pragmatic and stylistic indications about registers and specific uses, or be used as a more concrete and accessible complement to definitions, with an epilinguistic dimension.
Authentic examples that meet at least some of these requirements are very difficult to extract from corpora:

> Finding good examples in a mass of corpus data is labour-intensive. For all sorts of reasons, a majority of corpus sentences will not be suitable as they stand, so the lexicographer must either search out the best ones or modify corpus sentences which are promising but in some way flawed (Rundell & Kilgarriff 2011).

Kilgarriff et al. (2008) have developed a method to automatically collect sentences that are good candidate dictionary examples, using two criteria: readability (judged from sentence length and average word length; it penalizes sentences with infrequent words, more than one or two non-a-z characters, or anaphora) and informativeness (judged from the density of collocates in the sentence).

3.2 Knowledge Rich Contexts (KRCs) in Terminology

Knowledge Rich Contexts play a very important role in identifying terms in specialized texts because they show conceptual relationships between terms. It is within that framework that Ingrid Meyer defined Knowledge Rich Contexts as "a context indicating at least one item of domain knowledge that could be useful for conceptual analysis." (Meyer 2001: 281). These contexts are used in order to develop knowledge extraction tools for text-based terminology and ontology building (Condamines & Rebeyrolle 2001). Rich contexts for terminologists typically contain terms that are specific to the domain together with linguistic patterns that signal the conceptual relations between these terms as illustrated below in Meyer (2001):

(1) Compost is an organic material deliberately assembled for fast decomposition.
(2) Compost contains nutrients, nitrogen, potassium and phosphorus.

This type of information helps building networks of terms, generally focusing on hyperonyms (example 1) and meronyms (example 2).

3.3 KRCs for Translators: Conceptually *vs.* Linguistically Rich Contexts

Based on the type of information needed by translators as described in 2.1.2, and the type of information provided by dictionary examples and Knowledge-Rich Contexts as detailed in 3.1 and 3.2, we decided to extend the notion of KRC in our experiment, considering under this category two types of contexts: those that provide 'conceptual' information about a given term-called "Conceptually Rich Contexts" (CRCs) in our study, and those that provide 'linguistic' information about that term "Linguistically Rich Contexts" (LRC). In our experiment, the contexts which are neither conceptual nor linguistic are considered as *poor* (cf. Reimerink *et al.* 2010: 1934).

4 An Experiment Centered on "Good Contexts" within the CRISTAL Project

The main aims of the experiment were (*i*) to check that rich contexts extracted from corpora are useful to translators, (*ii*) to identify which types of rich contexts (CRCs or LRCs) are the most useful to them.
A pilot study was carried out in December 2013 at the University of Geneva (Switzerland). This allowed us to test the protocol and to make the necessary adjustments for the two experiments we conducted in March 2014: one at the Université Catholique de l'Ouest (Angers, France), and one at the University of Toulouse le Mirail (Toulouse, France). We will now describe the main aspects of the protocol designed for the experimentation.

4.1 Protocol

4.1.1 Participants

For both the pilot study and the two experiments, participants were all trainee translators[7].7 students from the Faculty of Translation and Interpretation of the University of Geneva took part in the pilot study. There were 4 Masters' students and 3 PhD students. As for the experiments in Angers and Toulouse, the participants (42 in total) were students in their final year of a Master's Translation program.

4.1.2 Translation Task

The participants were asked to translate a text from English into French (i.e. from L2 into L1 for most students). The text is around 150 words long, it is a popular-science text on volcanology entitled "Cinder Cones"[8]. It was chosen because it is well-structured (the two phases of the building of a cinder cone are described), because it contains a certain number of terms whose translation might be complex for a translator, even if they are not highly specialized (*basalt cinder cone*, *fountaining stage*…), and a number of syntactic patterns or collocations that are particularly tricky to translate (e.g. *bubble off*; transitive use of the verb *erupt*). Only one group out of the two was already a little familiar with the field of volcanology. The participants were allocated around 2 hours to translate the text, choose the relevant contexts and fill out an online questionnaire about the main translation difficulties and the use and usefulness of conventional resources and KRCs. Then several group interviews and a couple of one-to-one interviews were conducted.

