REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

PRINCIPLES OF AUTOMATIC DICTIONARY COMPILING IN COMPUTER TUTORING SYSTEMS

Specialty: 5704.11 – Theory of the Language

Branch of science: Philology

Aplicant: **Aynur Akhmad Alizade**

ABSTRACT

Dissertation for a degree PhD in Philology

Baku – 2023

The dissertation was completed in the Department of Computational Linguistics of the Institute of Linguistics named after. Nasimi of Azerbaijan National Academy of Sciences

Scientific adviser: Assoc. Prof. Doctor of Philological Sciences
Kamila Abdulla Valiyeva

Official opponents: Prof. Doctor of Philological Sciences
Ilham Mikayil Tahirov

Prof. Doctor of Philological Sciences
Zemfira Shahnazar Mammadova

Assoc. Prof. doctor of Philosophy on Philology
Mirvari Mammad Gasimova

Dissertation Council ED 1.06 of the Higher Attestation Commission under the President of Azerbaijan Republic of Institute of Linguistics named after Nasimi of National Academy of Sciences

Dissertation Council Chairman: Prof. Doctor of Philological Sciences
Nadir Balaoglan oglu Mammadli

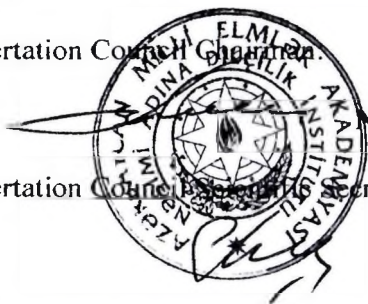Dissertation Council Scientific Secretary: Assoc. Prof. doctor of Philosophy on Philology
Sevinj Yusif kizi Mammadova

Chairman of the scientific seminar: Assoc. Prof. Doctor of Philological Sciences
Gulsum Israfil gizi Huseynova

INTRODUCTION

Relevance and degree of the Topic's scientific development.
Relevance and degree of scientific development of the topic. The
' rapid development of information technology in the 21st century has
created great opportunities for solving the problems of computational
linguistics. One of the problems of artificial intelligence in the world
linguistics directly opened wide horizons in various areas of
computational linguistics. Currently, in such developed countries as
the USA, France, Germany, etc., grandiose works are being carried
out on natural language processing, machine translation (MT),
learning and recognition of languages using a computer, speech
recognition, as well as systems of expert verification and man and
machine dialogue. This dissertation is devoted to the study of
language through a computer. And from this point of view, the
relevance of the dissertation corresponds to the requirements of the
day.

The Decree of the President of Azerbaijan Republic of dated
April 9, 2013 was of undeniable importance for the Azerbaijani
language at the present time. The decree on the "State Program on
the use of the Azerbaijani language in the context of globalization in
accordance with the requirements of the time and the development of
linguistics in the country" indicates the creation of a mechanism for
the development of the Azerbaijani language in the conditions of
globalization in accordance with the requirements of the time,
ensuring the participation of linguists in the creation of modern
information - communication technologies (Main objectives of the
State Program, paragraphs 3.1.1 and 3.1.6).

The State Program for the Development of Linguistics
recommends development of more intensively textbooks for
foreigners studying the Azerbaijani language and Azerbaijanis who
want to learn a foreign language, paying special attention to the
creation of relevant Internet resources, electronic and interactive
textbooks to ensure flexibility in the correct use of the Azerbaijani

3

INTRODUCTION

Relevance and degree of the Topic's scientific development.
Relevance and degree of scientific development of the topic. The rapid development of information technology in the 21st century has created great opportunities for solving the problems of computational linguistics. One of the problems of artificial intelligence in the world linguistics directly opened wide horizons in various areas of computational linguistics. Currently, in such developed countries as the USA, France, Germany, etc., grandiose works are being carried out on natural language processing, machine translation (MT), learning and recognition of languages using a computer, speech recognition, as well as systems of expert verification and man and machine dialogue. This dissertation is devoted to the study of language through a computer. And from this point of view, the relevance of the dissertation corresponds to the requirements of the day.

The Decree of the President of Azerbaijan Republic of dated April 9, 2013 was of undeniable importance for the Azerbaijani language at the present time. The decree on the "State Program on the use of the Azerbaijani language in the context of globalization in accordance with the requirements of the time and the development of linguistics in the country" indicates the creation of a mechanism for the development of the Azerbaijani language in the conditions of globalization in accordance with the requirements of the time, ensuring the participation of linguists in the creation of modern information - communication technologies (Main objectives of the State Program, paragraphs 3.1.1 and 3.1.6).

The State Program for the Development of Linguistics recommends development of more intensively textbooks for foreigners studying the Azerbaijani language and Azerbaijanis who want to learn a foreign language, paying special attention to the creation of relevant Internet resources, electronic and interactive textbooks to ensure flexibility in the correct use of the Azerbaijani

language (paragraphs 4.1.14; 4.1.15; 4.1.16).

The state program specifically indicates priority tasks in this area:

- Development of new improved spelling, explanatory, phraseological, terminological, translational, encyclopedic and frequency electronic dictionaries of the Azerbaijani language (6.4.2);

- Strengthening of activity on creation of Internet resources, electronic and interactive textbooks in the Azerbaijani language (6.4.3);

- Creation and development of Internet technologies, machine translation systems and other modern applied linguistic technologies to ensure wider use of the Azerbaijani language (6.4.7).

The requirements for science, technology and culture put forward the learning process to one of the leading places in all spheres of human activity. Increasingly important are the methods and means of training from the simplest standard methods of training to sophisticated and non-standard systems.

Modernization of the education system, considering the involvement of information and communication technologies (ICT), is necessary for modern conditions of informatization of society. Having studied the state of ICT using problem in the field of teaching foreign languages, we can conclude that the effectiveness of the use of computers depends not only on the methods and forms of application of these technologies, but also on how competently the researcher knows the methodology of working with them, considering the computer technologies he uses. educational programs. There are ample opportunities to solve professional problems for scientists and volunteers who wish to learn foreign languages, while using and relying on communication technologies as additional means. Therefore, in the context of the emergence of the "information society", the relevance of the topic lies in improving the efficiency of management in the field of scientific research, automating learning processes and introducing information and communication technologies (ICT) and expert tutoring systems (ETS), which are currently used in various fields and for different

goals of language learning.

One of the main elements of CTS is an analytical processor that acts as an intermediary between the user, the database and the knowledge base as part of the system. The classical structure of the analytical processor consists of three database blocks, a control block and a knowledge base.

Computer systems are used as a mechanism for representing knowledge, organizing a dialogue between the user and the system, which, at the request of the user, provides a course of reasoning when solving a learning problem in a form acceptable to the user being trained. The knowledge base, which is an integral part of the system, contains methods and procedures for managing the learning process, as well as a description of knowledge that is collected as a set of rules and facts in a formal form. With the help of the control block, the means of evaluating the effectiveness of training in the form of tests or other methods of monitoring the assimilation of the material are implemented.

