Human Language Technologies – The Baltic Perspective
A. Tavast et al. (Eds.)

doi:10.3233/978-1-61499-133-5-90

Geoinformational Database of Lithuanian Toponyms

Dalia KAČINAITĖ-VRUBLIAUSKIENĖ[1]

Abstract. Geoinformational database of Lithuanian Toponyms serves several goals: to foster scientific research and purposes of applied science (1), to satisfy practical needs of inhabitants (2), to preserve Lithuanian toponyms as a cultural heritage (3). It is the first database in Lithuania that jointly provides linguistic and geographic information about the toponyms.

Keywords: toponym, database, linguistic information, geographic information

Introduction

A Geoinformational Database of Lithuanian Toponyms is the first database that encompass information about linguistic units (toponyms) and geographical units (geographical objects) in Lithuania. The database includes information not only about existing settlements, rivers, and lakes, but the extinct toponyms that were collected from local people between the World Wars (1935–1940). Thus, the database provides an opportunity to approach the entirety of currently existing and historic toponyms of Lithuanian territory on the internet, that is, an easy access is provided to all the related information about authentic forms, accentuation, etymology, descriptions of toponyms to geographers and linguists worldwide.

1. General Information about the Project

The database[2] was started in 2007 when a governmental subsidy was acquired.

The database is created by The Institute of Lithuanian Language. During the period of 2007–2010 the project was carried out in participation with an Institute of Geology and Geography (later – with Centre of Natural Research). At present, The Institute of Lithuanian Language is collaborating with professional makers of linguistic and geographic databases.

[1] Dalia Kačinaitė-Vrubliauskienė: a research fellow, The Institute of Lithuanian Language, Department of Onomastics. P. Vileišio Street 5, LT-10308 Vilnius, Lithuania; e-mail: daliakacinas@yahoo.com.

[2] The database URL is htpp://lvvgdb.lki.lt, the database could also be reached through a website of The Institute of Lithuanian Language: http://www.lki.lt.

The toponyms are presented according to territorial-demographic principle. In a first stage (2007–2010) a database was created for Lithuanian cities, towns, township names and municipalities of Kalvarija, Kazlų Rūda, Marijampolė, Pagėgiai, Druskininkai bei Švenčionių region. A second stage (2011–2013) of a project is sponsored by a Council of Lithuanian Science according to a National Program of Lituanistic Development (2009–2015), (pact Nr. LIT-4-19). It encompasses data from municipalities of Ignalina, Birštonas and Rietavas.

1.1. The Sources of Data

As I have mentioned before, the database provides linguistic information that was collected form colloquial language i.e. toponyms that were collected from local people, and geographical information about the objects that the toponyms refer to. The toponyms and the objects are taken from already existing databases as well as from the archives of the Institute of Lithuanian Language.

For current toponyms and for borders of their geographic objects (cities, towns, villages), a graphic data (approximate borders of inhabited terrains included) of Lithuanian Republic addresses registry was used for the reference („© A Registry of Adresses of the Republic Lithuania"). Most of the rivers, lakes, forests, as well as their names are transferred from Vectoric database of a Cosmic view map M 1: 50000 („LTDBK50000 © National Land Office, 1996–2010 ") and from a georeference M 1:10000 database of Lithuanian territory („Georeference base of Lithuanian territory ©National Land Office, 2003–2010"). The cultural objects and their names were taken from a Database of Cultural Heritage Protection Department (© The Department of Cultural Heritage, as a part of the Ministry of Culture).

The aforementioned databases do not encompass all varieties of toponyms in colloquial language. They do not include some types of geographic objects and their names, as well as small rivers and lakes. Furthermore, a mass of extinct toponyms are still filed only in card indexes and forms of interwar and post-war period. The most important of them are stored in the Institute of Lithuanian Language: Sheets of proper names of Lithuanian land (1935–1937); Catalogue of the Commission for Surnames and Toponyms and files of oiconyms (1935–1938); Oiconym list of Vilnius region (1940–1945).

2. The Goals and Application of the Project

The database was started in order to achieve several goals: 1) to expand scientific and applied research of Lithuanian toponyms in easily accessible environment for scientists and general public (i. e. the Internet); 2) to reveal the importance of links between geographic objects and their toponyms in formation and genesis research; 3) to present an authentic and unpublished toponymy that was collected during an interwar and after-war period from a colloquial language; 4) to provide an opportunity to deepen regional, genesis, and formation research; 5) to teach and to promote the usage of authentic, correct toponyms; 6) to satisfy practical needs of Lithuanian Market Subjects and citizens; 7) to sustain toponyms as non-material cultural heritage.

