

Kamila A. Valiyeva

DOI: 10.25045/jpis.v07.i2.10

Institute of Linguistics after Nasimi of ANAS, Baku, Azerbaijan
kamilavaliyev@gmail.com

MODERN AREAS OF COMPUTATIONAL LINGUISTICS

The paper analyzes the main trends of computer linguistics - natural language processing, corpus linguistics, computer-aided dictionary compiling, computer-aided learning and recognition of languages, including machine translation issues.

Keywords: *computational linguistics, natural language processing, machine translation, corpus linguistics.*

Introduction

Computer Linguistics, which has been emerged at the intersection of computer science and linguistics, is one of the fastest growing areas of science in the modern era. As a branch of applied linguistics, it studies linguistic basics of computer science, and computer-aided modeling of thinking and language. Thus, it aims at addressing the issues related to the establishment and improvement of common language in the information presentation, including modeling of information systems and natural language interface, which is one of the main issues of artificial intelligence.

Developed in the 60s of the twentieth century, computer linguistics first focused on theoretical linguistic models. However, later the subject of the study has aimed at a broader direction - machine learning with the application of statistical methods, including texts processing, and a number of heuristic issues [1].

As we mentioned above, as computer linguistics develops mathematical model of natural languages, and, of course, the key points here are the use and development of language for multi-purpose computer systems. In this regard, computer linguistics tackles the following issues:

1. natural language processing - syntactic, morphologic, and semantic text analysis;
2. corpus linguistics (generation and use of e-corps of the texts);
3. electronic dictionary compiling (thesaurus, automated translation dictionaries, encyclopedias, orthographic, explanatory, terminological and field-specific dictionaries, and spelling dictionaries for automated error detection and etc.);
4. automated text translation systems (*Dilmanc, Google Translate* etc.);
5. fact extraction and text mining;
6. auto-summarizing - included into *Microsoft Word*;
7. knowledge (expert) systems;
8. dialogue systems;
9. optical character recognition (*OCR, FineReader*);
10. automated speech recognition;
11. automated speech synthesis;
12. development of information retrieval systems [2].

Natural language processing is a branch of artificial intelligence and mathematical linguistics, and studies computer-aided analysis and synthesis of any natural language. In short, interface language (human computer interaction) has been established for computer in the process of theoretical language processing from the point of view, which is not an easy task. Natural language perception (understanding) requires the rich knowledge about the world, hence one of the main issues in artificial intelligence is to introduce this “perception” to a computer. For example, the comprehension of the Azerbaijani text depends on word order, homonyms, synonyms, punctuation and emphasis. At this point, formal models of the listed criteria for text comprehension have to be developed.

Corpus Linguistics (CD) is an applied branch of linguistics that explores the text corpus and use. The term was introduced in 1960 with the emergence of corpuses in practice. "Linguistic corpus" implies a set of texts with certain principles and standards. Corpus generation aims at solving various linguistic issues (such as, graphic, grammatical and lexical text analysis).

The first major in the world linguistics - Braunov corpus, was established in 1960. According to the corpus model, the Russian frequency dictionary, containing more than one million words, was compiled by Zasorina [3]. The rapid development of computer technology has contributed to the development of large-scale national corps, as *British National Corpus* developed at the University of Birmingham, and Ershov's Machine fund at the former USSR. Currently, the Russian national corps contains 300 million word combinations [2].

The creation of a national corpus of the Azerbaijani language is still remains one of the issues to be resolved.

The problem solution: computer-aided dictionary compiling

The rapid development of computers has contributed to the automation of mental activity and opened broad prospects. Research carried out through computer released the linguists from the mechanical, tedious and labor-intensive works, such as material grouping, typifying, inventorying, and editing.

In the 50s of the last century, two major laboratories equipped with computers were established in Europe. One of them, located in Bezanson, is the lexicological analysis laboratory at the French dictionary learning center. Another is the laboratory of linguistics center on philological analysis, located in Italy. Moreover, in the Netherlands, a scientist F.de Tollener was conducted computer-aided lexicographical research [3].

Most of the Russian scientific centers have conducted many researches regarding computer application. These centers have developed different types of computer-aided dictionaries, such as the dictionaries of frequency of words used in the text, including concordances containing a list of specific expressions in the text, and the reverse-dictionaries listed in alphabetical order depending on the word endings, rhyme dictionaries, and automated dictionaries designed for machine translation system, and so on. The book "Electronic computing machines in linguistics" by Pines V.Y. and Mahmudov M. provides detailed information about the above-mentioned dictionaries [4].

