DLT CLAIMS NEW CONCEPTUAL BASIS FOR MACHINE TRANSLATION

One side of a corpus-based translation: translation units matched against a corpus
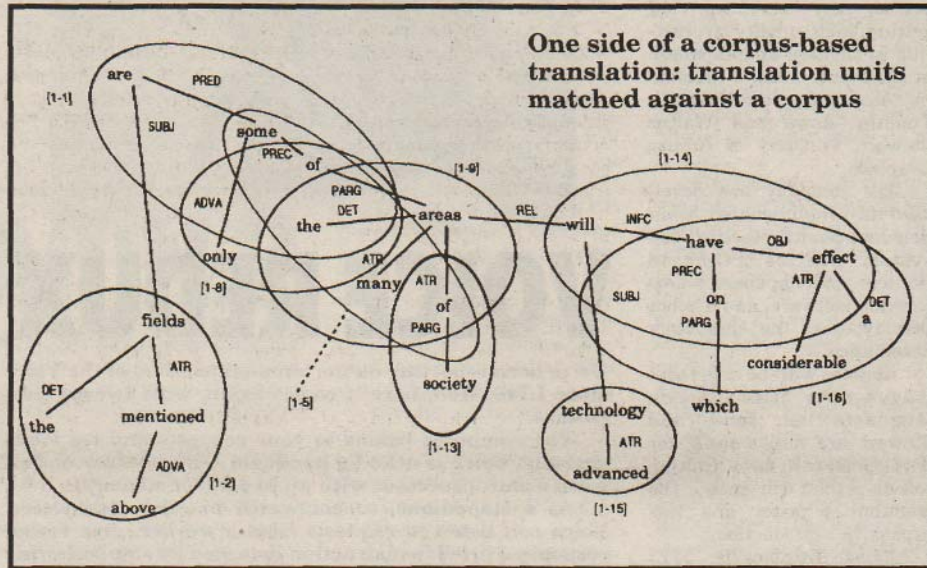
Dutch software developer BSO has announced its "invention" of what it claims to be a "new conceptual foundation for MT: the Bilingual Knowledge Bank (BKB)." The idea is to take a language pair, do away with separate grammars, dictionaries and semantic networks for each one, and replace them with a single, highly structured, corpus of bilingual text.

BKB is the fruit of four years of research by BSO's Distributed Language Translation (DLT) team, which has just completed an English-French prototype. The middleterm aim is to extend this into a twelve-language system – stored on CD-ROM to run on high-end micros and workstations – to be commercially launched no later than 1997.

DLT's accent is on automating not just translation but the development of the machine translation system itself. The team hopes to do this by processing existing bi- and multilingual corpuses, such as EC documents or Canada's parliamentary minutes, rather than compiling more conventional translation dictionaries and grammars.

This bears a resemblance to what Canadian "translatologist" Brian Harris recently advocated in LT/Electric Word (LT#7 and LT#10): "bitext," or using a kind of bilingual hypertext to scan your – or your machine's – previous experience before deciding how to translate a particular expression in a particular context.

Among the advantages claimed for the Bilingual Knowledge Bank are: easy customization (if you only want to translate software manuals, you build your BKB from existing manuals and their translations); full reversibility (either language can serve as the source language); and machine learning (runtime translation can be automatically added to the knowledge bank and thus influence subsequent output). A pilot implementation of the Bilingual Knowledge Base is scheduled for later this year.