4.1.3 Resources

Since we wanted the conditions of the experiment to be as close to a real-life context as possible for translators, [9] the participants had at their disposal the same type of resources as the ones they usually have when they translate in professional environment, i.e. various dictionaries and term banks. What made the experiment specific is that we added an extra resource, i.e. shortlisted contexts.

4.1.4 The Argos Interface

The participants used a customized interface, Argos, with four different windows (see figure 1): (*i*) one for the source-text, (*ii*) one for the target-text, (*iii*) one with several icons allowing access to term banks (*Termium*, *le Grand Dictionnaire Terminologique*), a specialized bilingual dictionary of earth sciences, a general bilingual dictionary, (*iv*) one with a list of shortlisted contexts.

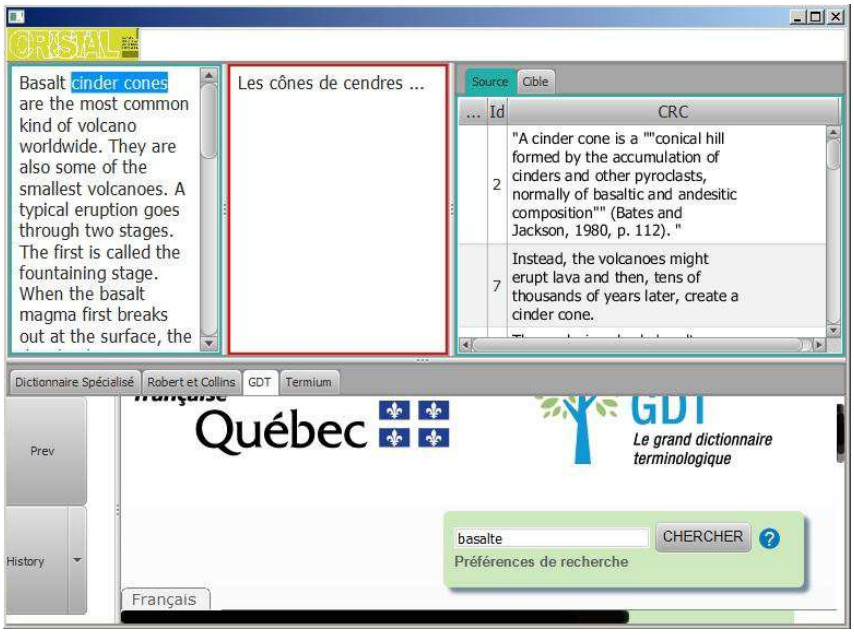

Figure 1: the Argos interface

4.1.5 The Contexts

Participants were provided with contexts in the source language, and contexts in the target language, which were

[7] This is the case in many empirical studies on translation: see for example Bowker 1998, Künzli 2001, and Varantola 1998.
[8] It was taken from *What's so hot about volcanoes?* by Wendell A. Duffield (2011), Mountain Press.
[9] This is what Ehrensberger-Dow & Massey (2008) call "ecological validity".

presented in random order. These contexts had been carefully chosen beforehand according to a classification devised by the team of linguists: details about the selection of contexts and the type of contexts provided will be discussed in the next section (4.2). During the translation process, the participants had to choose the contexts that had been most useful to them when translating by clicking on them. Once the translator had typed in a word (in the source language or the target language) in the search window for contexts, the target text window was blocked in order to ensure the translator chose at least one context, or explicitly chose that none was useful.

4.1.6 Extra Data Compiled

In addition to the final translations themselves and the answers to the online questionnaires, the data compiled comprise the following:

- screen-recordings performed through specific software[10]
- logs: all keyboard activity, as well as change of windows shifts, was recorded
- audio recordings of the two types of interviews[11] during which the participants were asked to give more detail about the usefulness of some given contexts.