It is worth to emphasize that it is the first electronic source of comprehensive geographic and linguistic data. Furthermore, it provides not only current, but historic data as well. This data is very unique and of great importance because it shows a natural state of toponymy that was not affected by collectivization, russification or irrigation processes. That provides a lot of data to linguistic, geographic or cultural studies. The data was not analyzed much, and was hardly accessible to researchers of Baltic studies, outside of Lithuania.

It is worth to note, that until now all research of toponym genesis and formation was conducted without referencing to its geographic object, its location. Thus, links between toponyms were not always accurately identified. The database provides coordinates of toponym objects on a map which should increase accuracy of aforementioned links.

The future results would not only be important by themselves. They would stimulate further research on toponyms: 1) regional toponymy, because objects are presented according to their current and interwar links; 2) formation analysis, because toponyms are already classified according to the formation type; 3) genesis analysis, because toponyms are already classified according to genesis and sources of genesis.

3. The Modes of Linguistic and Geographic and Data Classification

The geographic objects in a database are classified as follows: a settlement, a hydrography, a relief, a land revetment, an infrastructure, cultural heritage. Objects of each group are classified into types: a settlements into cities, towns, townships (unofficial status), villages, steadings, railways stations, estates, homesteads, parts of cities, parts of towns, parts of townships, parts of villages, parts of estates, etc.; hydrographic objects into lakes, rivers, streams, ponds, ditches, moats, trenches, bogs, swamps, springs, sources, quagmires, fords, shoals, hauls, parts of lakes, parts of rivers, gulfs, dams, swimming holes, bends of rivers, straits, water – troughs, etc.; relief objects into mountains, hills, downhills, taluses, scarps, exposures, dunes, caves, pits, combs, denes, mounds, valleys, hillocks, gravel pits, clay pits, islands, banks, peninsulas, plains etc.; land revetment objects into woods, forests, boscages, parks, cuttings, bushes, forests yards, gardens, parts of forests, parts of woods, outer woods, outskirts, trees, soils, unbroken soils, kitchen-gardens, pastures, fields, grasslands, meadows, strips of lands, sites, boundaries, stones, marshes, turbaries, etc.; objects of infrastructure into ways, roads, gravels-roads, bypaths, paths, bridges, streets, squares, parts of roads, etc.; cultural heritage objects into mounds, barrows, tumules, cemeteries, castles places, altars places, crosses, chapels-poles, stones, cults buildings, other buildings, birthplaces of importants persons, residences of importants persons, places of importants events etc. A search for geographic objects at the database may be done according to each of these layers.

Objects search is possible according to three status features: existing, extinct, unknown (interwar period data).

All existing geographic objects have a unique code and they are accurately indicated on a map, using M 1: 10000 Lithuanian territory georeferent base. The borders of existing settlements, lakes, and rivers are marked by lines, other types of borders are marked by dots. Contemporary (region, municipality, county, settlement)

and interwar (city, town, township, village, steading, other) period territorial dominion of objects is also indicated. A search may be based on any territorial unit.

All extinct objects or the objects of unknown status have their unique codes and are approximately located on a map (by a dot), using a georeference M 1:10000 base of Lithuania territory. One of the most difficult geographic issue that arose while creating the database was locating the objects on a map that were taken from historical sources of interwar and post-war period. Because the location of these objects was not clearly described or even absent (in that case, only a terrain is known). For that reason, an accuracy of location is indicated in a database (e.g. location is known; a terrain is known; a township is known). It is possible to search for objects using these accuracy indicators.

Additional information is provided about each geographic object. The database provides information about a size of the objects: a population, a warren, a length, a height, an age (of trees), an ambit, a weight (of stones), a diameter, a perimeter. The objects that are taken from other databases „bring with themselves" the information about their size, while others are described as it is provided in their sources. Furthermore, as much as possible, we try to attach photographs of settlements, lakes, rivers, hills, etc (See Figure 1.).

