

# *Hungarian Digital Toponym Registry<sup>1</sup>*

## *Results of a research programme*

Valéria Tóth

### 1. On the research programme titled ‘Hungarian Digital Toponym Registry’

1.1 Background. The ambition to compile a comprehensive registry of the toponyms of the Carpathian Basin has been present in Hungarian toponymic research since the 19th century. The collection of toponyms was launched at the national level in the 1960’s under the leadership of the Hungarian Academy of Sciences, with the movement reaching its peak in the 1970’s and early 1980’s. However, in the subsequent period of time the work slowed down soon to come to a complete halt. As a result of the nation-wide collection of toponyms (covering two-thirds of the country) huge toponym registers were published, from which we can draw two important lessons: 1. The toponyms they contain constitute a rich repository of the language and culture of the Hungarian nation and other peoples living in the Carpathian Basin. 2. The compiled toponymic registers remained mostly unexploited, as they failed to encourage theoretical processing. However, not only do these works provide precious source material to linguistics in a narrower sense, but other disciplines may also widely benefit from them. Furthermore, if – relying on the technical achievements of the modern age – the material of the toponymic registers is processed in a database and linked with tools of GIS, further possibilities open up, and the information that can be extracted from the database may benefit researchers of different disciplines (e. g. in the standardisation of toponyms or in the planning of landscape reconstruction).

Thus in this respect toponymic research in the Carpathian Basin has a double task: (1) linguistic processing of the huge amount of source material, which should make it possible (2) to resume after a long time the suspended collection of toponyms, applying identical principles and methods, yet this

---

<sup>1</sup> This work was carried out as part of Research Group on Hungarian Language History and Toponomastics (University of Debrecen / Hungarian Academy of Science).

time surpassing national borders to include the whole Carpathian Basin. It is a further fundamental requirement that the existing corpus of toponyms should be made available in an appropriate and durable form to researchers of linguistics and other disciplines.

1.2 Social and scientific significance of the programme. Besides the endeavour to serve scientific purposes, the Hungarian Digital Toponym Registry (HDTR) bears also considerable social, or we could also say, national-cultural significance. Names are, notably, such special linguistic elements that play an essential role in our everyday lives. Besides practical benefits, toponyms express our relationship to nation and culture, but they are also powerful emblems of our immediate ties to the local surroundings. Names, passed down to us from the past, are often reminiscent of archaic linguistic and cultural relations. Therefore it is not surprising that the versatility of the roles fulfilled by them attract the attention of several disciplines (in addition to linguistics, they fascinate researchers of history, historical geography and ethnography) and their study gave rise to an independent discipline, namely the truly interdisciplinary science of onomastics. This field of research has great accomplishments in surveying ancient culture and contemporary connections of the Carpathian Basin. This does not only apply to the Carpathian Basin itself, since names have been playing a universal role also in the shaping and preservation of national identity.

Besides obvious benefits drawn from the study of toponyms for research purposes, due to the tendency of globalisation, our contemporary life experiences an increasing need for the thorough study of names and their use in different areas of life. The use of names has been indispensable for more than a century in public administration, postal services or transportation, to which new dimensions have been added by mass communication and digital technologies (e. g. GPS). What is more, fire fighting, ambulance or natural disaster prevention services could not operate, if toponyms did not facilitate their spatial orientation.

Such increasing demand for names in our contemporary society is becoming more and more difficult to satisfy, since individual people's knowledge of the names in their surroundings is always necessarily fragmentary. The tension can be eased with the help of science, namely through the collection of names occurring both in oral communication and in historical sources, by sorting and categorising names into data stocks and by making them available to the public in the form of books and on-line databases.

Scientific analysis of the toponymicon may contribute also to the revelation of the past of inter-ethnic relations. In wider Hungarian terms, that is, in the regions beyond the borders of contemporary Hungary, names occupy a distinguished place in the language-political issues related to the use of the Hungarian language (to mention just one aspect, let us refer to the permanent skirmishes fired by the question of the official usage of toponyms). Under contemporary conditions (in the lack of toponym registers and collections), linguistics is barely able to provide professional support to the language-political situation. Improvement of the current status is in the fundamental interest of all Hungarians living in the Carpathian Basin. The use of names plays a vital role in the language politics of individual states, including the press, public administration and cartography, etc., which is one more reason for scientific research to compile authentic collections of name stocks and to make them available to society.

