Acoustic and dialectometric study of Hungarian dialects

Fruzsina Sára VARGHA

NKFI FK 138396 (2021)

1. Background, problems to be solved

The present project envisages the investigation of variation and change in Hungarian dialect pronunciation with acoustic analysis and dialectometry. Both methods are current and innovative in the field of dialectology, however, poorly integrated into Hungarian dialect research (cf. Boberg et al. 2018., with special reference to the 2nd section entitled "Methods"). The first steps for the adaptation of current research methods and theoretical approaches in Hungary are highly related to the previous computational dialectology projects and studies of the applicant.

Dialect atlases in general are based on impressionistic phonetic transcriptions of dialect speech. Although these transcriptions contain fine phonetic details (with the use of diacritics), they necessarily reflect the researchers' preconceptions and the already well-known patterns of dialect variation. The Atlas of North American English (Labov–Ash–Boberg 2006, ANAE) puts dialect geography on a new footing by the acoustic analysis of dialect recordings that eliminates the subjectivity typical of traditional transcriptions (for a critical review of the Atlas see Thomas 2016). A key concept of ANAE is to view vocalic changes, as reflected in the acoustic characteristics of vowels, interrelatedly, contrary to dialect atlases that represent variants of vowels separately on single maps. With the revelation of the nature of chain shifts – as Labov states it in the Preface – "A major aim of the Atlas is the reestablishment of the links between dialect geography and general linguistics" (2006: v). As ANAE focuses on cities, rather than rural areas, the overall picture it draws of American dialects can be refined and challenged with the more detailed investigation of specific dialect regions (see for example Fridland et al. 2016, 2017). Furthermore, the Atlas can be regarded as a baseline not only for the investigation of American dialects but, from a methodological and theoretical perspective, for any further research on dialect geography, including Hungarian dialect studies.

Until recently, little has been done to describe the acoustic characteristics of non-standard varieties of Hungarian and to use acoustic analysis as a tool to investigate dialect change. Nevertheless, the few examples are showing promising results: in the confirmation of the overall reliability of phonetic transcriptions (Presinszky 2016); in the analysis of vowel systems and ongoing change (Kocsis 2013, Mády 2015); in the maintenance of a dialect feature in the speech of in-migrants in Budapest (Kontra–Vargha 2014); in the comparison of the vowel systems of informants speaking different Hungarian dialects (Vargha 2013).

More than one million data instances from several traditional Hungarian dialect atlases have been already computerised (Vékás 2007) and a time-aligned corpus of dialect speech has been built up from the field recordings of the Atlas of Hungarian Dialects in a series of interrelated projects (e.g. NKFP 5/056/2004, OTKA K-73024, OTKA PD-108442, Bolyai János Research Grant), mostly in the last fifteen years with the key participation or leadership of the present applicant. Computational dialectology in Hungary is highly related to the development of a linguistic software called Bihalbocs, suitable for the time-aligned transcription, analysis, and automatic mapping of Hungarian dialect data. This tool, thanks to the development efforts of the present applicant, is unique in fully supporting the phonetic transcription standard of Hungarian dialectology, in addition, it has been proven to produce, in a highly efficient way, data sets suitable for further processing and analysis in a number of other research tools too. (This is a continuous, non-taxpayer-funded development made and supported by the interested researchers themselves, mainly the present applicant, in response to the evolving research needs.) The tool is used to build databases (oral corpora if the sound is available) and also to hold together the connected pieces of information. This way, the computerised, phonetically coded piece of data keeps its connections to the locality where it was collected, to the sociolinguistic metadata (such as the informant's age and gender), and also to the source recording it is extracted from. Thanks to the development efforts and the new methodology they generated (with the determinant role played in recent years by the present applicant), computational dialectology has become the main approach in Hungarian dialect research. The technology mentioned above is currently used at several universities in three different countries (HU, RO, SK) for the digitisation of already published dialect atlases and also for data processing in recent projects. A major advantage of this technology is the increased compatibility and convertibility (portability) of the constructed data structures. Time-aligned transcriptions are also convertible to ELAN .eaf files or Praat TextGrid files for further analysis. A simplified version of transcriptions can also be generated and published on the Internet (HTML files with embedded javascript, see for example at the host institution's website: https://nlp.nytud.hu/csango/). Moreover, it is possible to use Praat from within Bihalbocs: to measure formant values, the dedicated software generates Praat scripts on the fly and calls Praat through inline UNIX commands to execute the scripts in the background. This way, dialect research benefits from the flexibility and efficiency of Bihalbocs in dealing with data transcribed according to the Hungarian transcription standard and the transcribed sound files, as well as from the reliability of Praat, considered a well known linguistic tool, in acoustic measurements. To sum up, the dedicated software tool's main advantages are: support for standards and compatibility, integration of Praat functionality, and the possibility to incorporate – through additions to the source code – new methods and algorithms for data analyses of virtually unlimited complexity, in response to research needs. (The user is required to modify the source code for non-trivial tasks, which is not a problem in this project since the applicant is the developer.) For a more detailed overview of computational dialectology in Hungary see Vargha (2009, 2017) and Vékás (2007).

