

TRANSLATOR MTK: A LANGUAGE INDEPENDENT MACHINE TRANSLATION DEVELOPMENT ENVIRONMENT

G Deon Oosthuizen

Department of Computer Science, University of Pretoria
in collaboration with
EPI*USE Systems (Pty) Ltd, PO Box 36453, Menlo Park, 0102, South Africa.
Telephone: +27 12 348 2700 Fax: +27 12 348 2709
Email: deon.oosthuizen@epiuse.co.za

At the MACHINE TRANSLATION SUMMIT VI EPI*USE Systems (Pty) Ltd is launching a new version of its language independent machine translation engine TRANSLATOR MTK. The new version will include new capabilities that will add to the suite of excellent features this development aid has been known for.

1. A language independent transfer system

TRANSLATOR MTK is a powerful, language independent Machine Translation development environment. It is an ideal tool for organisations intending to build their own Machine Translation systems for new languages.

TRANSLATOR MTK is a feature-based transfer system:

- transfer rules consist of syntactic elements, each of which may consist of arbitrary number of features
- these features are altogether defined by and under the control of the grammar developer
- the features may differ from language to language
- the notation is, therefore, *totally* configurable.

The syntactic elements can also be of arbitrary levels of abstraction (e.g. phrase structures that are defined in terms of other structures), again altogether configurable by the developer. Similarly, the morphology and valid lexical constructs are definable by the developer to suit the peculiar needs of his/her language.

2. Platforms

TRANSLATOR MTK is platform independent. The kernel is written in C and C++ and runs under Windows 3.x™, Windows 95™, NT™ and UNIX, over a LAN and on the World Wide Web (it can be seen and tested at www.epiuse.co.za).

It has been integrated with MS Word 6.0™, MS Word 7.0™, WordPerfect™ and Trados™.

3. Background information

- More than fifty people worked on the system and its implementations over the past eight years.

- Close to a thousand licences of products, based on the system have been sold.
- Thousands of pages have been translated for government institutions as well as large corporations using products based on the system.
- About 50% to 100% improvement in productivity has been recorded.

TRANSLATOR MTK was developed in collaboration with the University of Pretoria, South Africa. It is the only commercial system catering for African as well as European languages.

4. Technical features

TRANSLATOR MTK has the following features:

- Semantic, syntactic, statistic and contextual analysis are all combined and comes to bear simultaneously during the transfer phases to counteract the combinatorial complexity introduced by the ambiguity inherent in natural language.
- User dictionaries allow for user-preferred translations, and enables the user to 'teach' the system new words as s/he goes along.
- Domain dictionaries can be given preference, thus allowing for domain specific translations.
- The system keeps track of the current context and changes it dynamically as the context of the text changes.
- It fully supports semantic subcategorisation.
- It combines top-down and bottom-up parsing.
- The system supports the translation of idiomatic expressions and their declensions.
- It supports interactive translation, i.e. where alternative translations are shown and substituted interactively.
- Hierarchical relationships between (syntactic and other) types can be defined - thus enabling inheritance.

The development kit associated with TRANSLATOR MTK has the following features:

- It includes a complete debugging facility that displays all phrase structures detected, all transfer rules that were tested and applied and all current alternatives that are considered by the system.
- It allows the system to learn automatically from simple changes made within phrases by the user.
- It enables the user to analyse the frequency of usage of transfer rules, and to eliminate subsections of rules to speed up translation, i.e. to trade quality for speed if necessary.
- It can display multiple potentially correct analyses of ambiguous sentences and help the user to construct new rules.

5. Languages successfully implemented so far

German/English	French/English	Portuguese/English	Spanish/English
Italian/English	Arabic/English	Afrikaans/English	Swahili/English
Tswana/English	NorthSotho/English	SouthSotho/English	English/Zulu
English/Xhosa			