



# The Use of Machine Translation System for Component Development of Adaptive Computer System for Individual Testing of Students' Knowledge

Alexander Fedosov<sup>(✉)</sup>, Dina Eliseeva, and Anastasia Karnaukhova

Russian State Social University, Moscow, Russia  
{alex\_fedosov, eliseeva.dy}@mail.ru,  
a.karnaukhova@gmail.com

**Abstract.** The article is devoted to the research of the modern machine translation systems as well as different aspects of their using while development and in the course of functioning of adaptive computer system for individual testing of students' knowledge. The contrastive analysis of computer-aided translation and automated translation is reported, computer-based translation in its three main modern sorts: based on rules, statistical and hybrid is estimated in the context of technological opportunities of application by both linguistics specialists and in the field of component development of learning management systems.

In the article the approaches to the problem of adaptation of computer system for individual testing of students in preparation of bachelors of engineering degrees are stated. The peculiarities of the use of machine translation systems for training course development, structuring of test material and in the process of realizing of adaptive algorithm for individual testing of students' educational progress are discussed.

**Keywords:** Machine translation system · Adaptive computer system · Adaptation · Computer testing

## 1 Introduction

Those times when a person knowing a foreign language was appreciated “as valuable as gold” whether being a member of special caste, connected by familial profession of translator, as in Carthage, or a polyglot, being proficient in several languages, as in the following centuries in Europe, recede into the past. The technologies which became habitual in the 21st century take root also into the area of the linguistic translation, bring a possibility of the modern person to the new level of development, allowing to carry out interaction by means of information and communication technologies [1].

A lot of modern researches in the field of the linguistic translation are focused on finding ways of implementation of technologies of its automation, making demands of search of practical, effective methods of implementation of the translations of large volumes of information, at the same time reducing time and financial expenditure.

The scope of the translation by means of the software is very high: common usage of so-called “translators” in the form of various mobile applications, translation systems which became integral parts of Internet browsers, the software for professionals and the audiovisual translation with the advent of which the new era in the area of the translation activity connected, first of all, with the film industry began (in particular, the screen translation).

Computer-aided translation is a translation of texts with the help of computer technology. Translation process is carried out by a man, and the software serves as a tool helping to make a translation for smaller time and with the best quality. The main idea of this process was stated in 1980 by Martin Kay: “By taking over what is mechanical and routine, it (computer) frees human beings for what is essentially human” [2].

So-called “translation memory” gained distribution. Translation Memory is a database which contains earlier translated texts, at the same time if in the set translation phrases or sentences with exact coincidence of earlier translated text from the database, the computer automatically sets it up in the text, and if the coincidence not of 100 percent, the software suggests a translator to make adjustment. The systems of this type allowed to accelerate the translation process, provided uniformity of the translation of terminology that permitted to apply them in the field of technical translation successfully. From the most widespread Translation Memory systems it is possible to distinguish the following ones: Déjà Vu of the company Atril; OmegaT (freely extended system under the license GNU); Trados (widespread in Russia from the T-service company); Star Transit; Wordfast (a set of macros for MS Word), etc [2–5].

Machine translation systems make a translation of texts in the automatic mode without participation of a man, at the same time the following types of forming of interaction are possible:

- a translation with pre-editing that is a man prepares the original text for processing by the computer program (simplifies phrases, eliminates possible discrepancies);
- a translation with post-editing that is the computer program makes a translation of the original text, and a man proofreads;
- a translation with inter-editing that is a man interferes with the work of the translation system in difficult cases of the translation;
- a mixed translation that is the use of systems of pre-editing and post-editing;
- a computer-aided translation that is the use of computer dictionaries in process of translation by a man;
- translation systems with division of labor that is a computer program translates only rigidly structured phrases, all the rest is done by a man.

