MemoQ: A New Approach to Computer-assisted Translation

István Lengyel, Balázs Kis, Gábor Ugray

Kilgray Translation Technologies
{istvan.lengyel;balazs.kis;gabor.ugray}@kilgray.com

Recent surveys and papers on the use of CAT tools among translators (Drugan 2004, Fulford-Granell-Zafra 2004, Somers 2003) have pointed out that only a few translators adopt a complete computer-aided workflow. However, considering translators as a community governed by similar rules and practices and implementing a holistic approach to the translation process yields additional savings which can only be exploited by means of an integrated translation scheme. Kilgray's ambitious MemoQ project – basically an intelligent, language-aware translation memory – will be the first practical implementation of the philosophy discussed here.

The translation market is interested in efficiency, and computer-aided translation is an efficiency tool. Nevertheless, most systems were developed in a modular way, and in the first stage of development, all of them attempted at satisfying the needs of an individual translator working in isolation, in an off-line environment. Later, the Internet challenged translators and changed the market conditions entirely. Translation memories expanded their operations to the Internet, but did not reflect the change in paradigm.

Translation is a project-based activity, and the workflow in a very early stage defines (1) the people working on the project, (2) the resources used within the project.

Translation companies hire project managers to provide for workflow-based efficiency. However, TMs do not take into account the workflow roles. The authors, together with Actiwise Consulting, have developed Forditas.net, a web and e-mail-based workflow automation system. This system – which spans the workflow – can be regarded the vertical network component, because it connects people playing different roles: the coordinator to the translator, the translator to the proofreader, and so on. However, the more we automate this process, from quotations to delivery and invoicing, the more efficient the system will get because of (1) the one-stop web storage of all project resources, preventing data loss, (2) the elimination of time spent on forwarding (waiting time is even more significant), (3) thee partial elimination of file names and full elimination of directory structures, resulting in less confusion.

The horizontal network component, on the other hand, connects users of the same role. Resource sharing and instant messaging contributes to consistency.

Fuzzy algorithms can also be complemented with language-sensitive parsing, providing a drastical increase in efficiency. In morphologically rich languages such as Hungarian, Spanish or Arabic, a language-sensitive operation as basic as word stemming can produce much better results. If we can even parse sentences, and create sentence skeletons (Kis-Gröbler-Hodász, 2004), grammar patterns can also be checked for. However, using intelligent parsing ruins the concept of a translation segment.

An intelligent translation memory, complemented with a multi-dimensional domain system, raises questions. How to keep the interface intuitive if we have at least three results which helps in assessing the quality of match: a fuzzy index, a domain match index, and a grammar index? The translator is only interested in one thing: the best order. Establishing a single composite index is a challenging task, and gives ground to further research.