4.2 The List of Contexts

The list of contexts provided to the participants was created with two principles in mind: we wanted to compile a large enough set of contexts, in order to limit the chance that the translator would look for contextual information on one particular word and get no results. At the same time, considering that the selection of contexts is a labor-intensive task, we wanted to limit ourselves to a reasonable number of words, and to keep only words for which our definition of Rich and Poor Contexts applies (see 3.3 above): in particular, the notion of Conceptually Rich Contexts (CRCs) is irrelevant for very familiar words that are not characteristic of the field of volcanology from which the text is drawn. This compromise was difficult to reach, so we took advantage of the pilot study to adjust and complement the list of terms that we had initially created.

4.2.1 Term Selection

For the pilot study in Geneva, we compiled a first set of contexts illustrating the use of:

- 7 words (noun, verbs and adjectives) in the source language, namely some lemmas from the text to be translated that we regarded as terminological units, or at least as words related to the field of volcanology (*basalt magma*, *blobs*, *cinder*, *cinder cone*, *fountaining*, *scoria*, *vesicles*)
- 11 words in the target language, selected among the possible equivalents of the corresponding source words.

We considered both simple and multiword units.
For example, contexts are provided for the word in bold type in the following sentence:

(3) As the **cinders** fall back to the Earth, they form layers that pile up into a cone-shaped hill.

One outcome of the pilot study was that the initial list of words proved to be very insufficient, especially in the target language: the logs compiled thanks to Argos (cf. 4.1.6) provided us with a much larger list corresponding to words that had actually been typed in the search window by the participants in order to get contexts. We thus decided to complement the first list with words that have been searched for by at least 2 participants. The final list for the experiments in Angers and Toulouse contains contexts for 22 English words and 41 French words.
In the same sentence as the one mentioned above, contexts were provided for four words (in bold type) instead of just one (i.e. 2 nouns, 1 verb, 1 adjective):

(4) As the **cinders** fall back to the Earth, they form **layers** that **pile up** into a **cone-shaped** hill.

4.2.2 Context Selection

The contexts were selected from several sources:

- we preferentially used a 800,000 word corpus of volcanology composed of specialized and popular science texts, which is used for the CRISTAL project as a whole,
- this source was complemented by a variety of web sites, giving priority where possible to texts dedicated to the presentation of volcanology to a wide audience.

We chose not to test the readability dimension of contexts (3.1.), so we only selected contexts that meet the criteria of readability (well-formed, not too long, with no anaphora elements, etc.). Contexts are one or two sentences long.
As explained before, our aim is to test whether the opposition between rich and poor contexts as defined in section 3 is relevant for the translation task. As a consequence, we annotated contexts according to this dimension. In figure 1, we give 3 examples illustrating (1) a linguistically rich context for the word *basalt* (with the presence of the term *basalt lava*),

[10] BBFlashback Express.
[11] In the one-to-one interview, the participant viewed part of the screen recording (the extracts corresponding to his translation of two terms for which contexts were provided: *cinder* and *fountaining stage*) and was asked to verbalize what he was doing, following the methodology used by Ehrenbersger-Dow and Massey (2008).

(2) a conceptually rich context providing a definition of the term, (3) a poor context. Note that linguistic and conceptual richness can combine in some contexts, which is not the case here. When possible, the conceptually rich contexts were classified into the following subcategories: definition, meronymy, hyponymy and co-hyponymy.

	Term	Rich or Poor	CRC	LRC	Type of CRC	Context
1	basalt	rich	no	yes	n/a	Shield volcanoes are made of thousands of thin basalt lava flows.
2	basalt	rich	yes	yes	def	Basalt is dark volcanic rock made up of small crystals and glass.
3	basalt	poor	no	no	n/a	When basalt enters water passively, it forms pillow basalt.

Figure 1: Examples of contexts for the word *basalt*.

We intended to balance the number of poor and rich contexts for each word. This proved impossible to achieve in many cases, since the great majority of contexts exhibit at least one relevant collocate. We collected 10 contexts per word, including no less than 2 poor contexts, totalling 222 contexts in English and 441 contexts in French.

4.3 First Results

We report here some preliminary observations about the results.
For the source language, 48% of the available contexts were chosen by at least one participant (108 contexts) *versus* 36% for the target language (152 contexts). This is an indication that the contexts are perceived as helpful, but the data are very dispersed, since about 40% of the selected contexts were chosen by only one participant in either language. Some terms are found several times in this list (*fountain*, *cinder cone*, *basalt magma*, and their French counterpart), showing specific translation problems.
To get a first picture of the results, we have chosen to focus on the 20 contexts that were selected most in either language. Each context was chosen between 4 and 14 times. The following examples show two of the most frequent ones.