## **2. Establishment of the HDTR, its current status and IT background**

2.1 Being faced with a pronounced and constantly increasing need for toponyms, it was high time for linguistic research to make appropriate steps to satisfy such need. At the Department of Hungarian Linguistics of the University of Debrecen two research groups are involved in the establishment of the Hungarian Digital Toponym Registry (HDTR). With this programme we undertook the creation of a huge data warehouse suitable for the accommodation and management of the entire toponymicon of the Hungarian language. The data warehouse is public and available to anyone online, providing information to both the general public and to researchers. As for the software, the fundamentals of the database are complete, with the uploading of linguistic material and toponyms in progress.

Also the selection of the database management software and the setting up of the structure for the accommodation of the data are completed: data stocks emerging during registration are imported into the 4D database management system. a) Geo-referencing and digitisation of the maps are carried out with the help of the GIS software ArcGIS and ArcView, while b) name data are recorded first in MS Word, then in MS Excel spreadsheets to be compatible with most popular database management systems ensuring a wider, general use. The overall popularity of these widely used software applications, namely,

enables us to involve university students into the first phase of data recording, as the use of these software applications does not require any special IT skills.

2.2 Since work is carried out in several phases, registration and entering into the database of the toponymicon of particular areas, has so far reached different levels of completion. As a first step, the registration of digitally available (or digitised) toponymic stocks is performed, which means that we progress from comitat to comitat, and within comitats, from settlement to settlement, registering all information comprised in published volumes of toponymic registers in MS Word or MS Excel spreadsheets. Uploading of the data takes place mostly with the co-operation of students from the University of Debrecen, but students and instructors from other institutes of higher education are also involved. Due to the wide co-operation, a large toponymic stock could be recorded in the database within a relatively short period of time. The data recording is currently in a phase of completion as follows. Almost the entire toponymic stock of three territorially connected counties of South-Western Hungary has gone through the mentioned work phase. Yet name stocks have been recorded also from other areas of the Hungarian language territory: besides some Western counties, also the names of several Northern and Eastern Hungarian settlements have been entered into the database. Currently, this adds up to a stock of approximately 300,000 name data having gone through the first phase of work.

Yet the mentioned stock becomes suitable to be imported into the database management system only through further work: definition of the standard linguistic form of name data and the selection of the lemma name (main variant or table-name) from the variants belonging to a particular object are also indispensable for the operation of some functions. Nevertheless, since these work phases require much more competence they are mostly performed by PhD students and assistance researchers. At present this part of the work has been concluded only with respect to the toponymicon of particular Southern-Hungarian areas, covering some 100,000 records in total. Considering name data and the related philological information, this is the name stock quantity that is suitable to be imported into the database. Nevertheless, this can be carried out only if the cartographic stock of the settlements of the respective area has been completed as well. Namely, the 4D database management system contains a GIS component, which enables also the map-like representation of the toponymic data.

Consequently, at present only a part of the material of the toponymic register, i.e. some 90,000 modern-age toponymic data are available through the Internet at the homepage of the Hungarian Digital Toponym Registry ([mdh.unideb.hu](http://mdh.unideb.hu)).

In the operation and development of 4D we are assisted by German and Hungarian professionals of IT and GIS, what is more, also the processing of certain areas (that include also German-speaking toponymic material) takes place within the framework of a German-Hungarian co-operation.

### 3. Opportunities offered by the HDTR

The following section contains an overview of the options that the HDTR offers to its users. The HDTR provides separate user interfaces to researchers and the general public. Yet currently, selection and search options function only via the general user interface. The version for researchers, which is meant to be suitable for more complex search and selection, as well as sorting based on different linguistic criteria, is yet in the development phase.

**3.1 Selection.** The selection function can be used to simultaneously view all toponyms belonging to a settlement that has already been entered into the database. There are two methods to do this: users can either opt for representation in the form of a map, or select the settlement name from a list. In the first case, by clicking on the map icon of the opening page, with the help of Google Earth, users will see the settlement names of the selected area; from here they can proceed by clicking on a particular settlement name to see all toponyms belonging to the respective name. (In such cases only one name version, i.e. the table name appears on the map.) Further basic information on the toponyms appearing on the map can be obtained by clicking on them. Using the Selection field, a list is displayed containing all toponymic data of the chosen settlement. (This function will also display the number of toponym variants in use on the particular settlement.) It is possible to click on any name version appearing in the list, and the user will be taken to a more comprehensive data page of the name containing basic information on the toponym and the settlement. (E.g. the data of the name are listed both in a dialectal or a written, literal form, marked with the exact year of the usage, adding possible versions with affixes or synonyms, etc.). With the help of the globe icon appearing on the data page a map-like representation of data can also be prompted.

