The Third International Conference "Problems of Cybernetics and Informatics"
September 6-8, 2010, Baku, Azerbaijan. Section #1 "Information and Communication Technologies"
www.pci2010.science.az/1/24.pdf

EXPERT SUPPORT SYSTEM OF MACHINE TRANSLATION WITHIN THE FRAME ENGLISH-AZERBAIJANI CONTEXT

Zarifa Guliyeva[1], Minara Manafli[2]

Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]*depart15@iit.ab.az*, [2]*minulya@rambler.ru*

As it is known databases (DB) are most widespread technology for gathering, storage and processing of the huge data objects. However the latter do not allow to structure the data stored in them on the basis of the relations which exist between the facts directly in the real environment. Expert systems being large achievement of modern computer facilities and artificial intellect methods represent the specialized computer system capable to accumulation and generalization of experience of highly skilled experts. They also model reasoning of the experts in some certain area, using the knowledge base (KB) for this purpose, containing the facts and rules from this area and some procedure of a logic conclusion.

Industrially operating systems of machine translation give low quality translation and consequently require post editing. Experimental machine translation systems give more qualitative translation that essentially reduces a share of participation of the person in translation process. So, developed expert support system of machine translation works in an experimental mode and in this connection it can be characterized as experimental bilingual system of MT which uses full morphology, the limited syntax and partial semantics of applied languages. At the given stage of realization it is characterized by following properties:

• Full integrality of the descriptions of source and target languages. The principle of description integrality means that the morphology, syntax and the dictionary are completely co-ordinated with each other by the type of linguistic information replaced in it, and that this linguistic information in all three components is registered absolutely uniformly that is in the same formal languages.

• Declarativeness of the of linguistic knowledge set that is their total independence from algorithm. Declarativeness of the linguistic information set has two advantages at a stage of experimental operation of the system. Declaratively set linguistic model is easy to correct during machine experiments provided that the system simultaneously with translation of each phrase gives out a detailed protocol of its reception. This element is partially written down in expert support system of machine translation.

• Standardized nature of working language description formats. Both of working languages English and Azerbaijani are described under uniform schemes.

• Directivity of the lingware on one subject domain. This property of the system is a direct consequence of incompleteness of linguistic models of working languages. Morphology and syntax of working languages being at the initial stage of design are intended for processing of various scientific and technical texts, that is the wide spectrum of forms and constructions are considered to be met.

• Filling of dictionary database without directivity on a certain subject domain takes place, as for check of rules as a part of the system, word forms of the general and neutral lexicon were chaotically filled in database. Considering the fact that that the entries included in the dictionary are filled enough and fundamental in other words are capable to provide the adequate analysis and translation of the given word. The higher translation quality is the more full and more basic is the model of language composing a linguistic component of machine translation system.

The expert support system of machine translation based on morphological and syntactic knowledge has been created for practical application of the automatic dictionary in the machine translation system, and also checks acknowledgement of theoretically developed principles.

Developed system is realized on the basis of program Delphi 7 applied to creation of control systems by databases and knowledge.

The dictionary volume in the realized version makes 3000 inputs on each language in the dictionary of the combined type. On the basis of the dictionary the database created for a concrete subject domain, however, at an initial stage is realized words of neutral lexicon for check of the rules which are a part of the knowledge base have been included. ESS uses certain initial lexicon, grammar, and also semantics for creation interlinguistic models - interlingva representation (IR) of various word-combinations within the limits of a simple sentence. The given system concerns a wide spectrum of the information systems dealing with processing of texts in natural languages, in particular, in dialogue systems, providing dialogue with databases and knowledge bases in rather free natural language.

Automatic dictionary as a foundation of database

Translation process from one language on another is reduced to overcoming of divergences between languages. In ESS (expert support system) interlanguage divergences partially are taking away on each of analysis stages, and basically at a transfer stage. Difficulty of divergence overcoming is caused by that it is difficult to find translation correspondences between significant elements of languages with various structure. The same correspondences can be coded by various means - morphological, lexical and syntactic. The establishment of translation correspondences at all language levels was a major principle of construction of the automatic dictionary functioning as a part of the machine translation system for any pair of languages. It lays not only in revealing of translation equivalent from one language on another, and also in establishment of correspondences for grammatical categories, syntactic contructions and morpho-syntactic functions of words, word-combinations and sentences of working pair languages.

The automatic dictionary as a part of expert system represents the storehouse of the information used for text processing on the basis of knowledge presented in transformation rules of recognition and generation of grammatical, phonetic and semantic language phenomena.

The automatic binary dictionary is developed as a part of the integrated translation system and used for performance of the following problems:

- serves as the basic tool of search (establishment) of lexical translation equivalents in ESS;
- for work in a dialogue mode the dictionary is integrated into the general lexicographic base of ESS and is the main informatively-directory base.

In ESS as in one of systems of automatic text processing, automatic dictionary is a source of the grammatical information necessary for work of algorithms of automatic morphological and syntactic analyses, and also for work of lemmatization algorithms and knowledge base rules. The latter ensure functioning of the dictionary at the performance of the named functions in any paradigmatic form of a word.

Classification and functions of transformation rules in the knowledge base

The knowledge base contains the information necessary for the decision of problems of demanded type, in the form of rules and the facts. The conclusion mechanism represents the general algorithm of problem decision realized, as a rule, in the form of the interpreter. Its application to the knowledge base on the concrete subject domain set by the expert and to the data about the current situation set by the user gives the decision of a demanded problem.

