

Andrea Hegedűs

**GEOLINGUISTIC SURVEYS IN THE HUNGARIAN
LANGUAGE AREA IN ROMANIA USING THE
UNITED ATLASES**

Abstract of the PhD thesis

Budapest

2008

1. Theme and aim of the thesis

In my paper I deal with the theoretical, technical and formal questions of – mostly with computer – uniting dialectal atlases collected in the Hungarian language area and do geographical linguistic analyses using the united language maps of the eastern Hungarian language area. Sometimes I extend my survey to the whole Hungarian language area in order to get a better view.

In the introductory part of my thesis I gave the reasons for choosing this theme: After the publication of the Atlas of Hungarian Dialects (AHD) the researchers planned further and comprehensive dialectological surveys in the Hungarian language area, but their hopes were not realized. The causes among other things were as follows: 1. experts who could analyse maps had not much time to deal with linguistic geography among their other tasks, 2. there are just a few young researchers who have interdisciplinary knowledge which is necessary to “sound” maps, 3. nowadays it is very complicated to see through, dispose and group ten or hundred thousand data by traditional means without computer. In my thesis – among other things – I sketch out the plan of the technical apparatus which in my opinion would increase the more frequent use of the published language atlases. This is a database with a program, its name could be: “United Computerized Hungarian Language Atlas” (hereinafter called UCHLA).

In the first part of my paper I write about the plans of creating and running UCHLA, about the practical advantage of the database and the software and other tasks which are necessary to create it. In the second part I show different linguistic geographical surveys that I made on the basis of united linguistic maps. These maps were made and united by me from atlases and from a linguistic geographical work that presents the Hungarian language area in Romania: the Atlas of Hungarian Dialects in Romania (AHDR), the Atlas of the Moldavian Csángó Dialect (AMCsD), the Linguistic Atlas of Szilágyság (LASz), Székely Geolinguistic Dictionary (SzGD), and the Atlas of the Dialect of Hétfalu (ADH), as well as the AHD which also has Romanian collection sites. At the end of the paper I touch upon the future tasks relating to uniting linguistic atlases.

II. The UCHLA

In the second chapter of my thesis I define what uniting the atlases means and touch on the sources of this idea then I summarize the guiding principles of UCHLA. The birth of the idea of uniting the atlases was helped by the followings:

1. The big atlases and the regional atlases complete each other well, however they act different parts in the scientific research.
2. Some of our atlases are too heavy and large to use.
3. Uniting will be an evident task in the case of atlases by the same collectors or editors (e.g. the Moldavian Csángó and the Székely atlases by the

team of the University of Kolozsvár or the regional atlases by Olga Penavin collected in the regions of the former Yugoslavia).

4. The storage and processing of the large quantity of data which amounts ten or hundred thousand units is very difficult without computer nowadays.

5. The idea of uniting the atlases was joined to the technological developments at the Department of Hungarian Historical Linguistic, Sociolinguistics and Dialectology (ELTE) from the 90's.

6. The future plans of creating digital dictionaries had a stimulating effect on uniting the atlases.

In the further part of the chapter I write about the guiding principles of UCHLA and the technological elements and tables supporting the creation.

According to my plans the database would contain obligatory and free elements. The obligatory elements are the followings: 1. The headwords of the maps, 2. the labels of the headwords, 3. the settlements where data was collected, 4. the name of the atlas where the data got from, 5. some hydrographical objects helping the orientation, 6. the countries where the settlements are, 7. the counties where the settlements are, 8. the dialectal regions where the settlements are, 9. the collected language data, 10. other comments on the collecting circumstances, 11. the codes of the language data (according to the dialectal characteristics). I consider the following tables optional: 1. other dialectal texts, 2. some information from settlement histories, 3. dialectal groups, 4. regions. The database would increase permanently as long as dialectal collections are carried on in the Hungarian language area and it could be accessible via internet. Its expansion would be controlled by experts.

The program operating the database would work depending on the type of the future map. It could create maps with simple writing technology (as the data of the maps take up lot of space the representation of this technology is very limited), visual maps and summarizing phenomenon maps. The creation of the latter could be realized already after the adequate processing of the word forms. In the case of visual maps the dialectal characteristics of the data are replaced with colorful circles, symbols or the combination of these. They are more informative and they take up less space on the display of computer than the written word forms. The summarizing phenomenon maps could be created only by the use of encoded data, diagrams, colorful circles, other demonstrative instruments show the analyzed phenomenon, its rates and spread in the searched area.