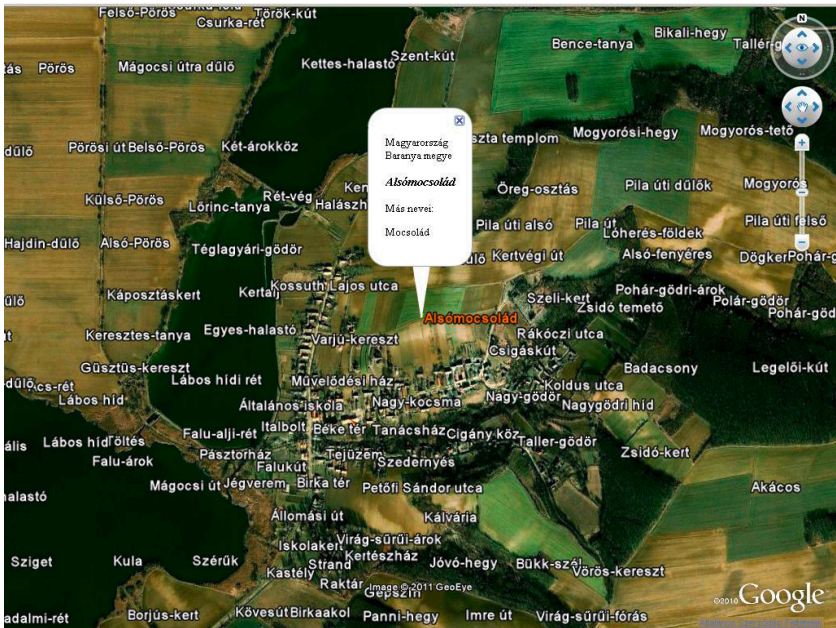


Fig. 1: All toponyms of the settlement Alsómocsolád in the HDTR.

3.2 Search. As opposed to the Selection function, the Search function serves queries regarding particular toponyms or parts of toponyms (i.e. lexemes constituting toponyms). Entering e.g. the toponym *Kenderföld* (referring to a ‘land where hemp is grown’) into the Search field, the results of the query will appear in two lists: the first list will contain the names that are structurally identical with the original name, that is, all names that have the form *Kenderföld* (we can see that at the moment the database contains 99 objects under this name). The second list will contain names that incorporate the original one: cases where e.g. an attribute is linked to it (*Alsó-Kenderföld* ‘lower hemp field’), or a posterior determinant is added (*Kenderföld utca* ‘street leading to the hemp field’), or possibly an inflection morpheme is inserted (*Kenderföldek* ‘fields where hemp is grown’). At the moment the database contains 108 such names. Thus with the help of these two lists users can obtain all toponyms in which the surveyed name form appears independently or as part of another name. With help of the globe icon appearing above the lists, users can also display their search hits in a graphical representation on Google Earth.

Fig. 2: Data page of the HDTR.

Searching for a lexeme or a structure (e.g. the name of a particular tree, a geographic common name or a postposition, etc.) will yield the same result.

In both cases, by clicking on any of the name forms appearing in the list the respective data page of the name appears in the next window, along with pieces of information and the possibility of map-like representation described earlier.

#### 4. Long-term objectives and plans

(a) The toponymic data and their representation on the maps clearly display certain deficiencies as well. Since collection of toponyms mostly took place in the 1970's, the processed registers reflect the status of toponyms at that time, without reference to the significant changes that have been taking place both in the settlements themselves and their names (the Democratic Turn of 1989 for example brought along a fundamental alteration of land ownership relations and, for different reasons, the transformation of several names). For the database to be up-to-date on the toponymicon of different areas, further on-the-spot collection work is necessary to complete, verify and update the material. In addition, the maps of some older toponym repositories are rather inaccurate due to technical constraints of the time, which can be clearly observed when compared to the maps of Google Earth. The review of local characteristics can also be performed on the spot. Of course, names whose exact place is not familiar to the name user community can occur also in the modern-age toponymicon. Therefore, names that cannot be localised accurately are marked in our maps at the level of the settlements only, with an additional comment in the information section referring to this fact. We believe, however, that from a linguistic or onomastic-analytical aspect linking names to settlements (without further detailed information) may in many cases provide sufficient information that has substantial relevance from a linguistic aspect.

