



## Active Application of Database Technologies in Lexicographic Practice

Djurayev M.K.

Termdu. Department Of "Applied Mathematics And Informatics"  
Teacher

### ABSTRACT

The article is devoted to the problems of creating semantic fields when expressing the semantics of lexical units in a database format, a parametric description of the meaning of a word. Database technologies are actively used in lexicographic practice. In the field of creating lexicographic databases, including the semantic description of some semantic units, it will be necessary to develop a system for automatic translation of a natural language, a machine translation system.

### Keywords:

Applied Linguistics, Lexicography, Linguistic Databases, Lexical-Semantic Field "Voice".

### Introduction

In modern linguistics, a lot of research is done to solve practical problems. Such research is related to the need to respond to the challenges of the external environment: society's information requests, problem-solving, and the development of new information and telecommunications technologies.

Nowadays, information is changing the way people think about their place in society. In the early stages of the development of the information society, it acquires the status of a source, and this inevitably leads to the need to store large amounts of information, which in turn leads to a new problem with its organization and systematization.

One of the problems with the need to create information systems in science, especially in linguistics, is that wider access to this information means more efficient actions, which allows for more efficient management of information flows. Therefore, it is important to develop interdisciplinary methods that allow analytical processing of information and the search for structural ways to store and process it effectively.

In this regard, database technologies (DB) are optimal, allowing the creation of arrays of stored and managed data systems using computer technology used to create and operate efficient information systems.

### Main Part

The purpose of this work is to review the research in the field of linguistic database and the possibility of applying this technology in lexicographic projects, as well as to provide options for such projects in the form of a lexicographic database reflecting the vocabulary of the Uzbek language. In projects, it is possible to form a language sound with semantics in exchange for providing a variant of such a project in the form of a lexicographic database, which reflects the dictionary of the Uzbek language with sound semantics.

Database technology is used in the process of creating traditional and electronic dictionaries. Dictionary databases of special and terminological dictionaries are being actively developed [1, 3, 6]. In linguistics, in order to develop this technology and create similar types of resources, it is necessary to solve the following tasks: 1) compilation of experimental material and initial analysis from creating

language level units (grammar, dictionaries, phonetic databases) to editing whole texts, this allows, on the one hand, to complete and define the model of the system, the language system, on the other hand, functional models of discursive fields and allows the formation of a model of the language system in general;

2) the task of finding new ways to correct and preserve language information, as well as to organize access to these materials;

3) the problem of finding new ways of processing materials to optimize research and achieve new results;

4) the task of verifying the results of research by referring to a large amount of material.

For example, the Russian scientist E.I. According to Yaroslavtseva, "World Languages", which is a universal grammar, can be used in linguistic typology to test assumptions about the typological similarity of computer database languages [7].

Today in Uzbekistan, the development of the theoretical basis for the creation of databases is being carried out for the first time in the framework of the creation of information retrieval systems for information and bibliographic searches. Work on linguistic resources will begin with the creation of a machine fund of the Uzbek language. In modern linguistics, computer representation of information based on database technology is used in different areas:

- ✓ in typological and comparative, phonetic, lexical, grammatical researches;
- ✓ lexicography;
- ✓ semantic research, including compilation thesaurus;
- ✓ for solving computational linguistics and applied linguistic problems (linguodidactical, automated translation, automatic speech detection and synthesis).

In the field of linguistics, in the current development of this technology, we can talk about two main types of databases:

1) full-text databases are documentary databases in which full texts related to a particular discursive field are provided.

2) real databases are facto-graphic type information systems that contain structured

information about different types of linguistic units. For example, the "Chronological morphemic-word-formation dictionary of the Russian language", the "EDGE" database for Russian adjectives, the phonetic database, etc. have been developed [1].

Despite active and effective research, the unfinished object of linguistics - language - is radically infinite and multidimensional due to a number of peculiarities:

- ✓ The elements of the language system are heterogeneous and quantitatively infinite;
- ✓ language structures are constructive, that is, a set of structured statements adapted to express the most complex mental structures is potentially infinite;
- ✓ language sign is dynamic - that is, changes occur in the process of its formal and meaningful aspects, which includes all language generations and provides variants of the sign or new elements of the system;
- ✓ In expressing the most difficult communicative meanings, language activity is closely interrelated with communicativeness;
- ✓ linguistic structures are related to cognitive processes and therefore it is very difficult to formalize linguistic semantics.

While each of these features speaks volumes about the infinite information potential of a language system, on the other hand, this information is extraordinarily complex, even at the level of individual linguistic elements. At the same time, the need for formal presentation of linguistic information to solve practical problems has led to an increase in research in this area. [8,9]

We will be able to talk about the theoretical and practical perspectives of using the database. Theoretical perspectives can be seen in the application of database technology in linguistic research, particularly in linguocognitive research. The use of this technology in the study of cognitive activity is related to the study of the

conceptualization of reality. The database provides information on the description of lexical semantics, methods of conceptualization of reality used in text analysis and machine translation systems, linguodidactics, etc.

