

Krista Varantola
Department of Translation Studies, University of Tampere

The Dictionary User as Decision Maker

Abstract

My paper deals with the type of lexical contextual information translators need when translating into L2. I argue that the best way to obtain this information is to use a separate, well-planned corpus text base linked to an electronic dictionary. The corpus examples will give the translator an idea of a word's semantic range and potential so that decisions can be made about the most suitable translation for any particular context.

In this way text corpora can be used to bridge the gap between the general-purpose, context-free dictionary information and the highly context-sensitive needs of the user.

1. The setting

In this article I shall discuss ways in which translators can benefit from the use of electronic textual data in addition to traditional dictionary information. The translators I have in mind are those who translate or produce text in a foreign language, i.e. their L2. I am dealing with advanced translators whose command of the L2 is very good and whose problems, therefore, are often very subtle and difficult to pin down to simple queries. From a different angle, it could be said that my emphasis is on how to bridge the gap between the relatively context-free information in dictionaries and the highly context-sensitive needs of the dictionary users.

Ideally, it has been said, translators should only translate into their native language and this is normally the case in literary translation. In non-fiction translation the situation is, however, very different. Translators not only translate texts into their L2, they also produce texts in L2 that are only loosely based on material provided in L1.

I shall not discuss in this context whether this practice is advisable or not and what should be done in the process to ensure the best possible outcome. I shall only note that translating into L2 is a fact of life for many translators, particularly those whose mother tongue is a so-called minor language.

2. Dictionaries as information sources

If we think of translators as dictionary users, they naturally need various types of lexical information in their work. At its simplest, it is a clearly definable L2 equivalent that is needed. At the other end of the scale it is a more or less indeterminable stretch of coherent language in a special context.

To find the lexical information they need, translators have at their disposal very different types of sources; bilingual dictionaries, monolingual learners' and native speakers' dictionaries, combinatory dictionaries, thesauri, new types of production dictionaries (e.g. the Longman Activator), encyclopedic sources etc.

The main problem for the translators, in addition to finding answers to their queries, is how to be sure that the decision they make, on the basis of the information provided, is correct and acceptable for the context. In other words, L1–L2 translators also need to find a great deal of supportive information in the dictionary source to compensate for their lack of native language competence in L2. Paradoxically, the more competent the L2 users are, the more critical and choosy they tend to be about the information they receive.

How could dictionaries then guarantee user satisfaction? Most dictionaries are still available in paper format only but the situation is changing. With the electronic format, the ways of querying multiply and with a resourceful query language the potential of the source becomes more fully and systematically available. The basic dilemma, that of context-free information in the source vs. the context-sensitive needs of the user, however, remains.

3. The dilemma

Typically, translators have always been looking for “parallel texts”, in addition to lexical dictionary information. Parallel texts, i.e. texts representing the same text type and dealing with the same subject matter in the target language, have been considered essential for a number of reasons. In special fields, they are used to get confirmation that the terms found in a special field dictionary are actually in use and mean what they are supposed to mean. More generally, they are used to get information about style, idiomatic usage, combinatory properties of words and also modern vocabulary and usage that has not yet reached the dictionaries.

The typical complaint about dictionary entries, on the other hand, is often that they do not give enough usage examples. They only give examples of what are perceived as prototypical and frequent uses. As a non-native speaker, however, you cannot be sure whether the prototypical uses can be extended to apply in the particular context you have in mind. In other words, as a translator you would like to make the decision on the basis of the information provided in the dictionary entry, but the examples are not varied or elaborate enough to give you the information about the semantic range of the expression that you would need to feel confident about your decision.

What are the alternatives if you do not find an appropriate example in a dictionary? You can take the risk, trust your intuition and use the word or collocation you have in mind or opt for something more secure even if it does

not sound as good as your first choice. Whatever happens, as a user, you are not satisfied with the answer that the dictionary has given to your query.

4. Text corpora as a solution

Could text corpora be used to bridge the gap between the general-purpose aims of the dictionary and the specific needs of the user?

Text corpora certainly help the lexicographers to choose more appropriate and illuminating examples. It is another matter whether the examples in the dictionary should remain "authentic" or whether they should be further modified by the lexicographers to better suit the user needs (Cf e.g. Minaeva 1992 or Laufer 1992). Rundell and Stock (1992:14) compare large computer corpora and citation banks and argue that corpora are strong on the prototypical uses whereas individually collected citations give equally important information on unusual and innovative uses.

Should electronic dictionaries then give an unlimited number of usage examples to suit every possible user need? It can certainly be argued that there is no point in giving the user an unsorted collection of examples. It is the lexicographer's duty to crystallize the corpus evidence into well-chosen examples and to give the user a good breakdown of the lexical item's behaviour, both its prototypical behaviour and its range, as well as its semantic mobility. But if this succinct breakdown is not enough for the L2 user, how could the text corpus then give more cues for the decision-making process?

5. Corpus information

Let me give a few examples of the type of information that advanced users could glean from large or customized corpora when typical dictionary entry information is not enough.

The word *cellular* is a prolific collocator in present-day English. According to dictionary examples, it is typically used together with *phone*, *telephone* and *radio*. In journalism, however, its use has greatly expanded recently and you can find it collocating with *technology*, *business*, *communication*, *counterparts*, *manufacturers* and perhaps less predictably with *giant* and *time*. *Cellular* has obviously become part of the general vocabulary and, at the same time, lost some of its semantic stringency and become more fuzzy-edged. It would, however, be difficult to capture its new semantic potential in a dictionary entry. On the other hand, an up-to-date, journalistic text corpus could implicitly give this information.