With these obviously longer-term tasks in mind, we launched an experimental toponym-updating collection work in South-Hungary in August 2011, in the framework of which we checked, completed and localised the data of 16 settlements with up-to-date information related to the names. With this experiment we aimed at developing the methodology of the working process,



on the one hand, and we intended to gain an estimate of the temporal, financial and human resource requirements of such tasks, on the other hand.

(b) At the same time, we consider it of high importance that the collection of toponyms should be facilitated also in areas where so far no (or only few) toponym registers have been published. In this respect we took the first steps in Hajdú-Bihar county (which includes Debrecen) by selecting the settlements of a smaller region. Work here as well was carried out with the help of undergraduate and PhD students. After collection work on the premises and the processing of the sources from public records, the compiled data stock will evidently be integrated into the HDTR and the new toponym registry will be published also in the form of a book, the first volume being expected to be printed in autumn 2013. We plan to proceed with the material of the rest of the regions of the county in the same manner.

(c) Scientific preoccupation with toponyms has been characterised right from the early beginnings by an intensive fascination with the names of ancient times, with special attention being focused at all times at the toponyms extracted from early charters. On these grounds we decided to dedicate a separate module within the HDTR to toponyms dating back to the period before 1350. (Setting the time limit for 1350 was somewhat arbitrary, but in fact our decision was primarily motivated by the fact that the named year is an important division line in Hungarian language history, what is more, it is often considered a landmark also in different dictionaries of language history, toponymy and historical geography.) Thus the HDTR includes two main modules: one of them comprises the contemporary toponym stock, covering the whole Modern Age (with toponymic data mainly of the 18th to 21st centuries), while the other module includes the early toponymicon focusing especially on the period prior to 1350; of course, the two modules are organically interlinked. (Their division can be executed in the easiest way by using the Year field of the database.) Since scientific interest in historical names is nowadays satisfied by several manuals, a great amount of primary sources are at hand to be used for the early Hungarian module of our registry. In this module the cartographic representation of the toponymic data are realised with the help of reconstructions of medieval maps projected (through georeferencing) onto Google Earth. Obviously, in this module the local characteristics of any given toponym can only be established at the level of settlements. The compilation of the early toponym registry has been launched and now it is ready for operation (with



more than 20,000 early toponym data).

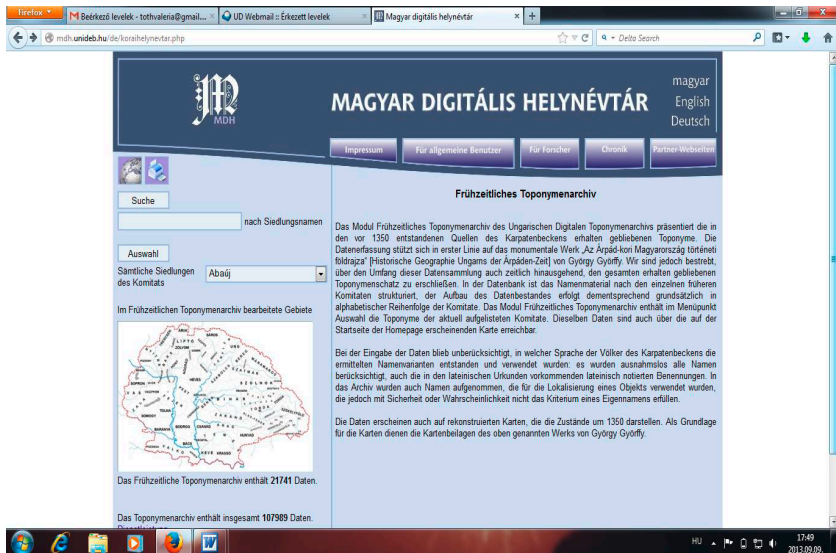


Fig. 3: The Modul of Early Toponym Registry.