In a case study, we will review only the computer-aided dictionary compiling in the Azerbaijani linguistics. Before touching upon this matter, it is worth providing brief information about a computer and its performance.

Obviously, "human-machine" system resolves the issues in the following steps [5].

1. problem statement;
2. mathematical formulation of the problem;
3. preparation of the material to be included in the computer (initial data), i.e., coding);
4. selecting appropriate methods to solve the problem;
5. developing an algorithm and its block-scheme;
6. programming the algorithm;
7. computer-aided problem solving;
8. analyzing the results.

These steps can be briefly characterized as follows:

"Mathematical statement of the problem" – data enumeration (in this case, materials depending on the nature of our issue), i.e. coding, and solution of its components.

Sequenced solution of the components completes the issue. Indeed, the most optimal solution method should be selected. For example, a typification method is typical for compiling dictionaries of frequency. An algorithm should be developed based on the selected method.

Many important frequency dictionaries of the Azerbaijani linguistics have been compiled with the help of mathematical and statistical methods. Obviously, frequency dictionaries record

the frequency of words and their use in the texts. A number of difficulties occur while compiling these dictionaries. First, the text should be selected in such a manner that literary language norms are adequately incorporated. It is not an easy process. Second, more complicated task is a selection of the overall volume of texts. The main source for the frequency dictionary of the Azerbaijani language is a newspaper. Hundred thousand words have been selected from random texts.

The use of frequency dictionaries in automated data processing - automated translation, summarizing, polygraph, forensic science, and in identifying the anonymous authors of the articles is of great importance.

As it is noted earlier, the first frequency dictionary was published in 1997 [6]. It was revised and published in 2004 [7]. In 2010, the frequency dictionary of the Azerbaijani language was published [8]. All the methods of the Azerbaijani language have been represented here. Linguistic texts analysis involved about 50 million words. Generally, the dictionary contained nearly 12,000 word roots. The frequency dictionary of the Azerbaijani language plays a role of an invaluable resource in solving many problems of general lexicology and language teaching as a statistical model.

Furthermore, the application of statistical methods in the study of historical monuments is of undeniable importance. The Polish scientist Chekanovsky Y. used this method for the detection of the kinship among the groups of Indo-European languages for the first time in 1927 in the world linguistics [9]. In 1948, the American researcher Svodesh M. studied dictionary fund of all languages using lexical and statistical method and concluded that, dictionary fund of all languages is changing slowly, at a steady pace [9]. In addition, statistical methods play an important role in the study of language typology.

Alphabetical-frequency and reverse-alphabetical list of the critical text of 1988 year issue of "The Book of Dede Korkut" (authors: Zeynalov F. and Alizadeh S.) was published with regard to the computer-aided study of historical monuments, including written monuments of the Azerbaijan linguistics [10]. The frequency list here covers ten thousand word forms. The words of the legend are listed in a condensed format according to its frequency of use.

The list of alphabetical-frequency the word forms are not arranged according to the frequency, but to the alphabet. The compiled list is of great importance for professionals involved in the language history.

Statistical and distributive analysis of the classical heritage of Fuzuli's poetry has also been explored. Of course, it is of great importance for revealing the stylistic features of classic literature authors, and for the identification of unknown authors of the works by comparing the frequency.

Moreover, "Reverse-dictionary of the Azerbaijani language" widely used in solving many problems of the Azerbaijan linguistics has been drawn up [11].

What is reverse-dictionary and how it differs from usual dictionaries (bilingual, explanatory, etc.)?

Reverse-dictionary differs from usual dictionaries in structure. In ordinary dictionaries, the words are arranged in alphabetical order according to the initial letters. While in reverse-dictionaries, they are arranged in the alphabetical order according to the last letters.

"Reverse-dictionary of the Azerbaijani language" has been compiled based on the "Spelling dictionary of the Azerbaijani language". Moreover, the words that are not included to the "Spelling dictionary of the Azerbaijani language" have been taken from the "Azerbaijani-Russian" dictionary and included to the reverse-dictionary. "Reverse-dictionary of the Azerbaijani language" is one of the smartest among the Turkic language dictionaries in terms of volume.

