STYLUS - the MT product line for Russian:
the current state

Svetlana Sokolova

PROject MT Ltd St. Petersburg Russia

common@promt.spb.su

Abstract

The current state of the machine translation system STYLUS is described. The system can produce smooth and accurate translation for more than 80% of source text in the domain chosen. The modular structure of the dictionaries gives the possibility of customising the system to personal needs. The grammar employed is based on ATN-like formalism.

1   Introduction

PROject MT Ltd is a relatively new company in Russia that specialises in designing machine translation systems. PROject MT's team consists of mathematicians, programmers and linguists who have differing experiences in this field. Some of these professionals are making their first steps in applied linguistics but others have been working in this field for more than ten years and were previously involved in various machine translation projects for state organisations.

Starting in 1989, PROject MT's team has worked on its own project to develop tools on PCs for personal translation. "Personal translation" means that the program can be used for limited text types with specialised dictionaries of approximately 20000-25000 entries.

The goal of this project was to design a software product that could consistently produce smooth and accurate translation and would combine high-level linguistic methods with state-of-the-art software design.

As the first stage of the project , a specialised system under the working name "PROMT" (PROgrammers' Machine Translation) was issued in 1991. It was specialised for software documentation translation from English into Russian and was developed by programmers and for programmers.

While working with this specialised system, the basis of techniques for the creation of multilingual translation systems were gradually developed.

The second stage of the project was focused on the application of these techniques to other language pairs and on the performance of the common programming tools. In 1992, PROject MT issued the first versions of the product line named STYLUS that were specialised for business correspondence translation: English-Russian, Russian-English and German-Russian. In 1993, Russian-German was issued. The Romance languages are now under development and the first edition of French-Russian, Spanish-Russian and Italian-Russian versions for business correspondence translation have been launched on the market.

At the current stage of the project, efforts are centred not only on improving the translation quality but also on setting up the STYLUS product line to work with the range of modern office software products on PCs such as Microsoft Word, WordPerfect, AmiPro etc.

2 The STYLUS architecture

The heart of the approach employed is to constitute the software products on a basis which provides consistent and accurate results of translation in the domain chosen without any pre-editing, allows the system to function in real time, and is open for adaptation by the user.

Instead of the generally accepted decomposition of an MT system into analysis, transfer, and generation processes, the STYLUS translators are constructed as a complex cybernetic system employing hierarchical architecture which relies on a hierarchy of objects processed. The layers of the STYLUS hierarchy correspond to units of text such as words, groups, clauses and sentences.

In this architecture, each layer of the translation process deals with the respective layer of the text units. For example, the words translation layer translates the source sentence as a sequence of lexical units to the target sequence of lexical units without any relation between these units. The group layer considers the source sentence as a sequence of groups without any relation between the groups accepted, and so on.

Co-ordination between the layer processes is performed on the basis of inherited and generated attributes of units or processes.

This "object-oriented" approach, being concentrated on analysis and synthesis according to the text unit layer, gives a unique opportunity for the translation process to be stable for any source text. Because, if a layer process cannot recognise the unit structure and cannot translate it respectively, the prior layer's results will be used instead of this layer's result while saving some unit attributes as undefined if necessary.

3 Grammar

For the different layers, different grammar models are employed:

- finite-state automata for the words layer

- ATN-like formalism for the group layer

- procedural frame tools for the clause layer

The ATN-like formalism is based on syntax-controlled translation grammars. This kind of formalism is productively employed for artificial programming languages. A syntax-controlled translation grammar does not imply using syntax attributes only. It handles source and target chains without constructing an intermediate structure representation and the "syntax" means only that there is a set of rules for the chains under consideration.

The group layer process translates the accepted source group chain into a target group chain and generates, not a group structure, but a group translation with group attributes which are inherited from the higher layer (clause translation) or are generated on the current layer.

The clause layer process operates with frame representations of clause structure using such slots as VERB, SUBJECT, OBJECT, INDIRECT OBJECT, DATIVE and COMPLEMENT. It fills in the slots and then performs the necessary transformation to surface expression of the structure recognised.

4   Lexicon

The structure of the lexicon is governed by the system architecture. Because each system is designed as a bilingual system, the dictionaries are also bilingual. Each entry includes: the word stem (or word stems collocation) of the source language, morphological, semantic and syntactic attributes employed for bilingual translation processes of each layer, and the list of translation equivalents that are presented with some target language attributes.

The important feature of the STYLUS translator is that each translator has an intellectual utility to update the system dictionary and to create user dictionaries.

The dictionary interface has been developed to make dictionary work easier for the user. It is a kind of expert system, which automatically generates stems and morphological classifications and also automates the encoding of entries.

The utility suggests, to the user, choices for the government models and for the encoding by analogy. It also gives an opportunity to check up on the results of his choice (by showing the proposed conjugation or declension). The utility frees the user from having to think about homographs which present the whole information according to the currently handled part of speech.

The supplied general dictionary consists of about 20 000 entries, and the specialised dictionaries consist of about 5 000 - 20 000 entries. They can be supplied at the user's request. The system can simultaneously employ dictionaries made by the user himself.

The possibility of adding a user-made dictionary to the system gives a flexible tool for customising STYLUS to work with any particular topic.

5   Requirements and specifications

Currently, STYLUS is implemented on PC-compatibles. It requires about 5 MB on the hard disk for one language pair (English-Russian, Russian-English), and 640KB RAM to run the system.

There are two different versions of STYLUS. The first one can run under MS-DOS (3.3 and higher) and the second one under MS Windows 3.1.

The STYLUS translators are highly productive machine translation systems. The translation speed reaches 150 words per second on a 486 PC. More than 80% of the output text can be used as the final results of translation.

There are dictionaries for business correspondence translation, aerospace engineering, software manuals and oil exploration.

6   Interface

The STYLUS translators can support various interface tasks: import of text from different text editors (Word, WordPerfect, Write), automating the reformatting of texts, the administration of translation jobs, post-editing, pre-editing, background translation and interactive translation.

The Windows version gives the possibility of using "dynamic data translation", which means the calling of the translator from another application using the Dynamic Data Exchange (DDE) mechanism.

7   Experience in the field

The STYLUS translation system first appeared on the market more than three years ago. Now there are more than 5 000 users of STYLUS in different countries. Many of them have widened their experience by upgrading the system from DOS to the Windows environment.

Practice shows that users are quite willing to use the STYLUS translators as a tool which reduces routine work. A few days of initial training are usually sufficient to enable a user to operate the system. Of course, many users do not have the linguistic background to immediately understand the interaction between grammar and lexicon to create their own dictionary, but the user-friendly expert system breaks down this barrier very quickly.

For specific questions which may occur when installing or using the product, a hotline has been installed. When a new version is issued, all users who have sent registration cards to the company are informed about features of the new version. Registered users can also purchase the next version at a significant discount.