Bi-text, a new concept in translation theory

by Professor Brian Harris

Brian Harris is a member of the staff of the School of Translators & Interpreters, University of Ottawa, and a member of the New Technologies Committee of the International Federation of Translators (FIT).

What is *bi-text*? (*Bi-text* is written with a hyphen here to help newcomers pronounce it, but the hyphen can be dropped.) I offer it to translatologists as a new concept in translation theory. But though it will be helpful for everyone who uses bi-text to understand the basic *notion,* translation practitioners will be more interested in the *applications* of it that are now becoming possible thanks to computers.

The concept
The concept of bi-text has three parts so far as I have worked it out: its primary nature, which is psycholinguistic; its adaptation to take advantage of computers; and its presentation — that is, the way it is displayed to us so that we can use it and work on it.

Let us begin by considering how translations are usually viewed. For every translation there is necessarily an original text (termed the *source text,* or ST for short) and there is the translation itself (the *target text,* or TT). The two lead semi-independent lives. Both share the same author and the same content; but the readerships are different. The latter differ to the extent that the readers of ST may even be unaware of the existence of TT and most of them would be incapable of reading it anyway; and conversely the readers of TT cannot read the original — if it were not so, there would be no point in doing translations. These statements are of course generalisations. There are always some bilinguals who are capable of reading both versions. Yet even for them the task is not facilitated by the way translations are usually presented to the eye. ST and TT are usually published in separate volumes, or on separate pages (sometimes, in Canada, different ways up!) so that making comparisons between the two requires a good deal of patience.

On the other hand, there is one person par excellence for whom, at least briefly, ST and TT are not separated but on the contrary are simultaneously present and intimately interconnected in his or her mind. That person is the *translator.* One way to describe bi-text, therefore — and this is a basic definition — is to say that it is ST and TT as they co-exist in the translator's mind at the moment of translating. At that point TT, still in the creation and growing process, has not yet gone its own way; but even later on, those of us who *are* bilingual and who want to study the translation can also 're-store' both versions, or at least parts of them, in our minds simultaneously and consider them together. It is a translation in this state, when it is not a separate ST and TT but the two sewn firmly together like a piece of cloth and its lining to be used as one fabric, which constitutes a *bi-text* for others as well as for the translator.

Another way of putting it is to say that a bi-text is not two texts but a single text in two dimensions, each of which is a language. However, neither this metaphor nor the one about the cloth and its lining is quite appropriate. To see why, we have to go back to the translator's mind. The image of the cloth and its lining, for instance, suggests something that is stitched together at the beginning and the end and around the edges but only loosely held together in the middle. The image of the text in two dimensions refers to the text as a whole. Yet translators do not translate whole texts at one fell swoop. They proceed a little at a time, and as they proceed each spurt, each segment forms a fragment of bi-text in their minds. Bi-text retains this structure when it is recorded on paper or in a computer: that is to say, not only is the whole text a bi-text but each segment combines ST and TT. So a better metaphor might be that of a roll of two laminated materials of different colours, whereby if you cut a piece out of either side you will always get with it the corresponding piece of material from the other side in the other colour.

The translator's working segments of text are called *translation units in* the writings on the subject. We can say, using this term, that retrieval of a translation unit of ST from a bi-text will always bring with it the corresponding unit of TT. People who do not know much about translation tend to think the translation units are individual words, but in fact they mostly consist of whole phrases and even clauses or sentences. Bi-text therefore binds together not the individual words of ST and TT but those somewhat longer segments. This is important for practical applications.

To semioticians, by the way, I submit that *bi-text* falls into the same paradigm as *intertext,* in that it is a construct of two or more related texts.

Computer implementation
Bi-text does not necessarily have to be produced in large amounts to be useful. Nevertheless, in this section I will explore the potential of large accumulations, because then the advantage of computer storage and processing becomes more marked, and exciting new possibilities are opened up.

