

TransType: Text Prediction for Translators

George Foster, Philippe Langlais, Elliott Macklovitch, and Guy Lapalme

RALI, Université de Montréal

{langlais,foster,lapalme}@IRO.UMontreal.ca

1 Introduction

The idea of using text prediction as a tool for translators was first proposed in (Church and Hovy, 1993). In such a tool, the system is given access to the source text under translation, as well as its current partial translation, and tries to predict what the translator will type next. Predictions can benefit the translator in several ways: by speeding typing, by suggesting translations, and by reducing the number of errors. Text prediction can be seen as a form of interactive machine translation (IMT), that, in contrast to traditional IMT based on Kay's original work (Kay, 1973), is unintrusive and gives the translator direct control over the target text. Its interactive nature gives it an advantage over other current translators' tools such as translation memories and post-edited MT.

We will demonstrate the TransType prototype for English to French translation developed at the Université de Montréal. This tool offers two different modes of text prediction, as described below.

2 User Interface

The user interface common to both predictions modes is shown in figure 1. It provides a graphical interface typical of translators' workstations, with a split screen for simultaneous display of source and target texts, and standard editing capabilities. To use it, a translator selects a sentence in the source text and begins typing its translation. After each character is typed, the system displays a short pop-up menu containing one or more suggestions for the next word or several words. These suggestions are continually adapted to conform to the translator's input. At any point, the translator has the option of

incorporating a suggestion into the target text with a special keystroke or mouse action, or ignoring the proposals and continuing to type normally.

3 Word Completion

The first prediction mode is word completion. To generate predictions in this mode, the prototype searches through a vocabulary of words and lexicalized multi-word units to find those that are most probable in the current context. Probabilities are assigned by a linear combination of a trigram language model and a variant of the IBM model 2 (Brown et al., 1993) for statistical translation. To speed up the search for the most probable predictions, the vocabulary is limited to a relatively small *active vocabulary* of words and units likely to appear in translations of the current source sentence, according to the IBM model 1. These techniques allow the prototype to operate in real time. An overview of the models, including the details of how multi-word units are induced automatically from parallel corpora, is given in (Langlais et al., 2000).

4 Multi-Word Predictions

In the second mode, TransType tries to anticipate the next several words the translator will type. In contrast to word completion, at most one proposal is placed in the pop-up menu in any given context.

A basic problem with multi-word predictions is that it is not clear how much text to predict in any given context. Longer predictions are more useful if correct, but in general they are less likely to be correct, more costly if incorrect, and take longer to generate. To balance these factors, we relied on an explicit user model that estimates the expected ben-

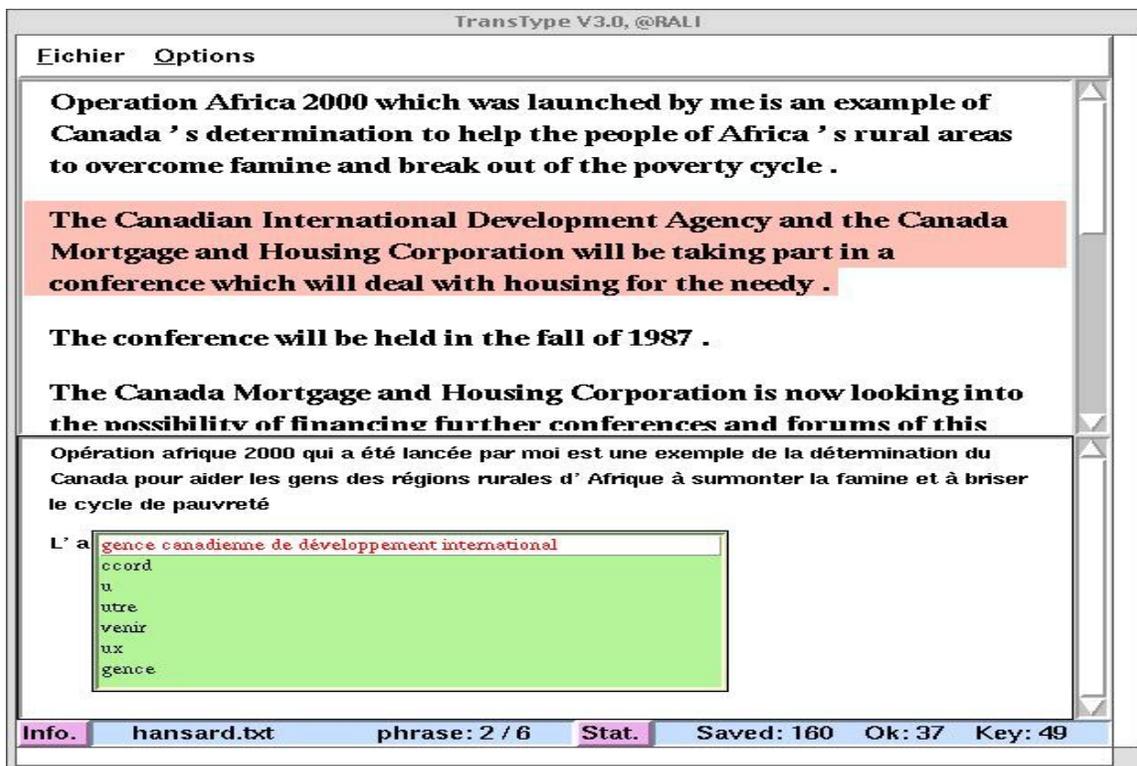


Figure 1: Screen dump for the TransType prototype. The source text is shown in the top half of the screen, and the target text is typed in the bottom half, with suggestions given by the menu at the cursor position.

enefit to the user that will result from each candidate prediction. This also takes into account the probabilities the model assigns in the current context. Qualitatively, the approach can be described as making fewer but better predictions: predictions will tend to be longer in contexts where the translation model is confident, shorter when it is less so, and absent altogether when it is very uncertain. More details are given in (Foster et al., 2002).

To be able to predict sequences of words very quickly (in the interval between successive user keystrokes), we used custom translation models, based on the maximum entropy principle, described in (Foster, 2000). These support very efficient Viterbi-style searches that, in combination with the active vocabulary described in the previous section allow us to make predictions of up to 5 words in real time. The maximum entropy models are more accurate than the linear models used for word completion, and faster than the noisy-channel models typically used for statistical MT.

We look forward to enhancing these techniques and pursuing other research avenues in the TransType 2 project, due to begin this year under the auspices of the European Commission's 5th Framework programme.

References

- Peter F. Brown, Stephen A. Della Pietra, Vincent Della J. Pietra, and Robert L. Mercer. 1993. The mathematics of Machine Translation: Parameter estimation. *Computational Linguistics*, 19(2):263–312, June.
- Kenneth W. Church and Eduard H. Hovy. 1993. Good applications for crummy machine translation. *Machine Translation*, 8:239–258.
- George Foster, Philippe Langlais, and Guy Lapalme. 2002. User-friendly text prediction for translators. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, PA.
- George Foster. 2000. 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.
- Philippe Langlais, George Foster, and Guy Lapalme. 2000. Unit completion for a computer-aided translation typing system. *Machine Translation*, 15(4):267–294, December.