

## Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study

Lynne Bowker

Volume 43, Number 4, décembre 1998

L'approche basée sur le corpus  
The Corpus-based Approach

URI: <https://id.erudit.org/iderudit/002134ar>  
DOI: <https://doi.org/10.7202/002134ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)  
1492-1421 (digital)

[Explore this journal](#)

Cite this article

Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta*, 43(4), 631–651.  
<https://doi.org/10.7202/002134ar>

Article abstract

This article reports on the results of an interesting experiment comparing two translations produced by a group of translator trainees. One translation was carried out with the use of conventional resources; the other with the aid of a specialised monolingual corpus. The results reveal that the corpus-aided translations were of higher quality in respect to subject field understanding, correct term choice, and idiomatic expression. The author observes that although she did not find any improvement with regard to grammar or register, the use of the corpus was not associated with poorer performance.

# USING SPECIALIZED MONOLINGUAL NATIVE-LANGUAGE CORPORA AS A TRANSLATION RESOURCE: A PILOT STUDY

LYNNE BOWKER  
*Dublin City University, Dublin, Ireland*

## **Résumé**

*Cet article rend compte des résultats d'une expérience comparant deux traductions effectuées par des apprentis traducteurs. La première traduction a été faite à l'aide d'outils conventionnels, alors que pour la seconde l'outil consistait en un corpus monolingue spécialisé. Les résultats montrent que les traductions réalisées à l'aide du corpus sont de meilleure qualité en ce qui a trait à la compréhension du domaine, à la sélection des termes et à l'utilisation d'expressions idiomatiques. L'auteur observe que, bien qu'elle n'ait pu noter d'amélioration côté grammaire et registre, l'utilisation du corpus ne peut pas non plus être associée à une baisse de la qualité du travail.*

## **Abstract**

*This article reports on the results of an interesting experiment comparing two translations produced by a group of translator trainees. One translation was carried out with the use of conventional resources; the other with the aid of a specialised monolingual corpus. The results reveal that the corpus-aided translations were of higher quality in respect to subject field understanding, correct term choice, and idiomatic expression. The author observes that although she did not find any improvement with regard to grammar or register, the use of the corpus was not associated with poorer performance.*

## **INTRODUCTION**

It is generally held that three of the most important criteria required to produce a high-quality translation are an understanding of the subject field, an excellent command of the target language, and a good knowledge of the source language. In this paper, we outline a pilot study that was conducted to determine if a specialized monolingual native-language corpus can be used as a resource to help student translators improve two of these competencies: subject-field understanding and specialized native-language competence. In part one, we briefly discuss why subject-field understanding and native-language competence are important for translation, and we consider the problems that arise when students do not have these competencies. In part two, we look to the discipline of corpus linguistics to see if it can provide any solutions for improving these competencies. In part three, we describe a pilot study which we conducted with 14 students at the School of Applied Language and Intercultural Studies at Dublin City University in Ireland: translations done using conventional resources were compared with those done using a specialized monolingual native-language corpus. In part four, we discuss the results of the experiment.

## 1. TRANSLATION COMPETENCIES

Good translators are generally viewed as having at least the following three competencies: a solid understanding of the subject matter treated in the source text; an excellent command of the target language (usually the translator's native language); and a good knowledge of the source language (usually a foreign language) (Sykes 1989: 35-39). In this study, we will consider only the first two of these skills (i.e. subject-field understanding and native-language competence), with a view to identifying the potential of a specialized monolingual native-language corpus as a resource for helping translators to improve these skills.

### 1.1. Subject-field Understanding

The fact that a translator needs to understand the source text in order to produce an accurate translation is obvious to anyone in the profession. Translators cannot simply translate words; one of the things they need in order to understand the text, and thereby translate it, is knowledge about the subject field. The importance of understanding the subject field has been acknowledged by many translation researchers (e.g. Folkart 1984: 230; Cormier 1991: 440). Miller (1993: 2) makes the following observations about a translator's need for subject-field understanding:

Translators are knowledge users [...] they need background information about the subject matter of the source text; given that they are not the audience whom the author had in mind when composing the original text, they probably do not have the specialized knowledge that was assumed of the intended readership.

### 1.2. Native-language Competence

With regard to target-language command, it is generally accepted that translators should translate out of their foreign languages and into their native language (e.g. Newmark 1988: 3; Baker 1992: 65).<sup>1</sup> This is one of the main reasons that many translator-training institutes train translators to translate out of foreign languages and into their native language. Paradoxically, however, the curricula at many of these institutes contain relatively few courses designed to help students refine or enhance their native-language skills.

### 1.3. When Competence is Lacking...

As a lecturer in translation at Dublin City University (DCU), I spend a lot of time correcting student translations, and I encounter many errors that result from both comprehension and production problems. For the most part, the comprehension errors result from a lack of understanding of the subject field, rather than a lack of understanding of the foreign language. A considerable number of production errors also occur; these errors result from a lack of competence in the specialized target language, which, in the case of my students, is always their native language (English).<sup>2</sup>

This raises the difficult question of how to resolve the situation. Translator training institutes are limited in what they can accomplish with regard to helping students become subject-field experts. Few translators have the luxury of being able to restrict themselves to translating texts in a single, sharply defined subject field. Given the infinite number of possible subjects which may require translation, it is a relatively common scenario for translators to be handed a text in an unfamiliar subject area, requiring them to become "instant mini-experts" by quickly acquiring a working knowledge of the key concepts in the field. This is further complicated by the fact that many texts cover multiple disciplines. It is therefore inconceivable for translator training institutes to try to teach students about all or even many subject fields; instead, we can only hope

to pass along some sound advice which stresses the importance of acquiring background knowledge about the subject field before beginning to translate.

With regard to improving native-language competence, the most obvious solution would seem to be to incorporate more courses in native-language skills into the program; however, this is not always pragmatically feasible. At DCU, students already have very full timetables: they are required to achieve competence in two foreign languages, and they spend the third year of their four-year program following courses at a foreign university. This leaves very little room for adding in extra courses in English-language skills.<sup>3</sup>

If some of the skills needed by our students cannot be easily built into the curriculum, we must try to find another way to assist them. One possible solution is to provide them with a resource<sup>4</sup> which can help them acquire background knowledge and provide them with examples of correct usage in context. To this end, we turned to the discipline of corpus linguistics.