This block, in turn, consists of a statistics subblock of correct and incorrect answers, a diagnostic test block, and a general testing subblock. The database is a block that implements the means of storing information about the subject area and the learning process. The database is represented by an automatic dictionary (AC), which is a repository of various types of formalized information necessary for more effective learning. The interface implements the means of audiovisual and sensory interaction between the system and the user.

The automatic dictionary (AC) is the main part of the database, which lists all used lexical units of the computer tutoring system, indicating syntactic, morphological and semantic features. AD can be used in the learning process, both as an independent program and as part of a learning system. AD, being the main component of the database, is a multifunctional subsystem with a branched structure.

Object and subject of the research. The object of the study is an automatic dictionary in a computer tutoring system (in the English-Azerbaijani language environment), the lexical-syntactic structure of the language within the framework of the task of their

5

formal description, which is necessary to build a program for automatic text processing. The subject of the research is the development of the optimal structure of the automatic dictionary and its functioning as part of the CTS.

Purpose and tasks. The purpose of the study is to develop principles for compiling the optimal structure of an automatic dictionary, the principles of its functioning as part of the CTS. In accordance with the purpose, the following tasks were set in the work:

- Analysis of the structure of existing computer tutoring systems to determine the requirements for building a tutoring system.

-Analysis of the purpose, structure and principles of functioning of the automatic dictionary in existing systems of tutoring a foreign language.

- Development of principles and methods for the formation of the optimal structure of an automatic dictionary in a computer tutoring system.

- Conducting a comparative analysis of modern and non-standard methods of tutoring a foreign language.

- Development of principles for using an automatic dictionary in an expert tutoring system.

- Selection of teaching methods that are optimal for automating the learning process.

- Formalization of data selected in the dictionary entry AD.

- Development of a lexical formal model of the AD in the tutoring system.

- Development of an expert system knowledge base based on lexical-syntactic and semantic information.

Research methods. The research methods are based on the use of the information-retrieval system approach, the technology of linguistic text processing and expert systems. Researchers in the field of teaching a foreign language, systematically moving from the simplest methods to more complex ones, are gradually approaching to such a representation of the tutoring system that meets modern standards. However, completely relying on the computer in the

6

learning process is not realistic even on modern computers. The fact is that any improved models provide the development of only certain language skills.

Defense provisions. According to the result of the provided research the following provisions were defined:

1. To provide information about the types of electronic dictionaries of classical computer tutoring systems (similarities and differences);

2. To provide information about dictionaries by region (by specialty) (to include in the linguistic database);

3. By morphological analysis, clarify to which part of speech the word forms included in the base and the corresponding features belong;

4. To determine combinations of selected word forms;

5. To explore the polysemy of word forms (antonyms, synonyms, phraseological expressions, terminological expressions, abbreviations and sentences);

6. To study the formal description of the structural composition of the dictionary entry of an automatic dictionary in a computer tutoring system;

7. To present the formalized data of the automatic dictionary in various functional blocks of the computer tutoring system;

8. To present the architecture of expert tutoring systems;

9. To present a contextual database in expert tutoring systems.

The scientific novelty of the work is as follows:

- theoretical foundations for constructing the optimal structure of an automatic dictionary have been developed;

- a methodology for constructing a contextual dictionary as part of an expert tutoring system, the dictionary entry of the system includes a wide range of formalized linguistic information necessary for the effective presentation of material, consolidation and control of acquired knowledge was proposed for the first time;

- the principles of selecting formal features and the methodology for creating a contextual dictionary by selecting formal features into dictionary entry for various parts of speech, which are

formalized linguistic information of a language pair (English and Azerbaijani), acceptable for storage in a database and processing in a knowledge base were proposed for the first time;

- a formal model of a dictionary entry for the contextual AD in the tutoring system has been developed;

- a new approach to building a database of an expert tutoring system based on a contextual dictionary is proposed;

- a new formal model of a multifunctional dictionary that can be transformed into various subspecies of dictionaries, and has the intellectual ability to collect information in the "pocket" subbase as an auxiliary tool that provides a more efficient process of learning vocabulary is proposed.

At present, the predominant part of learning aids is presented in the form of information not only in book form, but also in the form of information technology tools. For the effective representation of knowledge in the learning system, it is necessary to consider the introduction of ICT into the learning process itself. Language is a multifaceted phenomenon, in the study of which you need to use all the necessary skills, such as pronunciation, vocabulary, spelling, grammar and colloquial speech.

Despite the abundance of implemented gaming and developmental works in the field of learning a foreign language, it has not yet been possible to achieve perfection. The successful solution of these problems requires a constant search for new approaches and technologies.

The leading role in the computer tutoring system plays an automatic dictionary, which is developed as part of an integrated lexicographic database. According to a review of past studies, multilingual or very large linguistic databases play a major role in computational linguistics and information technology.

The methodology for constructing an automatic dictionary is based on the definition of the goals and objectives of the study course, including such features as trained audience, subject area, trained linguistic knowledge and skills in the selection of methods and techniques of teaching a foreign language. The methods should

be optimal in automating the learning process and aimed at developing certain knowledge and skills, creating a script for a computer tutoring system and its corresponding text database based on the content of the lesson material, selecting lemmas and building a dictionary of an automatic dictionary in a computer tutoring system, formalizing data, selected in the dictionary entry of an automatic dictionary for various parts of speech and on the presentation of formalized AD data in various functional blocks of a computer tutoring system.

Practical value and implementation of results. A method for constructing an expert system for teaching English vocabulary to researchers is proposed on the basis of the theoretical principles for the development of the optimal structure of the AD, proposed in the dissertation work. A wide range of the knowledge base of the expert system (ES) includes formalized linguistic information of linguistic units of dictionary entries, which makes it possible to increase the efficiency of the learning process. This system can be used to study all levels of English (from basic to advanced) in a scientific environment.

Approbation of work. Approbation and application of the results of the dissertation work. The main provisions of the dissertation work and the results of the study were reported and discussed at international and republican scientific conferences, articles were published in various scientific journals. 16 publications have been published on the topic of the dissertation; 8 articles, 8 reports in conference proceedings.

The name of the organization where the dissertation work was carried out. The dissertation work was completed at the Department of Translation of the Baku Eurasian University and was discussed at the Department of Computational Linguistics of the Institute of Linguistics named after. Nasimi of ANAS.

The total volume of the structural sections of the dissertation separately and the total volume with a mark. Introduction 8 pages; Chapter I 35 pages; Chapter II 35 pages; Chapter III 48 pages; conclusion 2 pages, list of references 30 pages,

list of abbreviations 2 pages, total volume 161 pages - 201,179 characters.

THE BASIC CONTENT OF THE RESEARCH

The **introduction** substantiates the relevance of the work, its scientific novelty, defines the object and subject, goals and objectives of the study, characterizes the scientific state of the problems under study, outlines the theoretical and practical significance of the research results, formulates the main provisions submitted for defense.