(5) Hawaiian Eruptions are types of volcanoes and types of eruptions wherein basaltic lava is normally thrown up the air in jets. This process is called **fountaining**.
(6) During an eruption of gas-rich magma, small **blobs** of magma are ejected.

Example 5 is a conceptually rich context, more specifically a definition. Example 6, which contains several collocations (*blobs of magma*, *blobs ejected*), is a linguistically-rich context.

	Rich contexts	CRCs (definitions)	LRCs
Source language			
All available contexts	69%	25% (11%)	52%
20 most selected contexts	90%	65% (60%)	25%
Target language			
All available contexts	70%	29% (14%)	49.5%
20 most selected contexts	90%	55% (40%)	35%

Table 2: Distribution of the contexts

Table 2 makes a comparison between this subset of contexts and all the contexts that were made available. First, this shows that the great majority of contexts that are considered as helpful by the participants are rich contexts (90%). Second, participants show a strong preference for conceptually-rich contexts, mainly definitions, as opposed to linguistically-rich contexts. Yet we can observe that if this overall pattern applies to both languages, there are some differences: the distribution between LRCs and CRCs is different when the users are exploring source (English) and target (French) contexts. They seem to give a higher priority to CRCs and definitions in the source language. This is consistent with the idea that the CRCs should provide help for the comprehension of terms and LRCs should be more useful when checking the usage of the words in the target language.
These are encouraging results: they confirm our assumption that rich contexts are seen as helpful by the participants, and they suggest differences in the way the translator uses these contexts in the source and target language. However, these first observations must be confirmed and complemented by the analysis of the entire set of data and the analysis of the replies to the questionnaires.

5 Conclusion

The notion of context is where lexicography, terminology and translation meet. Even though the specific needs of translators regarding their resources now seem quite clearly identified, addressing them still seems quite challenging. We hope the findings of the CRISTAL project will help tailor the tools according to the translator's profiles in one aspect, that of contextual information.
To fulfil that objective, we plan to explore in detail the considerable amount of data we have collected (around 60 hours of screen recordings, and just as much structured translation logs). The evidence gathered will enable us to answer the following questions:

- apart from definitions, do some sub-categories of KRCs play a specific role in the translation process?
- in which precise situations do translators give preference to KRCs over conventional resources such as

monolingual or bilingual dictionaries?

- what is the impact of the use of KRCs and the other resources on the translation quality of the 49 final translations

The main challenge the CRISTAL project plans to address in the near future is to devise a method to automatically retrieve the 'good contexts' whose main features will have then been identified.

6 References

Atkins, Beryl T. Sue. (1996). "Bilingual Dictionaries: Past, Present and Future." In: Euralex'96 Proceedings, Martin Gellerstam et al., ed. Gothenburg: Gothenburg University. 515-590.

Baker, M. (2006). Contextualization in translator- and interpreter-mediated events. *Journal of Pragmatics*, 38(3), pp. 323-337.

Bowker, L., (1998). « Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study ». *Meta*, 43(4), pp.631-651.

Bowker L., (2011). "Off the record and on the fly," *Corpus-based Translation Studies: Research and Applications* (Eds. A. Kruger, K. Wallmach and J. Munday). London/New York: Continuum, pp. 211-236.

Bowker L., (2012). "Meeting the needs of translators in the age of e-lexicography: Exploring the possibilities", in S. Granger, M. Paquot (eds) *Electronic Lexicography*, Oxford University Press, pp. 379-387.

Condamines, A., Rebeyrolle, J. (2001). Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CTKB): Method and Results. In D. Bourigault, C. Jacquemin & M.-C. L'Homme *Recent Advances in Computational Terminology*, John Benjamins, pp.127-148.

Durán Muñoz, I. (2010). Specialised lexicographical Resources: a Survey of Translators' needs. In S. Granger & M. Paquot (eds.), *eLexicography in the 21st century: New challenges, new applications, Proceedings of Elex2009. Cahiers du Cental*, vol. 7. Louvain-la-Neuve. Presses Universitaires de Louvain, pp. 55-66.

