The PANGLOSS MARK I MAT system

Robert Frederking, Ariel Cohen, Dean Grannes, Peter Cousseau and Sergei Nirenburg
Center for Machine Translation
Carnegie Mellon University
Pittsburgh, PA 15213 USA

The goal of the PANGLOSS project[1] is to develop a system which will, from the very beginning, produce high-quality translations of unconstrained text. This can only be attained currently by keeping the human in the translation loop, in our case via a software module called the AUGMENTOR. The main measure of progress in the development of the Pangloss system will therefore be the gradual decrease in need for user assistance, as the level of automation increases.

The analyzer used in the first version of PANGLOSS, PANGLOSS MARK I, is a version of the ULTRA Spanish analyzer from NMSU [Farwell 1990], while generation is carried out by the PENMAN generator from ISI [Mann 1983]. The Translator's Workstation (TWS) provides the user interface and the integration platform [Nirenburg 1992]. This paper focuses on this use of TWS as a substrate for PANGLOSS.

PANGLOSS operates in the following mode: a) a fully-automated translation of each full sentence is attempted; if it fails, then b) a fully-automated translation of smaller chunks of text is attempted (in the first PANGLOSS configuration, PANGLOSS MARK I, these were noun phrases); c) the material that does not get covered by noun phrases is treated in "word-for-word" mode, whereby translation suggestions for each word (or phrase) are sought in the system's MT lexicons, a machine-readable dictionary, and a set of user glossaries; d) The resulting list of translated noun phrases and translation suggestions for words and phrases is displayed in a special editor window of TWS, where the human user finalizes the translation. At stages a) and b) there is an option of the user being presented by the system with disambiguation questions via the AUGMENTOR. We provide an intelligent environment, the CMAT (Constituent Machine-Aided Translation) editor, for postediting. It allows the user to select, move, and delete words and phrases (*constituents*) quickly and easily, using dynamically-changing menus.

As can be seen in Figure 1, each constituent in the target window is surrounded by "≪" and "≫" characters. If the user clicks with the mouse anywhere within a constituent (between the "≪" and "≫" symbols), a CMAT menu for that constituent appears. It contains the word or phrase in the source text if available, the functions **Move** and **Delete**, and alternative translations of the word or phrase from the source text if any. Using these popup menus, the user moves, replaces, or deletes a constituent with a single mouse action, rapidly turning the list of translated words

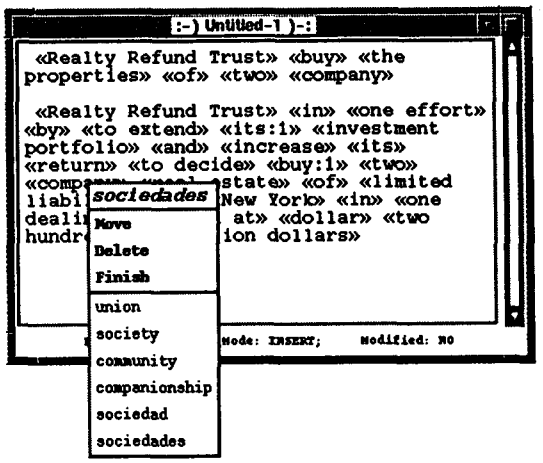

Figure 1: A typical CMAT menu

and phrases into a coherent, high-quality target language text. The user is not forced to use the CMAT editor at any particular time. Its use can be intermingled with other translation activities, according to the user's preferences.

While the above environment is useful as a substrate for a gradual shift to ever more automatic systems, it is also useful as a practical translator's tool. Many minor improvements of the tool itself are planned that should together result in a significant increase in the human translator's comfort and efficiency.

References

Farwell, D., Y. Wilks, 1990. ULTRA: a Multi-lingual Machine Translator. Memoranda in Computing and Cognitive Science MCCS-90-202, Computing Research Laboratory, New Mexico State University, Las Cruces, NM, USA.

Mann, W., 1983. An Overview of the Penman Text Generation System. In Proceedings of the Third AAAI Conference (261-265). Also available as USC/ Information Sciences Institute Research Report RR-83-114.

Nirenburg, S., P. Shell, A. Cohen, P. Cousseau, D. Grannes, C. McNeilly, 1992. Multi-purpose Development and Operation Environments for Natural Language Applications, In Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92), Trento, Italy.

[1] PANGLOSS is a joint project of the Center for Machine Translation at Carnegie Mellon University (CMU), the Computing Research Laboratory of New Mexico State University (NMSU), and the Information Sciences Institute of the University of Southern California (ISI).