(d) As mentioned earlier, the interface of the digital registry that is available over the Internet at the moment is meant primarily for the general public. Nevertheless, besides the ambition to satisfy the curiosity of the wider public, the HDTR is, above all, a database meant to serve purposes of scientific research. Therefore the version for researchers – that will include complex search and selection functions, the option to retrieve data on the basis of linguistic criteria and the possibility of cartographic representation – is among our top priorities. To be able to accomplish this plan, besides philological data, we need to complete the database with data of the linguistic analyses (e.g. information on name structure, denotatum type, semantic relations and etymology, etc.). Obviously, the interface designed for research purposes will offer much more complex functions than the one meant for the general public, yet in addition to the satisfaction of research needs, it will be made accessible also for non-academic queries. Yet more restricted needs of the general public will not be taken into consideration in the design of the interface.

## 5. Summary

Finally, I would like to add a few general closing remarks based on the conclusions of the first three years of the programme. Naturally, we are aware of the fact that – similarly to other large-scale enterprises – also the HDTR may encounter certain stumbling blocks: for example, excessive expectations may easily jeopardize the whole project. Therefore it is of utmost importance that right from the launch of the project we restrict ourselves to a clearly outlined framework and work processes that are transparent and organically inter-linked; naturally, the initially set limits may be modified in the course of the process to adapt to the experiences of the accomplished work. The human resources and infrastructural conditions required for purposes of the Hungarian Digital Toponym Registry are basically available, yet continuous financial support is needed for its implementation. We strive to finance these needs through applying for different calls for proposals. However, the crucial pre-condition for the success of the project is the widest possible co-operation between research communities of the Carpathian Basin in order to cover the entire Hungarian language territory.

[**Abstract:** Hungarian Digital Toponym Registry. Results of a research programme. – In the summer of 2010 a research programme was launched under the title *Hungarian Digital Toponym Registry* with the long-term objective of recording and analysing the complete stock of toponyms of the Carpathian Basin. The programme has been implemented with the co-operation of experts working at various Hungarian and foreign research facilities, and it aims at the exploration of the history of toponyms reaching back to ancient times up to our days. The Hungarian Digital Toponym Registry is suitable for research purposes and may at the same time be of interest to the general public. The database encompasses two separate modules: the Modern Toponym Registry, which basically contains 18th–20th century toponyms, and the Early Toponym Registry, dedicated predominantly to storing place names dating back to the times before 1350.

The IT background architecture is provided by the 4D database management system which also encompasses a GIS component thus making visualisation of toponymic data on maps possible: modern toponymic data are

projected onto the aerial photographs of Google Earth, while historic data are placed on reconstructed medieval maps.

My paper aims at describing the general objectives and research purposes of the Hungarian Digital Toponym Registry, as well as the results achieved so far. The Toponym Registry is available at <http://mdh.unideb.hu.>]

[**Abstract:** Das Digitale Ungarische Ortsverzeichnis. Ergebnisse eines Forschungsprojekts. – Im Sommer 2010 wurde unter dem Namen *Digitales Ungarisches Ortsverzeichnis* ein Forschungsvorhaben begonnen mit dem langfristigen Ziel der Aufnahme und Analyse des vollständigen Ortsnameninventars des Karpatenbeckens. Das Programm wurde in Zusammenarbeit mit Experten verschiedener ungarischer und ausländischer Einrichtungen aufgenommen, es möchte die Geschichte der Ortsnamen von den Anfängen bis heute untersuchen. Das *Digitale Ungarische Ortsverzeichnis* dient wissenschaftlichen Zwecken, kann aber gleichzeitig für ein allgemeines Publikum von Interesse sein. Die Datenbasis besteht aus zwei Teilen: das Moderne Ortsnamenverzeichnis enthält hauptsächlich Ortsnamen des 18. bis 20. Jahrhunderts, das Historische Ortsnamenverzeichnis erfasst Ortsnamen aus der Zeit bis 1350.

Die zugrundeliegende Software-Architektur wird durch das 4D Datenbank-Management-System realisiert, das unter anderem über eine GIS-Komponente verfügt und damit die Visualisierung toponymischer Daten auf Karten ermöglicht: moderne toponymische Daten werden auf Fotografien von Google Earth projiziert, während historische Daten auf rekonstruierten mittelalterlichen Karten eingetragen werden.

Dieser Beitrag möchte die allgemeinen und wissenschaftlichen Zielsetzungen des Digitalen Ungarischen Ortsverzeichnisses beschreiben und auf die bereits erreichten Resultate hinweisen. Das Ortsverzeichnis ist verfügbar unter <http://mdh.unideb.hu.>]