In the first chapter "The purpose of a computer tutoring system in the study of a foreign language" the chronology of development and classification of CTS, the requirements that are presented in the construction of CTS are investigated. It also considers the use of information technology in the process of teaching English as a foreign language, using modern and non-traditional methods. Among such methods, we can note such methods as the method of sequences, the communicative method, immersion in language, the method of silence, the method of suggestopedia, natural method, the method of physical response, the method of reading and telling stories, the method of teaching language Dogma, the method of the 25th frame, computer language learning, situational learning, guided practice and interactive practice. It presents the types and stages of the learning process. Further, this work gives brief information about such types of automatic dictionaries as: information retrieval thesaurus, frequency, translation dictionaries, concordance dictionaries, reverse dictionaries of word forms and quasi-bases, explanatory dictionaries and dictionaries for speech processors, dictionary components of knowledge bases and ES and their functions, as well as the paper presents the role of AD in CTS.

The use of computer technology in teaching foreign languages has changed approaches to the development of educational materials. The intensity of the educational process occurs with the help of

information technologies, which have a number of huge opportunities. Serving as a means for storing information, simulating and modeling the processes being studied, the computer individualizes and differentiates learning. With the practical use of Information and Communication Technologies (ICT), the student actively develops cognitive activity and discovers new knowledge. The students' independent cognition is also enhanced, the ability to independently replenish knowledge, to search for new information and navigate in this stream is formed. In order to modernize education in order to improve the quality of education, increase the availability of education and ensure the harmonious development of the individual, it is necessary to introduce ICT.

"CSE is a complex of software, hardware and educational tools that provide students with the material being studied, test the knowledge of students, interactive interaction between students and teachers in the learning process, as well as provide students with the opportunity to work independently to master the material being studied».[1]

The main task of the developers is to create the most productive CSR structure, which will have all the capabilities to implement the above stages of the learning process in an automated environment. As a result of research in the field of existing developments of computer training systems, a classification of CTS was developed, ranked according to various criteria and functional features:

In accordance with the **learning functions** performed by the system, CTS can be classified into the following types:

- Training CTS, in which students are offered only questions and tasks and the number of correctly and incorrectly solved tasks is counted;

- Controlling CTS are designed to test the skills and abilities of

[1] Зайцева, Л.В. Разработка и применение автоматизированных обучающих систем на базе ЭВМ / Л.В.Зайцева, Л.П.Новицкий, В.А.Грибкова. – Рига: Зинатие, – 1989. – с.17-22

the trainee before or during the tutoring;

- Informative CTS - tasks and questions that serve to organize a human-machine dialogue and manage the learning process, during which students are offered theoretical material;

- Simulation CTS - is a simulation model of training that simulates a real situation in which the student applies certain knowledge and skills;

- Gaming CTS, in which students are provided with the means provided by the program for the realization of opportunities related to the study of the world of play and activity in reality, which leads to the development of certain skills of the student;

- Diagnostic CTS are used to determine the level of knowledge of the student in the relevant area, periodically re-control of knowledge is carried out in order to identify poorly learned material, the control of knowledge of each user in a particular subject is individualized.

According to the **structural features** of the interaction between the tutoring system and the user, CTSs are divided into two basic classes:

- direct systems (without feedback), in which the student does not interact, but only receives information and performs tasks;

- circular systems (with feedback) aimed at establishing the level of knowledge of students in a certain period of the educational process with the possibility of developing certain skills and re-checking the level of knowledge.

In accordance with **the language skills being taught, the methodological purpose of certain linguistic aspects and types of learning activities,** CTS can be divided into the following types:

- for the development of articulation skills and the acquisition of a phonetic base;

- to study vocabulary (development of vocabulary base);

- to study grammar;

- to develop speech perception skills;

- for the development of academic writing;

- to develop speaking skills;

- to develop reading and spelling skills.

Criteria for creating CTS. As a result of the automatic systems analysis that currently exist, the main criteria for creating the most advanced training system that can meet most of the requirements of users have been identified. The established criteria include the following:

1) Age factor.

When creating an automatic tutoring program, it should be considered that in most cases they are aimed at a specific audience, such as children, teenagers, businessmen, scientists, economists and bank employees, as well as medical personnel. "There are learning systems of a generalized type that start learning from the very basics to complex lexical and grammatical structures suitable for independent study of a foreign language at any age." [2]

2) Purpose of use.

This criterion plays a functional role in the construction of the entire tutoring system. The creator of CTS must clearly define for teaching which aspect of the language the tutoring system will be created: grammar, vocabulary or spoken English, or the system will cover all these areas at the same time.

3) Social factor.

Specialized training courses allow you to focus on specific problems that often arise in a particular area. "For this purpose, special vocabulary corresponding to the selected area is included into the contextual dictionary of the automatic tutoring system."[3]

4) Availability and simplicity of the system.

The advantage of CTS lies in the simplicity of choosing the level of basic training of the user. Quickly acquiring the minimum vocabulary for everyday communication, expanding one's own knowledge, improving the skills of understanding live speech is possible only if the interface and CSR modules are easy to use and do not require the user to spend time and effort on studying the

---

[2] http://www.langust.ru/etc/lang_tut.shtml
[3] http://www.langust.ru/etc/lang_tut.shtml

program, but immediately encourage him to switch to the assigned task.

5) Individualization.

Individualization of learning when using a computer is also associated with the interactive nature of the work. With the help of a diagnostic test, the program can determine the level of knowledge of the student and, in accordance with this level, present theoretical material, questions and tasks, as well as tips and help throughout the training.

6) Interactivity.

The learning material can be presented in the form of interactive exercises and often consists of theoretical and practical sections on various aspects of a foreign language. The program can also provide feedback when the user performs certain operations: the student's successes or mistakes are noted, and, if necessary, hint messages are issued. *A computer program is much more convenient because it saves users from mindless cramming and makes the process of memorizing new material meaningful, turning it into a fun and exciting activity. In addition, many CDs have a powerful vocabulary base that you can use on your own.*[4]

"CTS are systems that help to master new material, control knowledge, and help teachers prepare educational material. The specificity of computer language learning is associated with the possibility of using not only ordinary teaching, but also applied,

---

4

http://www.langust.ru/etc/lang_tut.shtml;
http://gendocs.ru/v38853/лекции_по_информатике; https://studall.org/all-139353.html;
Богомолов, А.Н. Виртуальная языковая среда обучения «Русский язык дистанционно (по материалам СМИ): структура и содержание учебного контента // Методика: теория и практика обучения языку, – Москва: – с.28-31.

instrumental, gaming programs for learning purposes."[5] "The use of teaching systems for the purpose of teaching a language enables students to immerse themselves in the environment of the language being studied and work effectively on the formation of skills in various types of speech activity (reading, writing, listening, speaking) and studying various aspects of the language (phonetics, vocabulary, grammar)".[6] "Experts of such software products make up a specially designed CTS structure and the learning environment included in it. Remote access programs, high-quality graphics and sound, presentation tools are part of the learning environment and thus increase the effectiveness of computer-based language learning compared to using separate learning programs. At the first stages of using computers in language teaching, there were programs such as word processors, databases, various game programs."[7] "The modern period of development of computer linguadidactics puts forward new requirements and significantly expands the types of programs needed to create a computer learning environment within the CTS".[8]

CSR can be used as a means of representing knowledge, organizing a dialogue between the user and the system, which, at the request of the user, provides a line of reasoning when solving a learning problem in a form acceptable to the user being trained. The presence of the following functional blocks is mandatory in the CBS

---

[5]

Кулиева, З.Ю. Модели автоматических обучающих систем, их структура и классификация // – Bakı: İnformasiya texnologiyaları probleml?ri, – 2012. №1 (5), – s. 7-17.

