INTERLINEAR BITEXT

BITEXT AND INTERLIN-EAR WORDPROCESSING

Earlier this year, Brian Harris, of the University of Ottawa's School of Translators and Interpreters, introduced the concept of bitext, the computer storage of translated texts in juxtaposed source and target "translation units" (Language Monthly #54, p.8; Language Technology #7, p.41).

Here, in response to the enormous worldwide interest those articles aroused, Harris elaborates on his theme.

INTERLINEAR

Bitext is a simple enough idea: the splitting up of a translation's source text into "translation units" – independent phrases such as we translate in one fell swoop, but which provide individual words with enough context to illustrate their usage – and the attachment to each one of the corresponding unit of the target text.

The computer storage of bitext means that each time a segment of source or target text is retrieved, it will be accompanied by its translation in the other language. A real-life example is given below. Computer storage also makes it possible to search out interrelated fragments of text – words or sentences – in enormous quantities, and display them within seconds.

The most obvious and convenient way to enter bitext into the computer, or view it upon retrieval, is with the source text on one line and the target right above or below it. This gives bitext the visible form of old-fashioned "interlinear" translations.

ALIGNMENT

The benefits bitext offers the practicing translator lie mainly in its capacity as a reference tool. Provided you're using a wordprocessor, and your source text is stored in the computer, the whole of your output – along with that of other translators, if you're networked – can be stored as bitext and made available for copying or guidance whenever you have to translate a text similar to one you've done before.

With bitext, nothing in the translator's experience is forgotten or irretrievably filed away. The laborious compila-

tion of glossaries is no longer necessary, because examples of the translation of all the terms you've ever encountered can be assembled on demand. For the same reason, bitext offers terminologists an ideal tool for collecting raw material.

The only major technical problem involved in interlinear bitext is that of aligning source and target translation units, since as a rule they vary in length and position. After all, no two languages have exactly

the same structure or word order.

So far, I know of just one low-cost wordprocessor specifically designed to facilitate the entry and display of interlinear translations: Interlinear Text-processor (IT), the fruit of lengthy development at the Summer Institute of Linguistics (SIL) in Dallas, Texas. IT runs on PC/compatibles, and comes on two diskettes, 5 1/4" or 3 1/2", with a very complete and well-written manual.

IT displays the target text either split up in alignment with the source segments, or as normal continuous text. In addition, the package lets you add further "interlines" of re-

visions or annotations, which, like the source – or "baseline" – text can be made to appear and disappear at will. IT costs $60.

IT's only shortcoming is that it still does not include search and retrieve functions. But authors Garry F. Simons and Larry Versaw are working on one, which they've provisionally called Interlinear Text Search (ITS). Meanwhile, they themselves point out that other software is available for handling retrieval.

MORE FACILITIES

Once the search and retrieval problems have been solved, additional facilities of special use to researchers and teach-

---

"|" marks translation unit boundaries. A search for any word or combination of words in the source text (printed here in bold) retrieves the segment containing it together with the corresponding translation segment (roman).

This enables other translators, or the same translator at some future time, to perceive reusable translations like "unanimously confirms/est unanime dans sa confirmation du" and "awareness in Canada/sensibilisation du public canadien," which would not appear in the dictionaries or termbanks because they are context specific.

```
The Board of PAC unanimously confirms the PAC mandate and
Le Conseil est unanime dans sa confirmation du mandat et du

concept.                          | This includes support to long-
concept  fondamental  du  PAC.|  Le concept comprend un  appui au

term development through the strengthening of African NGOs;
développement  à  long  terme  par  le  renforcement  des  ONG

          | supporting African awareness in Canada
africains;|  un  appui  aux  activités  de  sensibilisation  du

          | which focusses on African abilities and
public  canadien|  qui  accentuent  les  forces  et  habiletés

strengths | and the root causes of current problems;        |
africaines|  et  examinent  les  causes  profondes  de  la  crise;|

encouraging partnerships based on a recognition and respect
l'encouragement  de  partenariats  basés  sur  le  respect  mutuel

for mutual roles| and confirmation of Africans as the agents
                    | et la reconnaissance que les Africain(e)s

of their own development;                          | supporting
sont  les  premiers  agents  de  leur  développement;|  l'appui  à

networking and linkage efforts both in Africa and Canada.  |
la  création  de  liens  et  de  réseaux  en  Afrique  et  au  Canada.|

PAC's emphasis is on African priorities                      |
Le  PAC  met  d'abord  l'accent  sur  les  priorités  des  Africain(e)s|

and on activities which evolve out of the African context.|
et  les  activités  qui  émanent  du  contexte  africain.              |

Networking, and linkage have been identified as priority areas
La  promotion  de  liens  et  la  formation  de  réseaux  sont  des

for PAC                          | and essential to developing
domaines  prioritaires  pour  le  PAC|  et  sont  essentiels  pour  la

true partnership relationships.              |
mise  en  oeuvre  de  relations  de  partenariat.|
```

ers can be added. When the source text has been pre-stored and the translator is using a wordprocessor, it becomes feasible not only to record the finished translation, but also to capture and collate all the intermediate drafting and revision the translator has done.

To this facility can be added the bells and whistles that would record the time taken on each translation unit, as well as interruptions, dictionary lookups – if the dictionaries are also on the computer – and backtrackings. Another suggestion for improving bitext – from Dutch MT developer Toon Witkam – is the color keying of links between semantic elements in the source and target texts.

Bitext therefore offers a complete recording tool for anyone who wants to investigate or diagnose translators' working methods, and one that would not intrude on the translators' normal work, since everything except the wordprocessor would operate in background mode.

The two basic components bitext offers are potentially at hand. It remains for systems builders to integrate them, and for computational translatologists to refine the retrieval and display strategies in order to provide translators with the most relevant information in the shortest possible time.

Brian Harris
Summer Institute of Linguistics
7500 W. Camp Wisdom Rd.
Dallas, TX 75236, U.S.A.
Tel: +1 (214) 298 3331