

THE TOOLS FOR ADAPTING THE MACHINE TRANSLATION SYSTEM TO INDIVIDUAL NEEDS

Svetlana Sokolova

PROject MT Ltd, P.O. box 632, 199053, St.Petersburg, Russia

ABSTRACT

Stylus is a translation software product for Russian language. The paper describes the structure of dictionary database and feature of the product that give possibility to adapt it to the individual translation needs.

INTRODUCTION

STYLUS is a machine translation system that can translate from Russian and into Russian for some European languages (English, German, French). It is a commercial product with large general-purpose dictionaries and specialized dictionaries in domains of computer science, business correspondence, law, financial documents, space engineering, mathematics, gas and oil industry, telecommunications, medicine, ecology, building and construction.

STYLUS is the registered trade mark of PROject MT, Ltd. PROject MT is a relatively new company in Russia that specializes in designing machine translation systems. The Project MT's team consists of mathematicians, programmers, and linguists who have been developing the STYLUS project since 1989. Now the company staff includes 32 employees.

STYLUS is designed for PC and runs under DOS and Windows. At the moment, STYLUS is the most widespread software for machine translation in Russia. We have more than 10000 registered users of the program. They are individuals, small and big companies such as Rolls Royce, AT&T, Siemens, Lockheed Corp., Chevron, NPO 'ENERGIA', and state organizations, for example, Russian Space Agency, US Navy, NASA, Administration of President of Russia, FBI, and so on.

What purpose does STYLUS serve and why is it so popular? The main idea of the project is to give the end user some tools for getting fast and quality translation.

These tools are designed to provide consistent and accurate results of translation without any pre-editing. They include also the adaptation utilities. The heart of the system is object-oriented architecture of translation algorithms based on the hierarchy of text units processed. The parsers for different layers are employed separately. They function on ATN-like formalisms and procedural frame tools. But STYLUS parsers can be a topic for another paper. Here I would like to concentrate on the interface features for user adaptation (see Fig.1).