Electronic dictionaries - different types of dictionaries in computer memory (compiled depending on the purpose). By the way, it should be noted that, depending on the purpose, electronic dictionaries differ from the computer dictionaries. Computer dictionaries are developed for the users in order to take advantage of computer applications. Various dictionaries are available

on the Internet, such as *ABBYY Lingvo*, *Poliglot (Azeri Dictionaries)*, *Lingvo*, thesaurus, ontology, thesaurus, spelling, reverse-dictionaries and so on.

Machine learning - a branch of artificial intelligence aimed at teaching and developing algorithm models. As an area of artificial intelligence, it was founded at the Dartmouth conference in 1958 with the report made by Solomonov R. on training without teacher, i.e. computer-aided training. Machine learning of mathematical analysis of algorithms has contributed to the emergence of computer training as part of theoretical computer science. This area includes image recognition issues: recognition of symbols, manuscripts and speech, text mining, computer-aided vision related to robotics.

Computer-aided language learning

Currently, computer-aided study of the various systematic languages is very popular in the language teaching. Hundreds electronic dictionaries, machine translation for various languages, teaching software systems, expert systems, and information retrieval systems have been put into operation so far. Here, we will briefly inform about the computer-aided learning of agglutinative languages. We will mainly focus on linguistic analysis of the text, which is one of the most important steps in computer-aided languages learning. The Azerbaijani language is given here as an example of agglutinative language.

By the way, let's clarify some details of the processes taking place during computer-aided learning of natural languages. What language learning tools are usually needed? Of course, first of all, vocabulary, grammar and conversation manual are very important.

When it comes to computer-aided language learning, the situation changes. It is important to introduce an electronic version of the dictionary, to reduce the grammar to a formal structure, and information-retrieval systems, dialogue systems, expert systems, dictionary sounding, including query system should be developed. In other words, linguistic provision and information technology issues have to be addressed, and we mainly focus on this matters. It should be noted that bilingual dictionaries designed for the learning systems are quite different from the usual ones. Usual vocabularies include words and their meanings. Whereas the dictionaries designed specifically for the mentioned system are not limited with the words and their meanings. They take into account the character of the given problems. For example, along with the words and their meanings for machine translation and language teaching, they include what part of speech it belongs to, sound decline (grapheme decline), ending with the letters g, k, t, homonymic syntactic and semantic features, the presence in a fixed word combinations, and verbs grouping (valency and being effective or ineffective). It should also be noted that since the electronic dictionaries are open systems, new words can be added her anytime.

The dictionaries designed for learning systems provide the image descriptions of the words and their sounding using visual tools. When it comes to the suffixes included into the database, they are grouped based on the parts of speech they belong to. Moreover, field-specific terminological dictionaries are included into the database as linguistic software.

Unlike the training courses, learning programs based on phonology provide the system of vowels and consonants, and the harmony rules expressed with sound examples in the form of a dialogue.

Learning program system based on lexicology mainly contains word creativity. The main emphasis is put on the formal description of the word synthesis in the Azerbaijani language [12]. For a formal description of the word synthesis, the harmony of the suffixes with the word forms, interaction of roots and suffixes, position of the suffixes relative to the roots, the harmony of the suffixes in the word positions, and the algorithm issues should be designed as the query-dialog.

Learning program system based on morphology includes the parts of speech (nouns, adjectives, pronouns, numerals, adverbs, and verbs). Of course, the data about the parts of speech are included into the computer memory in formalized form. At the same time, auxiliary parts of

speech: *goshma* and auxiliary names, connectors, particles, modal words, binding, exclamation, imitation words, imperative words, children's words, and rhythmic words are included into the linguistic database [13].

Learning program system based on phraseology includes the key factors proving the existence of the language units of the fixed phraseology and the classification of phraseology units [14].

Lexical analysis implies the separation of the texts into the paragraphs, sentences and words, the definition of the language of the text, the determination of the sentence types (with the help of key words and the alphabet), and the definition of the types of lexical expressions (jargons, spells, etc.). The realization of these issues on computer is not so easy.

Morphological analysis deals with splitting the words of the text into the differential fragments, and defining which part of speech they belong to, in short, automated definition of features.

Syntactical analysis defines the relationship between the words and sentences in the text and the position between them.