On the other hand, modern works on linguistic thesaurus and database modeling speak about the prospects of using the category of natural languages in logical-linguistic conceptual modeling. R.Yu. According to Kobrin, there are currently two approaches to creating databases: a) creating language models of subject areas; b) construction of algebraic-logical databases. However, the second approach prevails in solving practical problems. For all this, the author argues that linguistic modeling can be the basis of conceptual modeling. In particular, he solves such a problem on the basis of linguistic analysis of terminology, establishing a system of semantic relations within a given field [10]. Attempts to solve this problem are presented in the works of Russian scientists A.N. Baranova [6], A.S. Gerda [3], S.E. Ni-kitina [12].

This approach is promising in terms of modern conceptualization theories based on the anthropic principle that is consistently reflected in natural language.

For example, the database "Voice", which reflects the lexical-semantic field, allows you to present it in a form composed of this conceptual field, which is typical for Uzbek speakers, because the so-called lexical-semantic field for presenting material in such a database. it is necessary to carry out the thesaurus structure of the semantic field and parameterize the semantics of the lexical units that make up this field.

Thus, the creation of linguistic data systems in the form of databases is an urgent and urgent task. But at the same time it is very complicated. Let's take a look at the problems that arise in the conceptual design of lexicographic databases using the example of the Voice database.

The first step in creating any database is to create a system of information tasks, the solution of which is to create an information resource. The structure of the subject area displayed in the database and the organization

of its infographic scheme depends on this system.

As for linguistic sources, in addition to information needs, it is necessary to take into account that here a part of the language system functions as a subject [11], so it is already necessary to take into account descriptions and classifications. In particular, when it comes to vocabulary with sound semantics, first of all, it is necessary to define the boundaries of the lexical-semantic field, and secondly, to study its structure in terms of existing works in this field 'ladi.

In this case, material boundaries can be set for at least two reasons: on the one hand, the units that directly define the types of sounds, even if they are in different forms, are reflected in the database possible (e.g., a vocal verb indicates the execution of a physical and vocal action at the same time). On the other hand, all the units that contain the semantic component of "sound" can be displayed in the database, including the units that define the voice of a musical instrument and the units that define a person's voice (whistle).

The structural and functional features of the database and its information potential are directly related to the answers to these questions. In our opinion, in order to realize the most complete information potential, it is necessary to include a wide range of units in the projected source, taking into account their different positions.

Looking at the results of previous studies [10-11] allowed to present the following model of the lexical-semantic field "Voice": a) LSG (lexical-semantic field) of onomatopoeia (boom, explosion, ku-ku, meow); b) LSG verbs are vocal (thunder, scream, shout); c) names of local self-government bodies (screaming, crying, voice); g) LSG of action names, the result of which is sound (singing, opening the mouth, screaming); d) names of LSG artifacts by sound (hissing, screaming, screaming); e) names of local self-government bodies by sounds (whistling, whistling); f) local authorities of animal names (screaming, hissing); h) self-governing bodies of sound signals (loud, loud, muffled, noisy).

## Results And Discussion

The core of the database, as well as the core of the lexical-semantic field, is the oral lexicon, which reflects the dynamic side of the sound phenomenon. Oral vocabulary, which is the main object of this description, all other objects (LSG) are derived from some degree of verbal expression in this field.

The next step in creating a database is to design a data-logic database schema, which is a system of tables that displays the attributes of the objects described in its fields. The semantic, grammatical, stylistic parameters of individual lexical units are used as attributes. The next step in creating a database is to design a data-logic database schema, which is a system of tables that displays the attributes of the objects described in its fields. The semantic, grammatical, stylistic parameters of individual lexical units are used as attributes.

The difficulty of parameterizing the semantics of natural language units has been noted by almost all researchers. Local and foreign works in the field of semantics offer a wide range of methods and techniques, but a complete description of the content plan of linguistic units remains an unresolved issue to date.

Allows you to rely on samples developed and tested in traditional lexicographic practice. As in the traditional dictionary notation, the following are distinguished in the formation of the attribute parameters of the unit in the database: grammatical data - characteristic grammatical forms belonging to the part of speech; stylistic features - information about the stylistic features of the unit; contexts of unit activities are given. However, the field of interpretation or presentation of semantic information, unlike the traditional form, requires additional structure, and here the method of analysis of the main components allows to identify the main semantic components. Peripheral semaphores are identified based on the analysis of the distribution of mobile drugs. The structure of the semantics for the presentation in the database is still incomplete, in which case a single description of the individual groups of units is adopted, but in this case, care must be

taken to take into account the maximum number of parameters.