## **2 The Characteristics of Main Types of Machine Translation Systems**

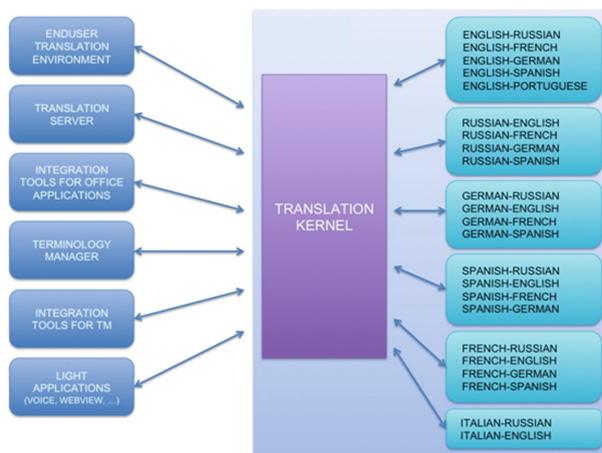
Today the problem of copyright protection in the Russian Federation is actually important more than ever. It is connected not so much with definite gaps and omissions in Russian legislation as mainly with low level of legal culture in the society in general.

It's obvious that the formation of legal culture firstly of youth is primary target, facing the system of education.

Nowadays there are three main types of machine translation systems: rule-based, statistical and hybrid.

**Rule-Based Machine Translation (RBMT)** implies machine translation systems on the basis of linguistic information on original and target languages. They consist of the bilingual dictionaries and grammars covering the main semantic, morphological, syntactic principles of each language [7, 8].

According to Fig. 1 [9] on the basis of these data the original text is being transformed to the target text sentence-by-sentence. Such approach to machine translation is also called classical. The principle of work of such systems is based on connection of structures of the texts (the original and conclusive result – final). These systems have several subspecies which are interlinguistic, transfer systems and the systems of the word-by-word translation.



**Fig. 1.** The main components of rule-based machine translation

Such systems are notable for translation accuracy, they have a possibility of setting for a certain topical area, but, at the same time, demand regular updating of databases, also they are expensive at a development and introduction stage.

The component of a system unites all translation engines for all languages which are supported by a system, and provides the interface between these modules and other modules of a system by means of API and also supports work with such objects of a system as the list of reserved words, translation patterns, mechanisms of protection, etc.

Let's look at the example of the PROMT Translation Software system (Fig. 2) work of such systems where all the translation engines have similar structure and generally consist of three main components which are the following [9]:

- a linguistic database which comprises linguistic data for a language pair: dictionaries, formal description of morphology, formal description of grammar;

- an editor of the dictionary which provides access to difficult dictionary information via the interface of the program;
- a translator which provides translation process.

The computer programming language which is used for development of the software is C++. At the same time for ensuring support of development of modules which functions will work in different operating systems (such as Win32, Linux, WinCE) the technological Single Source Technology method is used.

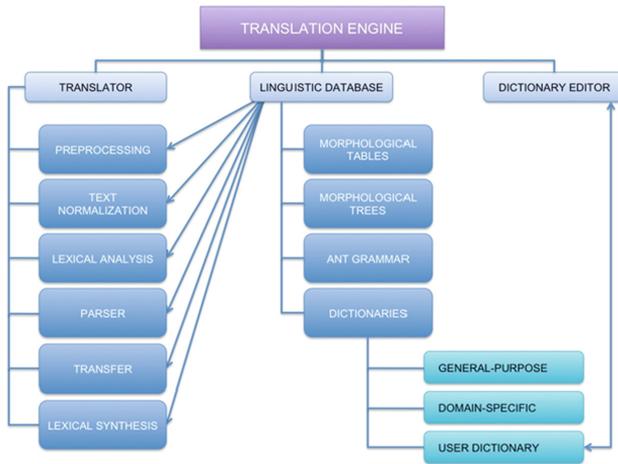


Fig. 2. The mechanism of translation into rule-based machine translation

**Statistical Machine Translation (SMT)** represents a kind of machine translation where the translation is generated on the basis of statistical models which parameters are derivative of the analysis of bilingual text corpus (big text fragments in original language). These systems are self-learning using at the same time earlier received statistics, and it is also not required to retrain a system introducing a new language [7, 10].

Such systems have rather simple setting for various topical areas. Besides, one of the essential advantages of statistical machine translation is rather high translation quality when performing a condition of existence of qualitative text corpuses and at long training of a system. These types of systems do not need presence of professional linguists; IT-engineer builds the system, at the same time his effort is minimized. A difficult mathematical apparatus is a disadvantage, and also for the training of a system it is necessary to have big parallel text corpuses, which essentially affects quality of the final translation. Besides training of a system takes rather long time and while integrating a new language it is necessary to carry out the analysis of a very large number of parallel corpuses.

