PREFACE

I first reported on the plans of the *Dictionary of the Hungarian Dialect from Moldova*[3] (MMTnySz.), on the beginning of the editorial process and the related practical and theoretical issues in 2004. At that time I could rely on the contributions of a large team of colleagues and PhD-students. After the first few years, however, the number of contributors significally decreased. The reason of this decrease later became more apparent: the potential contributors could not work on the project as a full-time job, since they each had their own research projects as well. Furthermore, the financial support obtained through grant proposals did not allow for creating full-time jobs within the constraints of the project. Even though I was not left alone, most of the editorial work had to be done by me, while – examining the linguistic data – I constantly observed the rich, problematic and heterogeneous nature of the vocabulary of this dialect. Consequently, I was constantly forced to make compromises – perhaps even at the expense of the scientific source value of the dictionary – in order to make it usable and useful not only to researchers, but to a larger audience as well. I was also compelled to deviate from my initial goal of giving the dictionary an "encyclopaedical" character. I had to realize: if I do not take account of the real circumstances, – just as the case of previous researchers – the dictionary would remain an ongoing project or would be left to the next generation, like it happened in the case of *Erdélyi magyar szótörténeti tár* (*Historical Dictionary of the Hungarian Language in Transylvania*, with 14 volumes, published between 1976–2014).

The previous objectives and partial results are well-known. Yrjö Wichmann first gathered dictionary material in Szabófalva (Săbăoani), in the winter of 1906–1907. These data were considered to be "northern Csángó" and were later published by Bálint Csűry and Artturi Kannisto in 1936 (CsángSz.), together with the totally different material from Hétfalu (Săcele) which was also considered to be Csángó. Bálint Csűry began to redact the "southern Csángó" dictionary based on the data he had collected in Bogdánfalva (Valea Seacă) between 1928–1931, but the work stopped with his early death in 1941. According to the plans, Attila Szabó T. would undertake the assignment as Csűry's former companion at one of his field trips in Moldova. However, this did not happen and Csűry's library and manuscripts were destroyed in the autumn of 1944 (Szabó T. 1959/1981: 604). In a study published in 1959, Attila Szabó T. restarted the project, and this time he used the data collected during a linguistic geography field trip that had covered the entire Hungarian-inhabited area of Moldova. His goal was to create an extensive dictionary of the Hungarian dialect of Moldova or to separately publish the dictionaries of the different sub-regions (1959/1981: 512). This could not be accomplished either, as he began the redaction of the *Historical Dictionary of the Hungarian Language in Transylvania.* Based on the same material, Gyula Márton redacted and published the dictionary of the Romanian loan-words of the Hungarian dialect from Moldova (Csrk.), and months before his death in 1976 he had started working on the dictionary of the dialect. At the begining of the century, before the extensive linguistic fieldwork covering the whole region, the goal was to publish the dictionaries of the postulated sub-regions (like the CsángSz.), however, later the goal became to create a dictionary based on the whole material.

---

3 Contrary to the local ethnographical translation practice, according to the English word usage, we use the term *Moldova* referring to a particular region of Romania as opposed to the country name *Moldavia*.

The Data

The *Dictionary of the Hungarian Dialect from Moldova* is also largely based on this material: the questionnaire data that can also be seen in the *Moldovan Csángó Dialect Atlas* (CsángA.) and the data related to the illustrative sentences gathered by the contributors of the atlas. The same archive contains the additional data gathered by Gyula Márton which were used as the basis of the works entitled *The Romanian Loan-Words of the Hungarian Dialect from Moldova* (Csrk.) and *Verb Stems and Verbal Inflections in the Csángó Dialect of Moldova* (MGyIge.). Most of these data may be considered authentic as they originate from the ethnographical material gathered at the time of the atlas fieldwork and from other partially published sources.

