

# APPLIED AND COMPUTATIONAL LINGUISTICS

## English Poetry Analysis with the Assistance of Database

Olena Bondaruk

Applied Linguistics Department, Lviv Polytechnic National  
University, UKRAINE, Lviv, S. Bandery street 12,  
E-mail: obondaruk@yahoo.com

*Abstract – the article illustrates the role of database application in linguistic researches, namely in master theses at the applied linguistics department of Lviv National Polytechnic University. The novelty of the analysis lies in exploring the linguistic phenomena using the database management systems. The analysis is exemplified by the exploration of database that is used to resolve the linguistic issues and to optimize the examination of lexical features of English poetry in the late 16th - early 17th centuries.*

Key words – applied linguistics, database, linguistics, master's thesis, poetry.

### I. Introduction

The applied linguistics research calls for the need to create the linguistically oriented databases to process the large volumes of texts and work with them. When students are writing their master's theses, they frequently prefer to work in the environment of database, namely when it is the applied part of the diploma, because the functions and means of this program (i.e. the conciseness of data presentation, the rate of its processing, the convenience of information usage, the low expenditures of labor) contribute to the qualitative and efficient fulfillment of the author's linguistic objectives.

The aim of the article is to illustrate the significant role of database application in the linguistic researches, especially when preparing the master's diploma papers at applied linguistics department (by the example of database development of English poetry of the late 16th – early 17th centuries).

Working with the texts, students have both to collect and save them for further analysis and to find lingual objects under research easily and quickly. Computer technologies of linguistic data arrangement allow students to examine the lingual units' relationship and to identify language patterns. Linguistic database illustrates an effective way to organize lingual units; it is used when scientists have to deal with large amount of information. The development of these databases is focused on their multiple and various application, i.e. the processing of language material by different parameters.

### II. The Database System as the Tool of Linguistic Analysis

The area of database systems has experienced sustained vigorous growth during the last three decades because of the strong pull of commercial applications and the incessant push of technology and research advances. Databases have also gained much recognition as a field for computer science research and university teaching [1: 1].

Databases and database systems are an essential component of life in modern society: most of us encounter several activities every day that involve interaction with a database.

A database is an organized collection of data for one or more purposes, usually in digital form. The data are typically organized to model relevant aspects of reality, in a way that supports processes requiring this information.

The term "database" refers both to the way its users view it, and to the logical and physical materialization of its data, content, in files, computer memory, and computer data storage. This definition is very general, and is independent of the technology used.

According to C. J. Date, the database organization is the computerized system of records storage and its central goal is to retain information and provide it to the user [2: 4]. The main purpose of the database is the reflection and storage of information and knowledge that appear in the process of resolutions of a certain number of tasks. Thus, the database system can be observed as the manifestation of objects, subjects, processes, and events in the environment of computer information system [3: 39]. All things that can be reflected in such a way compose the physical prototype of the database.

Generally, the word database is a computer term for a collection of information concerning a certain topic or business application. Databases help to organize this related information in a logical fashion for easy access and retrieval. A computer database is an automated version of the filing and retrieval functions of a manual paper filing system. It is a structured collection of records or related data about one or more subjects stored in a computer system. The structure is achieved by organizing the data according to a database model [4: 102].

The database may hold the information about the possibilities and methods of linguistic objects application in different communicative situations or in the different products of language activity, it also contains the opinions or assessment of these objects. Consequently, such information, on which you can make the inferences about language objects, is called knowledge, while the collection of this information presents the knowledge base or intelligent database [5: 34]. Thus, the database system is especially convenient for linguists in the process of language phenomena exploration.

A database has the following implicit properties:

- it represents some aspect of the real world, sometimes called the miniworld or the universe of discourse. Any changes in the miniworld are reflected in the database;
- it is a logically coherent collection of data with some inherent meaning. A random assortment of data cannot correctly be referred to as a database;
- it is designed, built, and populated with data for a specific purpose. It has an intended group of users and some preconceived applications in which these users are interested [6: 4].

In other words, a database has some source from which data is derived, some degree of interaction with events in the real world, and an audience that is actively interested

in its contents. In order for a database to be accurate and reliable at all times, it must be a true reflection of the miniworld that it represents.

The process of designing a database begins with an analysis of what information the database must hold and the relationships among components of that information. Often, the structure of the database, called the database schema, is specified in one of several languages or notations suitable for expressing designs. After due consideration, the design is committed to a form in which it can be input to a database management system, and the database takes on physical existence [7: 25].