## 2. CORPORA AND TRANSLATION STUDIES

The use of corpora in the discipline of translation studies is enjoying increasing popularity. Van Doorslaer (1995) investigates the requirements for compiling corpora that will be useful for descriptive translation studies. Researchers such as Baker (1996), Laviosa (1998) and Kenny (this volume) are using corpora to study the phenomenon of translation by carrying out descriptive studies which have resulted in the identification of a variety of universal features of translated texts. McEnery and Wilson (1993) discuss the potential of corpora in applications such as machine translation and other areas of natural language processing. Malmkjaer (this volume), Peters and Picchi (1998), and Danielsson and Ridings (1996) consider the potential of bilingual and parallel corpora as a source of translation equivalents.

While descriptive translation studies and computer applications for corpora are interesting and important areas of research, they cannot directly help us to remedy the problems outlined above.<sup>5</sup> The merits of using bilingual and parallel corpora for translation purposes are obvious; however, the problem here is that there are relatively few corpora of this type available (certainly not enough to cover the wide variety of subject fields that translators have to deal with), and it is more difficult to create this type of corpus because it needs to be aligned.<sup>6</sup> Therefore, we decided to conduct a pilot study to see whether a specialized monolingual native-language corpus, which is much simpler to compile, could be a useful translation resource.

## 3. PILOT STUDY

This pilot study consists of a translation experiment using a specialized monolingual native-language corpus that was carried out at Dublin City University in March 1997.

### 3.1. Hypotheses

We hypothesize that a monolingual native-language corpus comprised of texts in a specialized subject field can be a useful resource for translators translating into their native language. We consider it will be more useful than conventional resources (e.g. specialized monographs, journals, or monolingual dictionaries) in the following ways:

- a) *It will help translators to increase their understanding of the subject field and thereby make fewer errors resulting from lack of subject comprehension.*

Conventionally, translators have been encouraged to do some background reading in order to become familiar with a subject field before starting the translation. In this respect, an electronic corpus offers several advantages over a conventional printed corpus.

Firstly, it is often time-consuming to physically gather together a printed corpus (trips to various libraries and documentation centres, hours at the photocopier, etc.), with the result that translators often have time to gather only a handful of documents. An electronic corpus containing hundreds or even thousands of documents can often be compiled in a shorter time, thus giving the translator more information to consult. Furthermore, once compiled, an electronic corpus has the advantage of being reusable for other translations in the same subject area.

Once gathered, it takes longer to consult a conventional corpus. As observed by Miller (1993: 8):

Popular science texts and magazines are sometimes helpful in providing explanations of key concepts in a given field. Specialized publications — including books, textbooks, journals, manuals and papers — can be valuable sources of specific technical information. The main problem with these latter sources is determining and then locating the most appropriate items. The bibliography or references contained in the source text can sometimes help narrow down the search, but all too often the translator still finds himself rapidly scanning technical material, hoping to stumble onto a discussion of a relevant point.

In contrast, translators using an electronic corpus coupled with a corpus analysis tool can go directly to areas of text containing key words.<sup>7</sup> They can then read as much or as little of the discussion as desired — from a sentence, to a couple of paragraphs, to the entire text. Moreover, as mentioned above, they have a larger selection of texts available for consultation, so if one explanation is not particularly helpful, they can quickly move on to the next.

It is worth noting that specialized dictionaries are generally easy to find and consult; however, a common drawback here is that concepts are often treated in isolation, and it is difficult for the non-expert (i.e. the translator) to draw out the implicit relations between them in order to get an overview of the subject field as a whole.<sup>8</sup> In a cohesive text, such as that found in a corpus, it is generally easier to understand how the concepts in a subject field relate to one another.

**b) *It will help translators to make fewer errors related to target language production.***

Something often overlooked is that it is not enough for translators to understand the concepts and identify the equivalent terms: they also need information about how to properly use specialized terminology. One of the simplest and most useful types of usage information is collocational information. Collocations are combinations of words that, according to the conventions of a given language, are habitually associated (Sinclair 1991: 170). Some specialized dictionaries do provide limited collocational or contextual information; however, the translator inevitably wants or needs more. If translators really want to know what the usage conventions are for a term in the domain they are working in, professional literature is the best source for supplying the answer.

Putting a word in context means breathing life into it. Taking a word out of context is like stuffing an animal. If you want to know something about animals you may learn a lot from looking at a stuffed specimen. You may even learn more by dissecting it, but if you want to know about the behaviour of animals you must study them in their natural environment. If you want to know how words behave you must study them in their natural environment too, and the natural environment of words is text, context. Finding information in "live"

text means plodding through reams of books, newspaper, and more than anything else, professional literature. Not the kind of task that combines well with completing a translation on a tight deadline. (Roumen and van der Ster 1993: 215-216)

In addition to being time consuming, detecting linguistic patterns is actually more difficult when working with a conventional corpus: the translator may simply not notice a pattern when its occurrences are spread over several pages, or even several documents. The concordancing feature of a corpus analysis tool (see section 3.5), however, can quickly bring together all occurrences of a given pattern. Moreover, through statistical information, corpus analysis tools can also provide an indication as to the centrality of the pattern (i.e. whether it is just one author's idiosyncratic usage, or a well-accepted pattern in expert discourse).

### 3.2. Participants

When choosing participants for our experiment, we had several criteria in mind:

- a) We are involved in translator training and one of the aims of this study is to try and improve the quality of our students' translations; therefore, we wanted the participants to be students.
- b) However, because we wanted to test the potential of the corpus for improving specific skills (i.e. subject-field understanding and specialized native-language competence), we felt the participants should not be entirely novice, but rather should be familiar with the basic principles of translation and should have followed several courses in translation practice and specialized translation.
- c) Because one of our objectives is to see if the corpus can help translators improve their specialized native-language competence, we wanted all the participants to be native speakers of English.
- d) Because the experiment is not intended to help students improve their foreign language skills, we wanted to ensure that the participants already had a reasonably good grasp of the source language (in this case, French). We hoped that this would minimize the number of errors caused by linguistic comprehension problems.
- e) In order to reflect what we consider to be the prototypical situation for most translators, we felt that the participants should be comfortable with the general subject area, but should not be true experts on the specific subject matter of the text (see section 3.3).
- f) Because part of the experiment involved using an electronic corpus and corpus analysis tool, we wanted the participants to be computer literate and able to manipulate a corpus using a corpus analysis tool.

With these criteria in mind, we asked fourteen students from the Applied Computational Linguistics programme at DCU to participate in the experiment. We felt these students would be ideal candidates for the experiment for the following reasons: all of the students were native speakers of English; all were in their final year of undergraduate studies and had followed a minimum of 3 courses in translation and 6 courses in French language; all had spent a year at a university in either France or Belgium; all were familiar with the general field of computing, but none were experts in the subfield of optical scanning technology; all had followed a course in corpus linguistics where they had learned to manipulate corpora using a corpus analysis tool.