[6] Деркач, А.А. Педагогическая эвристика: Искусство овладения иностранным языком / А.А.Деркач, С.Ф.Щербак. – Москва: Педагогика, – 1991. – с.54-66.

[7] Петрушин, В.А. Архитектура экспертно-обучающих систем // Разработка и применение экспертно-обучающих систем: Сб. науч. Трудов, – Москва: НИИВШ, – 1989. – с. 7-18.

[8] Kudryavtsev, V.B. Modeling education process using Expert system / V.B.Kudryavtsev, А.С.Строгалов, В.В.Перетрухин [et al.] // Proceedings 2-nd Russian German Simposium, – Moscow: – 1996. – pp. 25-28.
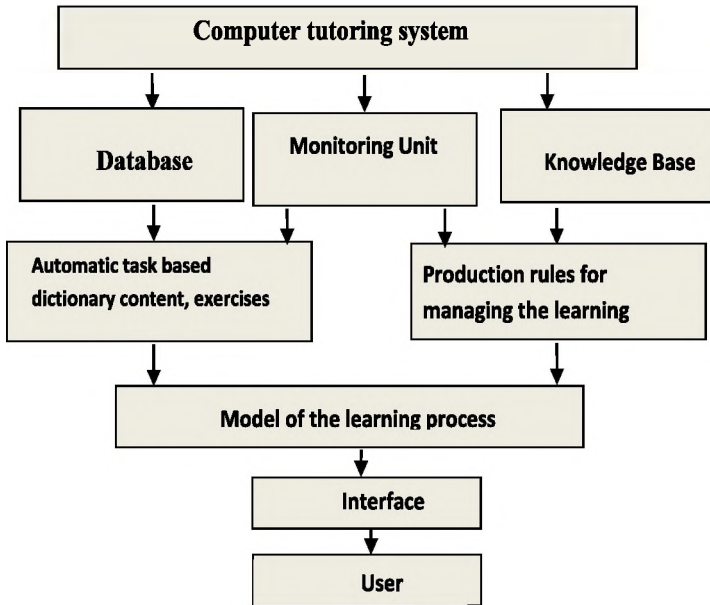
(Fig. 1):

Fig.1. CTS Structure

•**Knowledge base** is an integral part of the system containing a formal description of knowledge presented as a set of facts and production rules, as well as methods and procedures for managing the learning process;

• **Monitoring Unit** - implements the means of evaluating the effectiveness of training in the form of tests or another way to control the assimilation of the material. This block, in turn, consists of a subblock of statistics of correct and incorrect answers, a diagnostic test block, and a general testing subblock;

• **Database** is a block that implements the means of storing information about the subject area and the learning process. The database is represented by AD, which is a repository of various types of formalized information necessary for more effective learning, and

a sub-block of exercises, text banks, etc.

　　• **Interface** - implements the means of audio-visual, as well as sensory interaction of the system with the user, being a tool for visual representation of the successive stages of providing, consolidating and checking the learning material.[9]

　　Since the computer has recently become an integral part of lexicology, computer versions of traditional dictionaries have begun to appear. Being essentially electronic, but having different functions and purposes, the electronic dictionary, machine dictionary (MD), automatic dictionary (AD) and computer dictionary (CD) are all created using a computer. Automatic dictionaries are widely used and are used as independent programs. An automatic dictionary is a device that has a linguo-technical direction, which helps a person for various terminological work performed by a person, and for providing translation processes in the lexico-terminological sphere. It is possible to use an automatic dictionary as a database, which provide support for the work of other applications. Automatic dictionaries are classified into information retrieval thesauri, frequency, translation, concordances, reverse dictionaries of word forms and quasi-bases, explanatory dictionaries, dictionaries for speech processors, and dictionary components of the knowledge base

[9] Цытович, П.Л. Автоматизация проектирования обучающих систем на основе объектно-контейнерного подхода: / дисс. … кандидата технических наук / – Челябинск, 2000. – с.19-25; Богомолов, А.Н. Виртуальная языковая среда обучения «Русский язык дистанционно (по материалам СМИ): структура и содержание учебного контента // Методика: теория и практика обучения языку, – Москва: – с.28-31; Исаева, Г.Г. Педагогические условия подготовки будущего педагога профессионального обучения к использованию экспертных обучающих систем / Г.Г.Исаева, А.Ш.Бакмаев // – Махачкала: Вестник Учебно-методического объединения по профессионально-педагогическому образованию, – 2015. №2 (49), – с. 161-169; https://cyberleninka.ru/article/n/virtualnaya-yazykovaya-sreda-obucheniya-russkiy-yazyk-distantsionno-po-materialam-smi-struktura-i-soderzhanie-uchebnogo-kontenta; https://studall.org/all-139353.html

and expert systems.

The second chapter "Principles of building of an automatic dictionary" describes the principles of building AD in the tutoring system, and also defines the goals and objectives of the training course. The selection of methods and techniques for teaching English as a foreign language was carried out and a scenario of CTS and the corresponding text database was proposed. Also, the principles of selection of lemmas and the construction of a dictionary of AD in CTS are established, a thorough selection of formal features of the data selected in the dictionary entry of AD for various parts of speech is carried out. The following is a formal description of the structural composition of the AD dictionary entry in the CSR and formalized AS data in various functional blocks of the CTS.

The technique for constructing an automatic dictionary is based on the following steps:

1. Determination of the goals and objectives of the course of study, including such features as the target audience, the subject area and the trained linguistic knowledge and skills;

2. Selection of methods and techniques for teaching a foreign language that are optimal for automating the learning process aimed at developing certain knowledge and skills;

3. Creation of a script for a computer tutoring system and a corresponding text database based on the content of the lesson material;

4. Selection of lemmas and construction of a vocabulary of an automatic dictionary in a computer tutoring system;

5. Formalization of the data selected in the dictionary entry of the automatic dictionary for various parts of speech;

6. Representation of formalized data of an automatic dictionary in various functional blocks of a computer tutoring system.

Determination of the goals and objectives of the course of study

At the initial stages of the development of AD in CTS, it is necessary to establish the main linguodidactic tasks that will be performed by this tutoring system. In this work, the subject of

linguodidactics is the organization of the process of teaching the vocabulary of the English language, and, therefore, it is necessary to find answers to the following questions:

- selection of the course, determination of the purpose and reasons motivating the student to learn this language;

- defining the objectives of the developed course, aimed at developing certain language skills and abilities;

- selection of the techniques and methodology designed to present, consolidate the material and control the knowledge gained.