In order to construct the database of English poetry of the late 16th – early 17th centuries we have chosen the software that organizes the storage of data, it is known as a database management system (DBMS) Microsoft Access, because it is “the most common relational DBMS in Microsoft Windows that provides storage, classification and search of the data” [8: 283]. The DBMS is a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications. In Access, a database is the overall container for the data and associated objects. It is more than the collection of tables, however – a database is made up of all objects.

Defining a database involves specifying the data types, structures, and constraints of the data to be stored in the database. The database definition or descriptive information is also stored by the DBMS in the form of a database catalog or dictionary; it is called meta-data. Constructing the database is the process of storing the data on some storage medium that is controlled by the DBMS. Manipulating a database includes functions such as querying the database to retrieve specific data, updating the database to reflect changes in the miniworld, and generating reports from the data. Sharing a database allows multiple users and programs to access the database simultaneously [6: 5]. When the user is working in the database environment, he or she deals with the several objects, the main of them are tables, queries and forms. The tables are the foundation of the database that consists of columns (fields) and rows (records), according to the terminology of Microsoft. We can enter the data directly into the table, but it would be more convenient to use for this purpose another Microsoft Access object - the form. Moreover, the forms allow the client to view, edit and insert the data that is stored in the main database tables. Finally, the selection of required information from one or more tables depending on the user-defined criteria is automatically performed with the assistance of queries.

Database management systems are categorized according to the database model they support. A database management system, such as Access, FileMaker Pro, Oracle or SQL Server, provides user with the software tools needed to organize that data in a flexible manner. It includes facilities to add, modify or delete data from the database, ask questions (or queries) about the data stored in the database and produce reports summarizing selected contents.

Therefore, the databases permit to observe and analyze the data in different ways, taking into consideration various characteristics of examined units. This is very convenient for students while writing their master's theses.

### III. The Database Development as the Tool of Poetic Analysis

This article aims at describing the productivity and advantages of database system application in master's diploma papers by the example of developed database. This database represents the lingual features and images of real world that are inserted in English poetical texts of the late 16th – early 17th centuries (the period of Queen Elizabeth I). The main task was to decide what information the database will contain, which tables it will have, how these tables are related, and what fields those tables should include. Consequently, the elaborated database involves the list of Elizabethan poets' verses, the information about these authors (i.e. biography, photo, social status) and index of the analyzed images (“love”, “life”, “religion”, “grief”). The design of Microsoft Access database may consist of the following stages:

1) Determination of the purpose of database: what information we want to receive from the database. The developed database searches the images in analyzed verses, illustrates the lingual signs of the images, counts the frequency of the images' appearance in the verses, and gives the general information about the Elizabethan poets. Thus, the database needs to store facts about such subjects (the tables) as: Elizabethan poets' general information, list of analyzed verses, images that illustrate the man's world of life, verbal representation of the images in the verses and the facts the database needs to store about each subject (the fields in the tables) (Fig. 1). After opening Access 2000 environment we can see active window representing the core of our database. It shows the tables, queries and forms that exist within database.

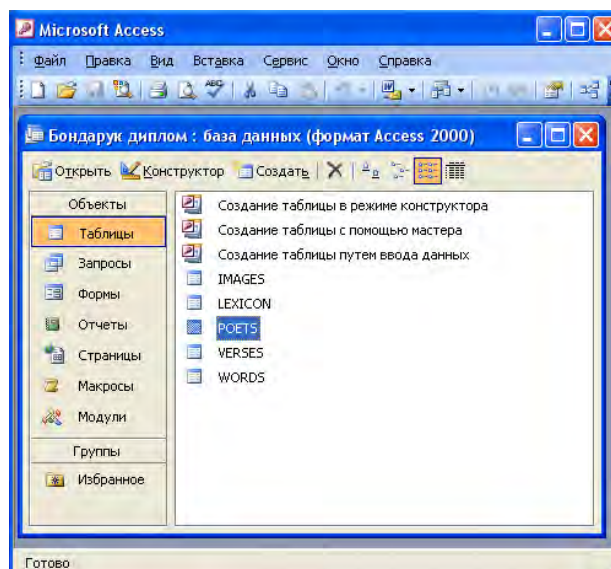


Fig. 1. Structure of the database

2. a) Search and organization of the information required: gathering of the information that will be recorded in the database.

b) Division of the information into tables: separation of the information items into major entities or subjects (Fig. 2). Each subject then becomes a table. A table should not contain duplicate information, and information

should not be duplicated between tables. When data are being entered in Access, a table stores it in logical groupings of similar data and the table's design organizes the information into rows and columns.

