TransType : A Target-Text Mediated Interactive Machine Translation System

Philippe Langlais, George Foster, Guy Lapalme, and Sébastien Sauvé
RALI, Université de Montréal
{langlais,foster,lapalme,sauvé,}@IRO.UMontreal.ca

1 Introduction

Interactive Machine Translation (IMT), where a human translator is assisted in real time by a MT system, is a promising idea for improving the productivity of translators. IMT first appeared as part of Kay's MIND system (Kay, 1973), where the user's role was to help the computer analyse the source text by answering questions about it. Most later work on IMT has followed in this vein, concentrating on streamlining the question-and-answer process in various ways.

TransType is based on the novel idea, first proposed in (Foster et al., 1996), of shifting the focus of interaction in IMT from the meaning of the source text to the form of the target text. In this approach, a translation emerges from a series of alternating contributions by the MT system and a human translator, with the translator's inputs serving as progressively informative constraints for the MT component. This relieves the translator of the burden of having to provide explicit analyses of the source text and allows him to translate naturally, assisted by the machine whenever possible.

Our current prototype for English to French translation embodies a simple version of this idea in which the machine tries to guess the next few words the translator will type. The translator is given access to the system's proposed completions and can incorporate them into the target text whenever desired.

2 Description

From a user's viewpoint, TransType operates as follows: a translator selects a sentence and begins typing its translation. After each character is typed, the system displays a new set of proposed completions in a pop-up menu. The top proposal can be accepted using a special key, and other proposals can be selected with the mouse. The translator may reject all proposals by continuing to type normally. Figure 1 shows a screen dump of the interface, which is implemented in Tcl/Tk, a multi-platform script language.

To generate completions, the system relies on a model for $p(w|\mathbf{h},\mathbf{s})$, the probability that a word w will follow a sequence of words $\mathbf{h}$ in the translation of a source sentence $\mathbf{s}$. This allows for the rapid retrieval of the most probable words in the current context. We model $p(w|\mathbf{h},\mathbf{s})$ using a linear combination of a trigram language model for $p(w|\mathbf{h})$ and the classical IBM translation model 2 for $p(w|\mathbf{s})$, as described in (Foster et al., 1997; Langlais and Foster, 2000). To predict more than one word, we allow w to range over certain multi-word units (Langlais et al., 2000a). We have also experimented with Maximum Entropy based models for $p(w|\mathbf{h},\mathbf{s})$ (Foster, 2000a; Foster, 2000b). The completion generator is written in C++ using a flexible object-oriented architecture which facilitates experimentation with different statistical models.

3 Evaluation

We have conducted a theoretical evaluation of TransType on a word completion task which assumes that a translator carefully observes each completion proposed by the system and accepts it as soon as it is correct. Under this optimistic scenario, we have shown that TransType allows for the production of a

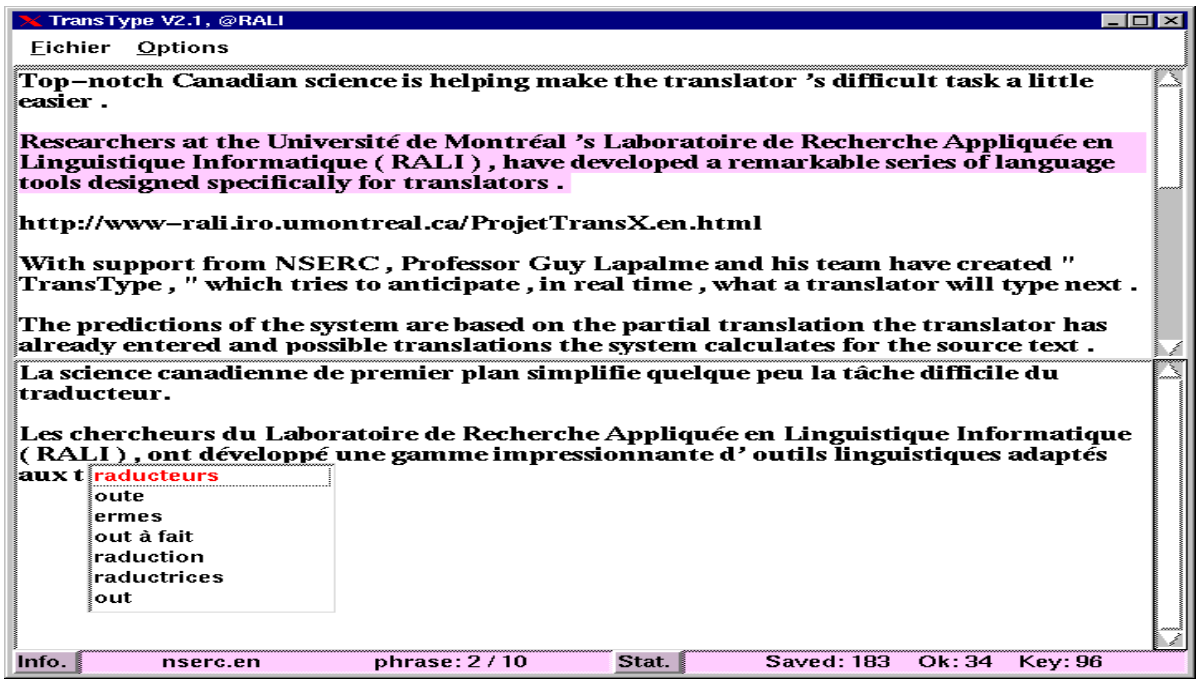

Figure 1: Example of an interaction in TransType with the source text in the top half of the screen. The target text is typed in the bottom half with suggestions given by the menu at the insertion point.

translation by typing less than a third of its characters. We have also conducted a series of more realistic tests involving actual translators (Langlais et al., 2000b). The results of this study, although much more modest, nonetheless support the validity of the approach and encourage further research.

Acknowledgements

TransType is a project funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). It also benefits from an industrial partnership with Machina Sapiens.

References

George Foster, Pierre Isabelle, and Pierre Plamondon. 1996. Word completion: A first step toward target-text mediated IMT. In *Proceedings of COLING 1996*, Copenhagen, Denmark, August.

George Foster, Pierre Isabelle, and Pierre Plamondon. 1997. Target-text Mediated Interactive Machine Translation. *Machine Translation*, 12:175–194.

George Foster. 2000a. Incorporating position information into a Maximum Entropy / Minimum Divergence translation model. In *Proceedings of the 4th Computational Natural Language Learning Workshop (CoNLL)*, Lisbon, Portugal, September. ACL SigNLL.

George Foster. 2000b. A Maximum Entropy / Minimum Divergence translation model. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL)*, Hong Kong, October.

Martin Kay. 1973. The MIND system. In R. Rustin, editor, *Natural Language Processing*, pages 155–188. Algorithmics Press, New York.

Ph. Langlais and G. Foster. 2000. Using context-dependent interpolation to combine statistical language and translation models for interactive MT. In *Content-Based Multimedia Information Access (RIAO)*, Paris, France, April.

Ph. Langlais, G. Foster, and G. Lapalme. 2000a. Unit completion for a computer-aided translation typing system. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*, Seattle, Washington, May.

Philippe Langlais, Sébastien Sauvé, George Foster, Elliott Macklovitch, and Guy Lapalme. 2000b. A comparison of theoretical and user-oriented evaluation procedures of a new type of interactive MT. In *Second International Conference On Language Resources and Evaluation (LREC)*, pages 641–648, Athens, Greece, June.