Bi-text is, to be sure, a kind of text, and we suppose that we are interested in exploiting large amounts of it. That leads us to converge on another relatively new concept relating to large collections of text, namely *hypertext.* So what is hypertext?

Hypertext is itself a complex concept. First let us consider the 'text' in hypertext. In my own computer I store running texts as 'documents' and data which fits into a predetermined and constantly repeated format, like bibliographies, as 'files'. Each text and each bibliography occupies its own document or file, and consequently I now have so many of them that I need yet another file just to list them all. Furthermore the computer can only process one such document or file at a time. This means that if I wish to search for a piece of text or information economically, I must start by naming for the computer which file or files it occurs in. Otherwise the search will have to be conducted through all the files, one by one. Hypertext breaks down the divisions between discrete, mutually

uncommunicating documents and files by treating all the texts and data files that are in memory, whether real memory or virtual memory, as one big text. Since the virtual memory includes the disks, the amount of text that can be processed together is very large. Let us call this large amount of virtually contiguous text the *hypertext base.*

The second essential feature of hypertext is that it enables us to link related items wherever they may occur throughout the hypertext base. It is fashionable on the current wave of 'artificial intelligence' to talk about hypertext linking "related concepts"; and so it can if we enter the indicators for the links ourselves, but that may be rather time-consuming. What the computer can do automatically, on the other hand, is connect up those time-honoured symbols for concepts, *words.*

Though hypertext has been dreamed of in computer circles for the past 30 years, it is only now coming into its own. The circumstances that have made it feasible are a combination of more sophisticated systems design, faster microcomputers, and a massive increase in the size of microcomputer memories. Today it is possible to take home a 20-pound portable computer with a 100-megabyte disk drive, enough to store some 10 years of an average translator's entire output and have it instantly available for reference.

Searching the hypertext
Now imagine a work station in which a translator has regularly stored all his translations with their originals in the form of bi-text, and together they make up a hypertext of several thousand pages. As an additional tool, there is also a thesaurus or synonym dictionary of the ST language stored. In this first scenario, our translator is stumped for the best translation of a certain ST word in the context. He interrogates the system, to find this word in the hypertext base. The hypertext 'search engine' is programmed in such a way that when it finds an occurrence of the word, it retrieves and displays the whole translation unit in which it occurs. More than that it will, at the extreme, look up the thesaurus or synonym dictionary and perform the same operation for the conceptually related words listed there. Unless the word in question is a very rare one — or a very new one — the search is likely to produce a listing of actual translation examples in contexts from the

translator's own previous work. In the view of some experienced translators, a 'contextual glossary' of his kind is the most practical kind of lexical aid. Note that in a way this is a *self-learning system,* because the more the translator translates the richer the contextual glossaries will *become automatically.* In this respect it contrasts with much other current software which requires the translator not only to translate but to make entries in a dictionary or terminology file as well. Note also that the translations retrieved in this way, since they are complete translation units, are not limited to word-for-word correspondences; this too is an advantage over the lexicons, or at any rate a valuable complement to them.

Next let us consider a rather more sophisticated scenario. This time it is not a single word or term that the translator needs help with but a whole translation unit, and it happens not to be a phrase that is conventional enough to have found a place in the dictionaries. The chances of finding *a* non-conventional unit in *exactly the* same form in the base are much slimmer than for single words, so instead we ask the search engine to look for a *similar* unit. What the translator can hope for by this method is not a ready-made translation of the problematic unit but bi-text segments that will be similar enough to help him towards his objective (the criteria of similarity to be applied in matching the units is another area that needs more thought and some experimentation).

Thus bi-text used in this way could help the translator by providing the following: translations of words in context; a memory-perfect exploitation of the translator's own previous experience; near-translations of non-conventional phraseology and even longer units.