ID Вірша	ID Поета	Назва вірша	Жанр
1	01P	Ah, Silly Pug, wert thou so Sore Afraid	couplet
4	01P	In Defiance of Fortune	heroic couplet
5	01P	Oh, Fortune!	rhymed stanza
6	01P	On Monsieur's Depart	rhymed stanza
7	01P	The Doubt of Future Foes	rhymed stanza
8	01P	When I Was Fair And Young	couplet
9	01P	Written in her French Psalter	heroic couplet
10	01P	Written on a Wall at Woodstock	rhymed stanza

Fig. 2. Table „Verses” with the main information about the Elizabethan poets’ works

c) Selection of the appropriate data types for fields. We can save space in database and improve join operations by choosing appropriate data types for fields. When defining a field, choose the smallest data type or field size that is appropriate for the data in the field.

3. Changing of the information items into columns: what information we want to store in each table. Each item becomes a field and is displayed as a column in the table. Each table contains information about the same subject, and each field in a table contains individual facts about the table's subject. When sketching out the fields for each table the following points are important:

- relate each field directly to the subject of the table;
- do not include derived or calculated data (data that is the result of expression);
- include all the information we need;
- store information in its smallest logical parts.

At this stage the suggested database contains 5 tables: “Poets” with the information about poets (name, bibliography, social status, photo), “Verses” with the list of analyzed verses (title of verse, literary style), “Images” with the list of images and their attributes, “Lexicon” with the examples of verbal representation of the images, “Words” with the list of attributes.

4. Specification of primary keys: choosing the primary key of each table. The primary key is a column that is used to uniquely identify each row. A key is an entity in a table that distinguishes one row of data from another. The key may be a single column, or it may consist of a group of columns that uniquely identifies a record. Tables can contain primary keys which differentiates records from one another. Primary keys can be an individual attribute, or a combination of attributes.

5. Establishment of the table's relationships: deciding of how the data in one table is related to the data in other tables. This database contains one-to-many relationships.

6. Entering the data for the database: recording of all the information that is gathered into the database.

7. Creation of the forms to make the data entry easier. Forms are made up of the fields that can be entered or viewed in edit mode. The one below is the main form which includes the references to the rest of forms (Fig. 3), thus informs us about all the data made of and give the opportunity of choices. From the perspective of daily use, forms are the most important objects we build in Microsoft

Office Access application because they are what users see and work with every time they run the application.

According to T. Konolli, when designing forms, it is necessary to place three types of objects on screen:

- labels and text box data-entry fields;
- special controls (multiple-line text boxes, option buttons, list boxes, check boxes, business graphs, and pictures);
- graphical objects to visually enhance them (color, lines, rectangles, three-dimensional effects) [4: 564].

Fig. 3. Main form of the database

8. Search of the data using queries. Query is used to extract information from a database. A query can select and define a group of records that fulfill a certain condition. For query development, Access offers a Query Designer, a graphical user interface that allows users to create queries without knowledge of the SQL programming language. In the Query Designer, users can show the datasources of the query (which can be tables or queries) and select the fields they want returned by clicking and dragging them into the grid. Joins can be created by clicking and dragging fields in tables to fields in other tables.

The most basic Select queries retrieve the records we specify from a table. We can choose the fields from a table to display, and specify the criteria for selecting records (Fig. 4, 5). In the most cases, while viewing the query results you can modify the data and update the original records. The first step in creating a query is to specify the table or tables to use and the fields to display.

Fig. 4. Result of search query on the attribute (lingual sign) of the image

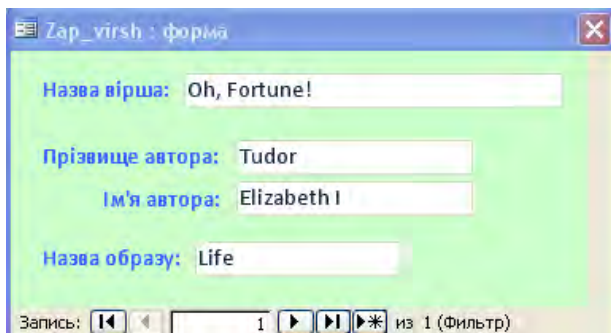


Fig. 5. Result of search query on the title of verse

9. Development of the forms that provide the most reasonable results of this research, for example, the information about the Elizabethan poets (Fig. 6). The elaborated database also illustrates the charts (Fig. 7).

