

Brief Guide to PROMT Machine Translation Technology

What is MT? Historical background

Machine translation is the technology for translation of texts of all sorts from one natural language to another automatically, without human intervention. Machine translation technology is based on the in-depth research in the field of linguistics and Natural Language Processing and the application of innovative computer technologies to the task of translating texts from one natural language to another. One of the earliest pursuits in computer science, MT has proved to be an elusive goal, but today a number of systems is available which produce an output which, if not perfect, is of sufficient quality to be useful in a number of specific domains.

Systems for automatic translation have been under development for 50 years - in fact, ever since the electronic computer was invented in the 1940s there has been research on their application for foreign language translations. For many years, the systems were based primarily on direct translations via bilingual dictionaries, with relatively little detailed analysis of syntactic structures. By the 1980s, however, advances in computational linguistics allowed much more sophisticated approaches, and a number of systems adopted an indirect approach to the task of translation. In these systems, texts of the source language are analyzed into abstract representations of 'meaning', involving successive programs for identifying word structure (morphology) and sentence structure (syntax) and for resolving problems of ambiguity.

The abstract representations are intended to be unambiguous and to provide the basis for the generation of texts into one or more target languages. There have in fact been two basic 'indirect' approaches. In one approach, based on the 'interlingua' principle the abstract representation is designed to be a kind of language-independent '**interlingua**', which can potentially serve as an intermediary between a large number of natural languages. Translation is therefore in two basic stages: from the source language into the interlingua, and from the interlingua into the target language. In the other, more common indirect approach (based on the '**transfer**' principle) the representation is converted first into an equivalent representation for the target language. Thus there are three basic stages: analysis of the input text into an abstract source representation, transfer to an abstract target representation, and generation into the output language.

Until the late 1980s, systems of all these kinds were developed, and it is true to say that all current commercially available systems are also classifiable into these three basic system types: direct, interlingual and 'transfer'. The best known of the MT systems for mainframe computers are in fact essentially of the 'transfer translation' type and **PROMT systems also perform the translation of the 'transfer' type**. PROMT translation systems are however improved versions of the type, unlike their predecessors, they are highly modular in construction and easily modifiable and extendable.

PROMT was established in 1991 and during the last 13 years PROMT highly qualified linguists and programmers have been working in close cooperation to develop unique linguistic technologies and software coding principles and create translation software that have no match in the world machine translation industry.

Now PROMT develop a wide range of machine translation tools that can work in PC, Mac, Internet or Intranet environment, either as an autonomous application (word processing for example) or as translation functions directly integrated in an application (Intranet, Internet, Word or Excel).

Advantages of MT

The main advantages of machine translation technology are as follows:

High speed. Machine translation can significantly reduce the time required to translate large volumes of text.

Cost savings. By reducing time normally spent on routine and recurrent translation processes, MT technology can reduce overall translation costs up to 50%.

Terminology consistency. The use of the extensive dictionaries, efficiently organized databases and preprogrammed linguistic algorithms to analyze the input language information provide for the significant gains in translation consistency.

Types of translation demands – what is MT used for?

When giving any general overview of the development and use of machine translation (MT) systems and translation tools, it is important to distinguish six basic types of translation demand.

Gisting. Since MT systems did not (and still cannot) produce high quality translations, some users have found that they can extract what they needed to know from the unedited output. They would rather have some translation, however poor, than no translation at all. With the coming of cheaper PC-based systems on the market, this type of use has grown rapidly and substantially.

MT as a component of information access systems. The integration of translation software into the following databases: (i) systems for the search and retrieval of full texts of documents from databases (generally electronic versions of journal articles in science, medicine and technology), or for the retrieval of bibliographic information; (ii) systems for extracting information (e.g. product details) from texts, in particular from newspaper reports; (iii) systems for summarizing texts; and (iv) systems for interrogating non-textual databases provide the users with the privileged access to information resources

MT for localization. Corporations today are challenged to communicate globally with customers, vendors, partners, and particularly internally with employees located in dispersed geographic regions. MT is used to deliver corporate information that was generated in one language to the speakers of the other languages and to localize company products on the new markets: the excellent performance of MT systems in high-volume translations of routine, recurrent publications make these systems unique localization tools.

MT for communication. The demand for translations of electronic texts on the Internet, such as Web pages, electronic mail and even electronic 'chat' lists, is developing rapidly. In this context, the possibility of human translation is out of the question. The need is for immediate translation in order to convey the basic content of messages, however poor the input.

MT for professional purposes. MT can contribute to the productivity and efficiency of the work of professional translators: special tools for terminology mining and an extensive set of individual user settings integrated in machine translation systems prove to be effective enablers for the professional translators.

MT as a part of automated translation processes. The integration of machine translation technologies within the framework of Translation Memory systems opens an important perspective for the development of fully automated technological chain that can be effectively customized to handle most complicated translation tasks.