Compiling the tables of the united settlement network and the united headword stock would improve the creation of the database to a high degree. In the appendix of my thesis there is a list of the settlements and headwords of the atlases collected and published in the Hungarian language area, and of course the settlements found in more atlases, as well as the headwords put down in different forms, but are relating to the same language reality.

I met the following difficulties while collecting the names of the settlements:

1. The names of the settlements occurred in different forms in different volumes, as they were not put down by the present administrative practice.

2. Different names were used for the same settlement in some atlases. This type occurs typically in the case of settlements beyond the borders, as the modern codification of place names relating to the Hungarian speaking areas is missing.

3. In the case of two settlements having the same name other points of reference help the localization and the lexical differentiation: e.g. the equivalents in a foreign language or the name of the county.

4. It happens that the foreign language equivalent of the settlement is not given correctly in certain volumes.

I specified and corrected the place names on the basis of the FNEsz or László Sebők's dictionary of toponyms (Dictionary of Hungarian place names beyond the borders. Budapest, 1997, Teleki László Foundation.)

After compiling a guide to the list of headwords and defining the real and reference headwords in the headword-list I concentrate on the difficulties occurred in compiling the list arranging them into 24 types. The following cases presented the biggest difficulties:

1. Spelling uncertainties: (e.g. in the case of compounds the same headword appears as an idiomatic expression in another volume). In general consulting the spelling dictionaries and dictionaries of Hungarian standard and occasionally dialectal dictionaries helped to determine the correct form of the headword.

2. Headwords in two different forms meaning the same in the standard language in two atlases: in this case the headword was formed according to the "principle of the bigger atlas", it means that the headword of the bigger atlas overwrites the one of the smaller one.

3. Two different headwords had relating to the same language reality, the one is less exact than the other: in this case of course the exact form becomes the real headword.

4. The different headwords mark the same denotation in two volumes, the one is usual in the standard language as well, the other in that dialect only: the standard form becomes the real headword.

5. The headword in the atlas is not informative enough for the computerized lexicographic system or the colloquial competence, so it has to be specified with different remarks in brackets or between appositions. In certain cases remarks can be left, although some volumes use these redundant information. In such a case the final headword appears without a remark.

6. Although the atlases do not indicate word class classification after a headword, it could be useful if the headword is a homonym.

7. If the headword occurs as an attributive compound in an atlas, it is worth forming it into a simple headword which is specified with remarks.

8. In the case of Romanian loanwords as headwords it is difficult to decide whether it has a standard or a regional standard equivalent.

9. I unified the headwords in the case of the parts of the plants: bean, maize, wheat, grain and grape: e.g. *tő* (a növényé) 'stock (of the plant)' or *töve* (a növénynek) 'stock (of the plant)', *szár* (a növényé) 'stem (of the plant)' or *szára* (a növénynek) 'stem (of the plant)' etc.

10. If two headwords relate practically to the same thing, but the one is a circumlocution, the other is an idiomatic expression and this appears in the data that way as well, I included them in two different rows.

The word list will become complete if not only real and reference headwords will be included in it, but also real reference words, that is those lexical, grammatical and phonological variants which are not headwords in the atlases but help finding the lexemes and word forms in them. There have not been made an index to most of our dialect atlases. While uniting first the index has to be made for the single atlases, then the matter has to be fitted into the common headword list. This big and complex task demands the initiation of a team, but the principles, the method of references, and the most important criteria could already be defined before the real work. I raise some points of view concerning this in my thesis. Some examples:

1. The points of view put into writing in the introduction part of the index of AHD can be considered as authoritative, other works, regional dictionaries less. But while compiling the index of AHD the small size was an important guideline, in the case of the database of AHDR or the united database morphological and important phonological variants have not to be left because the index can be published on CD or DVD as well.

2. It must be fundamental that the phonetic variants should be left from the index first of all.

3. The consonant reflexes being identical with those in the standard language (e.g. assimilation, fusion, falling, shortening guided by sound environment) do not create a form variant, so they has to be spellt according to the codified Hungarian orthography.