The most important aim of the present project is the acoustic analysis of the vowels of Hungarian dialects in a dialectometric framework, with the comparison of acoustic spatial structures to similarity matrices computed from computerised atlas data. The field recordings of the Atlas of Hungarian Dialects (made by the present project's host institution between 1960 and 1964 at 352 locations) constitute the main corpus of speech material for acoustic measurements. The recordings cover at least half of the Hungarian language area (as in the middle of the 20th century it was almost impossible to make recordings outside Hungary, we have sound material only from Hungary and former Czechoslovakia), focusing on the documentation of rural speech of non (or less) mobile, local respondents, which is ideal for a linguistic geography project that aims at comparing the vowel qualities of different dialects. As the informants were of different age groups, this corpus is suitable, to some extent, even for apparent time studies (as shown in Vargha 2007 for morphologic variables). The results of the measurements will be further compared to formant data from a limited number of present-day recordings (a follow-up project, conducted between 2007 and 2012, is one of the possible sources, more recent recordings will also be considered), thus providing real-time comparisons.

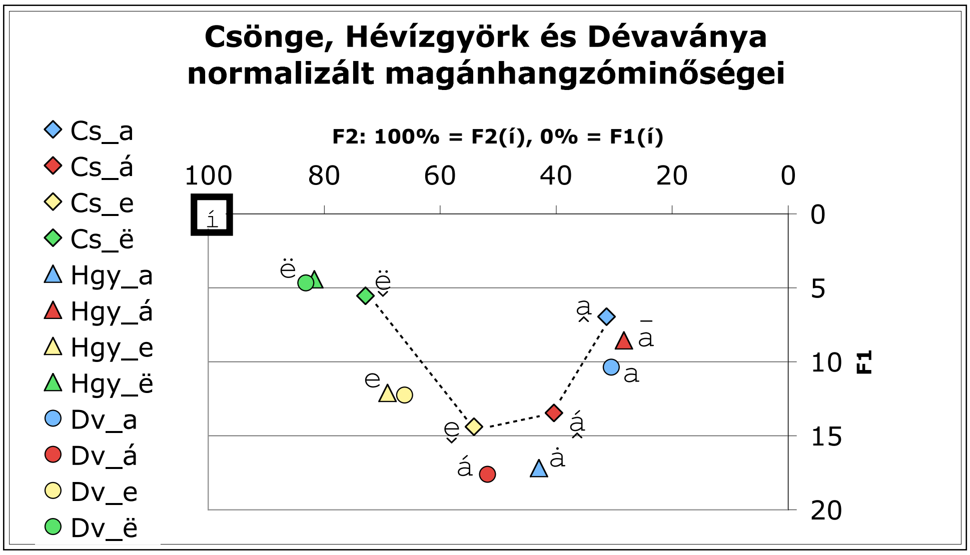


Figure 1: The arrangement of four vowels in the acoustic space in three different Hungarian dialects (Vargha 2013)

In a previous project (OTKA PD-108442, led by the present applicant, terminated with an 'Excellent' grade in 2016 and involving a number of collaborators) several computerised Hungarian dialect atlases have been submitted to dialectometric analysis (Vargha 2017). Comparing different linguistic similarity matrices, the effect of the narrowness of the phonetic transcription (i.e. the quantity of phonetic information contained in it) on the dialectometric analysis (Vargha 2018) and the relevance of linguistic levels (phonetic, phonological and lexical) in dialect comparison (Vargha 2015) have been tested. The effect of loanwords on dialectometry has also been investigated on the corpus of the Moldavian Csángó Dialect Atlas (Bodó–Vargha 2016). Different dialectometric techniques (the taxation method and the automatic string-edit-distance method based on Levenshtein's algorithm) have been compared on the integrated corpus of the Atlas of Hungarian Dialects and the Atlas of Hungarian Dialects in Romania (Kocsis–Vargha 2016). In the case of western dialects, the similarities between the geographical patterns of linguistic and onomasiological data have been investigated (Ditrói–Vargha 2016).

Besides the acoustic analysis of vowel qualities, the project will accomplish the diachronic dialectometric comparison of Hungarian dialects of Moldavia (a historical region in Romania). Dialectometry has already been used with success for the investigation of dialect change in apparent time studies (for an overview see Wieling–Nerbonne 2015: 251–253), and in the case of Dutch dialects the influence of standard Dutch on dialect change has also been investigated (Heeringa–Hinskens 2015). The present project plans a real-time dialectometric study on dialect change. The dialectometric analysis of the computerised version of the Moldavian Csángó Dialect Atlas will be compared to the analogous analysis of its follow-up project. In the former corpus data instances are linked only to the location where they were collected, while in the latter every data instance keeps its connection to the informant (Bodó et al. 2012.). As informants from different age groups have been interviewed, an apparent-time comparison will also be possible.

The dialectometric analysis of Moldavian Hungarian dialects will be completed with the study of a corpus of transcribed ethnographic interviews. The construction of the corpus started in 2020 with the participation of the present applicant at the Research Institute for Linguistics in an ongoing NKFIH-project (KKP-129921, research leader: Katalin É. Kiss). Linguistic variables will be selected, the usage of different variants will be mapped and compared to other dialect maps made from linguistic atlas data.

2. Hypothesis, key questions, goals of the project

The general goal of the project is the objective, comparative description of Hungarian vowel systems based on F1 and F2 measurements, the investigation of dialect change (in apparent-time and real-time), and the comparison of acoustic measurements and dialectometric analyses of phonetically transcribed atlas data. The main theoretical and methodological challenges include the research of the interrelatedness of vowel qualities, new in Hungarian dialectology, and the application of dialectometric methods to acoustic data.

