

*István Hoffmann (Debrecen, Hungary)*

***Linguistic Reconstruction — Ethnic Reconstruction\****

1. Both the modern and early history of the Carpathian Basin is characterized by a diversity of peoples and languages. Since this variety has influenced the life of both individuals and societies, it is understandable that the historical sciences have also focused on studying this issue. Direct sources referring to the ethnic composition of the Carpathian Basin describe only the past few centuries, but the interpretation and use of such data present numerous problems and challenges. Naturally, this is even more so with respect to earlier periods. A clear understanding of the past is hindered as much by the multiplicity of sources in one period as the lack of sources in others. An excellent researcher of the Middle Ages, GYULA KRISTÓ, emphasizes the almost infinite richness of source material with respect to the late Middle Ages (2003: 11). However, the main obstacle to a comprehensive study of the 10<sup>th</sup> century, when Hungarians arrived in the Carpathian Basin, is represented by a lack of source materials. Consequently, methodological questions concerning the use of secondary sources in scientific research are given priority.

It is not an easy task to decide which scientific discipline is most competent in discussing ethnic issues. KRISTÓ, in another work considering the opportunities for scientific research, calls attention to the fact that “the majority of criteria for distinguishing ethnic groups do not have a source and thus fall outside the scope of scientific study” (2000: 11). In examining the ethnic background of the early ages, he considers settlement history to be the most determinative, i.e., he analyses which people inhabited which region in specific periods of time in the Carpathian Basin (2000: 14). He considers history and linguistics the most important in this type of research. In the study cited, in which he maps the ethnic composition of the first third of the 11<sup>th</sup> century, i.e., the first decade of the Kingdom of Hungary, he relies solely on linguistic methodologies since “there is material only for one component, language” (2000: 11). In another, more comprehensive work, he completes a survey of non-Hungarian peoples in medieval Hungary by applying historical methods, “where written documents have priority over linguistic facts” (2003: 15). In his book KRISTÓ examines which ethnic groups are mentioned in the different sources, even

---

\* This work was carried out as part of the Research Group on Hungarian Language History and Toponomastics (University of Debrecen–Hungarian Academy of Sciences) as well as the project *International Scientific Cooperation for Exploring the Toponymic Systems in the Carpathian Basin* (ID: NRDI 128270, supported by National Research, Development and Innovation Fund, Hungary).



though (due to those noted above) they can obviously be linked with the languages used in the given era only indirectly. Namely, there were cases when certain groups of people called themselves *besenyő* or *kun*,<sup>1</sup> and even the surrounding groups of people referred to them as such, even though members of the community had long been speaking Hungarian. However, I do not wish to deny the correlation between linguistic situations and the names of different ethnic groups; I merely call attention to the methodological traps involved in equating or possibly even combining the two points, which could distort the plausibility of our conclusions on both levels.

A scientific analysis of multidisciplinary issues (including the subject examined here) poses other dangers as well concerning research methods. People working in different scientific fields sometimes rely on each other's findings to the extent that this reliance turns into a vicious circle, even though they might actually have the impression of moving forward in spiral progression. For example, a historian could, on the basis of linguistic research, claim that a certain area had a significant Slavic population only because there are many Slavic names in that area. The linguist, relying on the historian's findings of Slavic dominance, analyses the names accordingly and will prefer explanations leading to a Slavic origin even when other possibilities are equally plausible. The historians will consider this a confirmation of their findings and the vicious circle continues. Of course, this may apply to Turkic or Hungarian or any other language as well.

This erroneous methodological procedure is present in numerous aspects of folk etymology, casting doubt on the reliability of certain results as well. This risk, however, goes hand in hand with the complexity involved in researching eras characterized by a lack of sources. In what follows I would like to outline linguistic research methods which aim at the ethnic identification of earlier populations of the Carpathian Basin. Above all, I would like to do this by enforcing a critical approach through which I will not focus on specific details of a particular age, but rather on general issues related to key principles and methodology. Among the procedures used in this field there are certainly some that stand up to the expectations towards modern science; however, there are others that need to be modified. Even in some recent studies, one can regularly find theses that have long been refuted. Moreover, some new, so far unknown linguistic arguments can also be identified which have not been proven by historical linguistics or that may even be in sharp contrast with some of its basic tenets. I have limited my survey to toponyms, although certain linguists have studied personal names and even common nouns with similar objectives (e.g., KRISTÓ 2000: 27–41).



<sup>1</sup> Hungarian terms for Pecheneg and Kuman people (Turkic tribes).