### 3.3. Texts

Similarly, we had several criteria in mind when choosing the texts for translation.

- a) We felt that the texts should treat a subject field that falls within both our general area of interest and expertise (because we had to evaluate it), and also that of the students (as described above).
- b) We felt that the texts should be specialized, but not highly technical because the participants were students and not professional translators.
- c) Because we were relying on the good will of the students who volunteered their time to participate, we felt the texts should cover a subject coherently without exceeding 300 words in length.

Bearing these criteria in mind, we chose two extracts from an article on optical scanners which appeared in the French journal *Science et Vie Micro* in January 1991. This is a popular science magazine so although its contents are specialized, they are not highly technical. The first text (282 words) was a description of the way that a flatbed scanner operates. The second text (301 words) contained a brief description of three different types of scanners, and a discussion about the advantages and disadvantages of each. We felt these texts were suitable because they dealt with a general area with which the students were comfortable and familiar (i.e. computing), but it was not a subject that they had dealt with specifically in either their computing or translation classes, so the concepts, terms, and discourse style would not be immediately familiar to them, and therefore we anticipated that they would have to make use of the resources provided. Moreover, optical scanning is an area in which we have conducted terminological research in the past (Bowker 1995), so we felt competent to evaluate their work. The source texts used in the experiment can be found in appendix A.

### 3.4. Resources<sup>9</sup>

Any issues relating to the students' foreign-language skills were beyond the scope of this study; therefore, all students, whether using the conventional monolingual resources or the corpus, were allowed to use general bilingual dictionaries during the experiment. The dictionaries provided included Robert-Collins, Oxford-Hachette, and Harrap's.

#### 3.4.1. Conventional monolingual resources

The conventional monolingual resources which we allowed the students to consult included both lexicographic and non-lexicographic resources.

##### 3.4.1.1. Lexicographic resources

- Oxford English dictionary;
- five specialized dictionaries and glossaries relating to computing.

##### 3.4.1.2. Non-lexicographic resources

- an encyclopaedia of electronics;
- a user manual for an optical scanner;
- an article on scanners and optical character-recognition from a popularized computer journal;
- a monograph on desktop publishing.

### 3.4.2. *Electronic corpus*

Our electronic corpus was extracted from a series of commercially available CD-ROMs called *Computer Select* (Ziff-Davis Publishing, Computer Library, NY).<sup>10</sup> Each disc contains English-language articles from several hundred publications dealing with a wide range of computer-related topics.<sup>11</sup>

In Bowker (1996), we discussed several criteria that must be met in order to compile a specialized corpus that is balanced and representative for the purposes of terminology, and we feel that this can also be extended to the case of specialized translation. Table 1 outlines these criteria and evaluates the potential of *Computer Select* for meeting them. The potential of *Computer Select* for fulfilling the specific criterion *text type* is further elaborated in tables 2 and 3.

Although *Computer Select* does not perfectly meet all the criteria required to make a corpus a completely balanced and representative source for the purposes of specialized translation, we feel that it meets enough of these criteria to justify using it as a source for our corpus. As pointed out by Atkins *et al.* (1992: 6):

... we have found any corpus — however "unbalanced" — to be a source of information and indeed inspiration. Knowing that your corpus is unbalanced is what counts. It would be short-sighted to wait until one can scientifically balance a corpus before starting to use one, and hasty to dismiss the results of corpus analysis as "unreliable" or "irrelevant" simply because the corpus used cannot be proved to be balanced.

#### 3.4.2.1. Compilation procedure for the electronic corpus

As mentioned above, each *Computer Select* disc comprises thousands of articles taken from hundreds of journals dealing with a wide range of computer-related topics. All the articles in *Computer Select* are indexed according to key words.

We had previously done some terminology work in the subject field of optical scanners (Bowker 1995), so as a first step to compiling our corpus, we consulted our previous research to see what important terms we had identified in the field. Next, we used these terms as key words, instructing the computer to select all articles indexed according to these key words and to concatenate them into one file.<sup>14</sup> We repeated this step for each of the discs spanning the time period May 1989 to February 1995. As illustrated in table 4, the end result was four files which contained a total of approximately 1.5 million words and constituted almost 10MB of data.

### 3.5. **Corpus Analysis Tool**

The corpus analysis tool used in our experiment was WordSmith Tools, developed by Mike Scott. WordSmith provides the following features which we felt could be of use to translators:

- a) a concordancer, which finds and displays, in an easy-to-read format (e.g. KWIC — key word in context) all occurrences of a search term (and minor variations thereof);
- b) a collocation viewer, which allows users to see which words "go together;"
- c) frequency operations, which provide statistical information about the centrality of a pattern (i.e. whether it is just one author's idiosyncratic usage or an accepted pattern in expert discourse).

The students had been given training on how to use the tool during their corpus linguistics class, so they were familiar with these basic features; however, they had never previously tried to apply the tool to a translation situation.



<b>Criteria</b>	<b>Requirements for specialized translation</b>	<b>Positive features of <i>Computer Select</i></b>	<b>Drawbacks of <i>Computer Select</i></b>
<b><i>Corpus size</i></b>	large selection of texts (to determine if usage is widespread or idiosyncratic)	contains thousands of articles	
<b><i>Text size</i></b>	complete texts (so examples of usage or explanations of concepts are not cut short)	many articles are complete	some articles are only abstracts and so do not fully explain some concepts
<b><i>Text type</i></b>	mixture of instructional, expert, and popularized texts (to help translators achieve an understanding of the subject field and allow them to see different registers of usage)	contains a wide range of expert and popularized texts with a few instructional texts <sup>12</sup>	the number of instructional texts is relatively low
<b><i>Date of publication</i></b>	mainly recent texts, but also some older texts (older texts are useful because concepts are better explained when they first come out; recent texts are needed to reflect the state of the field at present)	contains a mixture of relatively current and older texts (May 1989 - Feb. 1995)	
<b><i>Author</i></b>	texts by a variety of authors (to determine if usage is widespread or idiosyncratic)	contains texts by thousands of different authors	
<b><i>Language</i></b>	preferably texts by authors writing in their native language (to show idiomatic usage)	contains only English-language articles which come primarily from journals originally published in English-speaking countries	cannot easily verify if the texts are written in the author's native language
<b><i>Culture</i></b>	texts written by authors with different cultural backgrounds (e.g. British, American, etc.) (to show appropriate regional usage)	most articles come from journals originally published in the US, though some are from the UK and Canada; one journal is originally from the Netherlands	most articles seem to be of US origin, cannot easily verify the cultural backgrounds of the authors

