POOLING AND THREADING IN TRANSLATION TERM BANKS

Merle D. Tenney
Automated Language Process. Systems, 750 North 200 West, Provo, Utah 84601 USA

Robert A. Russell
Weidner Communications, Inc. 1673 West 820 North, Provo, Utah 84601, USA

Bilingual and multilingual terminology data banks (term banks) have proven an effective aid to translation. The translation problem which has been most resistant to term bank assistance, however, is that of identifying precisely those translations which are appropriate for a given term in a SL text. This is compounded by the related problems of SL homonymy and polysemy and TL synonymy. The greatest challenge at the text level is dynamically specifying the lexicon used in the sublanguage of a document. The solution to this problem depends somewhat on whether terminology is held in one large bank, which promotes sharing and comprehensiveness, or in many smaller banks, which encourages independence and specialization.

The principle approach to the translation identification problem heretofore has consisted in filtering out irrelevant translations on the basis of such accompanying documentation as subject field, source indication, or grammatical code. Some organizations have developed initial capabilities for pooling or threading in order to deal with the sublanguage problem. Pooling refers to the creation of internally cohesive subinventories of a data base. Threading refers to the ability to specify (and access) a logical file made up of a sequence of pools, in which the logical keys are associated with physical records on a first-hit basis.

Our paper further elaborates those concepts and discusses their utility in term banks designed to support translation.

Pools may be organized on any functional principle. They may reflect the administrating body - from standards organization through company division to individual translator. Or they may reflect a domain of application - from technical field through company product line to chapter in operations manual. Some pools are created to override other more general pools, such as SL or TL dialect pools which preempt standard language pools.

The object of threading is the dynamic creation of logical files specifically tailored to the applications at hand. Important features of threading include the ability to access a select battery of pools, the ability to prioritize these pools by their task relevance, typically in a sequence from most specific to most general, and the ability, in interactive applications, to modify logical file records by granting read/write access to one or more leading pools in the operative sequence. Application programs utilize one of three forms of access: single access, which retrieves data solely from the first sequenced pool in which a term appears; multiple access, which makes data available from any of the sequenced pools in which a term appears; and composite access, a special case of multiple access which constructs a data complex from the full set of sequenced pools in which a term appears.

The principal motivation for pooling and threating lies in their use in direct aids to translation - text-related glossary generation and text editing with term look-up and other interactive aids. In addition, they extend the capabilities of maintenance programs, such as glossary publication and foreign language instruction.

They are some drawbacks to the use of pooling and threading: as pools are proliferated, the potential for redundant entries increases; the problems inherent in relating files of

distinct types are compounded; and the need for taxonomies, selectional codes and documentation, and heuristic mechanisms, though mitigated, is retained. On the other hand, use of these devices yields improvements in the identification of pertinent translations, in the flexibility of logical file construction, in the adaptability of available term banks, in the organization and management of terminological data, in the size requirements for diversified term banks, in the portability of existing terminology resources, and in the concomitant potential for added revenue to terminology holders.

References:

Brinkman, Karl-Heinz. 1981. "Machine Aids to Translation." META, Journal des traducteurs/Translator's Journal, 26, 67-75.

Goetschalckx, Jacques. 1979. "EURODICAUTOM." Translating and the Computer, ed. Barbara M. Snell, 71-75. Amsterdam: North-Holland.

Kittredge, Richard. 1978. "Textual Cohesion within Sublanguages: Implications for Automatic Analysis and Synthesis." Paper presented at the 7th International Conference on Computational Linguistics, Bergen, Norway, 14-18, Aug. 1978.

Masterman, Margaret. 1979. "The Essential Mechanism of Machine Translation." Paper read to the British Computer Society, Jan. 1979.

Schultz, Joachim. 1980. "A Terminology Data Bank for Translators (TEAM)." META, Journal des traducteurs/Translator's Journal, 25, 211-29.

Smith, Raoul N. 1978. "Computational Bilingual Lexicography: A la recherche du mot juste." Computer Support to Translation, ed. Jim Mathias, 84-112. Washington, D.C.: Foreign Broadcast Information Service.

Zgusta, Ladislav. 1973. "The Shape of the Dictionary for Mechanical Translation Purposes." Issues in Linguistics: Papers in Honor of Henry and Renée Kahane, ed. Braj B. Kachru, Robert B. Lees, Yakov Malkiel, Angelina Pietrangelli, and Sol Saporta, 912-22. Urbana, Ill.: University of Illinois Press.