Semantic analysis is the most complicated process, which determines the meaning and the content of the sentences in the text. This logic-based analysis defines the logical dependence among the words, and the problem given in this regard is formalized. Since the semantics is impossible to be formed through traditional methods, special quality expert systems should be developed. These systems define how properly the meaning is found and check its quality. It also depends on the creation of artificial intelligence systems.

The development of science in this regard is still pending. Effective use of search engines through available information technologies is not up to date yet.

Syntactic analysis of language teaching systems understood as the search for the main and second members of the Azerbaijani texts and their syntactic relationships. Of course, they should be reduced to a formal form for such a search in the dialogue system.

Automated text translation

As detailed information about the automated or machine translation (MT) is given in the Azerbaijan linguistics, we will focus on new trend of machine translation in the Azerbaijan linguistics - statistical machine translation, which has to be resolved [2].

Machine translation algorithms are established in accordance with two different principles in the literature – basing on special rules and statistics.

Statistical machine translation bases on the comparison of a large-scaled language pair. The language pair is an expression of the sentences of a text in one language with the respective sentences in another language. As if the text options written in two languages and their translation. The more language pairs are, the more accurate the compliance coefficient is, and consecutively the higher the quality of statistical interpretation.

In some cases, the hybrid machine translation is used. Hybrid machine translation uses various respective options of MT.

When it comes to the computer linguistics in Azerbaijan, mathematical methods have been used since the 60s of the last century - with the publication of the book "Mathematical linguistics in education" by Garayeva M. [15].

Research studies have been implemented on all branches (machine translation, the development of formal models of the Azerbaijani language, and language study with statistical methods) of mathematical linguistics founded by Garayeva M. In 1976, a new area of the Azerbaijani linguistics - a group of applied linguistics was found at the Institute of Linguistics after Nasimi. Under Pines's leadership, this group began to deal with the problems of applied linguistics, including machine translation (formal description of word synthesis, morphological analysis, automated syntactic analysis and synthesis of texts, automated text editing etc.), formal modeling (modeling the structure of verb forms in Turkish, automated synthesis model, etc.), and statistical analysis of ancient monuments with statistical methods. In 1979, Akhundov A. wrote

his work “Mathematical linguistics” [16]. His recently published work “About some features of the applying the structural and mathematical linguistics methods in the Turkic languages” [17] is also of great interest.

Azerbaijani scientists are successfully using above-mentioned methods in their research. Thus, Akhundov A., Valiyev A. and Melnikov G. worked on the axiomatic method [16-20]; Valiyeva K., Pines V., Mahmudov M., Amirov Z., Fatullayev A., Khalili A., Guliyeva Z. and Iskandarova N. – on modeling method [12, 21-27]; Melnikov G., Valiyeva K., Valiyev A. and Mahmudov M. – on theoretical linguistic methods [28-32]; Valiyeva K., and Mahmudov M. – on the use of the theory of sets, automats and algorithms [6, 7, 33, 34], Pines V., Valiyeva K., Mahmudov M., Rahmanov J., Sultanov V., Mammadova S. and Shikhiyeva I. – on the statistical method [35-40].

It should be noted that we will not be limited to the work done and will be involved in future research, such as oral speech recognition, computer-aided language training, development of video and electronic dictionaries and etc.

Computer-aided natural language recognition

As mentioned above, the importance of the human-machine relationships in regards to in-depth study of the structure of the natural languages, developing the most complete and important theory for the communication that reflects all aspects of the communicative act and the inevitability of involving the linguists to these issues to satisfy the users’ needs appeared at late 60s. Evidently, the main communication in man-machine relationship must be in the form of dialogue. Depending on the purpose, the dialogue can be at three levels: global, thematic and local levels. Global level communication is determined by the common characteristics of the given question; the structure of the thematic dialogue depends on the specific algorithmic solution of the problem, and finally, local level dialog involves the various stages of the mutual dialogue between the participants [41].

As for the computer-aided recognition of natural languages, speech recognition and text recognition should be noted here. Since both areas are studied separately and cover a wide scope, we will only provide a brief introduction to these recognitions.

For automated text recognition, naturally, first of all, linguistic support should be provided. Linguistic support includes the following issues.

1. Automated (machine-aided) phonological recognition of graphemes;
2. Automated (machine-aided) recognition of morphemes;
3. syntactic recognition of the chain of morphemes;

Grapheme level is the conversion of the respective alphabet of the language (graphemes of the text) into machine language (machine language alphabet consists of two figures: zero-0 and one-1), that is coding. The alphabet of the Azerbaijani language consists of 32 graphemes. So, 32 graphemes are expressed in the machine figures consisting of 0 and 1, i.e. encoded.