**Table 1**

An evaluation of *Computer Select* with regard to the criteria required for a corpus that is a balanced and representative source for specialized translation

Type of Publication	Instructional	Expert	Popularized
<b><i>Learned Journals</i></b> (e.g. AI Expert, IBM Journal of Research and Development)		X	
<b><i>Proceedings</i></b> (e.g. Proceedings of the IEEE)		X	
<b><i>Popular Journals</i></b> (e.g. Byte, MacUser, PC Magazine)			X
<b><i>Newspapers</i></b> (e.g. The New York Times, The Wall Street Journal)			X
<b><i>Newsletters/Bulletins</i></b> (e.g. Information Industry Bulletin, Soft*letter)			X

**Table 2**

The five basic types of publications found in *Computer Select* are classified according to the type of information they are intended to impart

Genres of texts	Instructional	Expert	Popularized
<b><i>Buyers' Guides</i></b>			X
<b><i>Columns</i></b>			X
<b><i>Company Profiles</i></b>			X
<b><i>Correction Notices</i></b>			X
<b><i>Cover Stories</i></b>			X
<b><i>Directories</i></b>			X
<b><i>Editorials</i></b>			X
<b><i>Evaluations</i></b>		X	X
<b><i>Glossaries</i></b>		X	
<b><i>Hardware Reviews</i></b>		X	X
<b><i>Industry Overviews</i></b>			X
<b><i>Interviews</i></b>		X	
<b><i>Letters to the Editor</i></b>			X
<b><i>Obituaries</i></b>			X
<b><i>Panel Discussions</i></b>		X	
<b><i>Product Announcements</i></b>			X
<b><i>Software Reviews</i></b>		X	X
<b><i>Technical</i></b> <sup>3</sup>		X	
<b><i>Tutorials</i></b>	X		

**Table 3**

The nineteen genres of texts identified by the publishers of *Computer Select* are classified according to the type of information they are intended to impart

File name	Total number of words	Total number of bytes
scan1	229,407	1,534,695
scan2	466,095	3,172,317
scan3	416,810	2,728,605
scan4	385,948	2,519,623
All files combined	1,498,260	9,955,240

**Table 4**

Total number of words and bytes in the scanner corpus

### 3.6. Experiment

For the purpose of this experiment, all students were translating out of their foreign language (French) into their native language (English). Students were presented with two different extracts from an article on optical scanners. They were asked to translate one of the texts using the following resources as necessary: a selection of general bilingual dictionaries and a selection of monolingual specialized lexicographic and non-lexicographic resources.<sup>15</sup> They were then asked to translate the second text using the following resources as necessary: a selection of general bilingual dictionaries and a specialized monolingual native-language corpus coupled with the WordSmith Tools corpus analysis tool. In an attempt to compensate for 1) any potential text-specific difficulties, and 2) differences in student ability, students 1-7 were asked to translate text i using the conventional resources and text ii using the corpus, whereas students 8-14 were asked to translate text ii using conventional resources and text i using the corpus.

The students were given two hours to translate each text. They were also asked to comment on the usefulness of the monolingual resources they had at their disposal (whether it be the conventional resources or the corpus).

### 3.7. Data Analysis

The aim of the pilot study was to determine whether a specialized monolingual native-language corpus would help translators to produce improved quality translations. Because our sample size was small in a number of respects (14 students, 2 texts, 1 subject field), we felt that we would not be able to make any definitive conclusions, but could only reasonably measure general trends. Therefore, translations were assessed for the following broad categories of errors: 1) comprehension errors, specifically errors resulting from a lack of comprehension of the subject field; 2) production errors including incorrect choice of term, non-idiomatic constructions, grammatical errors, and incorrect register.

#### 3.7.1. Improved subject comprehension: an example

The following example shows how the corpus has the potential to help students to acquire an increased understanding of the subject field. In text i, most of the students had difficulty grasping one of the concepts in the opening sentence of the text.

*quelle que soit leur sensibilité aux nuances, leur rapidité, leur précision,...*

which should logically be translated along the following lines:

*Regardless of such characteristics as colour-recognition capability, speed, precision,...*

The specialized dictionaries treated the concept *scanner* in general, but did not discuss the colour-recognition capability of scanners. The user manual made a passing reference to "black-and-white vs. colour scanners." The desktop publishing monograph contained a half-page discussion on the difference between "non-greyscale scanners, greyscale scanners, and colour scanners," but this section did not feature specifically in the table of contents and was therefore not easy to locate (though there were index references to each of the terms individually). The journal article treated the issue of colour recognition in some depth, but this discussion was towards the end of the article and was not visibly set off from the rest of the text. In other words, the necessary information was there, but it was not easy to find. Not surprisingly, none of the students using the conventional resources came close to rendering the concept correctly; in fact, some of them proposed rather peculiar translations which showed a definite lack of subject field understanding.

The corpus users, however, were able to go directly to those areas of the text that dealt with this subject. Even if students did not know the correct translation for *nuances*, a collocation search on the word form "sensitiv\*" (i.e. the translation of *sensible*) revealed that the following words were among those that appeared in its vicinity: colour (5), greyscale (4), shade\* (122), shading (8).<sup>16</sup> Students were able to read these particular contexts and achieve a somewhat better understanding of the subject field. As shown in table 5, three of these students came quite close to expressing the correct idea by referring to *shades* or *shading*, and another student actually referred to *colour*. Admittedly, none of them came up with a particularly elegant rendering for that specific concept, however, they did at least seem to have a better understanding of the idea that was being referred to in the source text.

1-14=student; D=dictionary user; C=corpus user; i=text i; +=subject understanding	
1-D-i	no matter how much attention to detail they pay...
2-D-i	no matter how sensitive they are...
3-D-i	even though their sensitivity to touch...
4-D-i	regardless of their sensitivity...
5-D-i	while adaptability...
6-D-i	no matter how good the resolution...
7-D-i	whatever their feeling to the suggestion...
8-C-i +	no matter their sensitivity to shading...
9-C-i +	despite differences in their sensitivity to shading...
10-C-i	whatever their sensitivity to detail...
11-C-i	regardless of how sensitive they are to differences...
12-C-i ++	whatever their sensitivity to colour...
13-C-i	whatever their sensitivity to small differences...
14-C-i +	whatever differences there may be in shade...