In our opinion, for units of the core part of the lexical-semantic field - for vowel verbs, the following parameters are relevant:

Value type (direct - metaphor - metonymic); Grammar information; Form of national language (literary language - dialect - vernacular - jargon); Sound type; Acoustic properties of sound; Situational sound features; Subject; Theme features; Type of action; Characteristics of movement; Assessment of connotative sound; Connotative assessment of the topic; Connotative evaluation of movement; Emotional meanings; Compatibility; Context.

It should be noted immediately that the connotative value is an arbitrary component of the semantics of the lexical units described, and if the evaluator is present, it is determined, and if the existence is uncertain, the "neutral" sign is set.

For example, consider the parametric description of the verb: lexeme: hum; Value type: direct; Grammar information: verb; Form of national language: literary language; Sound type: insect sound; Acoustic properties of sound: silent, low-pitched, quiet, noisy;

Identification of lexical and semantic variants can be done by entering key identifiers in digital format, which allows them to be described as separate records in the table, but the shape of the word also indicates the type of value and allows their identification.

In describing lexical-semantic variants, this structure allows us to observe changes in semantics. Consider the description of the lexical-semantic variant of the verb described above:

Lexeme: hum; Value type: metaphorical; Grammar information: verb; Form of national language: literary language; Sound type: inanimate sound; Acoustic properties of sound: silent, quiet, low-pitched, noisy; Situational sound features: repetitive; Topic: mechanism; Theme features; Type of action: ongoing performance; Features of movement: monotonous; Con-notative voice evaluation: neutral; Connotative assessment of the topic: neutral; Connotative value of motion: neutral;

Emotional meanings: no; Compatibility: motor noise;

In addition, modern database management systems are complemented by the integration of multimedia objects. This allows you to enter other types of information into the database: audio or video material that shows the type of sound, its subject, and the action associated with it.

Thus, the solution of the problem of conceptualization and structure of specialized semantic fields can be considered as a topical task of practical linguistics, on the one hand, and theoretical linguistics, on the other. Creating such sources allows you to quickly and easily get information about the structure of lexical units that make up a particular semantic field, their stylistic and grammatical features, semantic structure, metaphorical and metonymic models operating in this field. We believe that such a resource will be in demand both in research and in the teaching process. A similar database can be used as a lexicographic source in the educational process: "Lexicography", "Lexicology", "Linguoculturology", "Linguistic description of the world".

## References

1. Кочергина К.С. Применение технологии базы данных в лингвистических исследованиях // Российская академическая лексикография: современное состояние и перспективы развития : сб. науч. ст. по материалам Междунар. науч. конф., посвященной 70-летию выхода первого тома академического «Словаря современного русского литературного языка» / отв. ред. О.Н. Крылова, С.А. Мызников, М.Н. Приемышева, Е.В. Пурицкая. СПб. : Нестор-История, 2018. С. 158-169.
2. Trap-Jensen, L. (2018). Lexicography between NLP and Linguistics: Aspects of Theory and Practice. In Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana: Ljubljana University Press, Faculty of Arts, pp. 25–37.
3. Rambousek, A. & Horák, A. (2015). DEBWrite: Free Customizable Web-based Dictionary Writing System. In I. Kosem, M. Jakubiček, J. Kallas & S. Krek (eds.) Electronic lexicography in the 21st century: linking lexical data in the digital age. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 2015, pp. 443–451. ISBN 978-961-93594-3-3.
4. Кочергина К.С. Проектирование базы лингвистических данных (на материале оценочной лексики) // Материалы XIX Открытой конференции студентов-филологов (Санкт-Петербург, 18-22 апреля 2016 г.) / отв. ред. С.И. Монахов. СПб. : Изд-во СПбГУ, 2017. С. 170-174.
5. Гайдамакин Н. А. Автоматизированные информационные системы, базы и банки данных. Вводный курс: учеб. пособие. М.: Гелиос АРВ, 2002.
6. Pulatov A.K., Muhamedova S. Kompyuter lingvistikasi (o'quv qo'llanma). – Toshkent, 2014.
7. Rahimov A. Kompyuter lingvistika asoslari. -Toshkent: Akademnashr, 2011.
8. Кобрин Р.Ю. Лингвистическое описание терминологии как база концептуального моделирования в информационных системах: автореф. дис. ... д-ра филол. наук. Л., 2016.
9. Марчук Ю.Н. Компьютерная лингвистика. М.: АСТ: Восток– Запад, 2007.
10. Никитина С.Е. Семантический анализ языка науки: на материале лингвистики. 2-е изд. М.: Книжный дом «ЛИБРОКОМ», 2015.
11. Pulatov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011.