In his manuscripts, Gyula Márton reports on preparing the fieldwork of the CsángA. as follows: "at the beginning of the summer of 1949 the university [i.e., the Bolyai University; rector: Edgár Balogh, director of studies: Gyula Csehi] was able to finance a field trip during the summer holidays. ASzT. [Attila Szabó T., head of the department then] suggested that the department should start the linguistic geography fieldwork on the Moldovan Csángó dialect. I found out about this from Attila Szabó T., who contacted me to discuss the methods of the inquiry. As our previous fieldtrips with Mózsi [professor Mózes Gállfy] in the regions of Kalotaszeg (Țara Călatei) and Borsavölgy (Valea Borșei) made us realize that it is unnecessary to use the detailed questionnaire in every single village [...], I suggested [...] that, travelling by bicycles, two of us should cover the whole Csángó region using a shorter questionnaire, then based on our observations we should select the most representative spots where the other two researchers would use the detailed questionnaire […]. ASzT. accepted my proposal without reservations and began to create the shorter questionnaires. Regarding the devision of the tasks we agreed that ASzT. and I would use the short questionnaire, while Mózsi and Márta Vámszer […] would use the detailed questionnaire. ASzT. reported in detail on the collecting process in his articles published in the scientific journal MNy. and the fifth volume of MNyj. I would only like to add the following: As ASzT. had to put together the questionnaire in a hurry, it soon turned out that it needed completion. Thus, the questionnaire was revised multiple times, which was not in favour of the inquiry and the quality of the material. [...] The inquiry was not well prepared regarding several aspects (there was not enough time for it). […] Consequently, it was prolonged so much that during the verificatory inquiry made between 1956 and 1962 the interrogation had to be repeated based on the completed questionnaire." The following people took part in the research process of the CsángA. as well: Dezső Balogh, László Bura, Ferenc Gazda, László Murádin, Jenő Nagy and Pál Teiszler.

As it has been mentioned above, the material was gathered in the middle of the previous century, during the fieldwork of the CsángA. Initially I did not plan to include the material of the CsángSz. into this dictionary. However, it practically got included (with proper references), as the two corpora are strongly related and, more importantly, cross-checking the data gathered by the department and the material of the CsángSz. significantly helped the interpretation and verification of our material. The inclusion is also justified by the fact that the peculiar phonetic notation technique used by Wichmann and the resultant alphabetical order seem to cause difficulties to most researchers today. Therefore, Wichmann's data are also presented with the phonetic notation technique used in today's Hungarian dialectology. As far as the *Új magyar tájszótár* (*New Hungarian Dialectological Dictionary*) is concerned, even though it was of great help in verifying our data, a detailed cross-checking was omitted because of the time-consuming nature of the work and the significant overlap between the two corpora. Thus, none of the two dictionaries can be considered redundant.

To these two main sources – the CsángSz. from the beginning of the 20th century and the CsángA. from the middle of the 20th century – the following corpora, gathered in the second half of the century, were added: *Magyar néprajzi atlasz* (*Hungarian Ethnographical Atlas:* MNA.); *A romániai magyar nyelvjárások atlasza* (*Atlas of the Hungarian Dialects from Romania:* RMNYA.); the ethnographic material collected by dr. Károly Kós, Judit Szentimrei and dr. Jenő Nagy published in *Moldvai csángó népművészet* (*Moldovan Csángó Folk Art*: CsNm.); and some of Péter Halász' publications, especially the terminology of livestock breeding and crop production, the names of plants and the geographical names. The lexical data from the material collected by Bálint Csűry in Bogdánfalva (Valea Seacă) and by Ödön Balogh in Csügés (Ciugheş), as well as those from Sándor Bosnyák's material on folk believes had previously been introduced into the department archive. Klára Gazda, Tinka Nyisztor, Pál Pálfalvi, Ferenc Pozsony, Vilmos Tánczos and István Virt also contributed to the corpus with their own material.

Presumably, there are three basic directions in the vocabulary changes of the Hungarian dialect from Moldova: the disappearance of archaisms; an increase in language contact phenomena (i.e., Romanian loan-words and constructions based on Romanian structural patterns) due to the dominant status of the Romanian language and the constant language shift; and a sporadic increase in the passive or active knowledge (or even use) of the vocabulary of the Hungarian standard dialect due to population mobility and education as part of the ongoing language revitalization attempts. For obvious reasons, our dictionary does not intend to capture these multidirectional changes.

The Character of the Dictionary and the Temporal and Spatial Scope of the Material

The *Dictionary of the Hungarian Dialect from Moldova* is intended to be a dialect dictionary similar to *Szamosháti szótár* (*Dictionary of Szamoshát* by Bálint Csűry, published in 1936) in that it does not refer to the origin of the words. Despite the fact that the references to Gyula Márton's Csrk. indicate that the words in question are of Romanian origin, the dictionary includes a large number of words of Romanian origin that cannot be found in Csrk. (The Romanian semantic descriptions of certain words also indicate their Romanian origin.)

