

PANEL: VOICES OF EXPERIENCE: MT IN OPERATIONAL SETTINGS

Actually using MT is not just a matter of buying a system, installing it, and feeding text through. Real-world text comes with spelling mistakes, bad grammar, missing portions, etc. But perhaps the users don't care, and just want a rough idea of what the text is about. Aspects of the environment surrounding the system and the tasks for which the output is used can contribute significantly to the eventual success (or not) of the venture. This panel will share with the audience their experiences and advice on the real-world operational use of MT in the past, the present, and possibly their hopes for the future.

Moderator: Susan Armstrong, ISSCO University of Geneva

Panelists: Roberta Merchant, US Department of Defense

Kazunori Muraki, NEC Tokyo

Karin Spalink, Internationalization & Translation Services Inc.

Mike Tancelosky, Globalink/MicroTac

Michelle Vanni, Georgetown University and US Dept. of Defense

Michael Zarechnak, Georgetown University

The Cost of MT in Japan

Kazunori Muraki, NEC Tokyo

The usage of Machine Translation Software in Japan has increased substantially in the past two years. The sales of Software Packages for PC (in the unit price range of 100,000 to 200,000 yen) have exceeded 10,000 units annually. Previously, the sales of WS-based MT systems (unit price range 200,000 to 600,000 yen) were fewer than 1,000 units annually. This phenomenon may be viewed as a change of the typical user of machine translation from the office worker to the individual. According to recent findings, rather than a constant need for complete, large-volume translation work in specialized offices and businesses, what has largely increased is the occasional necessity of small-scale translation work. As for the MT systems of the future, it is expected not only that large amounts of data may be automatically translated at high speeds, but that the individual user may easily operate such systems.

Evaluating the Suitability of MT

Karin Spalink, Internationalization & Translation Services Inc.

The integration of MT into the translation production process more often than not turns out to be more problem ridden and less effective than anticipated. In most cases this outcome can be contributed to insufficient suitability evaluation.

An evaluation should never start with researching what is available but with what is needed. MT evaluation is a matching process between translation requirements such as language pairs, domains, text complexity levels and MT features such as hierarchical dictionaries, syntactical and file format capabilities. A second tier of evaluation parameters has to consider the possibility of source text adaptation to MT capabilities and the improvability of MT features in the direction

of the source text particularities. Another important factor is change management or in other words re-utilization of previous translations. The evaluation perspective is also based on the user environment, i.e., whether the user will be an individual translator or a translation agency or in-house translation department. For the latter two institutions an MT system not only has to fit the translation requirements, it also has to fit into the production process, it has to report statistics like number of entries per job, number of translation runs per job, number of file conversions and format recovery transactions and so on to the billing system.

The integration of MT into the production process is a multi-faceted task and only careful consideration of all aspects will give some guarantee of a successful outcome.

Project Cybertrans

Michelle Vanni, Georgetown University and US Department of Defense

Project Cybertrans is an effort aimed at evaluating the effectiveness of select commercial MT systems in handling specific language processing tasks. Such tasks include writing rough draft translations and identifying and prioritizing significant information in a large volume of text. Through contracts and in-house development, existing MT systems are being adapted to respond to the requirements of particular applications. A German MT system for highly structured text and a Spanish system for proper name identification have been delivered. A spell-checker and a multi-flow capability for simultaneous production of multiple translations has also been added.

Certain developers' assumptions have constituted roadblocks to our use of commercial systems. They involve capitalization, grammar, spelling, punctuation, formatting and dictionaries. For example, most systems cannot be used on all uppercase input or text in which diacritical markings are missing or variously encoded; they have difficulty handling telegraphic text, misspellings and non-conventional sentence boundaries, such as asterisks and bullet-markings; and, highly formatted text, such as outlines and forms often cause problems. Moreover, system dictionaries rarely include domain specific words and abbreviations.

Our multi-faceted approach involves contracting MT vendors to adapt their products to include batch-mode dictionary update capability, programming interfaces and UNIX operability. Domain dictionaries are generated from wordlists where possible and text preprocessors are developed to expand abbreviations and correct spelling. MT systems are integrated into the larger application environment and presentation systems are developed for human post-editing.

Our strategy reflects the often stated view that MT need neither be fully automatic nor of high quality to be of use to the end-user now. Our focus is on processing, integration, rapid prototyping and deployment to advance the state-of-the-practice. Our premise is that commercial MT systems are useful for language processing, not necessarily translation, *per se*. We thus focus on integrating these systems into other applications. By deploying currently available systems, we benefit from end-user feedback and from the early addressing of human factors. Our use of imperfect MT is a temporary measure. When available, better systems can and will be folded in.