4. In the entries of the lexical maps all simple, derived and compound words, all idiomatic words would be included, in the case of compounds and idiomatic words the second and the third (that is they are not at the beginning of the word) element as well. In the latter case the members of the compound would not relate to the headword, but to the word or the idiomatic expression.

5. It is important to indicate the meaning of the "dialect words by meaning", and real dialect words have to standardized . The standardization is not an easy task, I have not found exhaustive point of view system yet related to this.

6. Letters having different orthographic signs can be replaced with the grapheme most similar to them. The transcription of the phonologic divergences demands individual judgment.

In the first term of 2008 an occasional team began to edit the index of AHDR, but it is not read yet, because the expansion of the point of view system is in progress and the matters are not checked. In my thesis I processed some entries

referred by the team, partly to suggest the difficulty of the referring problems, partly to show my own points of view through examples.

In the further part of chapter II. I write about the program and the database connected to the program that I could use for the geolinguistic surveys. Then I touch on how could the united maps be researched and used. The geolinguistic software I used for my geolinguistic surveys run on Macintosh, its name is Bihalbocs and developed by Domokos Vékás. Visual and statistical maps can be generated by the software. In the first step the program works with uncoded data. In the next phase the user himself/herself carries out the encoding when he/she creates the maps organizing the data into different groups. He/she creates the so-called legend of the map with the textual drafting of the criteria of the groups. The map can be stored as an image file, moreover it can be printed.

The grouping of the data and their representation can be various. E.g. the *Erzsiékhez* united map shows six variants of grouping and presentation possibilities.

The use of the united maps can be complex. In general it can be used by synchronic dialectology but other linguistic studies, moreover by other branches of science. Partly it can be used in the partner fields of linguistics, e.g. in ethnographic, region historic and ethnic characteristic researches, partly in different linguistic disciplines: areal linguistics, historical linguistics, historical dialectology, onomastics, semantics, lexicography etc. The UCHLA would be connected with other databases having hundreds of language data, so their material would become comparable with each other.

III. Geolinguistic surveys

In my thesis a short glancing precedes the geolinguistic surveys at the former researches on the Hungarian language area in Romania. On the one hand I touch on the researchers at Kolozsvár, their atlas collections and analyzing geolinguistic works, on the other hand I write about the Romanian Hungarian geolinguistic researches at the Department of Hungarian Historical Linguistics, Sociolinguistics and Dialectology at ELTE: the publishing of AHDR, the “Csángó” geolinguistic researches, the Geolinguistic Workshop and Team. After that I touch upon the phonetic, morphological and lexical surveys. I generated the maps myself for almost all analyses, except the maps on the so-called “í-zés” (a dialectal phenomenon) which were made by Fruzsina Sára Vargha.

In the two phonetic surveys I consider the theme of the “í-zés” in Szilágyság (Transylvania). In the first analysis with the study of statistical maps started out from one of my former surveys, but expanding the corpus with more data I try to encircle the spreading of the “íé” diphthong occurring frequently in Szilágyság and showing apparently historical connection with the so-called “í-zés”. The expanded corpus does not give other solution, confirms the lessons of the first analysis.

I analyze 9 visual maps to the other “í-zés” theme. One from the 9 united maps shows the whole Hungarian language area with three focuses. “Í-zés” occurs in the root of the word in five maps, and three in the affix. The border of “í-zés” coincides with the one written in the previous survey. “The opening diphthong realisation” is not typical in the areas of “í-zés” applying in affix at all. Further surveys are necessary to clarify its causes.

I have made four morphological analyses. The first deals with the spread of the infinitive affix *-nyi* and *-nya*. According to the evidences of the maps made for the survey, the palatalized *-nyi* variant is not typical in the Romanian language area, its focus falls on the western and the “Palóc” area. The *-nya* variant which originally is added to the verbs containing velar *i* sounds has already been declining and its occurrence can be studied only on the infinitive form of some verbs. It is found in the language area in Hungary by and large where the *-nyi* variant appears. It is typical in the Romanian area in the whole Székelyföld, and frequent in Moldva as well, but appears also in Mezőség in less pots. This variant is decreasing, it can be found in some words in the temporary zone, it does not appear in others. Far-reaching conclusions relating to the result (spread, history) of the survey is not allowed to draw.