1. **How can the results of acoustic measurements be presented and compared on dialect maps?** F1 and F2 values will be normalised, and new mapping techniques will be applied or developed to illustrate dialectal differences in vowel articulation. Categorical (Labov–Ash–Boberg 2006) and progressive (Grieve et al. 2013, see also Figure 2) colouring techniques will be tested. New technics will also be developed, inspired by those applied in dialectometry (Vargha 2017, Figure 3).

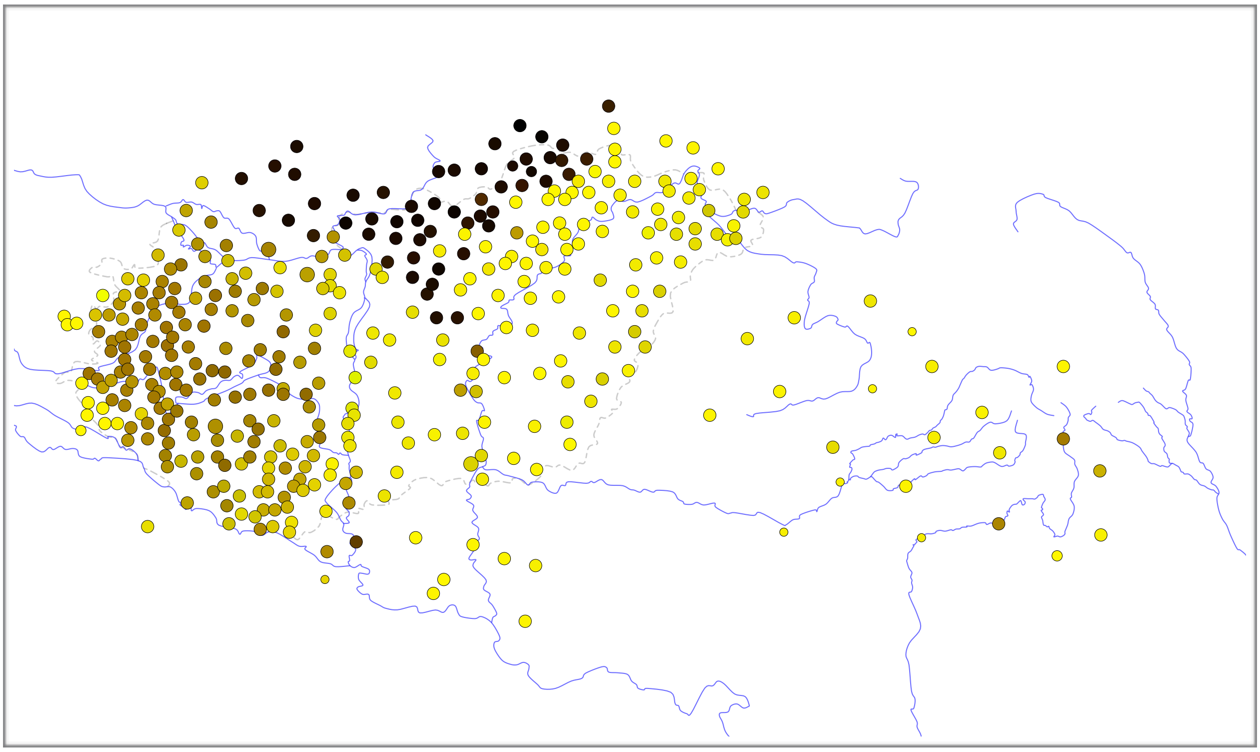


Figure 2: Estimated F1 (RGB red component) and F2 (RGB green component) values of the long vowel /a:/ at the investigation points of the Atlas of Hungarian Dialects. Values are calculated from the traditional phonetic transcription of atlas data, based on the correspondence of phonetic symbols and formant data already measured at a few locations (cf. Vargha 2020)

2. **Because of its objectivity, acoustic analysis is expected to reveal variability in a number of cases where traditional Hungarian dialectology could not.** One example is the case of the [e]-like vowels. In his quantitative study based on the analysis of the Atlas of Hungarian Dialects, Imre states that the dialect differences in the pronunciation of [e] (transcribed as *ë* in Hungarian dialect transcriptions) could not be investigated (1971: 272). In the atlas there is no considerable variation in the transcription of [e]-like sounds; in contrast to other vowels, they are hardly ever indicated to be more open or more closed with diacritics. However, acoustic comparison of three dialect speakers from different regions is showing

considerable differences in the realisations of [e]-like sounds (Figure 1). Variability and the change herein are also crucial to the understanding of the ongoing phonemic merger of /e/ and /ɛ/.

3. **What are the main characteristics of Hungarian vowel systems?** In Hungarian dialect studies, dialect vowels are almost never investigated interrelatedly. The project aims at revealing the interrelatedness of vowel qualities and determine typical patterns for the disposition of vowels in the acoustic space.
4. **How can the results of acoustic measurements be compared to the outcome of dialectometric analyses?** To what extent do the geographical patterns of acoustic analysis correlate with linguistic similarity computed from traditional atlas data? Acoustic similarity matrices will be calculated (cf. Leinonen 2010) and statistically compared to similarity matrices computed from computerised atlas data (Vargha 2017). New mapping techniques will be applied to present acoustic data in a visually comparable way to dialectometric maps (see Figures 3 and 4 as examples of MDS maps).
5. **How and to what extent have the acoustic characteristics of vowels changed in the last 60 years?** Does dialect pronunciation converge to standard Hungarian, if it does, does standard pronunciation have the same effect on every dialect? Which are the most long-standing characteristics of dialect vowel systems? The main theoretical question is the following: Is sound change systematic, affecting several vowels interrelatedly, or do vowels change individually due to extralinguistic factors, such as stigmatisation of some vowel qualities seen as typically dialectal?