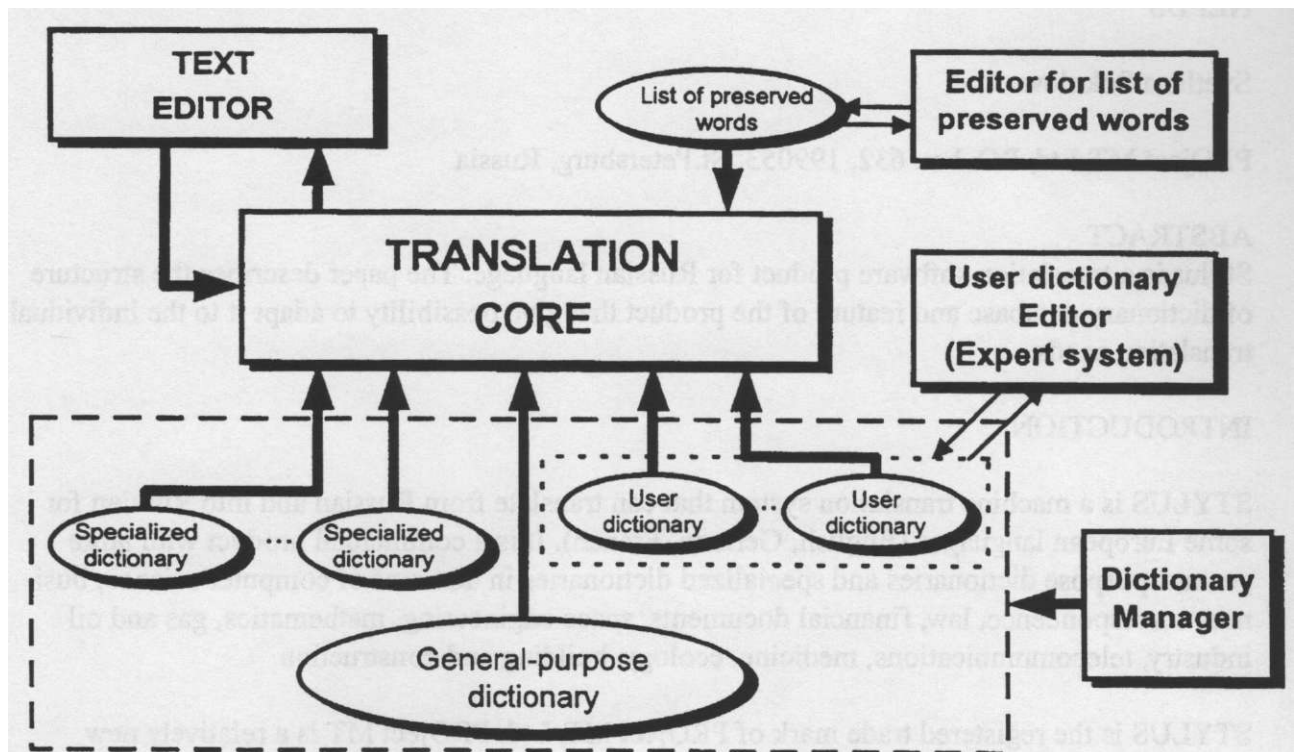


Fig.1 Structure of STYLUS Interface

STYLUS DICTIONARIES

Even if methods, used for the description of translation, allow to get translation of acceptable quality, it won't be enough: the natural language is an open system, and the translation quality depends not only on the correct analysis of a sentence, but also on the correct translation of regular word collocations and terms of domain. Therefore managing and updating the dictionaries are a key system feature for systems that are oriented to the end user. The linguistic data base of STYLUS contains the dictionaries of three different categories: general-purpose dictionaries, specialized dictionaries, and dictionaries, created by the user himself. The general-purpose dictionary contains, as a rule, about 50000 entries and is used as the basic dictionary of the system. It contains entries for the most frequent words of the source language. These entries can have rather complex collection of semantic and syntactic tags that are used for the translation algorithms.

A specialized dictionary contains the terms, specific to the corresponding field of knowledge. The volume of such a dictionary varies from 10000 up to 30000 stems. The user can connect or disconnect the specialized dictionaries in translation process. Connection of the specialized dictionary allows correct translation of the terms, and, besides, translation quality can be better, because in different specialized dictionaries the same lexical units may have different tags, that are employed for semantic or heuristic domain-sensitive algorithms, for example, for correct homonym resolution, predicate analysis, etc.. Specialized as well as general-purpose dictionaries are closed for user adaptation because they can content user-hidden information as well. It allows to prevent basic linguistic algorithms which employ this hidden information.

However, the user has a real possibility to adapt the system to his own texts and to create his own dictionaries for the system. The user dictionary can contain entries both for words and phrases. The dictionary manager provides a possibility to change priority for dictionary access,

to create a new user dictionary, to open user or specialized dictionary for looking through (see Fig. 2).