As for the automated recognition of morphemes, as the morphemes are composed of the graphemes’ chain, their coding is expressed with the sum of the respective graphemes. It is known that the root and suffix morphemes are distinguished. It should be noted that root and suffix morphemes must be included into the memory of the computer.

For syntactic recognition of the morphemes the phraseological expressions, encoded sintagmas of the fixed word phrases should also be included in the computer memory.

Along with the linguistic support, mathematical support should also be provided for computer-aided recognition of the natural languages. Of course, for mathematical support recognition applications should be based on the solution algorithm of the given problem.

Recognition programs of different system languages are currently available. The written texts are included in the computer through the recently developed scanner facilities, and then, the

texts are read with the use of recognition programs, and the proper operations are enabled in regard with the given problem.

Computer-aided recognition of the spoken language, as well as the oral speech differ from the recognition of the written texts. However, the similarities of solution method are manifested in the language approaching level. Here, instead of the graphemes, the phonemes should be sounded in order to include them into the linguistic database. On the one hand, they should be studied from the phraseology and acoustic aspects for their introduction, and on the other hand, all the data related to the phoneme system, including the key signs and status of phonemes, their perceptive and functional aspects and correlation should be included into the linguistic database. In addition, the allophonic variations and complexion of the phonemes, their types, orthoepy and recognition function should be considered when generating the linguistic support.

The syllables play a key role in the speech recognition. Therefore, the structure of syllables should be thoroughly studied and their formal models should be developed. One of the key factors in the speech recognition is the emphasis. According to the linguistics literature, it is clear that in different languages the emphasis depends on its function. For example, in the Russian language, each word has its own emphasis. Because, in this language, the emphasis has a 'culminative' function. However, as the emphasis carries the delimitative function in the Turkic languages, it is put on the final syllable of the word [42]. Speech recognition software defines the emphasis with the higher sounding of the latter syllable. In this context, as syllable models, intonation, and the rhyme of the words in the speech recognition is another research subject, we will not touch upon these problems.

It should be noted that the reading of ancient manuscripts is one of the interesting issues of artificial intelligence. The survey on the reading hieroglyphic writings of Maya can be cited as an example. Living in the American continent, and having its own writing system, these people have been deprived of their culture. They lost their books, and rich library was burned by the Spanish invaders [43]. Today this manuscript is being preserved in Dresden, Madrid and Paris. Scientists from several countries, especially the United States, France and Germany, have worked on deciphering this manuscript, but haven't achieved any result. In 1951, the Soviet scientist Knorozov Y. opened these manuscripts, i.e. brought the Mayan hieroglyphs to the same shape. Then, scientists encoded these hieroglyphs and included into the computer memory. In addition, the images of the Mayan manuscripts were also encoded and included into the computer memory. The goal here was not to find the meaning of separate Mayan hieroglyphs, but the whole meaning of parts of the text.

It is known that the definition of the frequency of allographs and syllables plays an important role when reading the manuscripts. In other words, for deciphering the Mayan manuscripts, the prevalence rate of different letter combinations (graphemes) must be determined, and then, the frequencies Mayan hieroglyphs should be calculated and obtained prevalence frequency should be compared. As the result of the comparison of the frequencies obtained based on the developed program, the meanings of the hieroglyphs of the Mayan manuscripts have been found. Thus, the ancient Mayan manuscripts have been decoded through the computer - one of the greatest achievements of cybernetics.

When exploring the scope of PC, inevitably, the question arises: is it possible to create artificial intelligence?

Before creating the artificial intelligence, that is, an "intelligent" machine "thinking" and "understanding" as a human, it is necessary to explore the essence of the human himself/herself. It should be noted that internal and external devices of the computer need to be improved most. These devices are unable to read the hand-written texts directly and to speak as a man yet. Though, any scanned text is now possible to be included into the computer memory. However, since the recognition programs are limited, it is impossible to read arbitrary texts. Bilateral relations through *Skype* program can be built over the Internet. Nevertheless, the problem of speech recognition to

handle the computer by natural language still remains unsolved. Extensive studies should be conducted on phonology, lexicology, syntax, grammar, semantics, and morphology on various language levels in order to teach the machine to talk to a man.