PROMT technology: how it works?

The translated text must be as correct as possible in the target language and give back the information contained in the text of origin - this is the most important challenge that the builders of MT systems face nowadays. To deliver high-quality and appropriate translation the providers of machine translation technologies should successfully tackle the following issues:

(i) **Dictionary volumes:** dictionaries are the largest components of a MT system in terms of the amount of information they held. If they are more than simple word lists (and they should be, if a system is to perform well), then they may well be the most expensive components to construct. The dictionaries of MT systems include not only the translation of words and collocations but also special morphological, grammatical and semantic attributes/tags which are used in the analysis and synthesis of texts.

(ii) **Translation modules** that include a sophisticated structure of linguistic algorithms describing morphological and grammatical structures of input and output languages.

(iii) A number of additional software tools for the creation and editing of custom user dictionaries and specialized dictionaries, terminology tools, linguistic editor, customization tools, friendly interface and post-editing tools.

PROMT machine translation systems provide effective solutions to all these problems due to the unique dictionary architecture and innovative linguistic algorithms realized in all PROMT products.

Multidimensional dictionary architecture. The dictionaries in existing MT systems are diverse in terms of formats, coverage, level of detail and precise formalism for lexical description. Different theories of linguistic representation can give rise to different views of the dictionary and different kinds of MT engine also put quite different requirements on the contents of the dictionary.

Since PROMT systems perform the translation of the 'transfer' type dictionaries of PROMT systems give information about source language items and their translations, target language items and their translation. The system uses a formal method of morphology description as a base of dictionary unit retrieval. In PROMT translation systems, morphological description is developed for all translation directions. This description is almost unique in its size. It contains 800 types of inflections for the Russian language, more than 300 types of inflections for German and French languages, and about 250 types of inflections for English. The variety of inflectional forms for every language is stored as tree structures thus providing not only for the effective way of lexical information storage, but also for the effective algorithm of morphological analysis.

This significant dictionary entry headword structure optimization reduces redundant and repeated information in the dictionary: it is no longer necessary to have separate dictionary entries for all related forms (conjugations, declensions and so on) because such information is now stored in the single entry. The morphological model of lexical information storage was also applied for the development of the advisory system for the users who create their personal dictionaries within the system frame. When the user creates new dictionary entry the system automatically extracts the word stem and assigns the appropriate type of inflection.

A new dictionary entry storage and association mechanism implemented in @prompt 7.0 provide for the further improvement in dictionary architecture. The innovative multi-dimensional approach is introduced for the structure of word description: every word or expression is provided with at least one active translation and besides they can be provided with multiple inactive translations within one part of speech. Active translation variants are used in the translation process directly while non-active translation variants can be browsed for the additional information on the meaning of the word.

Any non-active translation variant can be converted into an active one and vice versa. So, the multidimensional dictionary architecture makes it possible to enter unlimited number of translations from electronic and/or printed dictionaries, to reduce the time and effort for users to create and maintain dictionary entries and to work with the unlimited number of translation variants.

PROMT Linguistic algorithms. Instead of the common linguistic approach which assumes the implementation of sequential processes of sentence analysis and synthesis, the architecture of PROMT systems is based on the representation of translation procedures as "object-oriented" processes. Such approach allows for the application of various algorithms on different translation levels.

PROMT system translation algorithms are based on the hierarchical approach that provides for the subdivision of translation process into the interconnected procedures for the different units of linguistic analysis. The following levels are distinguished in the system: the lexical unit level, the group level, the simple sentence level and the compound sentence level. All these processes are interconnected and interact hierarchically according to text unit hierarchy, and also exchange synthesized and inherited attributes. This kind of algorithm arrangement allows for the use of different formal methods for the description of the algorithms on different levels.

Lexical unit description within a dictionary entry is closely interconnected with the structure of system algorithms and is configured not on the basis of "syntax-semantics" opposition, but rather on the basis of text component levels. Thus PROMT systems can work using incompletely described dictionary entries, which is very important when building dictionaries for users who are not familiar with the methods of linguistic analysis.

PROMT customization tools – how to improve the quality of MT?

MT systems work with natural language - a data set that is infinitely varying, ambiguous, and structurally complex. To translate adequately, an MT system must encode knowledge of hundreds of syntactic patterns, variations, and exceptions, as well as relationships among these patterns. Machine translation software should be provided with ever-changing vocabulary and specific semantic knowledge about the usage patterns of tens of thousands of words. The system is to ensure the accurate identification of the parts of speech and grammatical characteristics of words which may, in different contexts, be nouns, verbs, or adjectives, each having many possible translations. Translation also requires a vast store of knowledge about the world, the intent of the communication, and the subject matter.