**Table 5**

Proposed translations of the expression *quelle que soit leur sensibilité aux nuances...* The plus symbol (+) indicates which students came up with a rendering that seems to show a reasonable understanding of the concept

### 3.7.2. Improved term choice: an example

The following example shows how the corpus has the potential to help students find and use the correct terms. In text i, the term *vitre* is most properly translated as either *glass platen* or *scan bed*. The specialized dictionaries provided did not contain the term *glass platen*, and the translations given in the general bilingual dictionaries included *glass*, *pane of glass*, and *window*. The term *glass platen* did appear in the user guide, but was not in the index and was therefore difficult to find. It did not appear (to the best of my knowledge) in the desktop publishing monograph or the journal article. None of the students using the conventional resources used either of these terms, whereas three of the students using the corpus correctly used the term *glass platen*, which appeared in the corpus 41 times and ranked highly as one of the collocates of *glass*.

1-14=student; D=dictionary user; C=corpus user; i=text i; +=correct term used	
1-D-i	Glass
2-D-i	sheet of glass
3-D-i	pane of glass
4-D-i	Window
5-D-i	glass surface
6-D-i	Glass
7-D-i	Window
8-C-i +	glass platen
9-C-i	piece of glass
10-C-i +	glass platen
11-C-i	glass
12-C-i	screen
13-C-i	window
14-C-i +	glass platen

**Table 6**

Proposed equivalents for the term *vitre*. The plus symbol (+) indicates which students used one of the preferred terms

A similar example occurred with the term *scanner à plat*, which is properly translated as *flatbed scanner*. Of the students using the conventional resources, three translated the term improperly as *flat scanner*, even though the term *flatbed scanner* appeared several times in the monograph on desktop publishing. All of the students using the corpus came up with the correct term, and student (14) even made the following comment: "I was unsure of which spelling to use — *flatbed* or *flat-bed* — because I had seen both in the corpus. I looked up both terms in the frequency list and saw that *flatbed* occurred 1508 times and *flat-bed* only occurred 92 times, so I went with *flatbed*."

### 3.7.3. Improved idiomatic construction: an example

The following example shows how the corpus has the potential to help students create more idiomatic constructions. One of the phrases appearing in text i is: *photo-*

*diodes sensibles à la lumière*. Some of the specialized dictionaries contained the term *photodiode* and the desktop publishing monograph referred to *light-sensitive elements* (though not in the index, thereby making it difficult to locate). The journal article did not make any reference to *photodiodes*. I hypothesize that the majority of the students using the conventional resources verified in the dictionaries that *photodiode* was a term and then simply followed the syntax of the source text to produce a construction that, while grammatically correct, is not idiomatic according to the expert discourse. There were no instances in the corpus where this concept was expressed using the syntax pattern *photodiodes sensitive to the light*, and this may explain why all the students who had access to the corpus used one of the two more idiomatic constructions which appeared there: *light-sensitive photodiodes* (which appeared twice) or *photosensitive diodes* (which also appeared twice).

1-14=student; D=dictionary user; C=corpus user; i=text i; X=non-idiomatic	
1-D-i	light sensitive photodiodes
2-D-i X	photo sensors that are sensitive to this light
3-D-i X	photodiodes sensitive to the light
4-D-i X	photodiodes which are sensitive to light
5-D-i	light sensitive photodiodes
6-D-i X	laser diodes which are sensitive to this light
7-D-i X	photodiodes sensitive to this light
8-C-i	light-sensitive photodiodes
9-C-i	light-sensitive photodiodes
10-C-i	light-sensitive photodiodes
11-C-i	light-sensitive photodiodes
12-C-i	photosensitive diodes
13-C-i	light-sensitive photodiodes
14-C-i	photosensitive diodes

**Table 7**

Proposed translations of the expression *photodiodes sensibles à la lumière*. The X indicates a non-idiomatic construction

A similar example occurred with the expression *la tête de numérisation du scanner*, which is best translated as *scan head*, but could also be translated as *scanning head* or *scanner head*. Of the students using the conventional resources, four of them used an idiomatic construction, but three of them followed the French syntax and rendered the phrase as *head of the scanner*. There were no instances in the corpus where this concept was expressed using the syntax pattern *head of the scanner*, and all of the students using the corpus employed one of the idiomatic constructions.

### 3.8. General Trends Observed

A number of general trends came to light during the analysis of the data from this pilot study.

### 3.8.1. General trends in dictionary use vs. corpus use

We observed that for each of the five categories of error, the students using the corpus made fewer total errors than the students using the conventional resources. For three of the categories, namely subject field comprehension, term choice, and non-idiomatic construction, the improvements shown were reasonably significant. For the remaining two categories, grammatical error and incorrect register, the improvements were marginal. With regard to grammatical error, the majority of errors made by both the conventional resource users and the corpus users seemed to result from sloppiness (e.g. misused punctuation, incorrect capitalization), rather than from a lack of information in either resource. In the case of register, the problems seemed to be largely related to particular students, rather than to either of the available types of resources. Table 8 shows the breakdown of errors made by each student for each category in text i, and table 9 shows a similar breakdown for text ii. Table 10 summarizes the relative improvement made by corpus users over conventional resource users for each of the categories of error.

1-14=student; D=dictionary user; C=corpus user; i=text i	Comprehension errors	Production errors			
Students	subject comprehension error	incorrect term	non-idiomatic construction	grammatical error	incorrect register
1-D-i	3	4	3	5	1
2-D-i	3	4	4	1	0
3-D-i	2	6	7	3	0
4-D-i	5	6	4	1	0
5-D-i	4	5	3	3	0
6-D-i	2	2	5	5	2
7-D-i	2	6	7	3	0
<b>Total errors</b>	<b>21</b>	<b>33</b>	<b>33</b>	<b>21</b>	<b>3</b>
Average errors per student	3	4.71	4.71	3	0.42
8-C-i	2	4	1	4	0
9-C-i	2	3	7	1	1
10-C-i	2	4	0	2	0
11-C-i	2	2	5	2	0
12-C-i	1	2	4	3	0
13-C-i	3	4	5	4	0
14-C-i	1	4	0	3	0
<b>Total errors</b>	<b>13</b>	<b>23</b>	<b>22</b>	<b>19</b>	<b>1</b>
Average errors per student	1.86	3.29	3.14	2.71	0.14

**Table 8**

Errors made by each student in each category in text i.