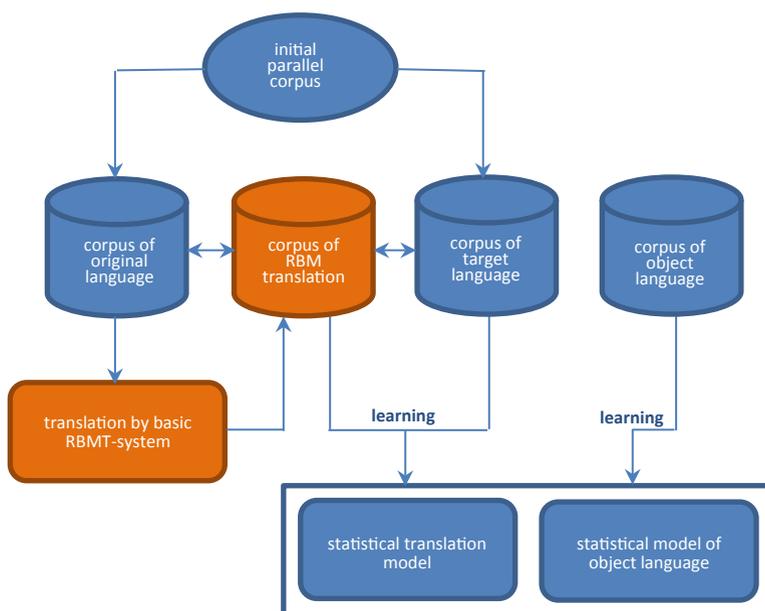
**Hybrid translation technology** assumes [11]:

- use of statistical methods for development of dictionary bases in the automatic way on the basis of parallel corporuses;
- making several possible translations at the lexical level and at the level of syntactic sentence structure of a target language;
- post-editing in the automatic mode and the choice of the best (the most probable) translation from possible ones on the basis of the language model constructed on a certain corpus of a target language.

Hybrid system (Fig. 3), as well as statistical one, go through a process of training at parallel data. This process can be divided into three stages:

- a translation of an initial part of the parallel corpus in original language by a basic RBMT translation engine;
- setting of statistical model of the translation from the computer language into the human one;
- setting of statistical model on the basis of the corpus of a target language.

Hybrid systems allow to unite all advantages of rule-based machine translation and the statistical machine translation systems. The systems of hybrid translation Abbyy Compreno (from the Russian company “ABBYY”), Google Translate, Yandex. Translate became widespread.



**Fig. 3.** The pattern of training of a hybrid system

One of successful developers in Russian market is the company PROMT. In particular, it develops the hybrid system based on introduction of statistical methods into the basic RBMT translation model. RBMT-system is complemented with two components:

- a module of statistical post-editing;
- a module of language patterns.

Statistical post-editing allows to smooth RBMT translation, bringing it closer to a natural language and at the same time keeping accurate structure of the output text. The language patterns are used for assessment of smoothness and grammatical correctness of the variants of translation generated by a hybrid system.

Having carried out the analysis of machine translation systems it is possible to draw a conclusion on existence of essential advantages of such software among which there are first of all the high speed of the linguistic translation of big texts, possibility of their use as a tool for realization of new models of teaching [1].

Besides in modern conditions at universal prevalence of the computer equipment it is possible to use such systems on various devices.

At undoubted convenience of machine translation systems the developers have good scope of work in respect of further development and technological improvement of the pointed systems. First of all the disadvantages connected with incomplete accounting of grammar rules and linguistic methods require correction which will allow to eliminate grammar and lexical errors in final texts.

### **3 The Use of Hybrid Machine Translation System at Realization of Adaptive Computer System of Testing of Students' Knowledge**

The development of education at the present stage is followed by continuous growth of the requirements to its results answering to the increasing inquiries of the society. It relates to personal development of the student, formation of his skills and competences belonging to socially important, defining further success of the person in all spheres of his activity.

One of the possible ways of realization of new models of the organization of the educational process giving opportunities for ensuring availability and high quality of education is widespread introduction of forms and methods of training, adaptive to individual opportunities of students including various forms of e-learning [12, 13].

Learning Management Systems (LMS) give rather flexible opportunities in management of educational process and allow to expand as much as possible the coverage of potential audience that is significant, in particular, at the organization of training of foreign-language students [14]. The use of standard LMS in higher education and vocational secondary education institutions creates opportunities for the following:

- activation of educational process by means of creation of the information educational environment of modern level combining rationally organized information and testing materials with the latest information technologies;

- creation of conditions for active interaction of students with the information educational environment in addition to traditional forms of education;
- personal capacity building of students and training their skills of independent work by its rational organization at the modern level of development of information technologies.