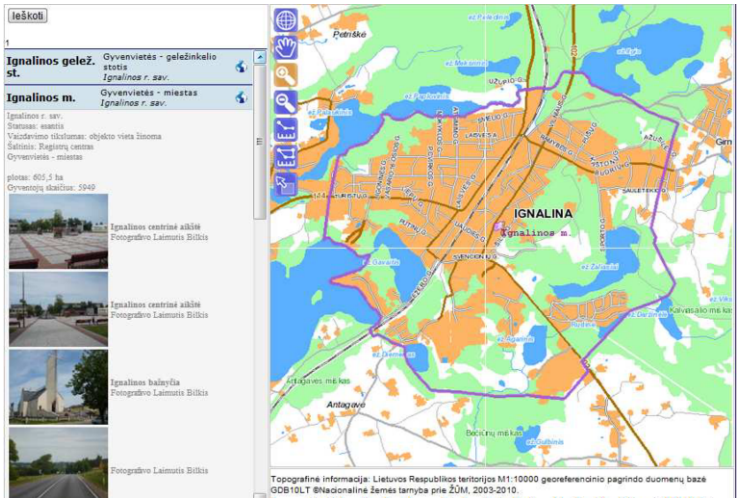

Figure 1. Geographic information about Ignalina town

The database provides toponyms of all geographic objects. The toponyms are analyzed in various linguistic aspects: an accentuated form is indicated, as well as gender; a number; an accentual paradigm; a genesis (Lithuanian, Latvian, Prussian, Curonian, Semigallian, Indo-European, Slavic, Germanic, Greek, Latin etc.); genesis based on a source (appellative, personal name, toponymic); a word-building (an original word, an inflection derivative, a suffix derivative, a prefix derivative, a compound). Furthermore, an additional information: accentuated cases, sound records (pronunciation of all cases); if data is accessible, a date of a first reference in historic sources is provided as well (usually about cities or larger towns) (See Figure 2).

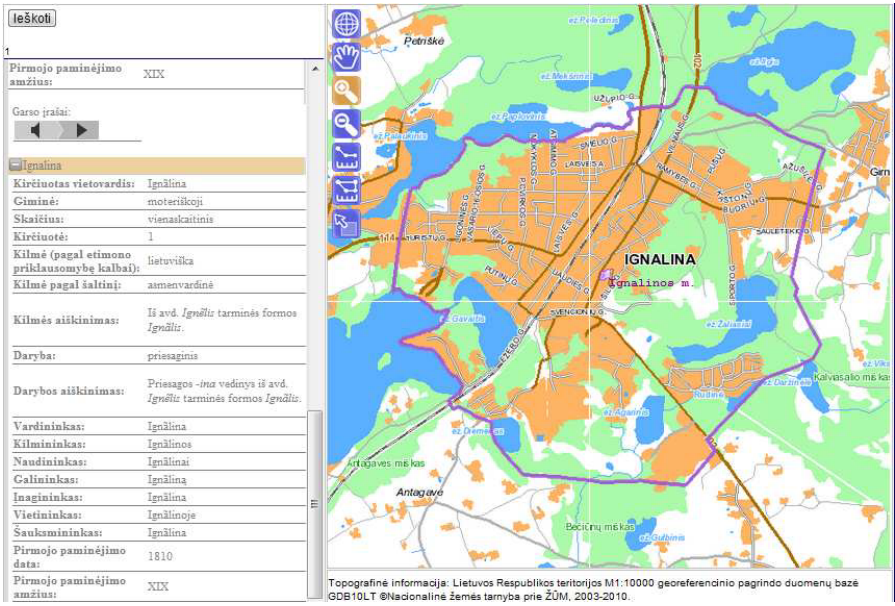

Figure 2. Linguistic information about Ignalina toponym

In order to keep data authentic and out of respect for persons and their hard work of recording data in an interwar and a postwar period we decided to provide some information about them: their names, professions, age, and a year of a record.

4. Technical Details of Data Inputting

Principles of data input and updating: 1) initial data is created in external systems. or existing data from geoinformational databases is used; 2) geographic data is used in ESRI Shape and dBASE formats. While geographic data is imported into system, a code of geoobject is created (geoobj.geokodas). Shape data is imported into geographic data tables (geo_taskai, geo_linijos, geo_plotai). Features that are linked with a type of an object are transferred from dBASE tables into table g_pozymiai; 3) a new linguistic record of a toponym is created in a system. A record may be linked with a geographic object; 4) information about geographic objects is administered in a system's geographic object editing module; 5) linguistic information of a toponym is created and administered in a toponym editing module; 6) toponyms and geographic objects may be linked in toponym and geographic editing modules; 7) sound records and pictures are edited in external systems. They are inserted into this system using relevant editing tools; 8) record of a toponym or geographic object may be deleted independently. In case of deletion a link between toponym and object is also eliminated.

Conclusions

In conclusion, this is (will be) the first large scope scientific database of Lithuanian toponyms on the Internet. All the data (authentic forms of toponyms, accentuation, etymology and description) will be easily accessible to linguists and geographers worldwide. It will be also a possibility (in case of repossessing a land or rebuilding settlements) to retrieve extinct toponyms and to start using it correctly.

A creation of Lithuanian toponyms geoinformational database is a long and demanding process. Despite all our efforts, it will take a huge amount of time to analyze all toponyms at hand. However, we are always looking for the ways to improve our process, to make it quicker, as well as to develop the structure of a database, to make it easier to use, more inviting.