Obviously, sounds is the spread of different frequency waves in the air. Different people may differently sound the same word in own way. Apparently, each individual's voice tone, timbre, intonation and so on differ. For this reason, it is very difficult for the computer to recognize the voice. Nevertheless, the human voice can be artificially synthesized by conversing the sound dances into the electronic ones. For the voice recognition, the voice is filtered and then compared to the etalon stored in computer memory. If any match occurs, the machine identifies that voice, otherwise it fails. Another method is based on the recognition of separate sounds, syllables and words.

When it comes to the computer "vision", each letter is coded and included into the computer's memory. Thus, the characters included into the computer's memory will be recognized only.

Thus, the creation of the "vision" and "conscious" computer does not mean the formation of the artificial intelligence yet. Intelligent computer that can be considered at that time, introduced him to solve a mathematical problem or a task in itself may pose.

In fact, there is no a general idea about the notion "conscious". Usually, the "conscious" means the creative ability or in other word, the ability to solve any problem differently from the previous experiences but with the vision of the interaction between the previous outlook and the surrounding objects. Actually, thinking is a broader concept, which combines a number of features as the perception of abstract and super abstract senses, teaching ability, character formation, decision-making and so forth.

A group of experts working in this field believes that the whole nature of the human thinking can be applied in the process of data processing. Therefore, these processes can be programmed in high performance computers and used. Thus, this computer, in principle, may be as a human in all spheres of intellectual activity. In-depth research work has already started for the creation of artificial intelligence.

Conclusion

Thus, the application of mathematical methods in linguistics is not irrational; it is needed to address certain issues. Universal programming languages (algorithmic languages) have been developed in order to implement the automated programming of the given problem. Due to these languages, the issues facing not only the natural languages, but also a number of humanitarian issues, including automated texts analysis in linguistics, machine translation, dictionary compiling can be programmed and solved.

It should also be noted that the recently published "Computer Linguistics" work by Mahmudov M. dedicated to the theoretical and experimental issues of linguistics has led to the resonance in the scientific society [42]. This monograph provides the comprehensive explanation of the national corpus of the Azerbaijani language, statistic lexicography, machine translation and formal linguistic analysis systems. Linguistic algorithms of explored formal analysis systems is of great importance in terms of language emergence and development. They can be widely used for retrieval systems, automated synthesis and analysis systems. Compiling the frequency and reverse dictionaries of the Azerbaijani language through computer, and statistical analysis of the historical monuments ("The Book of Dede Korkut" and Fuzuli's language) can provide useful facts to the linguistic studies.

Mahmudov M. rightly notes: "Computational Linguistics is a recently emerging science but with a great future, which will benefit the humanity. As it is not possible to imagine the future development of the society without computer, the future development trends and prospects of linguistics are also impossible without computational linguistics" [42].