Selection of teaching methods and techniques

Since the purpose of the course developed by the computer training system was to teach the vocabulary of a foreign language, a brief overview of the choice of methodology and techniques for teaching vocabulary is given below in Table 1:

Stage	Material presentation means	Material consolidation techniques
stage1	- usual images, 2-D and 3-D images (with literal reproduction of the words appearing on the screen by clicking on the image with a hyperlink to automatic dictionary) - short audio recording (1 min.) and videos without subtitles - animation of certain actions, concepts, and situations -videos with subtitles - texts with definite images - short dialogues with pictures - explanatory dictionary for each lesson (with explanations of the meanings) - the table of antonyms, synonyms, verbs and sustainable units	- the explanation of the meanings of proposed lexical units; -replacement of phrases with the words; - compilation of phrases and sentences with different meanings of needed words; - selection of synonyms and antonyms to the words used in different contexts; - translation of phrases and sentences including synonymous lexemes or different values of polysemic words from Russian into English and Vice versa; - words grouping according to their lexical meanings. - different kinds of substitute exercises, etc.
stage2	usual image without words - 2-D, 3-D images	question-and-response exercises including the techniques of vocabulary

	- audio recording (2-3 min.) and videos without subtitles - certain actions, concepts, situations animations -videos with subtitles - texts with image series - pictures with dialogue and text interpretation - explanatory dictionary for each lesson (with explanations of the meanings) - the table of antonyms, synonyms, verbs and sustainable units -lists of ready language models and samples.	usage in the speech with the aim of obtaining more or less deployed the statements of the answers to the questions; - interpretation of audio text context by the usage of certain words; - exercises demanding the usage of certain words in the description of the situation; - tasks where the learner is asked to use certain words to describe a situation; - making a dialogue according to the model; - making a dialogue according to the situation; - comparison of the story according to the beginning or with the usage of story line and certain words, etc.
stage 3	-usual image without words - 2-D, 3-D images - audio recording (3-4 min.) and videos without subtitles - certain actions, concepts, situations animations -videos with subtitles - texts with image series - pictures with dialogue and text interpretation - explanatory dictionary for each lesson (with explanations of the meanings) - the table of antonyms, synonyms, verbs and sustainable units - texts with hypertext links to various information sources	self-narration and narration about future profession, hobbies, sports, and etc.; - to talk about the delivered book, watched film and etc.; - to tell a story or write an essay on a given topic (native town nature, sport events, and etc.); - to tell about of a favorite writer, artist, composer's life and creation.

Table1. Selection of teaching methods and techniques

Creating a script for Computer Tutoring System

Writing a CTS script - A course script involves a detailed list of

course components and topics, as well as a preliminary description of the course structure that will be implemented in the future. Any training course you create has a hierarchical structure. Provided that the functions of the teacher are transferred to the computer, the work of the CTS can be represented as follows:

a) Diagnostic testing of the student is carried out - to determine his level of knowledge at the beginning of using the system, after which the system automatically starts working in the module corresponding to the knowledge level of the student;

b) The system can consist of any number of modules, depending on the program of the developed course and the information provided in it;

c) Modules, in turn, consist of several lessons, for example, in this case, each module consists of three lessons;

d) The lesson is divided into three sections of training, combining various linguodidactic goals, and, accordingly, various types of textual information, that can include descriptions of animation and video clips, illustrations, links between thematic sections, hypertext links.

The distribution of material means the creation of information sub-blocks containing a certain amount of data of the same type and content. It is necessary to consider the factor that multimedia information tools can be distributed in separate sub-blocks, however, their text version should also be included in the test sub-block. On fig. Figure 2 is given the model of information distribution in the CTS text base. The information provided in the system is distributed in multimedia, visual and text blocks. In turn, the multimedia block consists of video, audio, animation information subblocks, as well as their interactive animation subblock with various types of data included in them. The printed version of the linguistic information provided in videos and audio recordings is collected in the sub-block "transcript of video-audio data of multimedia" of the contextual base.
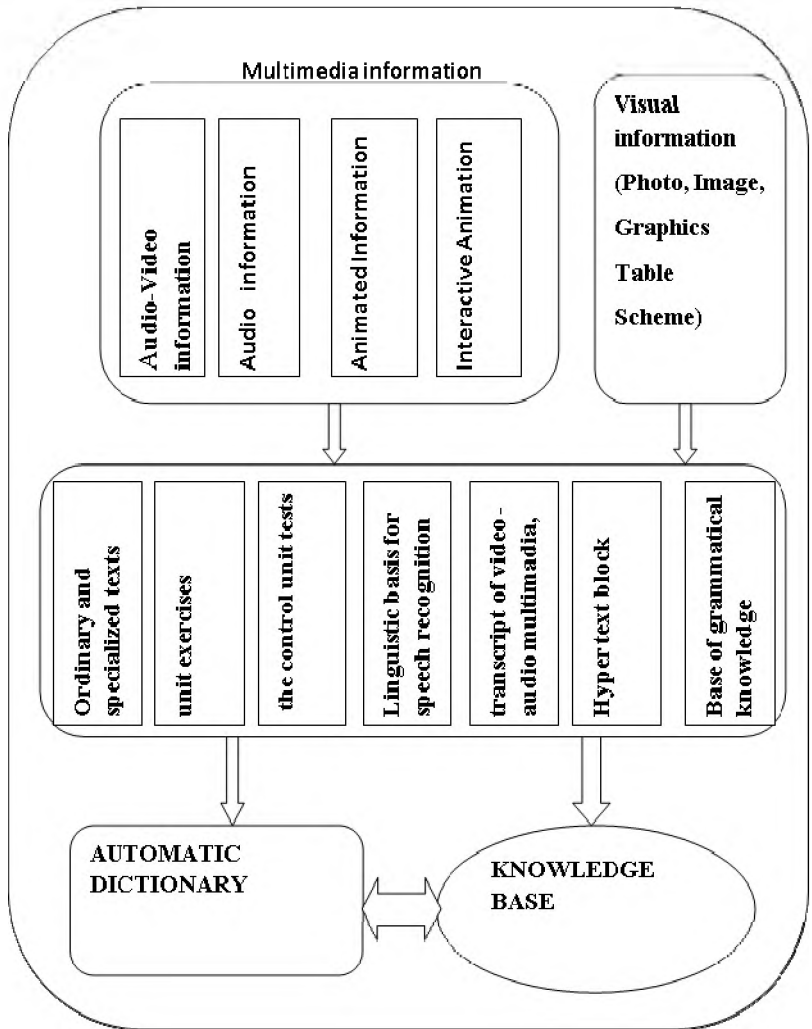
Fig.2 Information distribution model in the contextual database of CTS

The linguistic information of the visual subblock is also collected in a "hypertext block" in an identical way, where data in diagrams, tables, images and other types of visual aids enters a system of cross-referenced text pages. A hypertext base can be defined as a collection of documents containing both textual and audio-video data, interconnected by mutual links into a single text. As shown in Figure 2, all text information data (texts, dialogues, exercise block, test block, speech recognition block, etc.) are dispersed both in the AD and in the knowledge base that controls them.