While developing the electronic training courses realized in Learning Management Systems (LMS) and focused on studying of engineering disciplines the question of linguistic translation and structuring of educational content is exclusively relevant that assumes the solution of the following tasks.

1. The development of electronic training materials on the studied discipline in working languages with the use of functional lexis and the support on foreign-language training materials providing the necessary volume of knowledge of a subject (including an integrative component);
2. The development of test materials in working languages with the use of functional lexis on the studied discipline, providing modular check of level of the grasping both the theoretical and practical knowledge.

The disadvantage of traditional approach to the creation of information modules and their giving to the student is full copying of a teaching way when in group training each student is given the training material in the same invariable view. The modern information technologies, including means of computer linguistics by means of variable approach to creation and delivery of material are capable to consider the level of training and perception of the specific student and to align educational process with him flexibly [15].

The use of machine translation systems at realization of the electronic training courses realized in learning management systems allows:

- To activate self-study of information resources on the discipline in other languages;
- To use specific purposes dictionaries with the opportunity of their automatic configuration and connection while studying;
- To make a translation of training materials with maintaining the original structure of the initial document;
- To make a translation taking into account the topical area;
- To maintain the better translations, to create user's dictionaries;
- To create specialized glossaries on the discipline or a cycle of disciplines for the further use at machine translation.

At forming the chosen data file as training materials the information first of all has to be extremely structured. This problem is relevant because both the lack of necessary data, and a chaotic heap of the big, poorly connected among themselves information arrays are harmful to the organization of the educational process (which is especially realized in an electronic form).

The appliance of machine translation systems, in particular hybrid systems, when processing educational information in addition promotes the solution of a problem of structuring information.

For the division of information into separate components and for the establishment of necessary connections between them it is possible to use three basic types of structures:

1. parallel in which separate components are equivalent and are sorted according to any attribute;
2. hierarchic in which the treelike structure grows of a root where knots of the lower level depend only on one overlying knot;
3. network in which there can be connections between the elements of any levels.

The first principle (parallel) is logical to apply to creation of such information structures which are rather simple and also the information objects are a little or almost not connected among themselves.

The hierarchic principle along with existence of connections between the studied information objects reflects some orderliness existing between them. As the process of training is focused (directed) in time, it is the ideally most logical to organize hierarchical orderliness according to the temporary principle that is at higher level there are those objects which are studied earlier.

In real training materials the hierarchic principle should be broken in many cases, using references either between elements of one level or the reference to the previous knots, i.e. actually general structure of content is network.

All main resource materials on each studied discipline entered into LMS according to their destination usually are divided into two main units: information and testing. They can be complemented with other types of materials which play a supporting role in studying of a discipline.

The optimal structure of the information block of the training materials for a student course is the following: all block includes from 3 to 5 sections with size from 20 to 40 or 50 pages of the informative text; each section includes from 3 to 5 subsections with the size from 5 to 12 or 15 pages; each information subsection contains all necessary theoretical data and also examples of the solution of all necessary types of practical tasks. It is possible to accept as the module like an indivisible minimum unit of content the subsection or any logically finished parts [16].

As the electronic training materials are always built up on the basis of literary statement of material, for allocation of modules and sections the following agreements are offered:

1. to accept one paragraph having the independent semantic contents or several come in sequence paragraphs containing the general semantic filling as the module;
2. the information section has one chapter in which the finished part of the studied discipline or several interdependent heads come in sequence are completely opened.

The testing block is made by the control materials tied to modules. They can be generally divided into the following groups:

1. basic recitation in which knowledge of the concepts entered in the module and the main connections between them are checked;
2. advanced theoretical questions in which profound knowledge of theoretical material is checked;

3. typical tasks of the learnt material;
4. advanced tasks, etc.

In the course of individual testing it is possible to realize adaptability of the computer training system [17–19] according to the following attributes.