The general trend shows that for the categories of subject comprehension, term choice, and non-idiomatic expression, students using the corpus made significantly fewer errors than students using conventional resources. For the remaining two categories, namely grammatical errors and register, students using the corpus performed only marginally better than students using conventional resources

1-14=student; D=dictionary user; C=corpus user; ii=text ii	Comprehension errors	Production errors			
Students	Subject comprehension error	Incorrect term	non-idiomatic construction	Grammatical error	Incorrect register
8-D-ii	2	2	4	6	1
9-D-ii	3	6	7	4	2
10-D-ii	2	7	2	2	2
11-D-ii	3	2	4	3	0
12-D-ii	2	5	8	5	2
13-D-ii	2	4	4	3	1
14-D-ii	1	4	2	1	0
<b>Total errors</b>	<b>15</b>	<b>30</b>	<b>31</b>	<b>24</b>	<b>8</b>
Average errors per student	2.14	4.29	4.43	3.43	1.14
1-C-ii	1	2	3	4	3
2-C-ii	2	2	3	3	0
3-C-ii	1	3	5	4	0
4-C-ii	0	3	4	3	0
5-C-ii	2	2	3	3	0
6-C-ii	1	1	2	3	3
7-C-ii	2	3	4	3	0
<b>Total errors</b>	<b>9</b>	<b>16</b>	<b>24</b>	<b>23</b>	<b>6</b>
Average errors per student	1.23	2.29	3.43	3.29	0.86

**Table 9**  
Errors made by each student in each category in text ii.

The general trend shows that for the categories of subject comprehension, term choice, and non-idiomatic expression, students using the corpus made significantly fewer errors than students using conventional resources. For the remaining two categories, namely grammatical errors and register, students using the corpus performed only marginally better than students using conventional resources



	Errors made by dictionary users	Errors made by corpus users	Relative improvement
Text i			
Subject comprehension error	21	13	38%
Incorrect term	33	23	30%
Non-idiomatic construction	33	22	33%
Grammatical error	21	19	insignificant
Incorrect register	3	1	insignificant
Text ii			
Subject comprehension error	15	9	40%
Incorrect term	30	16	47%
Non-idiomatic construction	31	24	23%
Grammatical error	24	23	insignificant
Incorrect register	8	6	insignificant

**Table 10**

A summary of the relative improvement made by corpus users over users of conventional resources in the five different error categories

### *3.8.2. Individual student trends in dictionary use vs. corpus use*

In the case of individual students, the majority of students also showed either a general trend of improvement or no change when using the corpus. Only the work of student (13) showed a general tendency to decline when using the corpus. This is illustrated in table 11.

### **3.9. Student Comments**

In addition to doing the translation using either the conventional resources or the corpus, the students were asked to comment on the usefulness of these tools.<sup>17</sup> Several general patterns emerged from these comments.

#### *3.9.1. Comments about the conventional resources*

There were some feelings of dissatisfaction regarding the information provided by conventional resources. The most common frustration was that the type of information the students were looking for was not found in these resources,<sup>18</sup> specifically, information about different types of scanners:

- (8) "I didn't find the monolingual dictionaries very useful in this situation because they didn't really provide you with the specialized terms relating to scanners. The only information they provided was a very basic and broad definition of the concept. It never mentioned the types: hand-held, flat-bed, etc."
- (10) "I did not find the specialized dictionaries very helpful for this exercise. Some of the words I looked up were not listed (e.g. flatbed scanner) or the information was insufficient (e.g. scanner)."
- (11) "I didn't find the specialised dictionaries helpful really at all. I looked up 'scanner' thinking it would give the names of the different types but it didn't."
- (13) "I did not find the technical dictionaries to be of great assistance as the terms were either very broad (dictionary of computing terms) or very technical (like the engineering dictionary). I could not find any entry referring to the different types of scanner."

Two students (4, 5) indicated that they actually preferred working with the conventional resources because they were more used to this method of working:

- (4) "Maybe it's force of habit, but I felt I used the dictionaries more."
- (5) "I really didn't find the corpus tools much use; the dictionaries were much handier, quicker and easier to use. Then again, I may have been a little afraid of the corpus tools!"

Whenever a new method of working is introduced, there is always a learning curve

1st number=dictionary use; 2nd number=corpus use; *=more errors made using corpus					
student	Subject comprehension	Incorrect term	non-idiomatic construction	Grammatical error	Incorrect register
1	3-1	4-2	3-3	5-4	1-3*
2	3-2	4-2	4-3	1-3*	0-0
3	2-1	6-3	7-5	3-4*	0-0
4	5-0	6-3	4-4	1-3*	0-0
5	4-2	5-2	3-3	3-3	0-0
6	2-1	2-1	5-2	5-3	2-3*
7	2-2	6-3	7-4	3-3	0-0
8	2-2	2-4*	4-1	6-4	1-0
9	3-2	6-2	7-7	4-1	2-1
10	2-2	7-4	2-0	2-2	2-0
11	3-2	2-2	4-5*	3-2	0-0
12	2-1	5-2	8-4	5-3	2-0
13*	2-3*	4-4	4-5*	3-4*	1-0
14	1-1	4-4	2-0	1-3	0-0

**Table 11**

Difference in errors made by students when using dictionaries (first number) and corpus (second number)

involved. Hopefully, improved training on the use of the corpus analysis tool and more practice applying it in a translation environment would result in a reduction of this type of resistance to using the corpus. Another related point is that pre-conceived attitudes and expectations can have a significant impact on the acceptance of a new method of working. Student (14) made the following comment: "Although I was initially dubious of its possible use, the concordancer was in fact quite useful." Fortunately, this student was open-minded, but care must be taken to counter pre-conceived notions by properly educating translators about both the benefits and limitations of computer aids.<sup>19</sup>

### 3.9.2. Comments about the corpus

A number of the students (8, 10, 12, 14), indicated that they found it time-consuming to work with the corpus; however, student (10) did add that, "if one was more used to the system, it hopefully wouldn't be so time-consuming."

Those students who did find the corpus useful indicated that it was helpful in the following ways:

student (8) seemed to find the corpus useful for subject-field understanding, stating "I think that the concordances were useful for determining the exact sense of a particular word regarding the subject of scanners, considering I don't know much about them."

students (2, 6, 12, 14) found the concordancer useful for helping them to decide on an appropriate term (e.g. *hand-held scanner* as an equivalent for *scanner à main*).

students (2, 9, 14) found the frequency information useful for helping them to choose between two possible terms (e.g. *scanning mechanism* vs. *scanner mechanism*).

students (12, 13, 14) found the corpus useful as a means of verifying or confirming that a term did in fact exist or that it was being used correctly.