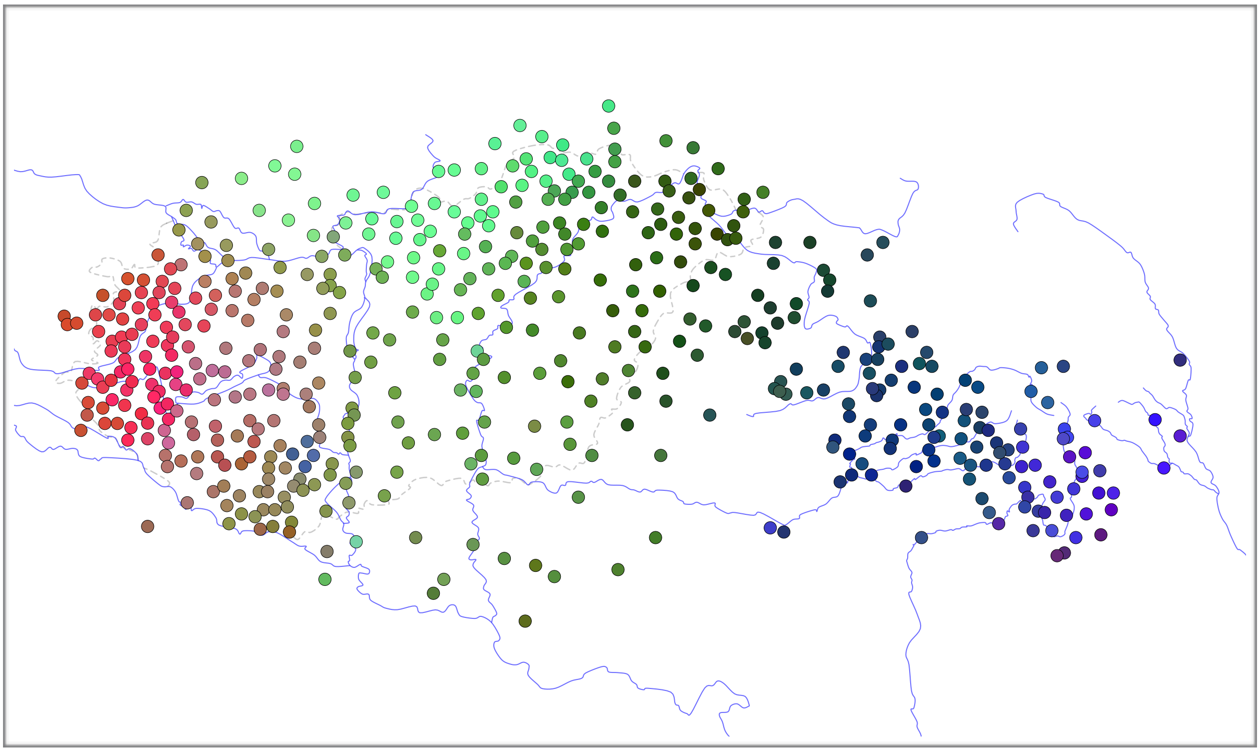
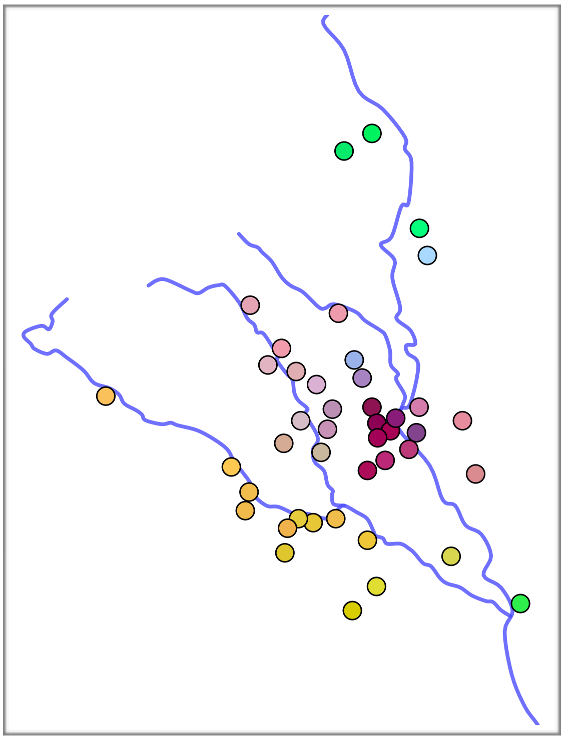


Figure 3: Multidimensional scaling map of the integrated dialectometric analysis of two atlases: the Atlas of Hungarian Dialects and the Atlas of Hungarian Dialects in Romania (Vargha 2017)

6. **What can be predicted for the future of Hungarian dialects?** The idea of "dialect death" is often envisioned in Hungarian dialect studies and rooted in a linguistic ideology that explains dialect change largely as a result of standardisation (cf. Kiss 2006: 132–134). Dialect research of other languages and some recent studies about linguistic variation in Hungarian point out that change cannot be treated simply as a movement towards standard forms (Britain 2009, Vargha 2013, Bodó 2012). The apparent-time study of older and younger speakers might reveal the nature and direction of ongoing changes and might as well shed light on the most characteristic phonetic features (that are less likely to change over time) of the different dialects. F1 and F2 values of older and younger speakers will be investigated, acoustic data will be compared statistically and the results will be presented on correlation maps.