Formalization of data

To obtain a universal and ideal AD structure, it is enough to bring together sociologically, semiotically, stylistically, linguistically and psychologically relevant properties of various structures of a dictionary entry. According to Denisov, an ideal universal dictionary entry should have at least the following components: *"1) a headword; 2) its formal characteristics (grammatical, orthoepic, spelling, etc.); 3) its semantization; 4) extracts from texts illustrating one or another formal or semantic feature of the headword; 5) indications of the "neighbors" of the headword in the lexical system of the language along different axes of the semantic space of the language; 6) referrals and certificates of a different nature of purpose. Each of these points has its own character of deployment."*[10]

---

[10] Денисов, П.Н. Лексика русского языка и принципы ее описания / П.Н.Денисов. – 2-е изд. – Москва: Русский язык, – 1993. – с.45-50

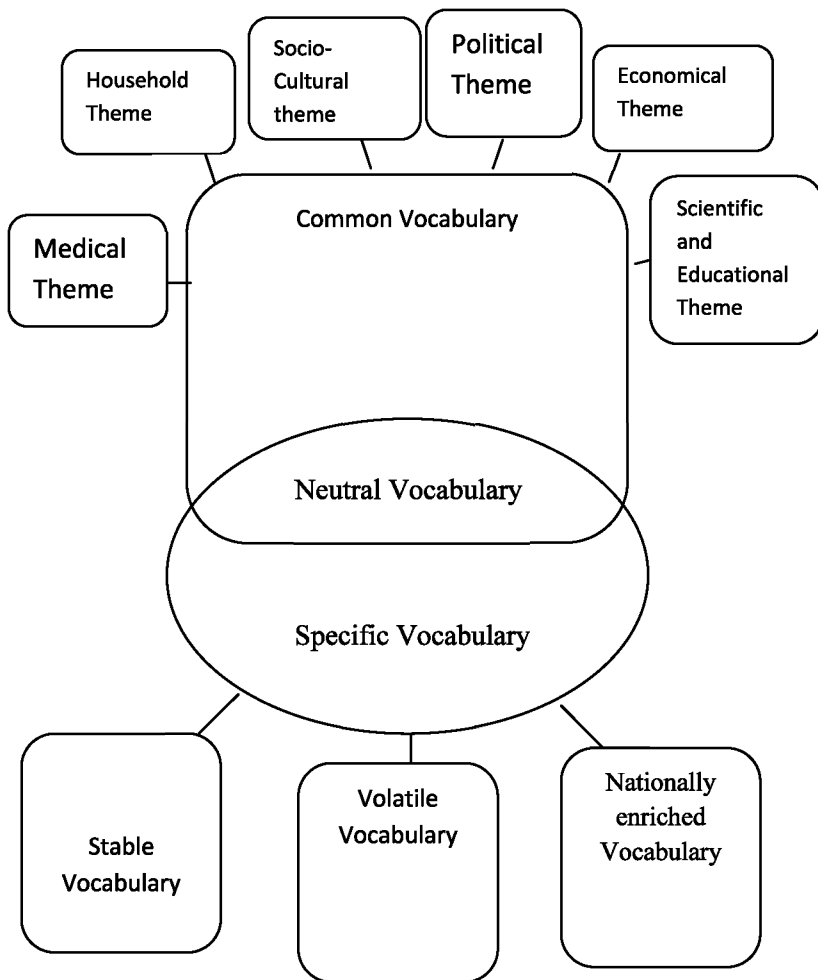# *Selection of lemmas and construction of an automatic dictionary*

Fig.3. Schematic representation of the language lexical system

The structure of a dictionary entry for an automatic dictionary in CSR consists of the following items:

1. Word form number;
2. Word form;
3. Transcription;
4. Code of the part of speech;
5. Code of the structural component of the contextual base;
6. Sound reproduction of the pronunciation of the word form;
7. Visual image of the word form;
8. Animated image;
9. Video image;
10. Interpretation of the meaning of the word form;
1.meaning1 (with an example of usage)
2.value2 (with usage example)
3. meaning3 (with an example of use)
11. Synonymous series;
12. Antonymic series;
13. Collocations with interpretation of meanings;
14. Transferable equivalents of collocations;
15. Phraseological combinations (idioms);
16. Translational equivalents of phraseological abbreviations;
17. Terminological combinations;
18. Translation equivalents of terminological combinations;
19. Abbreviations with interpretation of meanings;
20. Translation equivalents of abbreviations;
21. Examples of contextual use of all types of phrases;
22. Examples of contextual use of sentences.

Representation of formalized data

The diagram below shows the automatic dictionary sub-blocks into which the encoded InfoBase can be clustered.
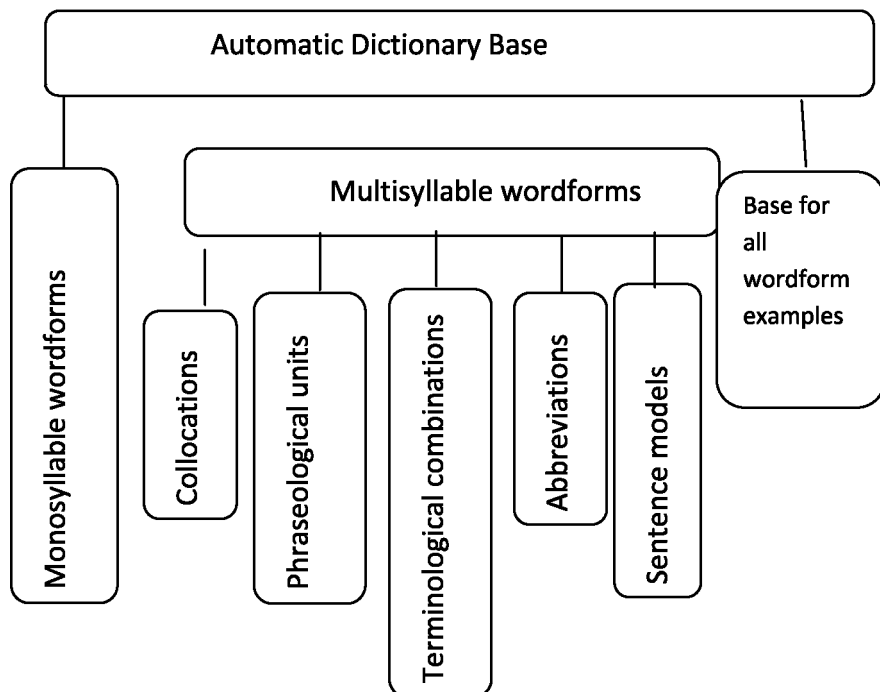
Рис.4. Схема подблоков автоматического словаря в компьютерной системе обучения

The third chapter "Application of Expert Tutoring Systems in linguistics" describes the principles for the development of an ETS. Expert Tutoring System is a complex of algorithms and programs, while in the process of implementation of this system, the established principles of building an AD is applied. In the same chapter, knowledge representations in the ES, the ETS test block, the diagnostic testing block and the contextual database in the system are established. It also gives the design of learning models and a model for the transformation of AD subspecies into ETS.