2. We may find examples for the use of linguistic data as sources in folk etymology in Hungarian scholarship as early as the 19<sup>th</sup>-century. These included both some naive explanations and progressive findings. But it was only in the 1920s and 1930s that two people transformed linguistic tools into real scientific methodology. One of these people was JÁNOS MELICH, who analysed the ethnic composition of the Carpathian Basin in the 10<sup>th</sup> century (1925–1929), when Hungarians settled there. The other was ISTVÁN KNIEZSA, who set out to map the linguistic-ethnic composition of Hungary in the 11<sup>th</sup> century (1938 and 1943–1944). The toponym corpus recorded in chronicles and documents served as their main source, however, since these appeared only later, they could not directly contribute to the analysis of this early period. For this reason, they reconstructed certain names they could not date back. In the course of this procedure they wished to determine the circumstances of name creation and also the morphological changes involved. In their work they relied on the methods and results of etymology, historical phonology and toponym typology. The former discipline was necessary for creating a connection with a particular language, the latter two for clarifying the questions concerning chronology, and the toponyms themselves were necessary for identifying the area itself.

In the recent past, GYULA KRISTÓ, taking up a similar task to KNIEZSA and criticising his approach, emphasized that KNIEZSA's work had been applied extensively but it was never developed further or criticized (2000: 3–4). There is some truth in this, since we see that KNIEZSA's methods and results—especially among non-linguists—have been accepted without reservation and used dogmatically even in recent research. At the same time, when we look at the relevant results of Hungarian historical linguistics and research into the etymology of names from the past half century, we may also identify various innovations. However, the fact that there has been no similarly comprehensive survey since KNIEZSA's work poses an obstacle to the wider distribution of recent findings.

3. The most important means of linguistic reconstruction with an objective of ethnic identification involves toponym etymology itself, which is inherently interested in the context the name was created in, as well as in the lexical and morphological elements used. JÁNOS MELICH developed such methods into an internationally acknowledged system. His etymologies were elaborated in detail, their main virtue being the extraordinarily precise implementation of historical phonology. He always analysed the names in their overall context and examined the entire Hungarian territory. However, since he was mainly interested in names of Slavic origin, he cited examples from these languages even from outside the Carpathian Basin.

Etymology has remained the leading approach within toponymy ever since and the generations of scholars after MELICH have improved on the methodology he had developed: DEZSŐ PAIS did so in the analysis of toponyms and personal



names together and ATTILA SZABÓ T. in the case of microtoponyms. LAJOS KISS has broadened this framework by asserting analogy, i.e. the onomato-systematical aspect, but he has also applied a novel principle by allowing multiple interpretations in his explanations. LORÁND BENKŐ has reinvented the richly detailed Melich-type etymologies elaborated within their diverse systems, but he did so with reference to a higher standard of historical linguistics applied today. One by one, he (re)considers all those etymologies which are considered crucial from a historical perspective but which are actually flawed and that display linguistic superstitions.

4. The methodology of etymological toponymy reveals which tasks are to be completed when we deal with toponym reconstruction for the purpose of ethnic identification. It also clarifies which requirements are to be met when compiling a survey similar to that of MELICH or KNEZSA.

Etymology traditionally focuses on the creator of the name and attempts to deduce the form in which the toponym has surfaced. This can often be assumed as a form similar or identical to already known occurrences, however, in certain cases it might be necessary to create a hypothetical reconstruction. Besides determining the time of name-creation, it is also important to pay greater attention to surviving data, since these refer to actual language use—even if not always directly. These forms show us the linguistic context the name was created in and this is of crucial importance from the perspective of the chronology of toponym-borrowing.

The etymologies should not be derived from individual pieces of data taken out of their context, a tradition often present in older etymological toponymy. This statement is valid for two reasons. First of all, the occurrence of a name—especially its first occurrence—must be analysed both within its own context and in comparison with the more recent historical records of the same name. The connection between these pieces of data is to be revealed as part of language reconstruction and light must be shed on the differences between data—differences which are due to language change as much as to orthographic or other discrepancies. Secondly, especially in the case of early data, it is vital to interpret them as parts of the source they occur in. For example, Hungarian elements (fragments) in a Latin document or narrative may mutually reflect upon one another from an orthographic perspective or by revealing the relationship between Latin and Hungarian elements, or possibly even concerning their name structural relations. It is of utmost importance that we use the most contemporary edition of sources and that we take into account the document's accurate and precise circumstances.