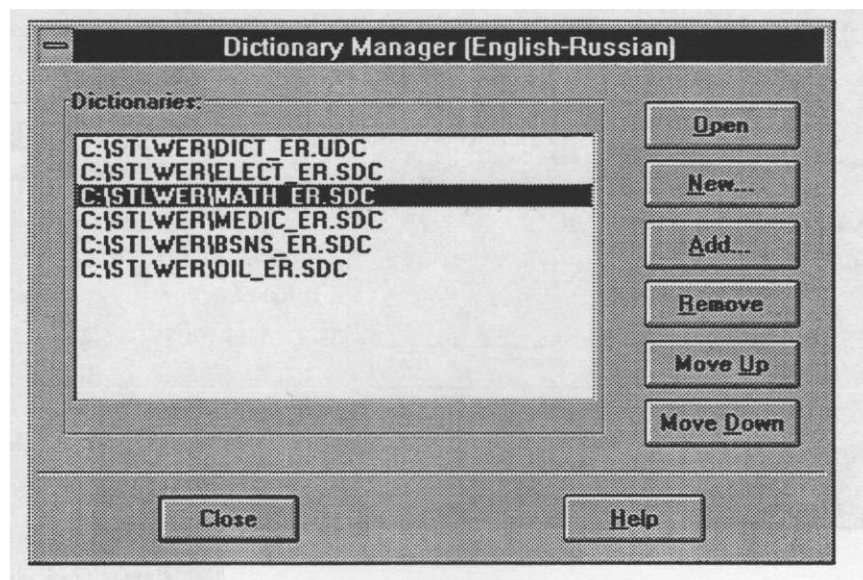


Fig.2 Dictionary Manager

When dictionaries are open, the user can copy an entry to his own dictionary by using "drag and drop" method, change translations and grammar tags of the entry at his own discretion, enter new words and phrases. It is possible to operate simultaneously with several dictionaries. In each session the dictionaries manager can set priority of interrogation of the dictionaries. This feature provides unique possibility to customize the system to a specific topic and domain.

ENTRY EDITOR

Each STYLUS dictionary is a bilingual dictionary, and entries in all the dictionaries have the identical structure. An entry includes a word stem, grammar description of this stem, and translations to the target language. Since the dictionary for machine translation has absolutely specific information and is not similar to the paper dictionary, there is a big problem how to open it to the user without explanation of all the details of linguistic methods employed. To provide friendly access to the entry, STYLUS system includes Entry Editor. This Editor is a kind of expert system that places linguistic information to a questionnaire, automatically forms a declination, gives a set of patterns to input entries (see Fig.3 and Fig 4).