The algorithm sequence of the Expert Learning System:

✓ The level of knowledge of the trainee is determined;

✓ Based on the current state of the student, the system selects the appropriate learning strategy, the next task is generated, requiring the response of the student;

✓ The student's answers or actions are compared with the reference solution in blocks of correct answers, and based on the differences, error diagnostics are performed;

✓ Incorrect answers are recorded in the block of erroneous answers;

✓ Based on the results of the diagnostics, the student is presented with additional material for re-training and fixing weak links.

Architecture of the Expert Tutoring System

The ETS architecture can change in accordance with the applied goals and objectives, but the structural components of all EDs are similar in basic criteria.

1. Database
2. Knowledge base
3. Interface
4. Analytical processor

The knowledge base contains metaknowledge and object knowledge in the form of facts and rules necessary for solving the problems. It occupies a large volume and plays the role of a thinking center, and using the information from the database, it makes attempts to imitate the function of the human brain in the learning process. The knowledge base, in turn, consists of many different sub-bases: a diagnostic test block, a training error fixer. The database is a repository of linguistic information.

The interface shows two types of information: demo information and non- demo information (Table2)

Interface Information	
Demonstrative information	No demonstrative information
Text data	ELS middleware
Automatic Dictionary	Lingua Didactic knowledge
Visual data (pictures, video and animation)	Encoded data
Tests	Learning models
Inferable results	Output rules
Tips	

Table2.

Figure 5 shows a model for the transformation of AD into subspecies or sub-base dictionaries of AD in the ETS of foreign vocabulary. The first five types of sub-base dictionaries have the usual structure and content, typical for machine dictionaries with all applications and functional properties. Other types of sub-dictionaries shown in the figure are typical for ETS in accordance with the functions performed. Thus, the module dictionary is a transformational subspecies of the AD in the ETS, which has in its composition all types of word forms used in various contexts of audio and video information knowledge that are part of one lesson of this ETS. For example: if this ETS consists of 15 modules (lessons), regardless of their subsequent forms and types of subdivisions, the module dictionary should contain all the word forms included in these 15 modules (lessons). The dictionary entry of this subspecies of the dictionary has a separate cell indicating the number of the lesson and a possible division, such as 10 / A1 – it means the 10th lesson, section A 1.
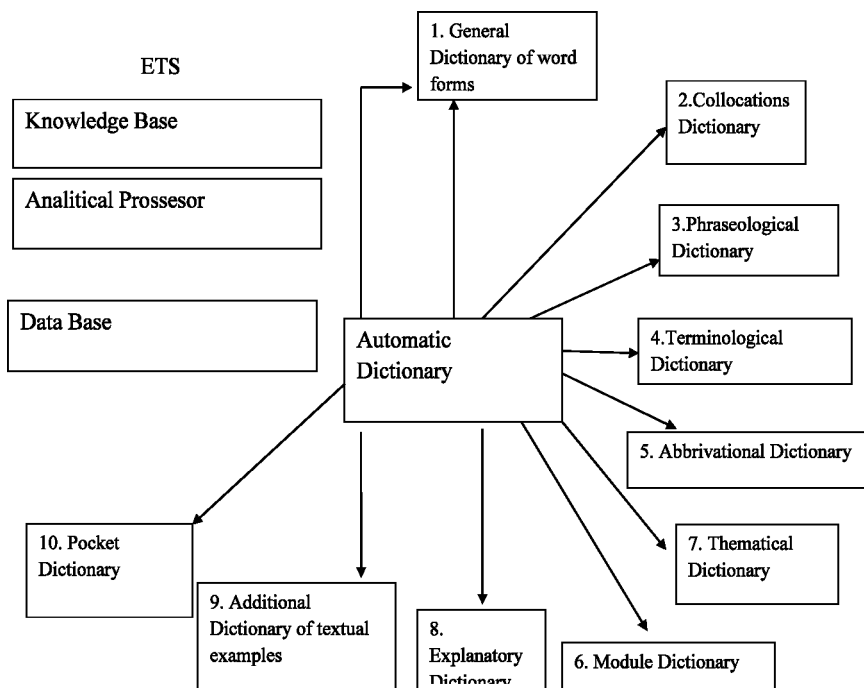
ETS

Knowledge Base

Analitical Prossesor

Data Base

Automatic Dictionary

1. General Dictionary of word forms

2.Collocations Dictionary

3.Phraseological Dictionary

4.Terminological Dictionary

5. Abbrivational Dictionary

7. Thematical Dictionary

6. Module Dictionary

8. Explanatory Dictionary

9. Additional Dictionary of textual examples

10. Pocket Dictionary

Рис.5. The model of AD subspecies transformation in ETS

In the "Conclusion" the following conclusions and generalizations on the basis of the conducted research and the results have been obtained.

The emergence of an "information society" and improved management efficiency in research and development can be achieved through the automation of learning processes and the introduction of information and communication technologies (ICT) and expert tutoring systems (ETS), currently used in various fields and for various purposes of language learning. The development of electronic and automatic dictionaries that allow you to create an order and system of lexicography has become one of the main tasks of applied linguistics. The introduction of a contextual dictionary

29

into applied linguistics, both for learning a foreign language and for translation services, is gaining momentum based on practical applications and modern research.

As a result of the researches carried out in this dissertation work the following conclusions are received:

1. A brief example of a linguistic database of an automatic dictionary in a computer tutoriing system has been practically developed. This dictionary includes tables of word forms, word meanings, synonyms, antonyms, collocations, phraseological expressions, terms, abbreviations, word examples, sentence examples and texts.

2. For the first time?, principles for the selection of formal features and a methodology for creating a contextual dictionary were proposed by selecting formal features into dictionary entries for various parts of speech, which represent formalized linguistic information of a language pair (English and Azerbaijani), where the ideal universal dictionary entries should include at least the following components: 1) headword; 2) its formal characteristics (grammatical, orthoepic, spelling, etc.); 3) its semantization; 4) extracts from texts illustrating one or another formal or semantic feature of the headword; 5) indications of the "neighbors" of the headword in the lexical system of the language along different axes of the semantic space of the language; 6) referrals and certificates of a different nature of purpose. In turn, this pair of languages is acceptable for storage in a database and processing in a knowledge base.

3. For the first time?, a methodology was proposed for constructing a contextual dictionary as part of an expert tutoring system, where dictionary entries include a wide range of formalized linguistic information: monosyllabic words (noun, adjective, verb), type 1 and type 2 collocations (bigrams, trigrams, 4-component) , phraseological units (functional nouns, bigrams, trigrams, 4-component), terminological combinations (TC1, synonymous TC, truncated TC), abbreviations (acronyms, compound words, mixed words), sentences (4 words, 5 words, 6 words) necessary for the

effective presentation of the material, consolidation and control of the acquired knowledge.