#### 4. CONCLUDING REMARKS

On the whole, the pilot study appears to have been successful in supporting our hypothesis that a specialized monolingual native-language corpus can be a useful resource for translators translating into their native language. Translators using a corpus have access to a greater amount of data, which they can interrogate more easily. The results of the study showed a general trend towards improved quality translation for the categories of subject-field understanding, correct term choice, and idiomatic expression. Although there was no significant improvement in the categories of grammatical error and incorrect register, we are happy to observe that the corpus did not contribute to a decline in performance either. We hope to eventually expand this study to include experiments involving other text types, other subject fields, and other language pairs to see if we can draw more definitive conclusions about the usefulness of a monolingual corpus as a translation resource.

#### Acknowledgements

I would like to thank the students who participated in this project, and also Dr. J. Williams (Dublin City University) and Dr. I. Meyer (University of Ottawa) who provided helpful feedback on an earlier version of this paper.

#### Notes

1. There are, of course, exceptions to this practice. For example, cases where it is desirable, for reasons of both national cultural identity and economic survival, to have translation out of a so-called minor language (e.g. Finnish) into a so-called major language (e.g. English), but where the volume of work exceeds the number of available majorlanguage native speakers (McAlester 1992: 292).
2. Discussions with colleagues at other universities in both Europe and North America have revealed that my experiences in this regard are not isolated incidents, nor are they limited to native English speakers.
3. At present, DCU students have only one course in English language skills which they follow during the first year of the translation program.
4. In addition to providing them with such resources, we should also endeavour to provide them with the skills necessary to compile similar resources of their own in the future.
5. The use of corpora in the discipline of computerassisted language learning (CALL) is well-established; however, the most common application is for the learning of foreign languages (e.g. Kettemann 1996), rather than for the finetuning of nativelanguage skills. Conventional CALL applications may be useful for helping translators improve their foreign language skills, but such investigations are beyond the scope of this study.
6. This statement describes the situation today; however, it may change in the future as more translators make use of tools such as translation memories to create bilingual, parallel corpora.

7. Tools are also being developed which can be used to conduct conceptual as well as linguistic searches (e.g. Text Analyzer (Kavanagh 1995)).
8. There are, of course, exceptions, such as the systematically organized Glossary of Computing Terms published by the British Computer Society (see bibliography for full reference).
9. The full references for the resources are listed in the bibliography.
10. One of our research partners, The Language Analysis and Knowledge Engineering (LAKE) Laboratory of the University of Ottawa, subscribes to *Computer Select*, and through them we obtained permission to extract and use corpora for research purposes.
11. The discs are updated monthly, and the disc issued in December of each year contains all the articles published during that calendar year.
12. See tables 2 and 3 for a breakdown of the text types found in *Computer Select*.
13. The articles classified in the genre *technical* were not technical manuals, but rather they seemed to be articles dealing with highly specialized subjects (e.g. "Polyester media development for inkjet printers" in *HewlettPackard Journal* or "Fuzzy cognitive maps model social systems" in *AI Expert*). The majority of these articles appeared in learned journals.
14. The creators of *Computer Select* have indexed each of the articles on the disc according to certain key words (i.e. topics). Users can instruct the computer to 1) locate all the articles that have been indexed according to a given key word (e.g. **scanning technology**), and 2) download these articles into a file on a hard disc.
15. These lexicographic and nonlexicographic resources were in paperbased form (see section 3.4.1).
16. The \* is a wildcard character that can be used to represent any letter or combination of letters (e.g. a search on "sensitiv\*" would find occurrences of both "sensitive" and "sensitivity"). The number in parentheses after the term indicates the number of occurrences of this term in the vicinity of "sensitiv\*."
17. Unfortunately, not all of the students did provide comments on the tools. No comments were received from three students (1, 3, 7); however, the remaining eleven students did provide some comments, in varying degrees of detail.
18. Interestingly, none of the students commented on the usefulness of any of the resources other than the dictionaries (i.e. the journal article, the user manual or the desktop publishing monograph). My observation from supervising the experiment is that they, like many students, relied heavily on dictionaries and made little effort to use the other resources provided.
19. The paradoxical combination of unrealistic expectations for machine translation and the fear of being replaced by a machinetranslation system means that many translators are particularly susceptible to being reluctant to incorporate computerassisted translation tools into their work practices.

## REFERENCES

- ATKINS, S., CLEAR, J. and N. OLSTER (1992): "Corpus Design Criteria", *Literary and Linguistic Computing*, 7 (1), pp. 1-16.
- BAKER, M. (1992): *In Other Words: A Coursebook on Translation*, London, Routledge.
- BAKER, M. (1996): "Corpus-based Translation Studies: the Challenges that Lie Ahead", H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, Amsterdam, John Benjamins, pp. 175-186.
- BOWKER, L. (1995): *A Multidimensional Approach to Classification in Terminology: Working Within a Computational Framework*, PhD Thesis, University of Manchester, UK.
- BOWKER, L. (1996): "Towards a Corpus-Based Approach to Terminography," *Terminology*, 3 (1), pp. 27-52.
- CORMIER, M. (1991): "Traduction de textes destinés à des spécialistes: Approche pédagogique," *Meta*, 36 (2/3), pp. 440-447.
- DANIELSSON, P. and D. RIDINGS (1996): "Corpus and Terminology: Software for the Translation Program at Göteborgs Universitet or Getting Students to do the Work", *Proceedings of Teaching and Language Corpora 1996 (TALC'96)*, Lancaster, University of Lancaster (UCREL), pp. 57-67.
- FOLKART, B. (1984): "A Thing-Bound Approach to the Practice and Teaching of Technical Translation", *Meta*, 29 (3), pp. 229-246.
- KAVANAGH, J. (1995): *The Text Analyzer: A Tool for Knowledge Acquisition from Texts*, MSc Thesis, University of Ottawa, Canada.
- KETTEMANN, B. (1996): "Concordancing in English Language Teaching", *Proceedings of Teaching and Language Corpora 1996 (TALC'96)*, Lancaster, University of Lancaster (UCREL), pp. 4-16.
- LAVIOSA, S. (1998): "The English Comparable Corpus: A Resource And A Methodology", L. Bowker, M. Cronin, D. Kenny and J. Pearson (Eds), *Unity in Diversity? Current Trends in Translation Studies*, Manchester, St. Jerome Publishing.
- MCALESTER, G. (1992): "Teaching Translation into a Foreign Language — Status, Scope and Aims", C. Dollerup and A. Loddegaard (Eds), *Teaching Translation and Interpreting: Training, Talent and Experience*, Amsterdam, John Benjamins, pp. 291-297.

