

For more information

- Ordinary mail:
José Esteban Lauzán
Atos Origin
Albarracín 25
28037 Madrid, Spain
- Phone: +34 91 214 8613
- Fax: +34 91 754 3252
- Email:
jfernando.esteban@atosorigin.com
- Web site: <http://tt2.sema.es>



Consortium Partners



<http://www.atosorigin.com/>



<http://www.rwth-aachen.de/>



INSTITUTO TECNOLÓGICO
DE INFORMÁTICA
UNIVERSIDAD POLITÉCNICA
DE VALENCIA

<http://www.iti.upv.es/>



RECHERCHE APPLIQUÉE EN
LINGUISTIQUE
INFORMATIQUE
UNIVERSITÉ DE MONTRÉAL

<http://www-rali.iro.umontreal.ca/>



CELER
SOLUCIONES

<http://www.celersol.com/>



<http://www.societe-gamma.com/>

Xerox Research Centre Europe

xrce

A Xerox Research & Technology Organization

<http://www.xrce.xerox.com/>



TransType2

Computer Assisted Translation

tt2.sema.es



IST-2001-32091

Project summary

Funding: Fifth Framework Programme, IST Programme. In Canada, by the International Cooperative Research and Development grant, NSERC - Canada and MDER - Québec

Action Line: III.3.2 *Natural and multilingual interactivity*

Duration: 36 months (March 2002 – February 2005)

Main themes: Interactive MT, Computer Aided MT

Languages: English, French, German and Spanish

Goals

The main objectives of TT2 are two-fold:

1. To provide a framework in which leading-edge research can be conducted in the area of data driven methods in natural language processing and, more specifically, translation.
2. To develop a practical application for that research which will help to solve a pressing social problem: how to meet the ever-growing demand for high-quality translation.

The research conducted in our project is applied to the development of an innovative, interactive machine translation system. TT2 facilitates the task of producing high-quality translations and making the translation task more cost-effective for human translators.

The TT2 prototype comprises six translation pairs: English ↔ Spanish, English ↔ French and English ↔ German.

Context

There is common agreement within the scientific community in distinguishing two major paradigms in the area of translation automation:

1. The Machine Translation paradigm
2. The Computer Assisted Translation paradigm

Both paradigms revolve around the idea of how the translator and the computer interact during the translation process.

Machine Translation systems aim at producing full translations, hence performing the function of the translator (i.e. replacing him or her). On the other hand, the CAT paradigm aims at using the computer to help the translator to produce better translations more efficiently.

TT2 can be described as an optimal combination of both paradigms, as it keeps the translator in control of the process and lets the machine provide suggestions via a data driven MT engine.

Approach

The TT2 approach proposes translation completions *while* the translator creates the target text. This approach acknowledges two basic facts:

1. Human translators are key to guaranteeing high quality output
2. Machines are key to speeding up the translation process and making it more efficient

Starting points

The results of two previous projects served as a starting point: TransType and EuTrans.

TransType (RALI Laboratory) developed a tool for speeding up and facilitating the work of translators by automatically suggesting translation completions. The system revises its calculations and provides new suggestions that are compatible with each translator's new input. The prototype was trained with the Canadian Hansard corpus for English to French translations.

EuTrans (ITI and RWTH) focused on the development of novel approaches for speech and text automated translation-using data driven methods:

- Finite-state automata language modeling
- Application to written and spoken translation in limited domains

Foreseen advances

TT2 combines different techniques:

- Probabilistic models
- Finite-state automata
- Incorporation of declarative linguistic knowledge

Prediction of translation completions is improved in two main aspects:

- Higher accuracy
- Larger completions (phrases, full sentences)

Productivity and usability is also evaluated in real working conditions with the incorporation of a speech modality.