4. Theoretical foundations for constructing the optimal structure of an automatic dictionary have been developed: word forms (morphemes, polysemantic words, synonyms, antonyms), phrases (stable combinations, lexical segments, idioms, prepositional combinations, collocations, phraseological phrases), sentences (standardized statements, sentence models (beginning .... end), proverbs), context (models of a paragraph, text (beginning ... end), introductory words, elements).

5. A formal model of a dictionary entry for the contextual AD in the tutoring system has been developed: word form number, word form, transcription, part of speech code, code of the structural component of the contextual base, sound reproduction of the pronunciation of the word form, visual image of the word form, animation image, video image, interpretation of the meaning of the word form: meaning1,2,3 (with examples of usage), synonymic series, antonymic series, collocations with interpretation of meanings, translation equivalents of collocations, phraseological combinations (idioms), translational equivalents of phraseological abbreviations, terminological combinations, translational equivalents of terminological combinations, abbreviations with interpretation of meanings, translational equivalents of abbreviations, examples of contextual use of all types of phrases, examples of contextual use of sentences.

6. A new approach to building a database of an expert tutoring system based on a contextual dictionary is proposed: entering the user's personal data; conducting diagnostic testing to test basic skills (accuracy, speaking, spelling, reading); comparative analysis of the results of diagnostic testing of the standard model of the student A level; determination of the student's model in accordance with the results of a comparative analysis; presentation of knowledge about software based on the selected model; fixation and analysis of the time spent, mistakes and correct answers made by the trainee when performing various tasks within the framework of one ETS module.

7. A new formal model of a multifunctional dictionary is proposed that can be transformed into various subspecies of dictionaries: a dictionary of word forms, a dictionary of collocations, a dictionary of phraseological units, a terminological dictionary, a dictionary of abbreviations, a lesson dictionary, a thematic dictionary, an explanatory dictionary, an additional dictionary of examples, its own dictionary with the intellectual ability to collect information to the "pocket" subbase as an aid to ensure a more efficient process of learning vocabulary.

8.Expert System constructing technique for teaching English vocabulary to researchers on the basis of theoretical principles for developing the optimal structure of the AD is proposed. This system can be used to study all levels of English in a scientific environment (from basic to advanced levels) and includes certain classifications of tasks it solves: Interpretation of initial data (diagnosing preliminary, step-by-step knowledge acquired in stages, diagnosing possible violations in the system, fixing errors to change the working strategy, diagnosing the frequency of repeated errors); Stage of building objects (building complex objects under certain conditions, assembling system blocks, creating production rules in the knowledge base, changing the course or working strategy of the system, synthesizing the data obtained to move to the next stage, planning an experiment of a curriculum, designing an action plan, building a description of an action from elementary operations); Monitoring stage (testing, monitoring system operation, detection of unforeseen situations, notification of unforeseen critical situations, monitoring of unplanned remarks for the speech base to recognize and fix errors in correct answers, monitoring the actions of the object).

The main content of the dissertation is reflected in the following published works of the author:

1. Глагольно-именная омонимия простых лексических единиц // – Bakı: Humanitar elmlərin öyrənilməsinin aktual problemləri, – 2006. №3, – s.43–46.

2. Информационные компьютерные технологии в обучении иностранного языка // – Bakı: Filologiya məsələləri, – 2011. №7. – s.94–99.

3. Требования к современным автоматическим системам // – Bakı: Filologiya məsələləri, – 2011. №9, – s.54–61.

4. Элементы когнитивного обучения при создании автоматических обучающих систем // – Bakı: Sivilizasiya, – 2012. №2, – s.150–154.

5. Dilin tədrisində mövcud üsulların təsnifatı // – Bakı: Tədqiqlər, – – 2013. № 1, – s.15–19.

6. ICT as motivation device in foreign language learning // "Xarici dillərin tədrisinin aktual problemləri". Respublika elmi-praktiki konfransının materialları, ADU, – Bakı: – 5–6 may, – 2016, – s.220–223.

7. Application of Intelligent tutoring system as a modern language learning device / "Demoge 17" 1st International Symposium on Modernization and Tradition in Language and Literature. Bildiriler Kitabı, – Karabuk: – 4–6 October, – 2017, – s.279–288.

8. Построение автоматического словаря в экспертной системе обучения научной лексике // "Turklang 2017". 5-ая Международная конференция по компьютерной обработке тюркских языков, – Казань: – 18–21 октября, – 2017, – с.22–32.

9. Роль автоматического словаря в компьютерной обучающей системе // Ağamusa Axundov və Türk filologiyası". Beynəlxalq konfrans, – Bakı: – 25–26 may, – 2017, – s.394–397.

10. İnformasiya-kompüter texnologiyalarının tətbiqi ilə xarici dilin tədrisində istifadə olunan üsulların təsnifatı // Uluslararası

Türk Dünyası Araştırmaları Simpoziumu, – Niğde: – 26–28 may, – 2017, – s.75–86.

11. Использование интеллектуальной обучающей системы как современного лингводидактического инструментария // Филологические науки. Вопросы теории и практики, – 2018, №1(79), часть 2, – с. 408–414.

12. Архитектура экспертной системы обучения // – Bakı: Tədqiqlər, – 2018. № 2, – s.102–113.

13. Principles of Automatic Dictionary compiling in Expert Tutoring System // 4th Language, Culture & Literature Symposium, – Antalya: – May 17-18, – 2018, – s. 97-106.

14. Avtomatik lüğətlər və Onların Öyrədici Kompüter Sistemlərində Təyinatları // VI Beynəlxalq Türk Dünyası Araşdırmaları simpoziumu, – 13-15 iyun, – 2019, – s. 7-12.

15. Применение Результатов Сравнительного Анализа при Создании Автоматических Систем Текстовой Обработки и Обучения Иностранному Языку // "Xarici Dillərin Aktual Problemləri". Azərbaycan Dillər Universiteti Filologiya Fakültəsi Nəzdindəki Xarici Dillər Kafedrasının Müəllim Gününə həsr olunmuş elmi seminarının tezisləri, – Bakı: – 5 oktyabr, – 2019, – s. 8-12.

16. The Use Of Expert Learning Systems In Linguistics // "Social and Cultural Transformations in the Context of Modern Globalism". SCTMG – 2020 International Scientific Conference. The European Proceedings of Social and Behavioural Sciences EpSBS. DOI: 10.15405/epsbs.2020.10.05.173. – p. 1311-1317

The defense of the dissertation will be held on 14 november 2023 at 13^{00}at the meeting of the ED 1.06 Dissertation Council operating under the Institute of Linguistics named after Nasimi of the Azerbaijan National Academy of Sciences.

Address: Institute of Linguistics named after Nasimi of the ANAS, fifth floor, AZ 1143, H.Javid avn., 115, Baku

The dissertation is available in the library of the Institute of Linguistics named after Nasimi of the Azerbaijan National Academy of Sciences.

Electronic versions of the dissertation and abstract are posted on the official website of the Institute of Linguistics named after Nasimi of the Azerbaijan National Academy of Sciences.

The abstract was sent to the necessary addresses on the 17 july 2023.