References

1. Toldova S.Y., Lyashevskaya O.N. Modern problems and tendencies of computational linguistics. 24th International Conference on Computational Linguistics COLING 2012, Mumbai // Linguistics Problems, 2014, No 1, pp. 120-145.
2. <http://www.wikipedia.org>
3. Shtindlova J. Application of mechanization and automation methods in lexicological study abroad. Automation in linguistics. Moscow, 1966, p. 240.
4. Pines V.Y., Mahmudov M.A. Electronic computing machines in linguistics, B., 1977, p.63.
5. Zubov A.V. Processing of natural language text in "man-machine" system. Speech Statistics and automated text analysis, Moscow, 1971, c. 286-434.4. VY Pines, Mahmudov MA Electronic counting machines, linguistics, B., 1977, p. 63.
6. Valiyeva K.A., Mahmudov M.A., Pines V.Y., Rakhmanov S. Frequency dictionary of the Azerbaijani newspaper language, B., 1997, p. 212.
7. Valiyeva K.A., Mahmudov M.A., Pines V.Y., Rakhmanov S. Frequency dictionary of the Azerbaijani newspaper language, B., 2004, p. 264.
8. Mammadov M.A., Fatullayev A., Mammadova S. et al. Frequency dictionary of the Azerbaijani language (word roots), B., 2010, Vol.I, p.464.
9. Methods of Mathematical Statistics and Modelling in the comparative history of linguistics, <http://www.yazikoznanie.ru/content/view/127/215>.
10. Valiyeva K.A., Mahmudov M.A., Pines V.Y. and et al. Statistical analysis of "The Book of Dede Korkut", B., 1999, p. 248.
11. Mahmudov M.A., Fatullayev A. Reverse-Dictionary of the Azerbaijani language, B., 2004, p. 258.
12. Valiyeva K.A. Automated text analysis and synthesis, B., 1996, p.158.
13. The Modern Azerbaijani language. B., 1981, p. 443.
14. Mirzaliyeva M.M. Phraseology of the Turkic languages. B., 2009, p.240
15. Garayeva M.S. Mathematical linguistics in education. (Justification of the problem of foreign language learning by model). Baku, 1964, p.55.
16. Akhundov A.A. Mathematical linguistics. B., 1979, p.79
17. Akhundov A.A. Selected works. Vol.II, B., 2012, p.464
18. Valiyev A.H. Passage accents of the Azerbaijani language. Doctoral thesis. B., 1974, p.400
19. Valiyev A.A. Passage accents of the Azerbaijani language. B., 2005, p.334
20. Melnikov G.P. Some methods of description and analysis of the vowel harmony in the modern Turkic languages // 1962, 6, pp.31-58.
21. Pines V.Y. Modeling of the structure of the Azerbaijani verb forms in connection with the problem of automated dictionary. 1970, p.19.
22. Mahmudov M.A. Development of the system of formal morphological analysis of the Turkic word forms (based on the Azerbaijani language), Baku, "Elm", 1982, p.26.
23. Amirov Z.M. Development of the formal models of the Azerbaijani language and establishment of linguistic processor based on them. Abstract of the thesis submitted for the scientific degree of PhD in technical sciences, B., 2006, p.19
24. Fatullayev A.B. Development and application of digit-modeling method of the Azerbaijani-English machine translation system. NDA, B., 2006, p.19
25. Khalili A.M. Development of formal grammar of the "Limited Azerbaijani language" as part of the knowledge base of the deductive machine, NDA, B., 2009, p.34
26. Kulieva Z.Y. Determination of the optimal structure of the automated dictionary and machine translation systems. B., 2011, p.46.

27. Iskandarova N.A. Software-frequency and quality-semantic analysis of the translation adequacy and its application to the translations of different languages. 2008, p.19.
28. Melnikov G.P. Systemology and linguistic aspects of cybernetics. Moscow, 1978, p.368.
29. Valiyeva K.A. A formal description of the synthesis of words, Moscow, 1971, p.20.
30. Valiyev A.G. Transitional dialects of the Azerbaijani language. ADD. B., 1975, p.65.
31. Mahmudov M.A. Automated processing system of the Azerbaijani texts, DDA, B., 1994, p.64
32. Valiyeva K.A., Mirzaliyeva M.M. Azerbaijani-English translation system (theoretical problems) Manuscript, 2015, p.245
33. Valiyeva K.A., Mammadova M.H. Automated texts editing. B., 2003, p.80
34. Valiyeva K.A., Mahmudov M.A. Linguo-statistics: Speculation and reality. B., Studies 1, 2000, pp. 56-63.
35. Valiyeva K.A., Mahmudov M.A. Pines V.Y. and et al. The statistical analysis of «The Book of Dede Korkut" (preliminary results), B., 1999, p.248
36. Valiyeva K.A., Mahmudov M.A., Sultanov V.S. alphabet-frequency glossary of Fuzuli's poetry. B., 2004, p.548
37. Rakhmanov J.F. Statistical and distributive analysis of the Azerbaijani text (at graphemes and phonemes level). Abstract of PhD thesis. B., 1988, p.23.
38. Mammadova M.H. Automated selection of lexicon in the information - retrieval thesaurus based on an analysis of the terminological dictionaries. Abstract of PhD thesis. M., 1984, p.20,
39. Mammadova M.H. Development of the terminological databank of the Azerbaijani language // Soviet Turkology, 1990, No2, pp.84-89.
40. Shikhiyeva I.K. Linguo-statistic features of The Book of Dede Korkut. Abstract of PhD thesis, B., 1995, p.12
41. Popov E.V. Artificial intelligence. Book. 1, Communication systems and expert systems. M., 1990, p.464
42. Veysalli F.Y. Elements of general and private linguistics. I part. B., 2011, p.357
43. The Great Soviet Encyclopedia, Vol. 26, Second Edition, 1954, p.652