

# TransType2 – An Innovative Computer-Assisted Translation System

**José Esteban and José Lorenzo**

Atos Origin  
Albarracín 25  
28037 Madrid, Spain  
jfernando.esteban@atosorigin.com  
jose.lorenzo@atosorigin.com

**Antonio S.  
Valderrábanos**

Bitext.com  
General Oráa 3  
28006 Madrid, Spain  
asv@bitext.com

**Guy Lapalme**

RALI Laboratory  
Université de Montréal  
C.P. 6128, Succ Centreville  
Montréal, Québec  
Canada H3C 3J7  
lapalme@iro.umontreal.ca

## Abstract

TT2 is an innovative tool for speeding up and facilitating the work of translators by automatically suggesting translation completions. Different versions of the system are being developed for English, French, Spanish and German by an international team of researchers from Europe and Canada. Two professional translation agencies are currently evaluating successive prototypes.

## 1 Introduction

TransType2 (TT2)<sup>1</sup> is an innovative tool for speeding up and facilitating the work of translators by automatically suggesting translation completions. The system uses probabilistic translation and language models to calculate completions that are compatible with translator's input and, furthermore, revises its suggestions in real time with each new character the translator enters. If the system provides a correct suggestion, the translator has only to accept it, thereby saving time in producing the target text. Otherwise, the translator ignores the system's suggestions and continues to type his or her intended translation. TT2 is based on a new Machine Assisted Translation paradigm that sits between fully automatic MT and translation memory in order to significantly increase translator productivity on non-repetitive texts. TT2 is unique in the way in which it combines the strengths of MT technology with the competence of the human translator.

The project is an extension of the TransType project that was developed from 1997 to 2000 by the RALI at Université de Montréal (Foster 1997, Langlais 2002), which demonstrated the interest of target text mediated computer aided translation.

Different versions of the system are being developed for English, French, Spanish and German (with English as the pivot). To ensure that TT2 corresponds to translators' needs, two professional translation agencies are currently evaluating successive prototypes. To date, translation technology has not been able to keep pace with the demand for high-quality translation. TT2 has the ability to significantly increase translator productivity and thus has enormous commercial potential.

TT2 is a RTD project funded by the European Commission under the Information Society Technologies Programme and includes five European partners:

**Atos Origin** (Spain): administrative and technical coordinator, system design and integration.

**Lehrstuhl für Informatik VI**, Computer Science Department, RWTH Aachen - University of Technology (Germany): statistical translation, speech recognition.

**Instituto Tecnológico de Informática**, Universidad Politécnica de Valencia, (Spain): finite-state techniques for translation and speech recognition.

**Xerox Research Centre Europe**, Grenoble (France): corpus provider and statistical translation modeling.

**Celer Soluciones**, Madrid (Spain): evaluation in the operational context of a translation bureau.

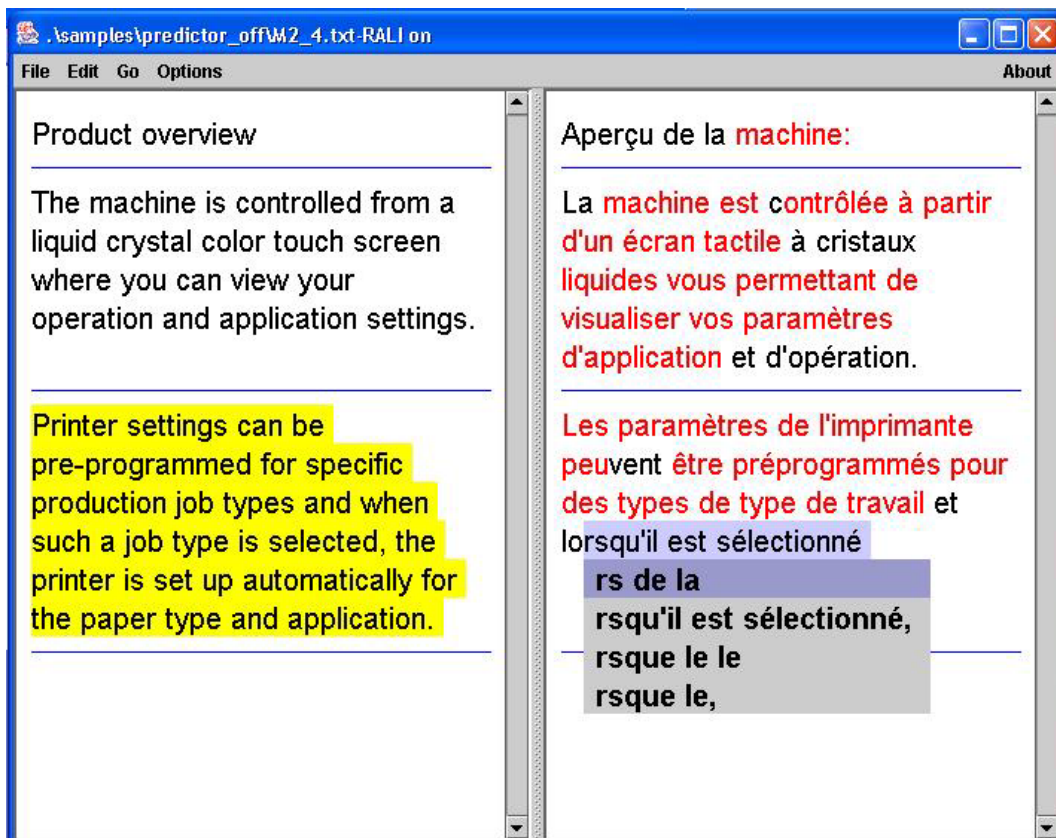
And two Canadian partners:

**RALI Laboratory**, University of Montreal (Canada): user-interface, statistical modeling, evaluation coordination.

**Société Gamma**, Ottawa (Canada): evaluation in the operational context of a translation bureau.

---

<sup>1</sup> For further details, see <http://tt2.sema.es>



**Figure 1.** User-view of TT2 with the source text on the left highlighting the sentence under translation. The translator types in the right pane in which TT2 suggests completions that appear in the menu in real-time. Completions can be accepted either by clicking an item from the menu or by the keyboard. This picture displays in red (appearing in gray in black and white) characters that have been suggested and accepted by the translator.

## 2 TT2 as seen by a translator

TransType is a tool that observes a translator as he or she is typing, tries to predict what will be typed next and displays its predictions to the user. The translator can incorporate these suggestions into the current target text if they are useful, or simply ignore them by continuing typing. The system will then adapt itself to the new text typed by the translator. The suggestions can potentially improve a translator's productivity both by speeding up the keying in of the target text and by contributing to the translation process itself. If the underlying machine translation technology is good enough, TransType2's contributions may reduce the need to consult conventional tools such as a bilingual dictionary, term bank, or translation memory.

The user interface (Figure 1) allows a real-time interaction with the output of the translation/language model to help a translator produce a translation. TransType2's main window is divided into two panes, one containing the source text and another containing the target text. The panes are displayed side by side, with their

contents divided into aligned segments. They are also synchronized, so that scrolling one moves the other in parallel. Many aspects of the main window's behavior and appearance, such as the orientation of the source and target panes, can be changed using the commands accessible from the menu or keyboard shortcuts.

The source pane is read-only in which the only operation allowed is the selection of a new sentence that triggers a new translation in the target window. The target window is a normal text editing window, except that after each character typed by the user, the system displays a pop-up menu of suggestions for completing the current input. If the user types a return or a tab, this suggestion is inserted in the text. Suggestions can be scrolled up or down with arrow keys or selected with the mouse. At initialization time, the user selects the prediction engine to be used according to one of six source-to-target translation pairs and one of the following domains: technical manuals, European Community official documents and official reports of the debates of the House of Commons of Canada (Hansards).

### 3 System Architecture

The TT2 system consists of two major subsystems that interact closely:

**user interface** (UI), written in Java, provides the typing and pointing modalities; a second UI supplements those with speech for operating the prototype via short commands uttered by the user. The user interface also produces a trace of all user-actions that can later be replayed by a special program or analyzed in order to evaluate the effectiveness of TransType2 both in terms of number of keystrokes needed for typing a translation and the various patterns of use.

**prediction engine** (PE), written in C/C++, of which there are multiple realizations available, several per language pair and specific domain (either technical documentation, EC official documents or Hansards). The translation engines developed by research partners are:

**RALI** (French↔English) is a maximum-entropy minimum-divergence translation model (Foster 2000) that proposes multiple completions for the next few words.

**ITI** (French↔English, Spanish↔English) are based on finite-state techniques (Cubel et al. 2003) and suggest a single completion of a whole sentence.

**RWTH** (French↔English, Spanish↔English, German↔English) are statistical based (Och et al. 2003) and suggest a single completion of a whole sentence.

The main communications between the UI and the PE are the following:

1. To initialize the PE, the UI calls a generic create method API function with the appropriate parameters required by each PE and checks its successful completion.
2. Once the user has selected the file he/she wants to work with, the UI produces a list of text segments (sentences) and displays them in the source text pane of the interface.
3. The selection of a source sentence is communicated to the PE by the UI. The sentence becomes the source text context prediction for the PE until the user selects another sentence.
4. The UI communicates to the PE every single modification of the target text: insertion/removal of a new character (letter, digit, punctuation sign or white space) and cursor movements within the target text. The UI communicates left-right one-character-at-a-time movements in the target text area. However, the PE does not take into account the text to the right of the cursor for making its predictions.
5. In response to the request, the PE initiates the search for completions that are eventually returned to the UI for their display.