The following morphological survey relates to the occurrence of the *-jik* variant of the genitive suffix of the third person plural. Six united maps demonstrate the spread of the phenomenon. The *-(j)ik* separates from the *-(j)ok* variants of the Big-Küküllő area along the river Olt and in Moldva, the other big block is found in Hungary, in the south of Transdanubia (Dunántúl). Fairly a characteristic picture takes shape concerning the territories of the phenomenon in spite of the little differences visible on the maps.

The third morphological analysis relates to the spread of the so-called “family adverbial suffixes” *-nott*, *-nól*, *-ni*. These suffixes are usual for the expression of outside local relations that is the adessive, ablative and allative. They are connected typically to personal names and common nouns meaning occupations, relating to groups and individuals as well. They do not appear in the standard language, meaning today other relations too, by today already occur and occur in more kinds of meanings, living together in some dialects with standard suffixes or rather other variants, or and they are often substituted for their standard equivalents. Some of our linguistic atlases contain the form of the outside local suffixes of *Erzsi*, *Józsi*, *Sándor*, *Ferenc* and *Mihály* to demonstrate and study this phenomenon. It turns out among others from the united maps that both *-nott* and *-nól* and *-ni* is used by and large in the same area meaning ‘-éknÁl’, ‘-éktÓl’ and ‘-ékhOz’, but they are related to both groups and people in some eastern “Palóc” settlements and a few in Szilágyság, Mezőség and Székelyföld. Otherwise their isogloss is not the same. The suffix *-ni* (which has only one form) from the three morphemes is used in the most extended circle. In some regions of Székelyföld it is the only form. The standard equivalents have already reduced the occurrence of the three suffixes, and in a few cases the regional and the standard variants live parallel

to each other. I touch on the many meanings of the suffixes and on other individual suffix formations as well.

The last subchapter deals with the spread of the non-assimilated and non-harmonized occurrence of the *-val* and *-vel* adverbial suffixes. Our language atlases contain many headwords with the suffix *-val*, *-vel* to illustrate phenomenon: stem ending in a vowel, stem ending in a consonant, and examples with front and back vowel quality. The united maps show the following picture: In general, the standard form was spread. The different archaic and new forms can be found in the edge of the language area. The non-assimilated form is typical in the “Palóc” and certain eastern and north-eastern dialects. Also the non-assimilated form is the most frequent in the so-called “Csíki Székely” region and also in Moldva, and it appears in the so-called “Hétfalusi Csángó” region. The non-harmonized forms are found exclusively in the western language area, in the region of Őrség, Hetés and Göcsej, the *-jel* variant of the suffix *-val*, *-vel* also appears in Austria at three settlement from four. There appeared almost exclusively the standard forms in Szilágyság. The bilabial voiced spirant which is the archaic variant of the *v* phoneme in the suffix appears on the eastern part of the language area, but it is a sporadic phenomenon. The labialized variant of the *v* phoneme occurs rather in Mezőség, and sporadically in a few settlements. The occurrence of the *-val*, *-vel* without *v* is frequent enough in the “Csángó” area in Moldva, that is a hiatus is formed between the stem and the suffix. Otherwise in Moldva the standards forms and the forms with labial *v* are typical.

I present three researches on the lexicology in my thesis. All the three verbs are so-called “animal onomatopoeic” showing the problem which is typical of the onomatopoeic verbs. They come up in extremely various forms, their segmentation is complicated, so the connection between the stems can not be easily demonstrated. In each map 20-30 lexemes can be separated. I tried to group the data on the basis of the beginning sounds of the word forms. The attempt proved successful: the various word beginnings separate well enough in the maps, the isoglosses can be designated relatively well, the historical connections of the settlements become visible, too. The theme deserves a further analysis.

IV. Look to the future

My thesis is not a finished unit but a part of a research of a longer distance, of a bigger work. At the end of the thesis I set some tasks and offers in connection with uniting the atlases for the future: e.g. some work phases can be made by university students as well and would be accomplished with little money but the result would mean a big help for the later analyses. I plan myself the completion of further tasks, first of all in connection with the compilation of the total settlement list and further geolinguistic surveys with the use of the united maps.