Figure 4: Multidimensional scaling of the linguistic distance matrix of Moldavian Hungarian dialects (Vargha 2017)



7. **Do the increasing mobility and the acceleration of language change lead to the alternation of linguistic similarity patterns in a small, but heterogenous dialect region, in Moldavia?** As loan words had already a significant effect on the dialectometric analysis in the older corpus (Bodó–Vargha 2016), we expect some increase in similarities especially in the case of neighbouring dialects with historically different dialect backgrounds (different home dialects and typically different vowel systems). The linguistic similarities based on the Moldavian Csángó Dialect Atlas are shown on Figure 4, with the mapping of the multidimensional scaling of the linguistic distance matrix.
8. **Do the attributes of informants have an effect on linguistic similarity relations in Moldavia?** In the follow-up study of the Moldavian Csángó Dialect Atlas data instances are linked to informants, contrary to the practice of traditional atlases where data instances keep their connection only to the investigation point. Several linguistic similarity matrices will be computed for different groups of informants to test the impact of mobility, education, age and gender on dialectometric relations.

3. Method

The vowel qualities of Hungarian dialects will be determined by the acoustic analysis of dialect speech, spontaneous speech data coming from field recordings. Field recordings (especially older ones) are not as ideal for acoustic analyses as laboratory interviews, but still usable if we pay more attention to the accuracy of measurements (cf. e.g. Brissos 2014, Brissos–Saramago 2014). F1 and F2 of stressed vowels will be measured with Praat, a well-known and widely used tool for acoustic measurement, and verified manually due to the mediocre quality of the recordings from the ’60s. The measurement of the first and the second formant is sufficient for the characterisation of Hungarian vowels (on the formant characteristics of standard Hungarian see Gósy 2004). In the case of monophthongs, ANAE standards (Labov et al. 2006) are to be followed, taking into consideration other studies on the acoustic features of Hungarian vowels (such as Gósy 2012, Gósy–Bóna 2014). Long vowels are often realised as diphthongs in Hungarian dialect speech, thus diphthongisation is a key question in the comparison and classification of Hungarian dialects (cf. Juhász 2011). At least three measurements will be taken for each token to describe upgliding and down-gliding diphthongs.

The measured values, including the length of the vowel, will be stored within time-aligned transcriptions of selected interview segments and will be ready to be searched for according to consonantal context, syllable type and various other criteria, and remain available (and convertible) for further analysis, normalisation or plotting. This way the acoustic data will be incorporated within the digital framework of previous Hungarian computational dialectology projects, thus the acoustic data will remain connected to every other piece of information already gathered. This level of integration favours efficiency. Results will also be presented on dialect maps. The use of Praat will facilitate comparisons with formant data of other projects that also rely on Praat. It will also be possible to automatically remake all formant measurements in all sound files with different Praat settings, as the temporal values of vowels (beginning, length) are to be recorded within the time-aligned transcriptions.

Men and women were interviewed in roughly equal numbers for the Atlas of Hungarian Dialects. To compare formant values of different speakers they need to be normalised. A formula inspired by Gerstman's normalisation method (1968) and explained in Vargha (2013) will be used. Average F1 and F2 values of long /i:/ sounds will be used as reference points for the highest possible F2 and the smallest possible F1 of a given speaker:

FxNorm = ((Fx-F1min) * 100) / (F2max - F1min)

As this formula (under the name Vargha 2013) is implemented in Visible Vowels (Heeringa–Van de Velde 2018), an online tool for formant normalisation developed at Fryske Akademy (https://www.visiblevowels.org), when enough data is available, several normalisation methods will be compared to find out the possible advantages of different normalisation technics.

The quality of the recordings is critical to make reliable measurements. Unfortunately, not all the tapes could be preserved in equally good quality until digitisation, thus not all the 352 locations can be investigated. 160 locations will be selected, taking into consideration geographical balance and the quality of the recordings. At least one speaker per site, producing enough data for analysis, will be chosen.

In the third year of the project at least ten locations will be carefully selected for the real-time investigation of changes in vowel articulation. All the dialect areas encompassed by the main corpus (the field recordings of the national atlas) will be represented. We will make apparent-time studies as well, based on either the old recordings or the recent ones. In the real-time comparisons we plan to rely mainly on sound materials collected by Hungarian dialectologists in the past years (a follow-up project, conducted between 2007 and 2012, is one of the possible sources, more recent recordings will also be considered). According to the research needs, fieldwork could also be conducted exceptionally at a maximum of three locations (in which case informed consent will be obtained prior to any involvement of the participants in the research).

Acoustic similarity matrices will be calculated (cf. Leinonen 2010) based on the normalised F1 and F2 data. The linguistic geographic patterns of the acoustic data will be compared to the outcome of dialectometric studies of the Atlas of Hungarian Dialects (Vargha 2017). The linguistic similarity patterns of atlas data will be calculated with the Levenshtein algorithm on at least 125 selected maps showing variation in vowel qualities, then compared to other dialectometric matrices and the linguistic similarity patterns of the acoustic data.