Knowledge grows out of knowledge process in difference from the information which basic characteristic is their value, the system and interpretability is inherent in knowledge, i.e. knowledge elements are connected among themselves and with ways of their use. As storage units – "elementary knowledge" – act not only as declarative data, but also varying from the expert to the expert. A problem of the linguist is formulation of linguistic rules – language and speech rules. Modeling of linguistic rules as base set predetermines corresponding architecture of expert system. .

The knowledge base carries out mechanism of ESS work on a basis of transformational (generating) grammar representing system of rules experimentally matching structural descriptions to sentences.

Any transformational grammar incorporates morphosyntactic, morphonological and a semantic component. The morphosyntactic component defines infinite set of abstract formal objects, each of which includes all information essential to one interpretation of a specific proposal. The morphonological component defines the phonetic form of the sentence generated by syntactic rules. It correlates the structure generated by a syntactic component, with phonetically represented signal. The semantic component defines semantic interpretation of the sentence. It correlates the structure generated by morphosyntactic component, with certain semantic representation.

In our research drawing up of knowledge base rules is based on a word compatibility technique with each other in the course of formation of their grammatical forms. In other words recognition of those or other grammatical forms of words in the sentence and revealing of their morphological and syntactic-semantic belonging is possible to carry out within the combinations of these word forms with each other. Syntactic division of the sentence into word-combinations with internal subordinate communications, is rare the coordination and an adjunction predetermines work of the syntactic block of the analysis.

Thus, knowledge base of expert system is presented by a set productive rules each of which consists of: antecedent (conditions) and consequent (result) [2]. On a simple language of the user the rule consists of the right and left part. Knowledge represents a complex corrected unification (tree-based) grammar which includes in the structure elements different grammars, such as: context-free grammar (CFG) which is providing the morphological analysis and synthesis and being a basis of analyzer, linear grammar (LG) and constituent grammar which is providing morphological parse and synthesis. So, elements of CFG formalize the description of language model as formal grammar with finite-state set. Elements of LG fix a sequence of chain objects of formal-language model that is the linear sentence structures of formal language model set in terms of grammatical classes of words. In the system "left to right" analysis strategy is applied: search of words, check of conditions, presence or absence of changes on conditions and addition of missing elements formally represent computer realization of finite-state grammar or CFG constructed on LG.

Formally transformation rules which constitute foundation of knowledge bases of expert system, and providing functioning of data and knowledge bases are possible to classify on:

- Recognition rules
- Generation rules
- Substitution rules.

Being based on the above-stated classification, accordingly, the knowledge base can be divided into 3 blocks which in turn continuously co-operate with a database as a part of the analyzer of expert system. For more clear notion of functioning of a database and the knowledge base we will give the stage-by-stage description of automatic processing of expert support system of machine translation.

1. The sentence is entered into the program on a source language where after the initial analysis passes in a database for identification of word forms or word-combinations available as a part of the automatic dictionary

2. Further the sentence passes in the knowledge base where by means of recognition rules there is a morphological analysis of word forms, an establishment of grammatical forms of words and branch of suffixes or other grammatical indicators from a basis. For example, it is possible to carry the description of plural nouns formation to recognition rules in the English language, coming to light by means of detection of those or other affix changes, at performance of the set conditions.

3. Then the established bases come back in a database where find the translation equivalents.

4. The following step: the found equivalents are substituted by substitution rules in syntactic chained structures. The way of representation of a syntactic sentence structure at which groupings (pieces) of the words connected with each other are allocated, is called as its system of components, and each piece in this system making the sentences. Statement rules are written down on the basis of variations of compatibility of parts of speech, and in an operating time of the syntactic block the established grammatical forms of components of the sentence pass to a stage of syntactic processing.

5. At last stage of text processing generation rules carry out synthesis of components of the sentence in translation language. In this block the generating grammar directly making and grammar of final conditions allows to solve a problem of the exhaustive description of set of correct sentences of language with elements of grammar, simultaneously эксплицитно having explicitly specified syntactic laws operating in it.

The developed expert support system of machine translation possesses following features:

1) it is focused on concrete area of expertise, in this case on the translation of texts from English into Azerbaijani;

2) it is capable to make conclusions of data: accurately formulated conditions of rules predetermine the qualities inherent to final automata.

3) it is capable to replenish on a course and as a result of work, covering more and more wide sets of knowledge;

4) it is based on a set of rules, including the practical rules formulated by the expert;

5) possesses practical value

References

[1] Mahmudov M.A. "Sistem of automatic text development". (in Azerbaijani)– Bakı, 2002. 242 p.
[2] Veliyeva K.A. "Avtomatic analysis and synthesis of the text " (in Azerbaijani) – ADD, B., 1996.156 p.
[3] Veliyeva K.A. "Problem of machine translation in turcology." (in Azerbaijani) – *Researches*, 1, B., 2002.
[4] Marchuk Y.N. "Translation modeling methods" M., Nauka, 1985. (in Russian)
[5] Makhmudov M.A. "System of turkish text automatic processing on lexico-morphological level" B., 1991.(in Azerbaijani)
[6] Melchuk I.A. "Theory experience of «Sense <-> Text » linguistic model." – M.: Nauka, 1974. (in Russian)
[7] Bruderer H. "The present state of Machine and Machine-Assisted Translation." – OLB. V.1.
[8] Kemal Oflazer. "Developing a morphological for Turkish// *Proc. of the NATO ASI on Language Engineering for Lesser-studied languages*. – NATO, ASI, July 2001, Ankara.