The interpretation of an individual piece of data, a linguistic reconstruction of a name therefore resembles a two-dimensional coordinate system. On one of

the axes, the name is interpreted within the internal system of relations of the source material itself, while on the other axis we indicate other data referring to the place in question. In a dual reference system like this we have to take into account every single factor related to the name, among which linguistic reconstruction mainly focuses on the morphological and semantic component. Morphological reconstruction primarily represents a phonological and morphological interpretation, or, depending on the circumstances, it may aim to reveal the functional-semantic structure of a name. What we mean by the interpretation of the meaning is the presentation of the denotative meaning of the toponym, and in an optimal situation a precise localization, but at the very least a discovery of the type of place denoted by the toponym. In the past, localization was not considered an organic part of etymological reconstruction, however, without it we cannot interpret certain data as part of a data sequence that may even lead up to the modern age, which would in turn add credibility to etymologies. Furthermore, the knowledge of the types of places signified by the toponyms in question is a key factor in analysing the creation and evolution of a name, and as such an indispensable piece of information in etymological analysis.

I believe that asserting onomatopoeic embeddedness is of crucial importance in etymological onomastics. An etymology that stands alone among name formulations and does not have analogous examples is less likely to be accepted than those supported by a whole range of similar name formations. This requirement of onomastics is a consequence of the fact that toponyms form a system where both the creation and the change of names can be described by well-defined regularities, or in other words, where the majority of names can be categorized.<sup>2</sup> We should, of course, enforce the typological aspect not only when analysing Hungarian-language toponyms, but also those of Slavic, Turkic and German origin incorporated into Hungarian. Here we may encounter certain serious flaws in etymologizing: Turkic studies, for example, when analysing early Turkic toponyms, barely give internal evidence for name systems and analogous ways of name formation.

We may also mention numerous false toponym etymologies, especially concerning ethnic reconstruction involving the early period of the 10–11<sup>th</sup> centuries, some of which are in sharp contrast with each other. In these cases, since they can provide additional information on an era otherwise poor in data, it is vital to weigh the new possibilities in light of our knowledge today, even if we cannot offer an unambiguous explanation. In cases like these, we are already making progress if we can exclude proposals that can obviously be

---

<sup>2</sup> This, of course, does not rule out the possibility of unique name formations in any name system, although certain details of the creation and change of these names show typical characteristics.



refuted and show other possibilities together with their advantages and disadvantages.<sup>3</sup>

However, even in the case of an ethnic definition that is based on linguistic reconstruction we do not necessarily have to put equal emphasis on proven etymologies. Onomastics can thus be represented on a probability scale. At one end of the scale those well-identified toponyms are located that occur in numerous parts of a language territory, which are rich in data and which can easily be categorized into a name type. At the other end of the scale are the toponyms that can only be localized with uncertainty, or not at all. These derive from a lack of sources and cannot be linked with appellatives, thus their linguistic-ethnic identification potential is much weaker than the others'. The certainty of etymology is increased if a name can be explained within the framework of its own system, for which we can provide illustrative examples—and counter-examples—from the group of toponyms derived from personal names. The origin of a personal name can be proven without a doubt if there is a document attesting to it, but there are very few examples of this. Another probability-increasing factor concerning personal names is if the personal name corresponding to the toponym is present in the early personal name sources of the Carpathian Basin. If, however, the name analogy is present only in a later personal name source of, for example, a Slavic language, the etymology obviously has less credibility, not to mention cases when the etymologist gives only the reconstructed form as evidence, a frequent practice in Turkic onomastics.

Since in the reconstruction of toponyms with the aim of ethnic identification—especially in clarifying chronological relations—certain phenomena of historical phonology have gained special status, we should briefly mention them as well. It is obvious that we need to rely on the most recent findings of historical phonology when analysing toponyms, but often not only can we not meet this requirement, but we also encounter other, more serious problems in this respect. To illustrate this, I would like to present a generally known phenomenon, the denasalization of Slavic nasal vowels, which has an important role as a criterion for determining the period of toponym-borrowing in KNEZSA's work. According to him, Hungarian names with alternation (vowel-nasal relationship) indicate a borrowing that took place before the 11<sup>th</sup> century and a Slavic-Hungarian mixed population. KNEZSA himself implied the territorial inequality of this sound

---

<sup>3</sup> LORÁND BENKŐ recently published several etymologies which weigh and refute already existing ones, admittedly without offering an alternative solution (for example 2003: 133–139, 168–180). With this he has, of course increased the number of unknown etymologies, but by excluding certain names from the list of words of Turkic, Slavic and other origin, he is giving way to a new and proper assessment of these names.



change, and recent Slavic studies have proven the same in Slovenian, which was at the time an important language in the Transdanubian region; change began to take place only in the 11<sup>th</sup> century, and was completed late, especially in the northern dialects. This circumstance has to be taken into account in the linguistic reconstruction of toponyms, similarly to the way KRISTÓ (2000: 9) modified the chronology used by KNIEZSA. At the same time, he did not remedy a mistake of KNIEZSA that has even more serious consequences: namely, KNIEZSA incorporated even Hungarian toponyms that can be traced back to Slavic personal names. However, the phonological form of names that belong to the *Berente*, *Döbrönte* type can at most point to an early borrowing of personal names, which could have been transformed into a Hungarian toponym at any time in later centuries. To consider such toponyms as borrowings from the 10–11<sup>th</sup> centuries solely based on the presence of the nasal vowel in Hungarian personal names would be misleading.