For the diachronic dialectometric analysis of Moldavian Hungarian dialects, 150 corresponding (thus comparable) questions will be selected from the old and the new corpus. Data instances will be compared automatically with the Levenshtein algorithm (Heeringa 2004, Vargha 2017). To analyse the data of the follow-up project new dialectometric techniques are to be elaborated because the data structure (data instances are linked to the informants) is different from the already analysed datasets (data instances are linked to the locations only). Similarity matrices will be compared statistically and on correlation maps.

To sum up, emphasis will be given to the use of well-established technologies. Praat will be used to make acoustic measurements, formant values will be normalised with Visible Vowels and R software will be used for statistical analysis. Bihalbocs, the dedicated software for doing Hungarian dialect research, will hold together transcriptions (supporting the phonetic transcription standard of Hungarian dialectology), formant values and statistical data, and will also be used for the cartographic and online display of the results. Other tools, such as Gabmap and Diatech will also be considered for complementary dialectometric investigations.

4. Expected results

We will launch a website for the project already in the first year and keep publishing online maps, plots, datasets, and the most important findings of the research. The results will be also presented at scholarly meetings, international conferences. Scholarly articles will be published each year along the research questions. Our findings will serve as a baseline for subsequent studies on Hungarian dialect variation. The methods applied will not only clarify and, to some extent, change what we know and think about Hungarian vowel qualities and their evolution, but will also take Hungarian dialectology closer to international standards.

The diachronic dialectometric analysis of the two Moldavian corpora will shed light on the dialect change in a bilingual community in the last phase of language shift. It may also constitute an example to follow methodologically for the diachronic comparison of atlas data for any language.

5. Research infrastructure (equipments, personnel, etc.)

The method for data processing and analysis to be used in the research has already been developed or under development in previous or ongoing projects. The recordings to be analysed from 352 locations are already digitised and available for the project. Computerised atlas data and the database of the Moldavian follow-up study are also available for the research. No special research tools are needed except for two computers running Mac OS X (for compatibility reasons) and data storage tools.

The present applicant is currently working as part-time research fellow in another NKFIH project (KKP-129921) at the host institution. If the present FK proposal is funded, she will work (full time) exclusively on the FK project. The present applicant has introduced dialectometry in Hungarian dialectology thanks also to a previous NKFIH research project she led (PD-108442, terminated with an 'Excellent' grade) and appropriate research tools she developed, porting quantitative methods to Hungarian computational dialectology while enhancing existing analysis and mapping techniques, thus reinforcing the recent prominence of this research area. She successfully involved a number of colleagues from different faculties in dialectometric or acoustic research. Linguistic technologies she developed are widely used and cited in Hungarian dialect research and also mentioned in Leinonen et al. 2016. Her formant normalisation method (Vargha 2013) was implemented in Visible Vowels in 2020. She has recently been involved in an ongoing initiative on dialect classification of languages in Europe (https://www.diacl.eu) as the author of the chapter about Hungarian dialect classification.

The other named participant, Zsuzsanna Kocsis, who recently obtained her PhD in historical linguistics, is a full-time researcher at the host institution, she will only receive an extra salary from the budget. She took part in a number of linguistic projects, including the above-mentioned NKFIH project (PD-108442).

The project will be funded exclusively by the budget provided by NKFIH. Based on experience from previous computational dialectology projects, the planned personnel will be sufficient for the accomplishment of the project's tasks, thanks also to the efficiency of the methods.

6. Changes compared to previously submitted versions of this proposal

The host institution now is the Research Institute for Linguistics, the traditional centre of Hungarian dialectology (about half a century ago the Institute's research team produced the Atlas of Hungarian Dialects and also made field recordings; the transcribed and computerised data of this national atlas as well as the digitised sound material from 352 of its investigation points constitute the main corpus for the present research). It is also worth noting that this is today the only Budapest-based institution to maintain computational dialectology as the main approach in Hungarian dialect research, as reflected in a recent paper, authored by the researchers of the host institute, including the present applicant (Oszkó et al. 2020).

Based on the useful advice of previous reviewers, an even greater emphasis will be given to the use of well-established technologies. Praat software will be used for the measurement of formant values, consequently, measurements made at some locations in a pilot study will be redone to keep the project coherent.

The project has been slightly restructured to better focus on each of the main tasks: the analysis of the field recordings from the 60s will be completed no later than the second year of the project, the apparent-time and real-time research will follow in the third year. The project was deemed undermanned in previous reviews, to address this issue and improve the feasibility of the project, we will investigate real-time changes in vowel articulation at fewer but carefully selected locations, still representing all the dialect areas encompassed by the main corpus (the field recordings of the national atlas).

As the budget planning rules changed in 2021 (prescribing a compulsory 10 percent for open access costs), we would not have enough funding for the previously planned publication of an acoustic atlas (printed version). We will rather start a homepage for the research in the first year and keep publishing online maps, plots, datasets, and the most important findings of the research. The publication of the results in a (diachronic acoustic) printed atlas might be postponed for a subsequent project.

Budgetary considerations should not affect negatively our previous commitments about the diachronic dialectometric analysis of Hungarian dialects of Moldavia, as these are considered endangered varieties in the last phase of language shift. This task will even be slightly expanded to include the study of a corpus of ethnographic interviews from the area, phonetically transcribed and published on the web in an ongoing NKFIH-project at the host institution (KKP-129921, research leader: Katalin É. Kiss).