Durán Muñoz, I. (2012). Meeting translators' needs: translation-oriented terminological management and applications. *Journal of Specialised Translation* (18), pp. 77-92.

Ehrensberger-Dow, M., Massey, G. (2008): Exploring Translation Competence by Triangulating Empirical Data. In *Studies in Translation*. 16, 1–20.

Halliday, M.A.K. (1999). "The notion of Context in Langue education). In M. Ghadessy (ed.) *Text and Context in Functional Linguistics*. Philadelphia: John Benjamins Publishing Company.

Josselin-Leray, A. (2005). *Place et rôle des terminologies des terminologies dans les dictionnaires unilingues et bilingues. Etude d'un domaine de spécialité : volcanologie*. Unpublished PhD thesis, Université Lyon 2, France.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. & Rychlý, P., (2008). GDEX: Automatically finding good dictionary examples in a corpus. In *proceedings of XIII EURALEX Congress*, E. Bernal & J. DeCesaris (Eds), Barcelona, Universitat Pompeu Fabra, pp.425-431.

Künzli, A. (2001). Experts vs. novices. L'utilisation de sources d'information pendant le processus de traduction. In *Meta*, 46, pp. 507-523.

Marshman, E., J., Gariépy & Harms, C. (2012). Helping translators manage terminological relations: Storing and using occurrences of terminological relations and lexical relation markers. *JoSTrans: Journal of Specialised Translation* 18, 30-56.

Melby, A.K. (2010). Context in translation: Definition, access and teamwork. *Translation & Interpreting*, 2 (2), pp. 1-15.

Meyer, I., (2001). Extracting Knowledge-Rich Contexts for terminography: A conceptual and methodological framework. In D. Bourigault, C. Jacquemin & M.-C. L'Homme *Recent Advances in Computational Terminology*, John Benjamins, pp. 279-302.

Reimerink, A., Garcia de Quesada, M. & Montero-Martinez, S. (2010). Contextual information in terminological knowledge bases: a multimodal approach, *Journal of Pragmatics*, 42(7), pp. 1928-1950.

Rey-Debove, J., (2005). Statut et fonction de l'exemple dans l'économie du dictionnaire. In M. Heinz (eds), *L'exemple lexicographique dans les dictionnaires français contemporains*. In Proceedings of the « Premières journées allemandes des dictionnaires » (Klingenberg am Main, 25-27 juin 2004). Max Niemeyer Verlag, Tübingen, pp.15-20

Roberts, R. P. (1994). Bilingual Dictionaries Prepared in Terms of Translators' Needs, *Proceedings of CTIC 3rd Conference, Translation in the Global Village*, Banff, CTIC, pp. 51-65.

Roberts, R.P. & Bosse-Andrieu J. (2006). Corpora and Translation. In Bowker L. (ed.), *Lexicography, Terminology, and Translation – Text-based Studies in Honour of Ingrid Meyer.* Ottawa : Presses de l'Université d'Ottawa, pp. 201-214.

Rogers M. & Ahmad K. (1998). The Translator and the Dictionary: Beyond Words? In: B.T. Sue Atkins (ed.) *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators.* Tübingen: Niemeyer (Lexicographica. Series Maior), pp.193-204.

Siepmann, D. (2005). *Discourse markers across languages: A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. Routledge.

Varantola, K. (1994). The Dictionary User as Decision Maker. Sixth Euralex conference proceedings, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands, pp. 606-611.

Varantola, K. (1998). Translators and their Use of Dictionaries. In B.T.S. Atkins (ed), *User Needs and User Habits. Studies of Dictionary Use by Language learners and Translators*, Tübingen: Niemeyer (Lexicographica. Series Maior), pp. 179–192.

Varantola, K. (2006). The Contextual Turn in Learning to Translate. In Bowker L. (ed.), *Lexicography, Terminology, and Translation – Text-based Studies in Honour of Ingrid Meyer.* Ottawa : Presses de l'Université d'Ottawa, pp. 215-226.