The material of the dictionary was collected in the 20th century and it has been partially published. The geographical area covered by the data is the region of Moldova inhabited by a constantly decreasing number of Hungarian speaking people also called Csángós.[4] In the middle of the previous century there were almost one hundred settlements of Hungarian speakers in the region; the data used by CsángA. were gathered in forty-three collecting spots. The Hungarian speaking region has become sporadic, it may be considered a geographically and typologically secluded speech island. Consequently, our dictionary is significantly different from the previous dialect dictionaries which exhaustively presented the vocabulary of a homogeneous region or settlement. The Hungarian dialect from Moldova is both temporally and typologically stratified: the different speakers, settlements and groups seem to represent different stages of some more and more accelerated linguistic processes. This heterogeneity does not allow us to draw a line between the local dialects and the standardized dialect of Hungarian, furthermore, sometimes separating the two languages,

---

4   The reasons why I avoid using the expression *csángó* and do not recommend its use as a scientific term either are explained in a separate article on the topic (Péntek 2014).

i.e. Hungarian and Romanian, causes difficulties as well. This made us choose specific editorial solutions, uncommon in the field of lexicography. For the same reason, in addition to the cross-references which are especially important in this work, we included a third volume besides the first two volumes containing the basic corpus. The purpose of the third volume, the standard Hungarian – Moldovan dialect part is to present the Moldovan equivalents of certain senses or notions.

Many researchers made some kind of inquiry in this region, and there have been various publications related to the Hungarian population of Moldova. The bibliography on the Hungarian population of Moldova contains 5,376 items (Ilyés–Pozsony–Tánczos eds. 2006), which in itself proves that it is hopeless to aim for exhausting this topic (in the case of dictionaries this is always impossible). Even after reducing our corpus – which mostly contains spoken language data – the redaction caused difficulties in several ways. The fact that the dialectological, ethnographical and folk poetry data were collected by many people made the redaction, the categorization and interpretation of the material even harder. This already renders the phonetic annotation and semantic interpretation of the data relative. In addition, the vocabulary and general style of the folk poetry material is noticeably different from that of the everyday spoken language, and the researchers' notations and interpretations only further increased the gap between these materials. In the circumstances of a general, active, Romanian-dominant bilingualism, even the bounderies of the two languages are hard to draw. This is why we considered it beneficial to separately enlist the occasionally used, mostly Romanian expressions brought forth by linguistic interference and frequent code-switching (I could not interpret some of these expressions). For the sake of simplicity, the two Romanian velar vowels present in some of the entries are annotated unconventionally using the Romanian letters *ă* and *î*.

The Entries

The heterogeneous vocabulary of the Hungarian dialect of Moldova also made it hard, almost impossible to set the entries according to the standard practice of Hungarian lexicography. In previous dialect dictionaries the entries were matched to standard Hungarian. In our case, it would have been deceptive to follow this principle: it would have reduced the prominence of the real dissemblances between the Moldovan dialect and standard Hungarian. In the case of our dictionary, neither the excessive approximation, nor the intentional, unnecessary differentiation would be felicitous, which is why there are entries in which the words or word-forms are separate lexemes only in the Moldovan dialect, and not in standard Hungarian. The mentioned differences apply to the level of phonology as well: e.g., the words with the prefix *össze-* figure under the dialectal version **ësszé**, with proper references; the verb *benéz* 'look in(side)' figures under the form *bénez* which is the common form in this dialect; other similar entries are **csürke** (standard Hungarian *csirke* 'chicken'), **esső** (standard Hungarian *eső* 'rain'), **faszujka** (*fuszulyka* 'bean'), **fazak** (standard Hungarian *fazék* 'pot'), **feteke** (standard Hungarian *fekete* 'black'), **kalán** (standard Hungarian *kanál* 'spoon'), **kócs** (standard Hungarian *kulcs* 'key'), etc. Words that have Romanian-like forms, too, and the relation between the versions is not quite clear, e.g., **bab** and bob, are listed under the same Hungarian entry. For lack of reliable criteria, there might be some inconsistencies regarding this practice, but hopefully these will not cause difficulties in using the dictionary. I also considered it advisable to list the transitive conjugational forms of the transitive verbs in the entries, e.g., **ad(ja)** 'give sg'; the verb-forms with the inflexion *-ik* in the case of verbs which are incompatible with this suffix in standard Hungarian, e.g., **akar(ik)** 'want', **áll(ik)** 'stand'; and the word-forms of nouns

with the possessive suffix in the cases where they are more frequent than the base forms, e.g., **agy(a)** '(sy's/sg's) brain'.

The compound words were considered separate entries, and their lexeme status was determined based on standard Hungarian, thus the preverb + verb constructions are all individual entries of the dictionary. However, as a compensating practice, in the entries of the base words (the second lexeme of the compound word) there are references to the initial lexemes of the compounds.

In addition to common nouns, proper nouns – personal names (surnames, forenames and nicknames), toponyms, etc. – are also included in the dictionary. Given the specific nature of the diminutive and hypocoristic forms present in the dialect, the researchers of the *Moldovan Csángó Dialect Atlas* paid special attention to forenames, thus they became an integral part of the material of this dictionary as well.