References

Boberg, Charles, John Nerbonne, Dominic Watt 2018. *The Handbook of Dialectology.* Wiley Blackwell.

Bodó Csanád 2014. Nyelvi ideológiák a magyar nyelvi változók kutatásában. *Magyar Nyelv* 110. 266–284.

Bodó, Csanád, Fruzsina S. Vargha, Domokos Vékás 2012. Classifications of Hungarian dialects in Moldavia In: Lehel Peti, Vilmos Tánczos eds. *Language Use, Attitudes, Strategies: linguistic identity and ethnicity in the Moldavian Csángó villages*. Romanian Institute for Research on National Minorities, Cluj-Napoca. 51-69.

Bodó Csanád – Vargha Fruzsina Sára 2016. Román kölcsönszói hatások A moldvai csángó nyelvjárás atlasza nyelvi hasonlósági viszonyaiban. In: Czetter Ibolya, Hajba Renáta, Tóth Péter eds. *VI. Dialektológiai Szimpozion: Szombathely, 2015. szeptember 2-4.* Nyugat-magyarországi Egyetem Savaria Egyetemi Központját; Nyitrai Konstantin Filozófus Egyetem Közép-európai Tanulmányok Kar, Szombathely/Nyitra. 167-178.

Brissos, Fernando 2014. New insights into Portuguese central-southern dialects: understanding their present and past forms through acoustic data from stressed vowels. *Journal of Portuguese Linguistics* 13: 63-115.

Brissos, Fernando – João Saramago 2014. O problema da diversidade dialectal do Centro-Sul português: informação perceptiva versus informação acústica. *Estudos de Lingüística Galega* 6: 53-80.

Britain, David 2009. One foot in the grave? Dialect death, dialect contact, and dialect birth in England. *International Journal of the Sociology of Language* 196/197: 121–155.

Ditrói Eszter – Vargha Fruzsina Sára 2016. Települések közti nyelvjárási és névföldrajzi hasonlósági viszonyok összevető elemzése Vas és Zala megyében. *Névtani Értesítő* 38: 113-130.

Fridland, Valerie, Tyler Kendall, Betsy E. Evans, Alicia Beckford Wassink eds. 2016. Speech in the Western States, Volume 1: The Coastal States. Publication of the American Dialect Society 101, Supplement to *American Speech* 91.

Fridland, Valerie, Alicia Beckford Wassink, Tyler Kendall, Betsy E. Evans eds. 2017. Speech in the Western States, Volume 2: The Mountain West. Publication of the American Dialect Society 102, Supplement to *American Speech* 92.

Gerstman, Louis J. 1968. Classification of self-normalized vowels. *IEEE Transactions of Audio Electroacoustics AU-16*: 78–80.

Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.

Gósy Mária 2012. Az alsóbb nyelvállású magyar magánhangzók formánsszerkezete. In: Gósy Mária ed. *Beszéd, adatbázis, kutatások.* Akadémiai Kiadó, Budapest. 43–66.

Gósy Mária – Bóna Judit 2014. Magánhangzók ejtése fiatalok és idősek spontán beszédében. *Magyar Nyelv* 110: 129–143.

Grieve, Jack, Dirk Speelman, Dirk Geeraerts 2013. A multivariate spatial analysis of vowel formants in American English. *Journal of Linguistic Geography* 1: 31–51.

Heeringa, Wilbert 2004. *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Groningen Dissertations in Linguistics 46.

Heeringa, Wilbert – Frans Hinskens 2015. Dialect change and its consequences for the Dutch dialect landscape. How much is due to the standard variety and how much is not? Journal of Linguistic Geography 3: 20–33.

Heeringa, Wilbert – Hans Van de Velde 2018. Visible Vowels: a Tool for the Visualization of Vowel Variation. In: *Proceedings CLARIN Annual Conference 2018, 8 - 10 October*, Pisa, Italy. CLARIN ERIC.

Juhász Dezső 2011. A magyar nyitódó kettőshangzók történetéről a tér és idő dimenziójában. In: Bakró-Nagy Marianne – Forgács Tamás (szerk.). A nyelvtörténeti kutatások újabb eredményei VI. Szeged: Szegedi Tudományegyetem, Magyar Nyelvészeti Tanszék. 123–128.

Kocsis Zsuzsanna 2013. Nyelvjárási ë hangok vizsgálata – különös tekintettel a labializáció mértékére. In: Kontra Miklós, Németh Miklós, Sinkovics Balázs eds. *Elmélet és empíria a szociolingvisztikában.* Gondolat, Budapest. 432–443.

Kocsis Zsuzsanna – Vargha Fruzsina Sára 2016. A magyar nyelvjárások atlasza és A romániai magyar nyelvjárások atlasza integrált dialektometriai elemzése. (= The integrated dialectometric analysis of the Atlas of Hungarian Dialects and the Atlas of Hungaraian Dialects in Romania) In: Czetter Ibolya, Hajba Renáta, Tóth Péter eds. *VI. Dialektológiai Szimpozion: Szombathely, 2015. szeptember 2-4.* Nyugat-magyarországi Egyetem Savaria Egyetemi Központját; Nyitrai Konstantin Filozófus Egyetem Közép-európai Tanulmányok Kar, Szombathely/Nyitra. 193–207.