What of the dictionaries? Would they be superseded? Not at all. They would simply be stored in the computer too and merged into the translator's hypertext, which, since it is in bi-text, I will henceforth call *hyper-bitext.*

Here then is my definition of hyper-bitext: it is bilingual hypertext stored in such a way that each retrievable segment consists of a segment in one language linked to a segment in the other language which has the same meaning. Why have I backed off from using 'translation' in the definition and employed the

periphrase "which has the same meaning" instead? It is because I foresee an extension of the concept to 'parallel texts' which express the same ideas as one another but are not strictly translations. That, however, is more difficult to implement and in any case it lies beyond translators' immediate needs.

The product of a search in a good hyper-bitext is likely to be an information overload rather than too little help. Users of the existing large term data banks are already familiar with this problem. Here above all, in the sorting, selection and ranking of the search output, we would need artificial intelligence and 'expert systems'.

Presentation
Earlier on I complained that examination of ST and TT together, that is to say as a bi-text, was made difficult by the way they are usually presented separately to the eye. In fact this inconvenience arises even for the translators themselves once the short segments of text on which they are actually working have dropped out of memory and they turn back to look them over: they spend a lot of effort switching their eyesight between two sheets of paper, or these days between a sheet of paper and computer screen, or even between two 'windows' on the screen, and trying to keep their place in both. Clearly this practical consideration as well as the notion of bi-text calls for a more closely welded 'display format'—to use the computerese for it.

There does exist a very old format for displaying translations, which comes about as close as one can to solving the problem. Translations in this format are called *interlinear translations,* which we can abbreviate to IT. Actually the term connotes a very literal kind of translation; but let us not dwell on that, because we are only concerned here with the typographical presentation. In this style of layout, each line of the printed TT text is interlaced between the corresponding line of ST. IT has long been useful for scholars — and for schoolboys in the days when cribs for Latin and Greek classics used to be printed this way—but never popular elsewhere because most people, of course, do *not* want to read ST and TT simultaneously. It was also a bit of a nightmare for typesetters.

Now, though, such constraints are melting away before the advance of

word processors. Already *interlinear word processors have* begun to make their appearance on the market. With the right software it requires no more than the touch of a button to split the ST and the translation which has been typed interlinearly between it into separate texts, or to reverse the operation and zipper the two back again into an *interlinear bi-text.*

There remains a serious problem with interlinear bi-text, that of *aligning* ST and TT, due to the well-known phenomenon that translations are not exactly the same length as their originals. Unless we do something about it, ST and TT get disturbingly out of step after the first line or two, and the problem increases if we wish to align, as we do, not just the lines but the translation units, which may be shorter or longer than a line. Obviously one could ask the translators to make some spacing adjustments on their word processing screens as they go along, but that would be so lacking in 'user-friendliness' that most translators would not be bothered. I am working on this problem.

Interlinear bi-text may be the most generally convenient 'display mode' of bi-text, but it is not the only one already in established use. In the case of bilingual dictionaries, every entry and sub-entry of which is a piece of bi-text, and so closely welded that ST loses its raison d'être without TT, the convention is that TT is printed immediately following each unit of ST and the distinction between the two is made by contrasting type fonts. Though dictionaries constitute a special kind of text — indeed we do not usually think of them as texts — they are clearly very important ones for translators and we should not neglect them, in designing bi-text applications.

Conclusion

There are types of bi-text and applications of the concept of bi-text which I cannot deal with in this first outline. One which has potential for empirical research on translation, and which also depends on computers for its practical implementation, is *incremental bi-text.* So there is more to follow. Meanwhile I have talked about *bitext* as a theoretical and psychological construct; and as a work tool in its forms of *hyper-bitext* and *interlinear bitext.* Both the construct and the work tools await developers.

In memoriam

More than a decade ago, long before hypertext became a buzz word, a project at the University of Ottawa input what was in effect a hyper-bitext of Canadian federal state law on to a mainframe computer and programmed, in the rudimentary way that was the state of the art in those days. This was the JURIVOC project in the Faculty of Civil Law. Unfortunately its visionary project leader, David Burke, died suddenly while still in his thirties; and despite the best efforts of its sponsor in the Faculty, Me Viateur Bergeron, the project was abandoned for lack of funding.

© Brian Harris, December 1987