The Structure of the Articles

Our intention was to create a dictionary that is usable and useful not only to dialectologists and linguists but also to a larger audience, which made us deviate from standard lexicographic practices. This increased the redundancy of the dictionary, especially regarding the necessary references and explanations.

The entry-words are always followed by phonetic versions of the words in italics which are listed using a simplified phonetic annotation technique in order to increase the functionality of the dictionary and to eliminate the inconsistencies of our sources. We did not deviate from our sources, but we did not follow them in every respect either (when an entry-word is also listed as a phonetic version, the swung dash ~ is used). The notations of the two types of *e* were, of course, kept in the dictionary (even in the entry-words, despite the inconsequencies of our sources in this respect), and we also indicated the use of the phonemes *ȧ*, *ā* and *ē* as well as the velar sounds using the Romanian letters *ă and î*, and the *ly*. However, we did not indicate the *aszt*-type consonant assimilations which are similar in the Moldovan dialect and standard Hungarian; their notation would in some cases cause difficulties in the interpretation of the sentences. Consequently, the dictionary cannot be used as the basis of phonetical analysis. The same applies to the inflected word-forms and the grammatical derivatives. Depending on the sources, some of these forms are listed in the dictionary, however, we did not aim for completeness, as the verbal paradigms have already been processed by Gyula Márton (MGyIge.), and the possibility of separately processing the noun paradigms is given based on the available sources (mainly texts). The corpus of our dictionary which is mainly based on questionnaire material did not allow for doing this properly.

The localization of the data was done generally or with reference to a specific settlement. The abbreviation "ált." indicates that the word or word-form in question occurs generally, independently of any geographical and typological distribution (it does not necessarily occur in each settlement or in every speaker's vocabulary). The dialectological material itself proves that in the case of the Moldovan region the internal geographical distribution is not of primary relevance, moreover, the differences between old and new forms are also of secondary importance; instead, the typological distribution, i.e. the differences between the more archaic forms of the Transylvanian Plain (*mez*) and the forms used in the Szekler version of the dialect (*sz*) are the key factor (Juhász 2001). For this reason, we totally disregarded the terms "northern" and "southern"; the abbreviations *sz* and *mez* refer to the two typological versions which may overlap in the same region or in some cases even within settlements. The abbreviation *sz* refers to the Szekler region and type, while *mez* refers to the

subtype of the dialect characteristic to the Transylvanian Plain; the abbreviation is noted using bold lettering if the word-form is truly characteristic to the specified type and region. The localizations regarding the settlements are signalled by the abbreviations of the relevant toponyms. The regional prevalency and the local versions of the words that are part of the CsángA. as well, are verifiable based on the references to the Atlas map-sheets given at the end of the articles. The semantic descriptions are given in Hungarian, Romanian and English, and in the case of plant names the Latin denominations are also listed. In the case of polysemantic words, the different meanings are indicated by Arabic numbers in the order of frequency determined based on the data.

The semantic descriptions are followed by a dash in italic lettering (–) which introduces the illustrative sentences related to the respective semantic descriptions, with the corresponding localizations. The order of the illustrative sentences usually follow the numeration of the collecting spots used in the CsángA. The vertical lines (|) between the illustrative sentences indicate that the sentences were noted down in the same settlement, but in different contexts and on different occasions; the double vertical lines (||) separate the sentences originating from different settlements. The semantic descriptions and the corresponding illustrative sentences are followed by the listing of the diminutive, respectively iterative forms and derivatives (introduced by the abbreviations **Kics.** 'diminutive form, hypocorism' or **Gyak.** 'iterative form') with illustrative sentences and sometimes interpretations, depending on the necessities and possibilities. The abbreviation **Sz.** introduces syntagms related to the entry-word. The abbreviaton **Kif.** mostly introduces verbal expressions, their semantic descriptions and the corresponding illustrative sentences. The abbreviation **Fr.** refers to idioms in the wide sense of the term: forms of greeting or swearing, idiomatic phrases, proverbs, riddles, locutions used in folk-tales, and other similar elements. The toponyms which contain common nouns and which originate from Péter Halász' publications (HPHn.) are last presented, and they are introduced by the abbreviation **Hn**. At the end of the articles, using their common abbreviations, we listed the basic sources, which is meant to help researchers in their further documentation. The abbreviation **Nr.** indicates the ethnographic relevance or aspects of the data, which is why it may occur in multiple parts of the same article.

The cross-references indicated by the sign → are either included in the articles or presented separately. These are functionally heterogenous: they may refer to an article as the main phonetic version of the item in question; in the case of the second lexemes of compound words they refer to the initial lexemes; in the case of the subordinate structural parts of syntagms or parts of expressions they refer to the corresponding entry, etc.

János Péntek