6. As part of the general exit procedure, the UI calls a generic destroy method API function with the appropriate parameters required by each PE and checks its successful completion.

All communication exchanges between the UI and the PE are initiated by the UI, while the PE is in charge of responding by doing some actual work. This is particularly the case in 5 (producing a list of completions), while the others are more of an informative nature (cases 3 and 4) or can hardly be considered communication exchanges at all: cases 1, 2 (loading a text file and producing a list of sentences) and 6 (termination).

Prediction engines and the speech recognizers are developed and tested under an operating platform (Linux) different than the one chosen for user testing (MS Windows). This duality implies that prediction engines and speech recognizers, while developed under Linux, should be able to run under Windows. The users (i.e. the two translation bureaus) voiced early in the project that TT2 system should run at least under Windows, although preferably it should also run under Linux. TT2 runs currently on both platforms, the dissemination and awareness of the TT2 prototype are broader, and go further than the initial objectives proposed inside the IST project.

Given that developers of the prediction engines and speech recognizers were in favor of using C/C++ as their principal programming language, two practical alternatives were discussed:

- Write code without operating platform dependencies and according to standards, that would allow compilers for both platforms to build functionally equivalent binary versions.
- Employ tools that lessen to a certain extent the requirement of written C/C++ platform independent code, while allowing the porting of code from the Linux to the Windows platform. This was the preferred option and the three PE's actually make use of one of such tool: Cygwin<sup>2</sup>. Cygwin provides a C/C++ compiler for the Windows platform and a library (cygwin1.dll) that gives support to Linux/Unix operating system services under the Windows environment.

The partners responsible for developing the user interface have opted for JAVA as the programming language because of its graphical user capabilities, in particular its text components, which are fully configurable and compatible with external C/C++ programs. This option solves the portability problem, since the resulting code will run under any JAVA-enabled operating system.

---

<sup>2</sup> <http://www.cygwin.com/>

## 4 System requirements

Generally speaking, running the TT2 system demands a high-end personal computer or workstation in order to be able to provide translation completions in real-time and also to be able to incorporate multi-modal user input.

The minimum user equipment is a high-end personal computer running under Windows with a minimum of 1GB of RAM; however, 2 GB of RAM and Windows XP Professional operating system is preferable. If a Linux operating system is used, the kernel version must be 2.4.20 or higher. It is also required to have installed the Java 2 Runtime Environment, preferably version 1.3.1\_09. To produce the PE, cygwin1.dll version 1.5.5-1 is required.

The interface requirements of both scenarios include standard keyboard and mouse equipment; video display capable of resolutions of 1024x768 pixels or higher and voice input hardware (microphone, a headset preferably, and sound card) if the optional speech recognition module is used.

## 5 Evaluation

TT2 is based on the premise that we can improve the productivity of translators by reducing the number of keystrokes needed for entering a translation. Professionals at two translation bureaus are currently testing the prototypes. Even though translators are not used to working with this kind of environment, some of them need about 50% less keystrokes to enter a translation and can thus produce a translation faster. Many user interface improvements suggested by the translators will be included in the next prototypes.

## 6 Conclusion

TT2 is the outcome of a successful cooperation between European countries and Canada to develop an innovative approach to machine aided translation. It is based on advances in statistical machine translation research and on a seamless integration in a word processing environment of the same type as the one currently used by translators.

## 7 Acknowledgements

TT2 is a RTD project funded by the European Commission under the Information Society Technologies Programme (IST-2001-32091). In Canada it is funded by the National Science and Engineering Research Council and the Ministère du Développement Économique et Régional du Québec (Mission Recherche).

## References

- E. Cubel, J. González, A. Lagarda, F. Casacuberta, A. Juan and E. Vidal. *Adapting finite-state translation to the TransType2 project*. Proceedings of the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Applications Workshop Dublin City University Joint Conference, Ireland, 2003.
- Foster G., Isabelle P., Plamondon P. *Target-Text Mediated Interactive Machine Translation*, Machine Translation, 12:1-2, 175-194, 1997.
- Foster G., *A Maximum Entropy / Minimum Divergence Translation Model*, Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 37-42, Hong-Kong, October 2000.
- Philippe Langlais, Guy Lapalme and Marie Loranger. *TransType: Development-Evaluation Cycles to Boost Translator's Productivity*. Machine Translation (Special Issue on Embedded MT Systems), vol. 17, num. 2, pp. 77-98, Feb 2002.
- F.J. Och, R. Zens, H. Ney. *Efficient Search for Interactive Statistical Machine Translation*. Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Budapest, Hungary, pp. 387-393, April 2003.
- Antonio S. Valderrábanos, José Esteban and Luis Iraola. *TransType2 - A New Paradigm for Translation Automation*. MT Summit 2003, New Orleans, USA.