- MCENERY, T. and A. WILSON (1993): "Corpora and Translation: Uses and Future Prospects", Technical Report from the Unit for Computer Research on the English Language (UCREL), University of Lancaster.
- MILLER, D.R. (1993): *Towards Knowledge-Base Systems for Translators*, MA Thesis, University of Ottawa, Canada.
- NEWMARK, P. (1988): *A Textbook of Translation*, New York, Prentice-Hall.
- PETERS, C. and E. PICCHI (1998): "Bilingual Reference Corpora for Translators and Translation Studies", L. Bowker, M. Cronin, D. Kenny and J. Pearson (Eds), *Unity in Diversity? Current Trends in Translation Studies*, Manchester, St. Jerome Publishing.
- ROUMEN, R. and T. VAN DER STER (1993): "CONTEXT: A New Concept in Computer-Aided Translation", *TKE'93: Terminology and Knowledge Engineering*, Frankfurt, Indeks Verlag, pp. 215-220.
- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- SYKES, J.B. (1989): "Chapter 3: The Intellectual Tools Employed", C. Picken (Ed.), *The Translator's Handbook (2nd ed.)*, London, Aslib, pp. 35-41.
- VAN DOORSLAER, L. (1995b): "Quantitative and Qualitative Aspects of Corpus Selection in Translation Studies", *Target*, 7 (2), pp. 245-260.

#### Conventional resources consulted during the experiment

- British Computer Society (1991): *A Glossary of Computing Terms (7th ed.)*, London, Pitman.
- GIBILISCO, S. and N. SCLATER (Eds) (1990): *Encyclopedia of Electronics*, Blue Ridge Summit, PA, TAB.
- GLOVER, G. (1990): *Image Scanning for Desktop Publishers*, Blue Ridge Summit, PA, Windcrest.
- GUNTON, T. (1990): *A Dictionary of Information Technology and Computer Science*, Oxford, Blackwell.
- Hewlett-Packard co. (1993): *Hp Deskscan II User's Guide*.
- JOSEPH, C. (1995): "Scan-o-rama", *The Mac*, Vol. 3, No. 7 (July 1995), pp. 63-67.
- MCCLAIN, G. (Ed.) (1994): *21st-Century Dictionary of Computer Terms*, New York, Laurel.
- NADAR, J.C. (Ed.) (1995): *Prentice Hall's Illustrated Dictionary of Computing (2nd ed.)*, New York, Prentice Hall.
- SPENCER, D.D. (Ed.) (1995): *The Computer Dictionary*, Ormond Beach, FL, Camelot.

## APPENDIX 1

### SOURCE TEXT # 1

**Instructions:** Translate the following text as if it were to be published in a semi-specialized journal.

#### Le scanner

Toutefois, quelle que soit leur sensibilité aux nuances, leur rapidité, leur précision, ou la manière dont la tête et le papier se déplacent, tous les scanners reposent sur le même principe, exposé ici en prenant l'exemple d'un scanner à plat.

1) Une source de lumière éclaire le document, placé contre une vitre située au-dessus du mécanisme du scanner. La lumière est plus ou moins réfléchiée, selon la densité de chaque zone du document. Un espace blanc la réfléchit au maximum, un espace noir au minimum.

2) La tête de numérisation du scanner se déplace à travers la page, la décomposant en points de mesure distants de moins d'un dixième de millimètre, et capte à chaque point la lumière réfléchiée.

3) Dans le scanner à plat, la lumière circule à travers un système de miroirs qui doivent pivoter continuellement afin de garder le faisceau aligné sur une lentille.

4) La lentille concentre le faisceau lumineux sur des photodiodes sensibles à la lumière qui traduisent l'intensité lumineuse en courant électrique. Plus il y a de lumière réfléchiée, plus il y a de voltage.

5) Un circuit électronique associe alors à chaque mesure analogique de voltage un pixel particulier, représentant une zone noire ou blanche, ou d'un certain niveau de gris, ou encore un certain teinte, selon le degré de sophistication du scanner. Le processus se répète à chaque point analysé, à raison de 300 points par pouce pour la plupart des scanners au marché (voire plus pour certains).

6) L'information numérique est envoyée au programme de contrôle du scanner, exécuté par le micro-ordinateur, qui la mémorise dans un format de fichier que peuvent reconnaître les logiciels graphiques ou de reconnaissance de caractères.

**Source:** JÉRÔME, M. (1991): "Le scanner", *Science et Vie Micro* janvier, pp. 150-151.

### SOURCE TEXT # 2

**Instructions:** Translate the following text as if it were to be published in a semi-specialized journal.

#### Le scanner

Il existe trois types de scanners, différenciés essentiellement par la manière dont la page qui contient l'image, et la tête de numérisation du scanner qui la lit, se déplacent l'une par rapport à l'autre. Dans un scanner à rouleau, le document est entraîné devant la tête de numérisation qui se déplace latéralement, comme une tête d'imprimante matricielle. Dans un scanner à plat, le document est posé sur une vitre, comme sur un photocopieur, la tête de numérisation se déplaçant à la fois sur la largeur et sur la longueur de la page analysée. Quant au scanner à main, il requiert comme son nom l'indique la main de l'utilisateur, qui balaye de haut en bas le document à numériser.

Chaque méthode comporte ses avantages et ses inconvénients. Le scanner à plat exige toute une série de miroirs pour renvoyer l'image qui sera captée par la tête mobile vers une lentille qui la dirigera à son tour vers un capteur. Comme aucun dispositif à miroir n'est parfait, chaque fois que l'image est reflétée, elle subit une déformation. En revanche, l'avantage du scanner à plat est qu'il peut balayer des documents de grandes dimensions ou très épais, des livres notamment.

Inversement, si avec un scanner à rouleau l'image est captée avec plus de fidélité, on est ici limité à la numérisation de documents peu épais. Ce type de scanner est sans doute le meilleur pour des applications telles que la reconnaissance de caractères, pour laquelle la précision est extrêmement importante.

Le scanner à main est quant à lui un compromis. Il est capable de balayer des pages de livres, mais sa "fenêtre" de numérisation est beaucoup plus petite que celle des deux autres catégories de scanners. De plus, la qualité du résultat dépend essentiellement de la fermeté de la main de l'utilisateur.

**Source:** JÉRÔME, M. (1991): "Le scanner", *Science et Vie Micro*, janvier, pp. 150-151.