Kontra, Miklós – Fruzsina S. Vargha 2014. Are there speakers of the /ɛ/ vs. /e/ dialect in Budapest? In: Barysevich, Alena – D'Arcy, Alexandra – Heap, David eds. *Proceedings of Methods XIV. Papers from the*

Fourteenth International Conference on Methods in Dialectology, 2011. Peter Lang Edition, Frankfurt am Main. 14–24.

Labov, William, Sharon Ash, Charles Boberg 2006. The Atlas of North American English: Phonetics, Phonology and Sound Change. Mouton de Gruyter, Berlin/New York.

Leinonen, Therese 2010. *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.* Groningen Dissertations in Linguistics 83.

Leinonen, Therese, Çağrı Çöltekin, John Nerbonne 2016. Using Gabmap. *Lingua* 178. (Spec. issue on linguistic infrastructure edited by Jan Odijk) 71–83.

Mády Katalain 2015. *A fogadodban sok a vendég*: hosszú magánhangzók egy mezőségi nyelvjárásban. In: É. Kiss Katalin – Hegedűs Attila – Pintér Lilla eds. *Nyelvelmélet és dialektológia 3.* PPKE BTK Elméleti Nyelvészeti Tanszék – Magyar Nyelvészeti Tanszék, Piliscsaba. 196–210.

Presinszky Károly 2016. Az á előtti illabiális à-zás a Csallóközben. *Magyar Nyelv* 112: 218–227.

Thomas, Erik R. 2016. The Atlas of North American English and its impacts on approaches to dialect geography. *Journal of Sociolinguistics* 20: 489–497.

Oszkó Beatrix, Prószéky Gábor, Vargha Fruzsina Sára 2020. Nyelvjáráskutatás, számítógépes módszerek, mesterséges intelligencia. (= Dilaect research, computational methods, artificial intelligence) In: Péntek János – Visy Zsolt szerk. *Hagyomány, változatosság és változás a székely nyelvjárásban.* Pro Énlaka Alapítvány – Nemzetstratégiai Kutatóintézet – Pécsi Tudományegyetem, Énlaka. 87–106.

Vargha Fruzsina Sára 2007. Nyelvi változók A magyar nyelvjárások atlasza hangfelvételeiben. (= Linguistic variables in the recordings of the Atlas of Hungarian Dialects) In: Guttmann Miklós – Molnár Zoltán szerk. *V. Dialektológiai Szimpozion.* Szombathely, 279–289.

Vargha, Fruzsina S. 2009. The New Oral Corpus and Related Talking Maps of Hungarian Dialects from the 1960s. SIDG Congress, Maribor, September 14. 2009. http://frufru.web.elte.hu/eloadasok/Vargha_Maribor2009.pdf

Vargha Fruzsina Sára 2013. Az adatok kor és nem szerinti megoszlása A magyar nyelvjárások atlasza ellenőrző gyűjtésében. (= Age and gender effects on the answers to the questionnaires recorded for the Atlas of Hungarian Dialects) In: Kontra Miklós – Németh Miklós – Sinkovics Balázs eds. *Elmélet és empíria a szociolingvisztikában: Válogatás a 17. Élőnyelvi Konferencia (Szeged, 2012. augusztus 30. – szeptember 1.) előadásaiból.* Budapest, 2013. 459–474

Vargha Fruzsina Sára 2013. A hangzó adat szerepe a magyar dialektológiában (= The role of sound data in Hungarian dialectology). In: Szoták Szilvia, Vargha Fruzsina Sára szerk. Változó nyelv, nyelvváltozatok, területiség: A VII. Hungarológiai Kongresszus nyelvészeti tanulmányai. Kolozsvár. 194–204.

Vargha Fruzsina Sára 2015. Lexikai, fonológiai, fonetikai stabilitás (és relevancia) a magyar nyelvjárásokban. (= Lexical, phololological, phonetic stability (and relevance) in Hungarian dialects) In: É. Kiss Katalin – Hegedűs Attila – Pintér Lilla eds. *Nyelvelmélet és dialektológia 3.* PPKE BTK Elméleti Nyelvészeti Tanszék – Magyar Nyelvészeti Tanszék, Piliscsaba. 246–261.

Vargha Fruzsina Sára 2017. *A nyelvi hasonlóság földrajzi mintázatai. Magyar nyelvjárások dialektometriai elemzése* (= Geographical patterns of linguistic similarity. Dialectometric analysis of Hungarian dialects). Magyar Nyelvtudományi Társaság, Budapest. 2017.

Vargha, Fruzsina S. 2018. The impact of phonetic information in dialectometry – a case study of Hungarian dialect atlases. *Dialectologia* 21: 185–208.

Vargha Fruzsina Sára 2020. Magánhangzó-minőségek meghatározása nyelvatlaszadatok alapján (= Determining vowel qualities from transcribed atlas data) *Magyar Nyelv* 116: 442–461. https://www.c3.hu/~magyarnyelv/20-4/VarghaFS_MNy_20-4.pdf

Vékás Domokos 2007. Számítógépes dialektológia (= Computational dialectology). In: Guttmann Miklós – Molnár Zoltán eds. V. Dialektológiai Szimpozion. Berzsenyi Dániel Főiskola, Szombathely. 289–293.

Wieling, Martijn – John Nerbonne 2015. Advances in Dialectometry. *Annual Review of Linguistics* 1: 243–264.