In a more specialised context, when compiling a glossary on EC terms in connection with the huge translation project of EC directives, it was noticed at TNC (The Swedish Centre for Technical Terminology) that translators were not only interested in the actual terms but also in learning the standard equivalents of multiword expressions like *temporary employment*

relationship, open market economy with free competition and formulaic stretches of language such as *by way of derogation from*. In special and standardized contexts, it is important even in L2–L1 translation to know what the accepted “norm” translations are for various frequently occurring word combinations even if, strictly speaking, they do not qualify as special terms in the field.

Other examples of multiword expressions that have made it to the English–Swedish glossary are: *be rendered operational, be of analytical purity, commitment by way of guarantee, declare contrary to, have an infringement established, immediate and full recognition, operate under a flag of convenience, suspension of the exercise of the voting rights, take redressive action against, undisclosed aid from public funds*.

This kind of information could also be made available in special parallel text corpora of EC texts and it is clear that the best way of searching for equivalents for these types of units would be an electronic parallel text base with a sophisticated concordancing program.

The prototype version of COBUILD on CD–ROM had, as one of its databases, a text corpus of five million words (word bank). Unfortunately, in the prototype version, it was not possible to search the word bank for collocations unless the collocation was also a dictionary entry in the COBUILD dictionary. The system has a fixed starting point in the main dictionary and the users cannot therefore tailor their searches according to their needs but must follow an established search hierarchy and search routines. For example, it is not possible to find usage examples of *information age* because it is not an entry in the dictionary although there are usage examples of this collocation in the word bank corpus. Unfortunately, the final version has the same weakness. It has not been considered necessary to make the database more user–friendly and leave the decision about what the user wants to look up – to the user.

I followed up, by means of a corpus use experiment, the students’ dictionary look–up processes after a dictionary use experiment that was conducted at Tampere University (cf. Varantola & Atkins, in preparation). I checked in the Oxford MicroConcord corpora of academic and newspaper texts (altogether about 2 million words) whether the students could have found the answers to their questions from concordanced corpus examples.

For example, a number of students wanted to know whether they could use the phrase *information society* in an L1–L2 translation. The corpus did not provide any examples of *information society* but had a few examples of *information age* that could have helped them.

Some students hesitated between *aches* and *pains* in the combination *neck and shoulder* The dictionaries did not help them but corpus examples would have told them that *aches* is rarer and usually occurs only in the fixed phrase of *aches and pains* whereas there were a number of examples of *pain* in combinations like *National Back Pain Association, pain control, pain killer* etc. Quite a few students wondered whether they could use the phrase

environmental hygiene as a name of a university study programme. They could not have found a direct answer in the corpus but they could have found expressions like *London School of Hygiene, professor of hygiene* and examples of *environmental* collocating with *organization, policy, services, health, sanity, safety* etc.

6. The lesson

The corpus examples naturally give much more information about the range of a word or phrase, about typical genre- or style-based preferences than is possible in a dictionary entry. For translators, the existence of a well-balanced and versatile text corpus, in addition to traditional dictionaries and dictionary information, would be a welcome help in their decision making process. They would no longer need to look up detailed context-sensitive information in a dictionary entry but would have a more appropriate source at their disposal for these types of queries. Now they feel frustrated from the start because they know that it is very unlikely that "their example" is included in a typical dictionary entry. On the other hand, lexicographers would feel better because users would no longer expect them to satisfy all their lexical needs and guess in advance the various possible starting points of look-up processes. When necessary, the textual database would provide the context-dependent lexical information for the users. The translators would save time because they could go straight to the text corpus with their queries of a word's collocational and semantic behaviour. In this way they could also have their hunches confirmed or refuted, either directly or through collating the data from a number of examples.

Ideally then, the translators would, in addition to typical dictionary data, have recourse to text corpora that could be sorted according to a variety of criteria such as domain, style, text type and also concordanced in sophisticated ways.

Now that a number of corpora of various types and sizes are becoming available to individual users it is certainly worthwhile to study how text bases could help the translators or writers to make up their minds in L2 and be confident about their choices. The key notion is that when decisions are made at this level of competence, the actual decision making should be left to the database users, but the data should provide versatile and adequate cues so that the users could make soundly based choices for the particular textual contexts they are working with.

References

- Laufer, B. 1992 "Corpus-based versus lexicographer examples in comprehension and production of new words". In H. Tommola & K. Varantola, T Salmi-Tolonen & J Schopp (Eds.) *EURALEX '92 Proceedings I-II:71-76*
- Minaeva, L. 1992. "Dictionary Examples: Friends or foes?" In H. Tommola & K. Varantola, T Salmi-Tolonen & J Schopp (Eds.) *EURALEX '92 Proceedings I-II:77-80*.
- Rundell, M. & P. Stock. 1992 "The corpus revolution". *English Today* 30:9-14.

- TNC. 1993. *EGs ord och uttryck. EC words and expressions. Engelsk-svensk ordlista över EGs regelspråk*. Solna.
- Varantola, K. 1993. "Technical vocabulary and learners' dictionaries". In *English Far and Wide: A Festschrift for Inna Koskenniemi*. *Annales Universitatis Turkuensis*, B:197, 249–260.
- Varantola, K. and B.T.S. Atkins. "Steps in the dictionary look-up process". In preparation.