There is an enduring perception that MT is not yet "good enough" for commercial use. Adopters of MT need comprehensive, easy-to-use tools to improve the quality of their translations. The tools must be accessible to non-developers who know the languages and the business terminology for their company. Among the handful of commercial MT systems available today, PROMT is the one who has tackled the quality issue effectively and successfully: we have developed customization tools that significantly improve the quality of the translation. PROMT products include the following features that contribute to the high-quality translation.

Topic Templates are used to thematically categorize and prioritize several dictionaries, translation preprocessors, and preserved words (non-translatable words) in subject domains. Topic templates include a set of customizable user-specific settings created according to the subject domain. The user can save topic templates and implement them when translating texts on specific topics - this may significantly improve the quality of translation, ensure the consistency of special terms usage and make translation process more effective. The software/system installation process automatically generates a General Topic template with the name General. After the user installs additional specialized dictionaries, other topic template categories are displayed and available. The topic templates are customizable, the user can also create topic template categories with specialized dictionaries installed.

The integration of **Dictionary Editor** in PROMT products provides users with the dictionary editing options that are unique in its range and flexibility: users can view, create and edit information in custom user dictionaries, work with any selected word or word combination, work with source and target translation entries in one or several dictionaries, view and edit extensive grammatical and morphological information, link bidirectional dictionaries, insert and import words from text files. When creating new dictionary entries users can enter not only the translation equivalent itself, but also grammatical information on the word including inflexion and government details that ensure the correct translation of the word. **Dictionary Editor** also enables users to store a list of words that don't require translation - this list may include the company trademarks and product names that can be transliterated to ensure the consistency of corporate branding policy.

Specialized dictionaries can be installed to improve the quality of translation of texts on particular topics. All specialized dictionaries are included with pre-defined topic templates.

Associated Memory manager is used to directly save the source text segments and equivalent translations into AM database for later use. The user can create, delete and manually edit AM translation bases and perform search and replace on database entries.

Professional tools for PROMT dictionaries allows the users to check the state, receive statistical information and view or edit dictionary properties, compare and merge dictionaries of the same translation direction.

PROMT Ltd (Headquarters)
16 Birzhevaya Liniya
Vasilyevsky Island
Saint-Petersburg
199034 Russia

Tel.: +7-812-331 75 40
Fax: +7-812-327 44 83
E-Mail: info@e-prompt.com
Internet: www.e-prompt.com

PROMT (Moscow office)
6a Letnikovskaja Street
Moscow
115114 Russia

Tel.: +7-095-580 48 48
+7-095-509 35 94
E-Mail: moscow@prompt.ru
Internet: www.prompt.ru

PROMT GmbH
Eiffestr. 632
20537 Hamburg

Tel.: +49-40-219 01 140
Fax: +49-40-219 01 143
E-Mail: info@prompt.de
Internet: www.prompt.de

PROMT Terminology Manager (PROMT TerM) provides users with an automated tool for terminology extraction, mining and management.

With all these PROMT effective customization tools MT technology can yield significant cost and time savings, as well as improved consistency in translations.

Synopsis

PROMT (www.e-prompt.com) is a world leading provider of machine translation technologies. The company was established in 1991 and its expertise spans over 13 years of machine translation technologies development. Combining the innovation and technological perfection, PROMT has developed the revolutionary machine translation (MT) technology and became the technological leader in MT industry. PROMT core technology powers customized translation solutions that bring together in-depth linguistic knowledge and cutting-edge computer technologies.

Nowadays PROMT provides the full range of translation solutions: machine translation systems and services, dictionaries, translation memory systems, data mining systems for all kinds of business, professionals and Internet/Intranet. Our policy of continuous innovation in software and service delivery has achieved market share growth in all these sectors. New translation software product line @prompt consists of a wide range of sophisticated software products and solutions that leverage PROMT 13 plus years of enhancements. @prompt translation software is rated as the most sophisticated on the market and ensures the highest possible quality and efficiency of translation processes.

Our customers appreciate our commitment to the continued research and investment in the fields of linguistics and Natural Language Processing: we have customers spanning consumer, SME and enterprise markets, including global industry leaders Xerox, SAP, Lukoil, General Motors, Volvo Car International, Deloitte&Touch and others.

PROMT mission is to enable global business by building new generation MT systems that provide high-quality translation.

PROMT Ltd (Headquarters)
16 Birzhevaya Liniya
Vasilyevsky Island
Saint-Petersburg
199034 Russia

Tel.: +7-812-331 75 40
Fax: +7-812-327 44 83

E-Mail: info@e-prompt.com
Internet: www.e-prompt.com

PROMT (Moscow office)
6a Letnikovskaja Street
Moscow
115114 Russia

Tel.: +7-095-580 48 48
+7-095-509 35 94

E-Mail: moscow@prompt.ru
Internet: www.prompt.ru

PROMT GmbH
Eiffestr. 632
20537 Hamburg

Tel.: +49-40-219 01 140
Fax: +49-40-219 01 143

E-Mail: info@prompt.de
Internet: www.prompt.de

Free Demo: www.online-translator.com