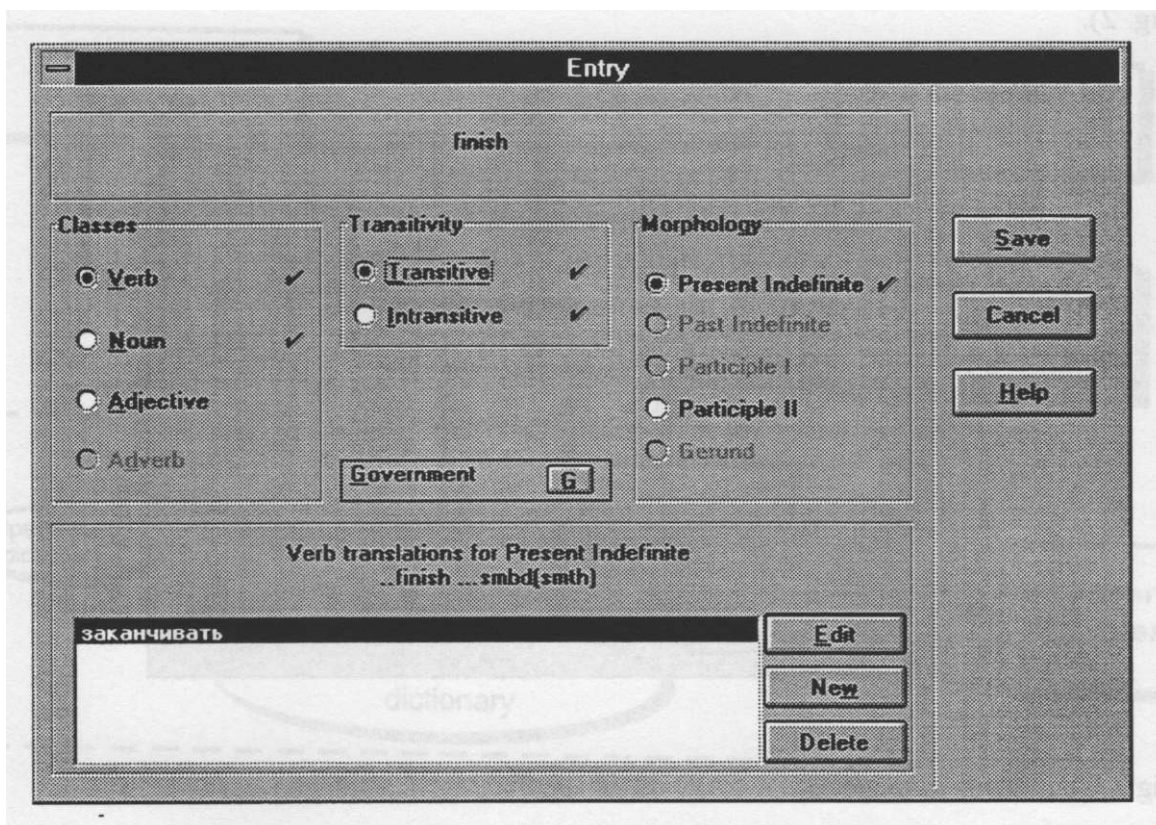


Fig.3 Window of Entry Editor

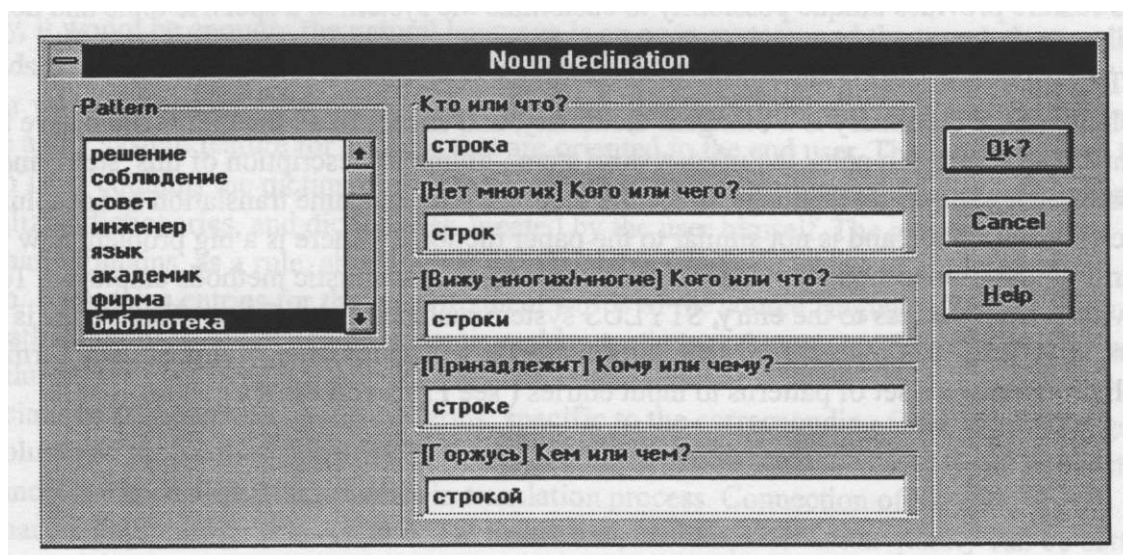


Fig.4 Automatically generated declination

There are two modes for operating with Entry Editor: Beginner and Expert. If the user chooses the 'Beginner' mode, then his interaction with system will be minimal. As the Expert he can actively interfere in updating process, for example, add the government himself, change semantic information, correct automatically produced wordforms. In this case the user should be

familiar both with source and target language grammar. Automatic declination is a very important feature of the system, because STYLUS employs full morphology description for all the languages proceeded: 800 morphological types for Russian, 230 morphological types for English, more the 300 morphological types for German and French. Having this automatic declination enables to define a stem for the entry very fast and reduces routine work for dictionary adaptation.

TEXT EDITOR AND LIST OF PRESERVED WORDS

One more interface feature to make better translation is a special text editor inside the system. This editor has some important possibilities: to switch on the transliteration procedure for unknown words(very helpful for names), to create the list of preserved words, and to scroll simultaneously both source text and target text, if a big text file has been translated. Usually these features are very fruitful. For example, when translating English software manuals into Russian many terms and command names can coincide with commonly used words such as Windows, Copy, Apple, etc. It is not necessary to enter these names into the dictionary. The best solution is to save these names in List of preserved words with a minimal information asked to confirm before saving.

SUMMARY

STYLUS translation system first appeared on Russian market four years ago. Now there are many users of STYLUS in different countries. The user's experience allows to make a conclusion that STYLUS suits both the professional translators and users without high-level linguistic background.