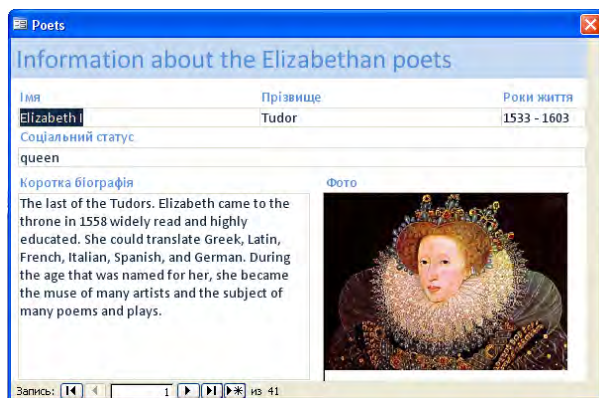


Fig. 6. Illustration of poet (Elizabeth I)

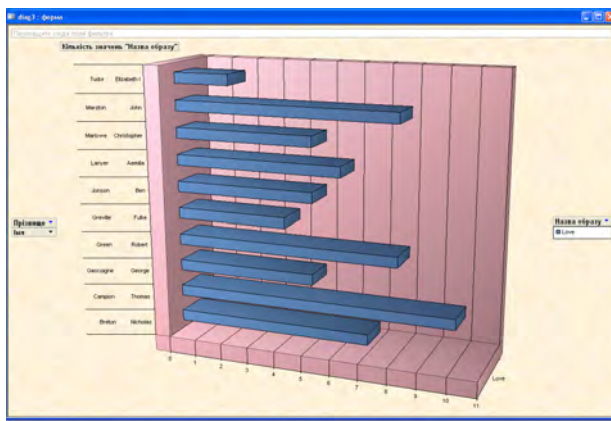


Fig. 7. Chart of frequency of images usage by the certain authors (image of LOVE)

The developed database offers a range of benefits to the potential users providing information on the list of main images in the verses of English poets during the late 16th – early 17th centuries and their verbal representation. In this research database fulfills the following linguistic tasks: to characterize each author by the image; to illustrate the image of each verse; to identify the frequency of each image; to

determine the attribute of the image. Due to using Microsoft Access forms the database provides a quick and easy way to modify and insert records.

## Conclusion

The database is the convenient and effective tool for the linguistic information structuring. It can be applied in various fields of linguistics, where is a demand to analyze a large number of lingual units or to make their classification. The analysis means of Microsoft Access allow the user to sort out the necessary material according to the determined criteria very quickly and to obtain the quantitative data. Consequently, the developed database helps to illustrate that each verse of the poet is personal, it reflects the experience, feelings, and thought of the author, that is the reason why we should be attentive in the process of verse analysis. The database is the best medium for storing the necessary material in large quantities; it can create the connections among the data types that facilitate the search of the results and can also illustrate the relationships between linguistic units. These facts confirm the diversification of the database program possibilities that are necessary for students of applied linguistics department of Lviv National Polytechnic University.

## References

- [1] C. Zaniolo, S.Ceri, Ch. Faloutsos, R. T. Snodgrass, V. S. Subrahmanian, R. Zicari, "Advanced Database Systems". San Francisco; California: Morgan Kaufmann Publishers, 1997, 574 p.
- [2] C. J. Date, "An introduction in database systems", 8th ed. Boston: Addison-Wesley Publishing Company, 2003, 1024 p.
- [3] A. Yu. Berko, O. M. Veres, "Teoretychni osnovy baz danykh: konspekt lekcii dlya studentiv Instytutu kompyuternykh nauk ta informatsiynykh tehnolohiy" [The theoretical background of database: the lecture notes for students of the Computer science and information technologies department]. Lviv: Vydavnytstvo Lvivskoi politekhniki Publ., 2007, 190 s.
- [4] T. Konolli, K. Behh, "Bazy danykh. Proektirovanie, realizatsia i soprovozhdenie. Teoriya i praktika" [The database: planning, realization and accompaniment. The theory and practice]. – M.: Vil'yams Publ, 2003, 1440 s.
- [5] Ye. A. Karpilovs'ka, "Vstup do kompyuternoii lnhvistyky" [Entry into the computer linguistics]. D.: Yuho-Vostok Publ., 2003, 183 s.
- [6] R. Elmasri, S. B. Navathe, "Fundamentals of Database Systems", 6th ed. Boston: Addison-Wesley Publishing Company, 2010, 1172 p.
- [7] J. D. Ullman, J. Widom, "A first course in database systems". New Jersey: Prentice Hall, 1997, 470 p.
- [8] T. M. Conolly, C. Begg, "Database systems: a practical approach to design, implementation and management", 4th ed. London: Pearson Education Limited, 2005, 1374 p.