1. Changing of strategy of recitation depending on answers of the student. It is a traditional type of adaptability; it can be realized due to special structure of an algorithm of recitation.
2. According to the purpose the testing is divided into:
  - educational testing conducted independently for assessment of knowledge by the student himself;
  - control testing carried out under the supervision of the teacher for objective assessment of actual level of knowledge; the important special cases are progress check and final check.
3. According to the depth of recitation that is the depth of knowledge check (one-level, two-level, three-level according to the number of test tasks).
4. According to the volume of material the testing is divided into:
  - subsection (module);
  - section (in particular progress (midpoint) check) or several sections;
  - the whole discipline (in particular final check).

It is possible to realize the adaptability on the second attribute by means of a way of the choice of test material and delivery of results to a student and a teacher. During the educational testing the checked modules are “ordered” by the student himself and the results after each question are told only him. During the control assessment the checked modules are set by a teacher, the final result is told both a student and a teacher.

Here it is possible to use the results of processing of test material by machine translation systems taking into account the level of language training of the student.

It is possible to provide adaptability on the first and third signs by means of application of special separate basic algorithms considering quantity of stages of recitation at the minimum number of basic questions at one stage which are 2.

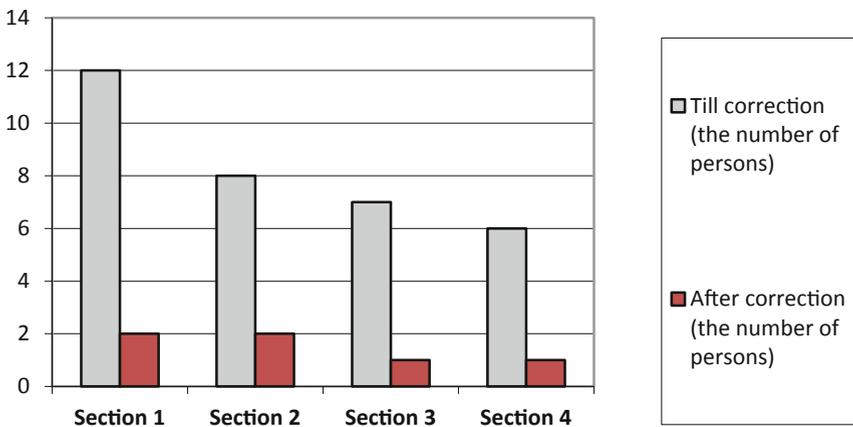
It is possible to realize adaptability on the fourth attribute by large-scale increase in numbers of tests at basic option of an algorithm. Adaptability on this attribute is also can be realized on the basis of inclusion in the checked material of the foreign-language elements and tasks of progress and final check demanding application of machine translation systems.

The offered approach and the corresponding adaptive algorithm of an examination are stated in detail in [13, 20] and in the maximum degree consider essential characteristics of computer process of knowledge check and allow to build an individual learning path of students when using various interactive means, including machine translation systems.

A skilled and experimental examination of a technique of adaptive computer testing with application of the machine translation system during the work with training materials when teaching foreign-language students on the basis of RSSU which showed increase in motivation of educational activity of students of experimental

group, improvement of quality of the acquired knowledge of the studied discipline, increase in level of professional knowledge and skills of independent work with foreign-language training materials was carried out.

The positive reviews were received not only from students of experimental group, but also from the teachers training students at the subsequent related disciplines who noticed more high standard of knowledge of the students of a bachelor degree who mastered a course with use of an adaptive computer system of testing of knowledge and also increase in the general positive working spirit when studying a course of discipline by the developed technique in experimental groups was noted that was expressed in aspiration of students to assimilation of a training material of a course within studies, bigger concentration and accuracy when performing laboratory and test tasks (Fig. 4).



**Fig. 4.** Change in the number of students failing by sections of the course in the experimental group

Thus, the results of the experiment confirmed efficiency of application of adaptive computer testing of knowledge for improvement of quality of professional education of foreign-language students of a bachelor degree.

#### 4 Conclusion and Further Research

The methodical approaches to application of an adaptive computer system for testing offered by authors gives the chance to look in a new way at a problem of development of student courses in a format of interactive training, to adapt the process of computer testing of students' knowledge at the same time according to several parameters and to reach higher educational rates when training students, building an individual learning path.

The use of machine translation systems in course of realization of electronic training courses allows to solve a problem of structuring training materials and

development of test materials for realization of an adaptive algorithm of individual testing more effectively that allows to introduce it widely in process of training of students of engineering specialties including foreign-language students.