It may modify our understanding of historical phonology related to the linguistic reconstruction of toponyms even more than the above-mentioned details is if in the future we not only look for the already known phonological changes, but also pay more attention to the characteristic forms of phonological adaptation, simultaneously with shedding light on the unique phonotactic features of names.

A survey of the linguistic-sociological context forms an organic part of the linguistic reconstruction of early toponyms. Steps in this direction have hardly been taken in Hungarian toponymy even though determining the pragmatic value of source data could be considered a prerequisite to its use with the objective of ethnic identification. From the point of view of interpreting data, this aspect appears as a third axis in the above-mentioned coordinate system: here we try to account for such questions as the extent to which toponyms used in documents reflect the language relations of the given area, and to what extent they can be connected to the writer of the text, etc. The position of names may also be analysed from the perspective of the circumstances of name creation and in this respect we may discover significant differences between the name-sociological value of the names of different types of referents (e.g., natural places and human settlements). This means that we should observe the role of a person and that of a community in a different manner. Within this topic an analysis of prestige relations of the languages spoken in Hungary in the Árpád Era is necessary, including among others issues of bilingualism. Unfortunately, our knowledge is still incomplete not only with respect to the general onomastic background of early Hungarian name-borrowing, but also concerning the modern age relations in the Carpathian Basin; this is true even if we can now see the events from the past more clearly in light of current research. I believe the sociology of historical toponomastics as a research field can add new perspectives to linguistic reconstruction with special attention to the reconstruction of ethnicity.



### References

- BENKŐ, LORÁND 2003. *Beszélnek a múlt nevei. Tanulmányok az Árpád-kori tulajdonnevekről.* [Names of the past speak to us. Studies of the proper names of the Árpád Era.] Budapest, Akadémiai Kiadó.
- KNIEZSA, ISTVÁN 1938. Magyarország népei a XI.-ik században. [Hungary and its peoples in the 11<sup>th</sup> century.] In: SERÉDI JUSZTINIÁN szerk. *Emlékkönyv Szent István király halálának kilencszázadik évfordulóján.* Budapest, Magyar Tudományos Akadémia. II, 365–472.
- KNIEZSA, ISTVÁN 1943–1944. Keletmagyarország helynevei. [Toponyms of Eastern Hungary.] In: DEÉR, JÓZSEF–GÁLDI, LÁSZLÓ eds. *Magyarok és románok I–II.* Budapest, Akadémiai Kiadó. I, 111–313.
- KRISTÓ, GYULA 2000. *Magyarország népei Szent István korában.* [Hungary and its peoples in St. Stephen age.] *Századok* 134: 3–44.
- KRISTÓ, GYULA 2003. *Nem magyar népek a középkori Magyarországon.* [Non-Hungarian peoples in medieval Hungary.] Budapest, Lucidus Kiadó.
- MELICH, JÁNOS 1925–1929. *A honfoglalás kori Magyarország.* [Hungary in the age of conquest.] Budapest, Akadémiai Kiadó.

### Abstract

It has been a central question in Hungarian scholarly research for a long time which peoples the Hungarians encountered at the time of the conquest of the Carpathian Basin in the 10<sup>th</sup> century and which peoples, languages they were in contact with later on. The most extensive findings in this respect were presented by researchers with the help of the analysis of old toponyms, with the works of János Melich and István Kniezsa in the 1920s and 1930s considered to be outstanding in this respect. Their results, however, need to be reviewed in light of recent theories and methods in language and toponym history. In this process we need to be aware that language and ethnicity interact with each other in a complex manner (just as in the past) and linguistic analysis may only strive to explore linguistic relations. The traditional etymological studies are replaced by toponym reconstruction, which pays more attention to context of name data within the source, while it also interprets particular data as elements of series of data referring to the same place. Analyses in historical phonology with a modern approach play a key role in this but aspects of toponym typology and historical socioonomastics cannot be missed either during such studies.

**Keywords:** old Hungarian toponyms, relationship of language and ethnicity, languages and ethnicities of the Carpathian Basin in the middle ages, toponym etymology, toponym reconstruction