The authors see the direction of researches in further integration of machine translation systems and the adaptive computer system of individual assessment of students' knowledge as elements of the virtual educational environment of the university; introduction of machine translation systems, for example, in the form of the program agent, into various applications of educational appointment, including the applications for the automated diagnostics of students' knowledge; expansion of use of mobile platforms.

## References

1. Shevchuk, V.N.: Information Technologies in Translation. Electronic Resources of the Translator-2. Zebra E, Moscow (2013). (in Russian)
2. Kay, M.: The proper place of men and machines in language translation. *Mach. Transl.* **12**, 3–23 (1997). <https://doi.org/10.1023/A:1007911416676>. Research report CSL-80-11, Xerox Palo Alto Research Center, Palo Alto, CA. It is reprinted in 1997
3. Baranov, A.N.: Introduction to Applied Linguistics. Editorial URSS, Moscow (2001). (in Russian)
4. Zubov, A.V.: Information Technologies in Linguistics. Academy, Moscow (2004). (in Russian)
5. Marchuk, Y.: Computational Linguistics. AST Vostok-Zapad, Moscow (2007). (in Russian)
6. Tissen, Y.: The internet in work of the translator. *World Transl.* **2**, 45–62 (2000). (in Russian)
7. Sreelekha, S., Bhattacharyya, P., Malathi, D.: Statistical vs. rule-based machine translation: a comparative study on Indian languages. In: Dash, S.S., Das, S., Panigrahi, B.K. (eds.) *International Conference on Intelligent Computing and Applications. AISC*, vol. 632, pp. 663–675. Springer, Singapore (2018). [https://doi.org/10.1007/978-981-10-5520-1\\_59](https://doi.org/10.1007/978-981-10-5520-1_59)
8. Wilks, Y.: *Machine Translation. Its Scope and Limits*. Springer, Heidelberg (2009). <https://doi.org/10.1007/978-0-387-72774-5>
9. PROMT. Automated Translation Solutions. <http://www.promt.ru>
10. Koehn, P.: *Statistical Machine Translation*. Cambridge University Press, Cambridge (2009). <https://doi.org/10.1017/CBO9780511815829>
11. Xuan, H.W., Li, W., Tang, G.Y.: An advanced review of hybrid machine translation (HMT). *Procedia Eng.* (2012). <https://doi.org/10.1016/j.proeng.2012.01.432>
12. Zenkina, S.V., Sharonova, O.V.: Forms, tools and technologies of interactive educational interaction in distance learning. *Inform. Educ.* **4**(273), 16–19 (2016). (in Russian)
13. Eliseeva, D.Yu., Fedosov, A.Yu., Mnatsakanyan, O.L., Grigoreva, S.V., Dmitrieva, T.V.: Application of Adaptive Computer Testing as a Means of Interactive Learning for Building an Individual Educational Trajectory of a Student. In: *NORDSCI Conference on Social Sciences*, pp. 43–59. SALMA CONSULT LTD, Sofia, (2018). <https://doi.org/10.32008/NORDSCI2018/B1/V1/4>
14. Lapchik, M.P.: The thorny path of E-technologies in education. *Inform. Educ.* **8**(257), 3–11 (2014). (in Russian)
15. Baker, F.B., Kim, S.-H.: *The Basics of Item Response Theory Using R*. Springer, Heidelberg (2017). <https://doi.org/10.1007/978-3-319-54205-8>

16. Gdansky, N.I., Rysin, M.L., Levanov, D.N., Altimentova, D.Yu.: Adaptive Models of Dialogue in the Computer Training Systems: Monograph. RSSU publishing, Moscow (2013). (in Russian)
17. Olsen, J.B., Maynes, D.D., Slawson, D., Ho, K.: Comparison and equating of paper-administered, computer-administered, and computerized adaptive tests of achievement. In: Paper Presented at the Annual Meeting of the American Educational Research Association. San Francisco CA (1986)
18. Wainer, H. (ed.): Computerized Adaptive Testing: A Primer, 2nd edn. Lawrence Erlbaum, London (2000)
19. Weiss, D.J.: New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing. Academic Press, New York (1983)
20. Altimentova, D.Yu., Fedosov, A.Yu.: Improvement of quality of bachelors with application of computer testing and rational modes of correction of knowledge. *Pedagog. Inform.* **1**, 45–53 (2017). (in Russian)